University of Alberta

Identification and Characterization of Differentially Expressed Proteins During Dormancy Acquisition in White Spruce (*Picea glauca*)

by

Leonardo Miguel Galindo González

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

in

Plant Biology

Department of Biological Sciences

Edmonton, Alberta
Spring, 2009

ABSTRACT

To cope with adverse conditions, plants go into a dormant stage in which meristem growth stops. Perennial plants enter a dormant state as one of the adaptations evolved to withstand overwintering. For overwintering plants, the metabolic changes that take place during dormancy acquisition largely determine the plant's status for spring regrowth. To study the changes of *Picea glauca* during dormancy acquisition, I compared the protein profiles of woody stems from two-year old seedlings undergoing active growth to those of seedlings entering dormancy.

Proteome analysis using two-dimensional gel electrophoresis and liquid chromatography – tandem mass spectrometry showed that there is an increase in proteins related to stress response, carbon metabolism and energy status. Chitinases were prominent among the differentially expressed proteins, and were selected for *in silico* characterization using data mining tools. The analyses suggested that these chitinases could have a storage, catalytic or antifreeze function, or a combination of these functions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

.

LIST OF ABBREVIATIONS

ABA	Abscisic acid
AEBSF	4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride
BCA	Bicinchoninic acid
BSA	Bovine serum albumin
CATH	Class architecture topology homologous
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CBD	Chitin binding domain
CRD	Cysteine-rich domain
CTE	Carboxy-terminal extension
Da	Daltons
DIECA	Diethyldithiocarbamic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
ESTs	Expressed sequence tags
FSA	Fraction of surface area
FDR	False discovery rate
GCAT	White spruce gene catalog
G3P	Glyceraldehyde 3-phosphate
IAA	Indole acetic acid
IEF	Isoelectrofocusing
IPG buffer	Carrier ampholyte mixture
KEGG	Kyoto encyclopedia of genes and genomes
LB	Liquid broth
LPB	Long pass blue
LPG	Long pass green
MP	Maximum parsimony
ME	Minimum evolution

MUSCLE	Multiple sequence comparison by log expectation
MW	Molecular weight
NJ	Neighbor joining
NL	Non-linear
ORC	Origin recognition complex
ORF	Opening reading frame
OSCs	Ordered surface carbons
pI	Isoelectric point
PDB	Protein data bank
PPP	Pentose phosphate pathway
PR	Pathogenesis related
PVPP	Polyvinylpolypyrrolidone
ROS	Reactive oxygen species
SD	Short days
SSAP	Structure sequence alignment of proteins
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TSA	Total surface area
UPGMA	Unweighted paired group method with arithmetic mean
VHs	Volt hours

1.0 INTRODUCTION

1.1 White Spruce (*Picea glauca* [Moench] Voss)

The genus *Picea* (spruces) has more than 30 species distributed worldwide. Eastern China and Japan have the greatest number of species of the genus, and apparently this region is the center of origin (Ogilvie 1972). Many of the species cross-hybridize. For example white spruce (*Picea glauca*) and Engelmann spruce (*Picea engelmannii*) hybridize naturally in their overlapped ranges in British Columbia, and along with their hybrid they are collectively called interior spruce (Grossnickle 2000).

White spruce is an evergreen, cone-shaped tree widely distributed in North America and Canada (Figure 1.1), mostly through coniferous forests in temperate and boreal regions, and is usually located in low to mid elevations (Grossnickle 2000). Structural characteristics include dense, four-sided, stiff blue-green needles that release a strong smell when crushed. The twigs are hairless with a pale whitish-grey to yellow color; the bark is scaly and grayish-brown in the outside; cones are slender and cylindrical with a light brown color and they usually open in late summer or autumn before falling on the ground to grow usually during the next season. The root system is shallow and so the trees can easily be blown down (http://www.lrconline.com/Extension_Notes_English/pdf/wht_sprce.pdf). White spruce grows better in moist sandy loams, but it is able to withstand a large range of conditions including moisture, pH (4.5-7.5), pollution and high salt (Ritchie 1959). It can live up to 300 years and be as tall as 30 meters with a maximum diameter of 90 centimeters (http://www.lrconline.com/Extension_Notes_English/pdf/wht_sprce.pdf).

White spruce's wood is light, soft and moderately strong, and therefore is an important source of pulpwood and construction grade lumber, thanks to its strong long fibers. The product can be made into a variety of papers, fiber molded products and boards (http://www.lrconline.com/Extension_Notes_English/pdf/wht_sprce.pdf). Additionally, it also gives food and shelter to many animals. Its seeds are an important

source of food for birds and small mammals like squirrels, and it can also provide shelter in the winter months for animals like voles and hares.

Most of spruce's characteristics make it a widely distributed species in high latitudes, which coupled with is high demand for lumber and pulp, make it an economically important tree in Canada. On a national level, white spruce constitutes 26% of the softwood stock and 20% of its total forest inventory (http://www.saskforestcentre.ca/uploaded/White_Spruce_tech_sheet.pdf). According to the (Government of Canada, 2007 - http://foretscanada.rncan.gc.ca/rpt), Canada has 402.1 million hectares of forest and other wooded land which represents around 10% of the world's forest cover and 30% of the boreal forest. The forest industry focused on pulp, lumber, paper and wood products contributes 3% to the gross domestic product of the country and comprises 10% of exports. In 1994 it was estimated that Canada planted 436 million seedlings of spruce alone from a total of 695 millions planted that year (Grossnickle 2000). Additional value of trees is expressed in the habitat they provide to wildlife, the economic stability of communities by supplying jobs, space for recreation, and the stability that they can give under global climatic changes, acting as carbon sinks.

1.2 Overwintering overview

Winter in high latitudes is characterized by a suite of limiting conditions for plant growth and so plants need to halt development and prepare for the potential biotic and abiotic stresses that come with the season. Big changes in the physiology of the tree are triggered by pronounced changes in factors like temperature, light and nutrient availability.
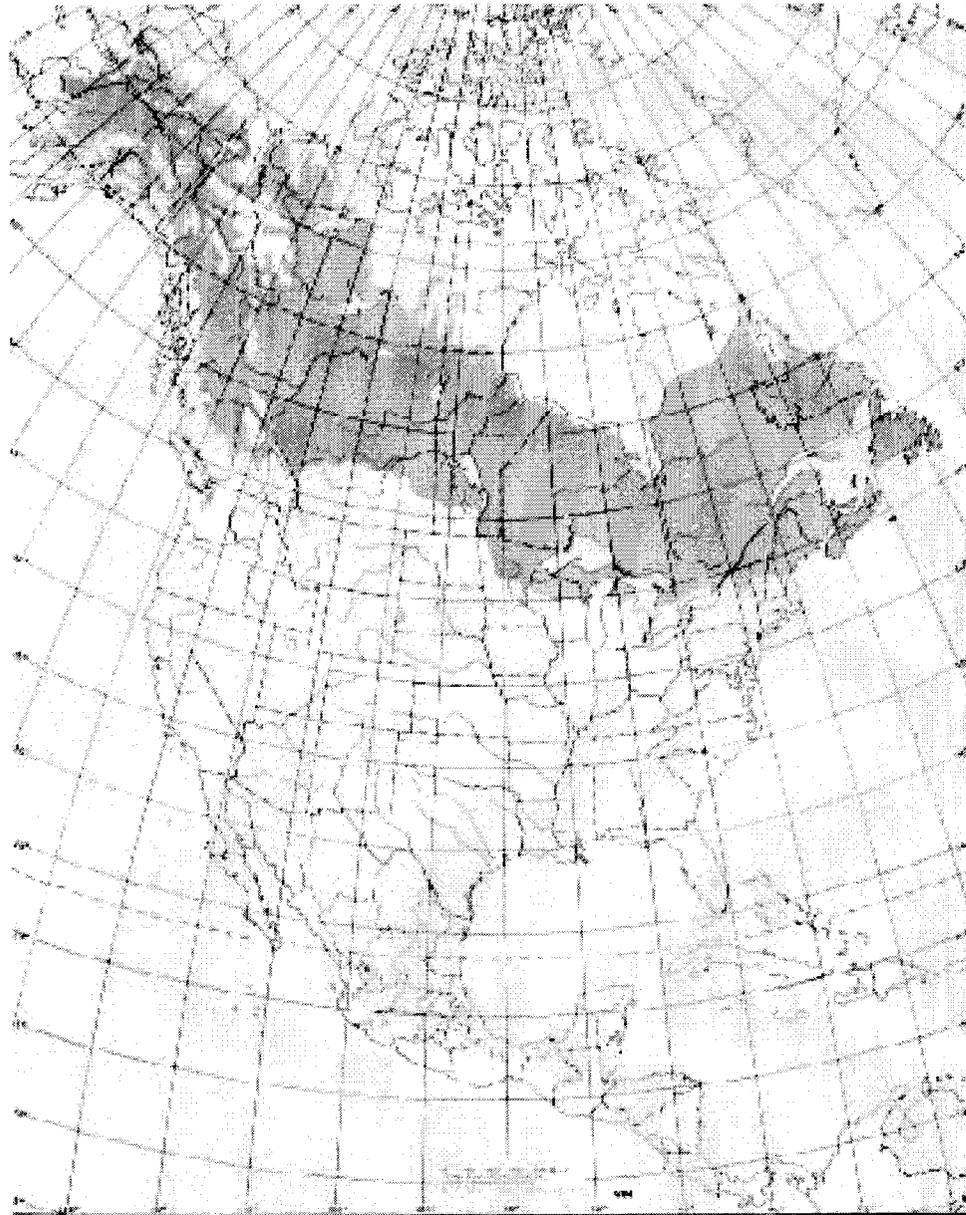
Figure 1.1. White spruce distribution in North America (Burns and Honkala 1990)

1.2.1 Cold temperature and associated stresses

Plants have different mechanisms to acclimate to low temperature. While cold acclimation allows plants to rapidly adjust to low, non-freezing temperatures, a more extensive change has to occur, for example, in trees in high latitudes, which develop freezing tolerance or cold hardiness as a way to withstand temperatures well below the 0°C mark (Welling and Palva 2006).

Ice formation can either occur inside of the cell, or outside in the apoplastic space. Decreases in temperature going into winter are generally gradual, and the plant can usually develop the cold acclimation processes in parallel and intracellular ice formation is rare. In plants adapted to low temperature, for intracellular ice to form, the decrease in temperature has to be fast (10°C per minute), which seldom happens in nature (Glerum 1985). However, under normal conditions, below the freezing point water starts freezing outside the cells, which draws the intracellular water to the outside causing cell dehydration. Dehydration is believed to be the biggest factor in freeze-induced injury (Glerum *et al.* 1985; Welling and Palva 2006) and brings along problems in the cell membrane, protein denaturation and increase in toxic solutes.

So upon cold and dehydration stresses, the plant needs to synthesize a series of compounds to stabilize the cells. Welling and Palva (2006) did a broad review of the compound and proteins synthesized upon cold acclimation in trees. Among these are the cryoprotectants which include proteins and sugars; sugars for example can change the structure of the membranes or cytoplasm by creating a glass-like substance which makes the cells highly viscous and stable; at the same time sugars can act as osmolytes, decreasing the likelihood of water leaving the cell. While certain proteins like chaperonins can bind proteins to guarantee the correct folding and therefore physiological function, others like antifreeze proteins, impair extracellular ice growth. Some of the cryoprotective proteins also protect membranes, and along with enzymes like desaturases, guarantee the maintenance of fluidity and association between lipids and membrane proteins. Additionally, an over excitation of the photosynthetic complex can result in the generation of reactive oxygen species (ROS), which results also in the synthesis of detoxification enzymes.

Fowler and Thomashow (2002) showed that in *Arabidopsis thaliana* cold acclimation response shows genes that act as cryoprotectants like the cold responsive (COR) and late embryogenesis abundant proteins (LEA), which are also highly hydrophilic. Enzymes involved in sugar synthesis (sucrose and galactinol synthase), which can stabilize the cell and work as osmotins, are also part of a long term response. A light stress responsive gene known as early light-induced protein (ELIP1) seems to be expressed as a means to photoprotect pigments. Cold temperature stress is also coupled with a decrease in turnover rate of the photosystem II, causing an over excitation, which in turns results in the formation of reactive oxygen species (ROS). In this sense, ROS responsive genes like gluthatione S-transferases are activated to help with detoxification. Besides these changes, a large number of transcription factors, kinases and phosphatases are also upregulated, as a mean to control the transduction cascades. Part of the responses were said to be mediated by the upregulation of the C-repeat binding factors (CBFs) also known as drought responsive element binding proteins (DREBs). These transcription factors play an important role in the cold response, by binding c-repeat boxes in the promoters of cold and drought responsive genes. As in Arabidopsis, the CBF regulon and the specific responses to cold are also functional in some tree species (Welling and Palva 2006).

As aforementioned an imbalance between electron transport due to cold temperature can result in detrimental consequences for the plant. Over-excitation energy of the photosystem II can be diverted by non-photochemical quenching, reducing energy transfer to photosystem II or by transferring some energy to photosystem I (Hüner *et al.* 1998). In the long term plants can also adjust the levels of the electron consuming sinks (carbon reduction and oxidation cycles). The adjustments of consuming sink enzymes also result in the possibility of generation of molecules like sucrose, necessary to withstand cold-related stresses (Stitt and Hurry 2002). Therefore upon cold stress, increases in the enzymes for carbon fixation are accompanied by a capacity for increased sucrose synthesis.

The changes occurring in cold acclimation are generally the result of low temperature, but factors like short photoperiod, the modulation of hormones like ABA and water stress can also trigger, at least in part, the same process (Lee and Chen, 1993).

Due to this fact, cold hardiness is often linked with other processes like growth cessation, bud set and the development of dormancy (Howe *et al.* 2003), which makes it sometimes difficult to separate the response to each process. However, growth cessation often preceeds the development of cold hardiness. In species with indeterminate growth, where preexisting and neoformed stem units elongate during the growth season, the growth cessation and bud set largely depends on the shortening of the day. In contrast, in species of determinate growth such as white spruce, where elongation comes primarily from preformed units in the bud of the previous growing season, bud set depends on the stem unit and their maturation. Either way, the cold acclimation process starts developing in response to shorter days, and the increasing exposure to low temperatures (Howe *et al.* 2003)

1.2.2 Nitrogen storage reserves and seasonal nitrogen cycling

Nitrogen is an essential component of plant metabolic processes. Its assimilation and recycling is closely related to the ability of the plant to grow and develop, and therefore soil nitrogen availability can be a factor limiting plant growth. That is why the nitrogen that is released from different metabolic processes in the plant during the growing season tends to be recycled for future use. In trees, processes like photorespiration, the phenypropanoid pathway and one-carbon metabolism require the recycling of the inorganic nitrogen liberated (Cantón *et al.* 2005). Plants also mobilize nitrogen according to source-sink relationships, meaning that nitrogen will be mobilized to sinks that demand it, like actively growing young leaves, shoots or fruits. However, seasonal nitrogen mobilization from leaves to perennial tissues is a requisite in perennial temperate plants, to retain and conserve nutrients that will help with regrowth during spring until the full photosynthetic apparatus is capable of producing the molecules for continuous growth. In this sense, resorption of nitrogen to be stored over-winter constitutes a component of nitrogen use efficiency (reviewed in Coleman *et al.* 2004; Cooke and Weih 2005). The nitrogen is deposited in several proteins, some of which are

recognized mainly for an exclusive nitrogen storage function and are widely known as vegetative storage proteins (VSPs). These proteins are described in the next section.

1.2.2.1 Vegetative storage proteins

The vegetative storage proteins (VSPs) are a heterogeneous group of proteins that serve a nitrogen storing function. In a strict sense, a storage protein is usually a protein which is mainly used for nitrogen reserve, has no enzymatic function and is broken down when the reserve is mobilized for reuse (Staswick 1994). However in a wider sense, these proteins can be synthesized upon resource surplus and can be used for diverse growth processes (Cooke and Weih 2005). Additionally, many vegetative storage proteins in this wider sense can still conserve their enzymatic function, as is the case of the alfalfa (*Medicago sativa*) VSP which has chitinase activity (Meuriot *et al.* 2004); the VSPs from soybean (*Glycine max*) which have phosphatase activity (DeWald *et al.* 1992); or the lipoxygenase of soybean (Grimes *et al.* 1993).

The initial identification of VSPs from alfalfa was made by the detection of seasonal changes of polypetides that accumulated during autumn and were hydrolyzed in spring (reviewed in Avice *et al.* 2003). However upon application of methyl jasmonate, VSPs of this plant also accumulated in taproots (Meuriot *et al.* 2004). This effect was attributed to the chemical having a direct effect on the transduction regulating VSP expression, and in changing the N partitioning, leading to a preferential allocation in the storage organs. It has been suggested that upon wounding signals from pathogens and insects, the plant releases linoleic acid which can be transformed into jasmonic acid, creating a signal transduction that results in VSP expression (Creelman and Mullet 1997 in Meuriot *et al.* 2004); a chitinase acting as a VSP then can serve a double function, being a transient storage protein and an enzyme against pathogens and insects. Another sequence found to have a transient storage function is a banana chitinase-like protein (Peumans *et al.* 2002). This protein accounted for up to 40% of the protein during fruit maturation, but a chitinase catalytic activity could not be detected.

In poplar (*Populus* spp.), a major storage protein of 32 kDa corresponding to bark storage protein (BSP) was present over winter and disappeared in the spring months, and a peak of transcript expression was seen likewise during September (Clausen and Apel 1991). The BSP transcripts accumulated upon short days (Coleman *et al.* 1992; Zhu and Coleman 2001a) and were demonstrated to be controlled by phytochrome (Zhu and Coleman 2001b). BSP transcripts were also shown to accumulate upon increased nitrogen supply (Coleman *et al.* 1994; Zhu and Coleman 2001a) along with *win4* transcripts, a wound inducible gene which accumulated in leaves, in contrast to BSP, that accumulated in the bark. Both of these genes appear to be closely related and also respond to wounding, BSP being again preferentially expressed in stems while *win4* is expressed preferentially in leaves (Davis *et al.* 1993).

BSP was immunolocalized in vacuoles of the ray xylem cells. The seasonal accumulation of the proteins bodies in small vacuoles between lipid droplets and amyloplasts had been also characterized before (Sauter *et al.* 1989) and VSPs in several trees have been shown to accumulate not only in xylem ray cells but also in the phloem parenchyma and cambial cells (reviewed in Stepien *et al.* 1994). After the characterization of this storage protein in poplar, a similar protein of the same molecular weight that followed a winter accumulation pattern was characterized in willow (*Salix microstachya*) (Wetzel and Greenwood 1991). The protein was accumulated in the cambium, phloem parenchyma and xylem ray cells, and was shown to be a glycoprotein, like poplar VSPs (Langheinrich and Tischner 1991). Glycosylation has been encountered in numerous VSPs and its related to protein thermostability and increased carbon accumulation (reviewed in Stepien *et al.* 1994). Seasonal accumulations of proteins have also been studied in several softwoods, showing changes in the accumulation of apparent storage proteins in some of them in their bark but not in their needles (Wetzel and Greenwood 1989). In interior spruce and Douglas fir (*Pseudostuga menziesii*), polypeptides close to 30 kDa also showed seasonal accumulation in buds, shoots and roots, but not in leaves (Robert *et al.* 1991). Accumulation of the proteins in interior spruce buds was later correlated to an increase in frost hardening of seedlings (Binnie *et al.* 1994). Bud accumulation was also seen for BSP in hybrid poplar (*Populus*

trichoparpa * *Populus deltoides*), but other related members like WIN4 and *pni288* did not show this pattern of seasonal accumulation.

In summary, VSPs can be triggered seasonally, in response to source-sink nitrogen relationships, and by response to several abiotic factors (reviewed in Avice *et al.* 2003, Staswick 1994). Their importance lies in their function as transient nitrogen storage compounds which plants can use once metabolism requires them.

1.2.3 Dormancy

1.2.3.1 Dormancy defined

Although dormancy was previously defined as the temporary cessation of visible growth of plant structures with meristems (Lang 1987), this can be expanded by saying that the meristems of the plant are impaired from resuming growth under favorable conditions (Rohde *et al.* 2007). Dormancy in plants can be imposed in different structures like seeds, bulbs, buds, tubers, and the whole plant, and can be dependent on life span and distribution of the plant. In an evolutionary context, it is hypothesized that dormancy evolved as a means to withstand extreme environmental conditions like heat, cold and drought (Hilhorst and Koornneef, 2007). In fact dormancy in perennial plants (plants that live for more than two years) like trees, is very much related to the concept of avoidance of harsh conditions; these plants can also stop growth in response to seasons, which is also related to changes in the environment (Rohde and Bhalerao 2007).

The phenomenon of dormancy is often further divided into three categories: (1) ecodormancy, when the dormant state is imposed by the changing and unfavorable conditions of the environment; in this scenario plant meristems cease activity in response to signals like short photoperiod, cold or drought stress, but the plant could resume growth if conditions are suitable again; (2) paradormancy where dormancy in the specific plant structure is imposed by the influence of a different structure within the plant and outside the dormant tissue, as in the case of lateral buds; and (3) endodormancy (true

dormancy) where the dormancy is imposed from within the dormant structure or tissue (Lang 1987). This later condition happens with increasing abiotic stress conditions in winter that eventually make the plant insensitive to positive environmental cues; in fact cells in this state seem to be symplastically isolated, which would stop growth promoting signals (Rohde and Bhalerao 2007). Breaking the endodormant state of a plant generally requires chilling for a specific time, and in certain cases this is accompanied by the effects of temperature at bud set, clinal variation and day length (Søgaard *et al.* 2008).

Although dormancy is not the only mechanism to withstand adverse conditions, it seems to be very effective in protecting the plant and guaranteeing its continuity and functioning. In fact, the dormant state in plants does not mean that molecular mechanisms in cells become completely inactive.

1.2.3.2 Dormancy in trees

In temperate and boreal trees, adaptations have evolved to cope with the seasonal environmental changes of higher latitudes. Woody plants in these regions exhibit a well defined dormancy stage interrupting their annual growth cycles as an adaptation to overwinter in high latitudes where climate does not allow for active growth during winter. Although many organs in the plant can become dormant, it is in the buds where dormancy is most evident. The apical meristem from trees generates a growing shoot tip during the active phase of the annual growth cycle, which in turn produces a specialized protective structure called the bud as part of the changes that are made for overwintering. The full bud is not dormant as a whole, but it does contain dormant tissues.

In most temperate forest tree species such as white spruce, dormancy in the apical meristems has three stages: quiescence, which occurs in mid summer where growth is controlled by environmental factors; rest, which generally starts in autumn and is controlled by bud physiology; and postdormancy, where growth is again controlled by the environment (Grossnickle 2000). These three stages are similar to the eco and endodormancy described above, and apparently in white spruce approximately 6 weeks at 5°C fulfills the chilling requirement to move to post-dormancy. However, chilling is not

an absolute requirement in some species such as Norway spruce, because limited bud flush can occur after a long delay without chilling (Soogard *et al.* 2008). Chilling, however, allows a rapid and synchronous bud flush. Therefore, according to the definition that we have given for true dormancy, Norway spruce may not exhibit true dormancy, but it does exhibit endodormancy in the practical sense of the word.

While dormancy is very well studied in buds, fewer studies have been done so far in cambial meristems. The onset of dormancy triggered by environmental signals seems to be similar between buds and cambial meristems, as is the common process of growth cessation. However as opposed to buds, there are no major development anatomical changes that can be correlated with the molecular changes of dormancy in the cambium (Schrader *et al.* 2004). Cambial meristems also go through the process of cold acclimation, increasing tolerance to stresses, accumulation of storage compounds and regulation of hormonal growth signals as buds do, although the exact onset of the genes may differ. Transcriptional studies by Schrader *et al.* (2004), Druart *et al.* (2007) and Park *et al.* (2008) have demonstrated how the changes in photoperiod, temperature and hormones result in the responses mentioned above.

Many molecular changes can however be easily associated with visible cellular changes in the cambial meristem cells. Growth cessation of cells, along with thickening of cell walls can be observed for example in aspen during autumn (Druart *et al.* 2007), and can be related to the changes in transcription profiles. The accumulation of storage nitrogen reserves in protein vacuolar bodies has been identified in several trees in the ray cells and bark parenchyma (Stepien *et al.* 1994). The diversity of changes and responses at the cell and molecular level will be outlined in the following sections.

1.2.3.3 Factors that influence dormancy in trees

It has been known for a while that the main factor influencing dormancy in trees native to cool temperate climates is day length (photoperiod). In woody plants, photoperiodic response studies were already performed in the 50's (Wareing 1956) and surveys on over 60 species showed a clear response of growth cessation when plants

where exposed to short photoperiod (in general 8 hours of daylight and 16 hours of night). The specific response of each plant to enter the state of dormancy depends on a combination of factors like age of the plants, species and geographical distribution. Photoperiod control is determined at the molecular level and is mediated by the main daylength detecting antennas, the phytochromes. Phytochromes are proteins coupled to chromophores that can detect changes in day length thanks to changes in red and far-red light typical of day and night respectively (Rüdiger and Thummler 1994). The active form, Pfr, is generated by excitation with red light and induces a conformational change that activates signal transduction cascades (Roux 1994), which result in specific gene responses. Experiments with phytochromes have shown a clear control of the dormancy response for numerous plants. Probably the best evidence supporting a role for phytochrome in dormancy imposition comes from experiments with transgenic hybrid aspen (*Populus tremula* x *alba*) overexpressing an oat phytochrome, in which the transgenic plants failed to halt growth even under short day conditions (Olsen *et al.* 1997). Other physiological experiments with forest tree species also suggest a role for phytochrome in dormancy induction. In black cottonwood (*Populus trichocarpa*) grown under short day conditions, interruptions with red light prevented the cessation of internode elongation which is part of the growth cessation process in dormancy (Howe *et al.* 1996). Similarly in blue spruce (*Picea pungens*), dormancy is prevented when night periods conducive to dormancy are interrupted by red light night breaks, which emulate daylight (Young and Hanover 1977); in Norway spruce (*Picea abies*) there is a latitudinal requirement for R/FR (red/far red) ratio to stop elongation (Clapham *et al.* 1998). Together, this evidence shows the role of phytochromes in controlling growth cessation.

Different ecotypes of the European white birch *Betula pendula* experienced growth cessation, bud set and an increase in freezing tolerance in response to short days (Li *et al.* 2005); the dormancy development, growth cessation and cold acclimation were all earlier events in the northern ecotype showing a latitudinal cline in the photoperiodic responses. In poplar more than 50% of light-dependent genes from a bud formation microarray experiment were up-regulated when plants were changed from long day to short day conditions (Ruttink *et al.* 2007); the study showed that although there might be

a relationship of some of the genes to dormancy induction, a direct control of these genes in dormancy was not established.

A relatively new important discovery has been made on the controlling mechanism downstream in the light signal transduction chain. This response seems to be mediated by the gene FT (FLOWERING LOCUS T), which was previously only linked to time of flowering, but was shown to be involved in growth cessation and bud set in *Populus* trees upon downregulation (Böhlenius *et al.* 2006). A homolog in Norway spruce (*PaFT4*) is also involved in bud set but the gene is instead upregulated (Gyllenstrand *et al.* 2007). In fact 32% of flowering-related microarray tested genes in poplar were overexpressed during bud formation and dormancy induction (Ruttink *et al.* 2007), which although limited, suggest some overlap in regulation between the processes.

Photoperiod seems to be the main environmental cue that triggers dormancy in trees like poplar which have indeterminate growth, and although in determinate growth species like spruce the change in photoperiod may be sufficient for dormancy induction, it is not strictly necessary (Rohde *et al.* 2000). Overall, photoperiodic control is one of the main mechanisms to go into the dormant stage, but other regulatory signals can have different levels of influence as mentioned below.

Another factor that seems important in dormancy is the relative changes in hormone composition and their regulation. Hormones like gibberellins, auxins and ABA are known plant growth regulators. For example, it was demonstrated that when the GA synthesis inhibitor paclobutrazol was used in hybrid aspen, stem growth stopped and when the inhibitor was combined with low temperature, bud set was consequently induced (Mølmann *et al.* 2005). At the molecular level there are several genes that control the transduction and synthesis of the hormone. In response to short day photoperiod, GA-20 oxidase, which constitutes a key enzyme for gibberellin biosynthesis, was downregulated and followed by decreased level of GA_1 (a gibberellin that controls elongation) in late expanding leaves of hybrid aspen (Eriksson and Moritz, 2002); in the same way GA repressors are highly up-regulated in buds subjected to short day treatments (Ruttink *et al.* 2007), indicating a strong control through phytochromes. However, contrary to what was expected, the study by Eriksson and Moritz (2002) showed slight up-regulation of the GA-20 oxidase for internodes and no change in GA_1

levels under short days, while Olsen *et al.* (1997) showed that levels of these and other gibberellins where reduced in hybrid aspen for the same kind of tissue for wild type plants exposed to short days; however, the upregulation in the enzyme of the former study does not necessarily mean more production of gibberellin or additional growth since post-transcriptional control is always possible. GA-20 oxidase genes were also investigated in the cambial meristems of aspen, and no down-regulation was detected (Druart *et al.* 2007), supporting other kinds of control for giberrellin production or post-transcriptional regulation. In fact, since plasmodesmata of the cells are blocked during endodormancy, even if gibberellins were produced, part of their signaling would be impaired (Rohde and Bhalerao 2007). Additionally up-regulation of negative regulators of GA signaling, have also been reported (Druart *et al.* 2007), so another point of control independent of GA synthesis could also contribute to growth cessation.

Despite these mixed patterns there seems to be a definite link between GA production, sensing and GA mediated regulation, and changes conducive to dormancy, so some conclusions can be drawn. It can be said that a lower level of GA is related with growth cessation since GA is widely known as a growth promoter; and there is an inherent link between light control by phytochromes and the production of gibberellins; finally, the absence of regulation of GA biosynthetic genes is not the only way to regulate GA signaling for triggering growth cessation, since additional mechanisms like insensitivity to the hormone could also account for modulation.

Although the role of auxins is not completely understood yet, auxins have been recognized as key hormones in maintaining apical dominance, a process that is related to paradormancy (reviewed in Rohde *et al.* 2000), and it is known that during dormancy basipetal transport of auxin is reduced and the cambium becomes insensitive to auxin (Schrader *et al.* 2003; Schrader *et al.* 2004). Studies in European aspen (*Populus tremula*) showed that indeed specific efflux auxin carriers (*PttPIN* genes) were down-regulated during dormancy in the cambial meristem (Schrader *et al.* 2003; Schrader *et al.* 2004). In the second study, the auxin transporter *PttPIN1 and PttPIN2* decreased their levels upon dormancy, however, not all the transporters were repressed so additional mechanisms should account for the auxin modulation. Schrader *et al.* (2004) also investigated the genes involved in auxin response. Some auxin inducible genes (*Aux/IAA*), involved in

repression of auxin response factors (*ARFs*), were downregulated but most *Aux/IAA* and *ARFs* were unaltered, and no changes were either visible in the proteins that mediate *Aux/IAA* degradation, indicating that this was not the main cause for changes in auxin sensitivity. In *Eucommia ulmoides* trees, the pattern seemed somehow more clear at least for the hormonal fluctuation; using immunolocalization and HPLC over one year in cambium and mature vascular tissues it was shown that the auxin IAA had its peaks over the growing season (spring and summer) and decreased during fall and winter (Mwange *et al.* 2005). These pronounced changes in auxin content were mainly viewed in the cambial region and are in agreement with the role of auxins in cambial cell division and differentiation. Furthermore, the putative auxin receptor ABP1, has also showed transcription and protein patterns that are positively correlated with the IAA oscillations (Hou *et al.* 2006). Auxins therefore, are considered to be promoters of cambial growth and their associated sensitivity seems to decrease upon dormancy.

ABA, which has been linked to numerous stress and growth processes in plants, is also associated with dormancy. Previous studies have shown the relationship of ABA to promote and maintain seed dormancy, and this field has been recently reviewed (Finlkelstein *et al.* 2008). But studies focused on winter dormancy have also shown a role for ABA. In a study of short day induction towards dormancy in buds of poplar, ABA response was studied using over-expressor and anti-sense poplar lines of the ABI3 transcription factor which seems to control many ABA related processes. The study showed that ectopic and putative targets of the over-expressor and wild type lines were related to the adaptive responses and to a lesser extent to light signal transduction regulators (Ruttink *et al.* 2007). These changes apparently precede growth cessation and the putative changes are said to be involved in bud formation. This latter fact is supported by the expression of not only biosynthetic but also signal transduction genes of ABA prior to bud set. However, at least in this case, the ABA response is mediated in its majority by short day photoperiod, since most of the ectopic and putative target gene differential expression depended on short days (Ruttink *et al.* 2007).

In aspen, ABA levels also increase in the cambial meristem due to short days in conjunction with low temperature but after growth cessation, and it was speculated that in this case ABA also responds downstream of the short day induction (Druart *et al.* 2007).

This showed that the effect of ABA is not only dependent on photoperiod but that low temperatures also have an effect on hormone regulation. In silver birch (*Betula pendula*) ecotypes subjected to short days, increased their shoot apex ABA levels with a slight difference of higher accumulation for the northern ecotypes (Li *et al.* 2005), but when low temperatures were imposed after short days the levels of the hormone started fluctuating, confirming that ABA changes are influenced also by temperature. In fact it is evident that ABA by itself may not trigger the whole dormancy process in plants like white birch, where external ABA applied to the ecotypes caused an increase in freezing tolerance but no growth cessation (Li *et al.* 2003). In conclusion, ABA seems to have an important role in growth cessation but possibly its function at multiple plant levels might obscure its role in dormancy.

Temperature seems to be another critical factor in the onset of dormancy. In temperate and boreal climates, trees have to prepare to withstand harsh winter conditions. When days shorten and temperatures start falling in autumn, trees start acquiring freezing tolerance. As the exposure to increasingly lower temperature continues, the plants develop frost hardiness and at some point are also able to break endodormancy when the plant accumulates a specific amount of time at low temperatures, which secures a correct period of overwintering (reviewed in Welling and Palva, 2006). However it seems that while a chilling requirement is necessary in many temperate and boreal plants to break dormancy, a decrease in temperature during fall is not the main or only factor in determining the changes towards dormancy. For example in aspen the cessation of cell division precedes suboptimal temperatures for growth (Druart *et al.* 2007), showing that photoperiod is more relevant in controlling dormancy. A combination of the gibberellin inhibitor paclobutrazol and low temperature are effective in triggering bud set and cold acclimation but not dormancy under long day non-inductive conditions in hybrid aspens (Mølmann *et al.* 2005). This means that temperature can not override the phytochrome system for the onset of dormancy. Also, in different latitudinal ecotypes of dogwood (*Cornus sericea*) a specific number of short days induces dormancy according to the latitudude of origin (more are needed for the southern ecotypes), but when ecotypes from Utah and NW territories were subjected only to low temperature, endodormancy was only induced in the northern ecotype (Svendsen *et al.* 2007). Control of bud set in poplar

hybrids is strongly dependant on photoperiod, but temperature seems to also play a big role too (Howe *et al.* 2000). So apparently a combination of these factors determines the onset of dormancy changes and the generation of frost hardiness. However, in some trees of temperate regions like apples and pears temperature had an absolute control over growth cessation and dormancy, and photoperiod apparently had no influence (Heide and Prestrud, 2005), which seems to be quite different to the predominant photoperiod control for most trees of temperate regions. In conclusion temperature seems to be a factor contributing to dormancy changes but its influence depends largely on other factors like hormone modulation and photoperiod.

1.2.3.4 Physiological, cellular and molecular changes during dormancy induction

During dormancy, trees have to withstand harsh conditions like cold temperature and reduced photoperiods, and this means restructuring their physiological and metabolic processes through changes in their cellular machinery and components. The transition to dormancy occurs throughout the plant but is usually centered and more important in the meristems that suspend growth. While cell division and growth stop, many molecular processes continue to adapt to the new conditions. The series of changes include restructuring of the carbon metabolism, strategies against the cold weather, activation of stress response mechanisms and accumulation of storage compounds that can be reused upon regrowth in spring.

In annual plants, where one growth cycle equals one life cycle, the use of nutrient reserves takes place during the lifespan of the plant. However in perennials, which have multiple growth cycles, some of the resources contained in parts of the plant that are shed at the termination of a growth cycle are recycled back into the perennating parts of the plant (i.e. those parts of the plant that remain during dormancy) at the termination of one cycle (Okubo 2000). These resources are then available to be used in the following growing season, principally to drive early growth before maximum photosynthesis is attained (Cooke and Weih, 2005). In cambial meristems of black cottonwood, 25% of all the soluble protein during winter comprised the two major forms (32 and 36 kDa

polypeptides) of the bark storage protein (BSP) (Langheinrich and Tischner, 1991), and the 32 kDa BSP of eastern cottonwood (*Populus deltoides*) constituted up to 62% of the total soluble protein upon 24 days of short day induction. At the transcript level, *BSP* accounted for 20% of all ESTs in a dormant cambium library (Schrader *et al.* 2004). These facts outline the importance of storing nitrogen in the form of specialized proteins which can be later broken down to use their amino acids for new protein synthesis or other metabolic processes. In peach (*Prunnus persica*), VSPs account for 13% of the mobilized protein and they seem to be the first nitrogen reserves to be mobilized. The reason for mobilization could be related to source-sink relations since it is apparent that new growth in buds and shoots requires nitrogen reserves; in poplar, the degradation of BSPs may depend solely on bud burst and therefore some form of communication (maybe hormonal) must work between the apex and the stored proteins in the bark (Coleman *et al.* 1993). Seasonal accumulations of VSPs in other trees have been studied also in willow (Wetzel and Greenwood, 1991), peach (Gomez and Faurobert, 2002), and diverse conifers (Wetzel and Greenwood 1989). The accumulation of these proteins seems to be located to protein storage vacuoles known as protein bodies which are generally located in the phloem parenchyma, xylem ray cells and cambial region (Clausen and Apel 1991; Sauter *et al.* 1989; Stepien *et al.* 1994), and therefore, most of these storage compounds have to be directed to the vacuole upon synthesis.

As temperatures decline, cells have to make morphological adaptations related to cold, and drought. Changes in the organelles, plasma membrane and cell wall are therefore another major adjustment that is necessary to cope with this the stresses of winter. Inside the cell, vacuoles and the endoplasmic reticulum are restructured (Sauter and Cleve 1994) to cope with the demand of vesicles for storage of proteins and lipids. At the same time, the plasma membrane has to maintain its fluidity and be protected against the effect of lower temperatures, and the cell wall has to acquire plasticity so the cell can shrink to maintain turgor pressure upon dehydration. In poplar cambial meristems, enzymes related to these processes are upregulated (Park *et al.* 2008) and along with the sugars constitute a protective role for the cells.

At the same time, adjustments in carbon metabolism are required. Since the amount of light is reduced, so is the fixation of carbon and the production of energy, and

in consequence the plant prepares carbon reserves (like starch) that can be later used to produce energy for other processes. Starch breakdown and the activation of glycolytic enzymes in a later stage may indicate the need for these sources of energy once photosynthesis becomes impaired during dormancy (Druart *et al.* 2007). Before a major repression of photosynthesis happens, one of the first responses to cooler temperatures is an over-exitation of the photosynthetic apparatus; consequently a series of adjustments have to be made in the electron sinks like the Calvin cycle (reviewed in Hüner *et al.* 1998); plants do this by increasing the transcript levels of enzymes from the Calvin cycle to fix more carbon, which at the same time results in the possibility of increased sucrose synthesis. Sucrose synthesis is then upregulated during cold acclimation but there has to be a delicate balance between its synthesis and the carbon fixation performed by the Calvin cycle (Stitt and Hurry, 2002). The main product of the Calvin cycle, glyceraldehyde 3-phosphate (G3P), is used to regenerate ribulose 1,5-biphosphate but also can be converted to sucrose. So although sucrose synthesis is needed, too much synthesis can deplete G3P intermediates for the Calvin cycle.

Sugars like sucrose play important roles in maintaining the osmotic balance and protecting the cells against the effects of cold temperatures. Carbohydrates in general are considered to be a good source of cryoprotectants and allow avoidance of drought. During winter, the concentration of soluble carbohydrates in shoots of European deciduous oaks (*Quercus robus* and *Quercus pubescens*) increases as a way to keep the cell's osmotic balance (Morin *et al.* 2007). Additionally, as described above (section 1.2), high production of sugars can change the structure of the cytoplasm into a gel-like glass structure that stabilizes the cell against further water loss and cold temperature (reviewed in Welling and Palva 2006).

Along with sugar synthesis, the cells synthesize a series of genes involved in responses to biotic and abiotic stresses. For example, dehydration in plant cells results from the increase accumulation of dry matter by storage compounds (proteins and sugars) and by the loss of water to formation of extracellular ice that comes with low temperatures (reviewed in Welling and Palva 2006). In response to this, plants may activate the production of sugars and antifreeze proteins. Antifreeze proteins inhibit increases in ice formation and can actually have additional roles in defense against

psychrophilic pathogens, since some of these proteins have been shown to be chitinases which can cleave chitin of insects and fungal cell walls (Griffith and Yaish 2004).

Low temperatures can also cause additional internal damage to cells. Low temperatures can produce protein denaturation and therefore proteins that allow correct protein folding and protein protection are also needed during this period. Chaperones are one class of proteins conferring cryoprotection (reviewed in Welling and Palva 2006). Additionally, photo-oxidative damage can also result from the effects of low temperature, and detoxification genes for the generated reactive oxygen have been shown to be activated too (Druart *et al.* 2007).

Schrader *et al.* (2004) observed a large increase in genes implicated in cell rescue and metabolism related to drought and cold stress in the cambial meristems of European aspen during the transition to dormancy. In addition to molecular changes associated with hormone action and photoperiod perception (discussed above), specific changes have also been detected in cell wall restructuring enzymes like glucanases, and stress responsive genes like dehydrins and chitinases. Genes like chitinases and glucanases have been shown to have dual roles (Griffith *et al.* 1995) since they posses antifreeze activity (cryoprotection) but also conserve their properties for chitin and sugar cleavage. Cryoprotectants along with oxidative stress response genes were also found to be an important component of aspen dormant cambial meristems in the study by Druart *et al.* (2007).

Dehydrins are by far the most studied proteins in relation to dormancy and concomitant cold acclimation process (reviewed in Welling and Palva, 2006). Increased dehydrin synthesis via drought and dehydration responsive transcription factors is a logical consequence of the cellular dehydration during autumn and winter, and these proteins function not only as cryoprotectants (antifreeze activity) but as osmoregulators (to control water loss through thorugh a decreased water potential). In downy birch (*Betula pubescens*) controlled experiments showed that dehydrin *BpuDhn1* was controlled by a combination of short day and low non-freezing temperatures while *BpuDhn2* was controlled mainly by low temperature (Welling *et al.* 2004). In peach, characterized dehydrins accumulate in bark in response to low temperatures (*PpDhn1*) or water deficit (*PpDhn2*) but a strong association with short days seems to be lacking in

PpDhn1. The promoters of these genes seem to have transcription factor binding sites for Drought Responsive Element Binding Factors (DREBs), and ABA response elements (Wisniewski *et al.* 2006), further demonstrating their mechanism of control by drought and cold signal transduction genes.

Some more specific changes also take place during the transition to dormancy. As previously mentioned, the molecular response is also largely mediated by light, temperature and hormones. In hybrid poplars, following these initial signals, ethylene metabolism seems critical before the cessation of internode activation, and the pathway may be active as a result of low sugar levels at the initiation of short days (Ruttink *et al.* 2007). In this study, there was also evidence of metabolic control via chromatin remodeling, by genes highly regulated just one week after short photoperiod treatment started; this remodeling control has also been inferred for the cambial meristems of aspen (Druart *et al.* 2007) and suggests additional regulation for downstream genes through mechanisms like histone deacetylation.

Finally, while most growth processes stop, meristematic identity has to be maintained. Genes are activated accordingly to maintain meristematic identity until dormancy actually starts, and cell cycle genes like cyclins and cyclin-dependent kinases have been shown to be downregulated (Ruttink *et al.* 2007). In the vascular cambium of hybrid aspen, the transcript downregulation of two PttCDKs (cyclin dependent kinases) did not occur until after 4 weeks of short days (Espinosa-Ruiz *et al.* 2004). However, there was a decline in kinase activity, showing a possible temporary arrest of the machinery, perhaps via post transcriptional control. Usually protein levels only decline completely when plants enter a full state of endodormancy, which reflects true dormancy. Although a logical change in response to growth cessation would be a general shutdown of core cell cycle genes, not all the machinery of the cell cycle was down-regulated in the cambium of European aspen (Druart *et al.* 2007; Schrader *et al.* 2004); so it was speculated that these genes could either be post-transcriptionally regulated or be used for rapid reactivation in spring.

Although photoperiodic control is the starting point for regulation of many processes, until recently, the downstream regulators were not known. A novel discovery was made when the *Populus* ortholog from flowering control in *Arabidopsis thaliana*

21

(*FT*), was shown to be related to control of growth cessation and bud set by using over-expressing transgenics and interference RNA (Bohlenius *et al.* 2006). This gene has now too been related to control of bud set in Norway spruce (Gyllenstrand *et al.* 2007), but the expression levels of the gene increase as opposed to the *Populus* homologous gene which decrease. *FT* in necessary for control of flowering and is regulated by the gene *CO* in response to long days. When the ortholog of this latter gene is shifted from long to short days, its peak of expression changes from the end of the day to the night which impairs *CO* from positively regulating *FT*. This demonstrated the dependence of expression due to photoperiod. Since phytochromes, *CO* and *FT* act in the same signal transduction network, and since *FT* expression in poplar seems to be negatively regulated when growth cessation and bud set take place, these circadian genes can be said to be involved in these processes. Other circadian genes have also been shown to be modulated in buds of poplar upon short day induction; however, a clear relationship between genes involved in flowering and bud set could not be established (Ruttink *et al.* 2007).

In conclusion, the response upon dormancy comprises major changes in specific metabolic routes that allow an overall adjustment to cope with the major changes of winter.

1.3 The present study

In boreal and temperate regions, active growth – i.e. the time between bud break and bud set, or the phase complementary to dormancy – largely determines the productivity of the tree (Howe *et al.* 2003). Seedling growth depends on the potential for growth of the tree itself and also on the environmental conditions, and at the same time the potential for growth depends on morphological and physiological characteristics (Grossnickle 2000). Without understanding the molecular basis of the physiological response of trees, silvicultural practices may not be as successful, and increases in yield will not be possible.

The present study aims to identify changes that occur at the protein level during dormancy acquisition in white spruce. Although there are published studies on the

changes in the transcriptome undergoing dormancy in the model tree poplar (Schrader *et al.* 2004; Druart *et al.* 2007; Ruttink *et al.* 2007; Park *et al.* 2008), there is little information on the protein changes associated with dormancy acquisition in forest trees, particularly in spruce. In poplar, changes in two-dimensional protein profiles upon short days were assessed for buds, leaves and bark (Jeknic and Chen, 1999); although no identification or characterization of the differentially-expressed proteins was performed in this study, increases in the number of proteins expressed after 10 weeks of short days in bark was evident. Similarly, descriptive studies followed the protein profiles of interior spruce through a time course using one-dimensional gels to detect major changes in polypeptide accumulation in buds, shoots, leaves and roots (Roberts *et al.* 1991) and in buds (Binnie *et al.* 1994). The main outcome of the studies showed the accumulation of proteins close to 30 kDa in weight, which were inferred to be vegetative storage proteins, but these proteins were not further characterized either.

Most deciduous trees mobilize their nitrogen from their leaves to the woody tissues before leaf abscission during fall; but spruce is an evergreen conifer which keeps its leaves for multiple years before abscission. Studies by Roberts *et al.* (1991) and Binnie *et al.* (1994) showed an actual accumulation of proteins which agrees with the seasonal accumulation of storage-like proteins in spruce. Ultrastructural studies have also shown accumulation of proteins in bark over winter for larch (*Larix* decidua), Scots pine (*Pinus sylvestris*) and eastern white pine (*Pinus strobus*) (Wetzel and Greenwood 1989). Furthermore, when deciduous and evergreen peach varieties were analyzed for seasonal changes of proteins and frost hardiness, it was found that both varieties accumulated seasonal proteins in their xylem and bark, although with a lower accumulation in tissues of the evergreen variety (Arora *et al.* 1992). Since it is apparent that even evergreens change their protein accumulation patterns, it is very possible that some of those changes result in accumulation of vegetative storage proteins, or could possibly be the result of a need to increase frost hardiness as was hypothesized by these studies.

If changes in protein profiles during dormancy acquisition are true for conifers, the molecular changes observed in other trees related to hormone activation or repression, stress responses, adaptation to cold and modifications of the carbon metabolism, should also be expected as a result to changes in photoperiod.

In this study we compared white spruce stems undergoing active growth to trees that have been placed in short days for 10 weeks and are still not-fully dormant. By examining changes in protein profiles between the two conditions we were able to identify major relevant changes in the process of dormancy acquisition. The changes in the abundance of individual proteins allowed us to get a look at important metabolic processes that need to be modified to overcome winters, and we could also identify specific relevant proteins (putative chitinases) that may have a central role in storage or hardening. These relevant proteins were further investigated to infer this and other possible functions that may lead the way into future research.

2.0 MATERIALS AND METHODS

2.1 Plant material and experimental design

White spruce (*Picea glauca* [Moench] Voss) seedlings in their second growth cycle were used for the two dimensional gel electrophoresis analyses. Dormant seedlings derived from independent families had completed one growth cycle and were obtained from Nathalie Isabel (Canadian Forest Service Laurentian Forestry Center, QC, Canada), and were shipped from Valcartier, QC to the University of Alberta. Seedlings were transplanted into 3.7 L pots while dormant, and then were grown in growth chambers under optimal conditions in long days (16 hours of full light and 20 °C constant temperature, and 50 to 60 % relative humidity) for approximately 8 weeks, i.e. the time required to nearly complete the active phase of a growth cycle. Plants were fertilized weekly with 1 g/L of 15-30-15 fertilizer (150 ppm of N) for the first two weeks to encourage root growth, then weekly with 0.5 g/L of 20-20-20 weekly (100 ppm of N) until the onset of the experiment. Shortly before seedlings were to begin bud formation – i.e. near the end of the active period of growth – the photoperiod was changed to short day conditions (8 hours of full light and 20 °C constant temperature) to induce bud formation. While under short day conditions, plants were fertilized weekly with 0.5 g/L of 8-20-30 (40 ppm of N); this lower N concentration was used to ensure rapid and synchronous bud formation. The day that the photoperiod was switched from long days to short days was designated Day 0. A complete randomized block design was used for each independent experiment; all plants were visually inspected on Day 0 to ensure that only healthy plants that showed no sign of bud formation were used in the experiments. Plants were sampled at 0 day, 1 day, 3 days, 7 days, 14 days, 28 days and 10 weeks after the switch from long day to short day conditions, but the experiment concentrates on the changes between 0 short days (0d SD, from now on), that acts as control, and 10 weeks of short days (10wk SD, from now on). In total, six independent experiments were conducted. Three of these independent experiments were used in this study (S008, S011, and S012) for all the two dimensional gel electrophoresis analyses.

For the one dimensional gel electrophoresis analyses, white spruce seedlings in their first growth cycle obtained from Coast to Coast Nurseries (Smoky Lake AB) were used (experiment S002). Plants were grown in basement growth chambers at U of A under 18 °C, 16 hour days, 50-60 % humidity. At Day 0, plants were switched to 18 °C, 8 hour days, 50-60 % relative humidity. Plants were then sampled at 0 day, 3 days, 7 days, 14 days, 25 days, 12 weeks after the switch to short days, then submitted to 3 months of cold temperature, and sampled again 15 and 23 days after switching them to long days.

For protein profiling, we used lignified whole stems from the previous year's growth (10 cms approximately). Each sample comprised a pool of stem material from three replicate trees grown in the same experiment (Figure 2.1). Comparison was performed between samples for the control and treatment (0d SD and 10wk SD) in three independent biological replicates (experiments), and performing 3 pseudoreplicates (technical replicates) at the gel level (explained below). Collection of the tissue was performed by removing a segment of the stem from above the hypocotyl to just below the current year's growth, cutting the branches off and dividing the stem in pieces of approximately 5 cm long. The pieces were then plunged into a magenta box containing liquid nitrogen using forceps, and needles removed using forceps. The stems were then transferred to a labeled 50 mL conical tube containing liquid nitrogen; once the stems from all three trees had been harvested into the same tube, the liquid nitrogen was poured off and the tube transferred immediately to a dewar of liquid nitrogen before transfer to a -80 °C freezer.

0d SD   **10wk SD**

Exp S008

Exp S011   Each sample was
                                                            made of three pooled
                                                            stem sections with
Exp S012   secondary growth

Exp S008

Exp S011   One extraction per
                                                            sample was used to run
                                                            three technical replicate
Exp S012   gels in each
                                                            treatment x experiment
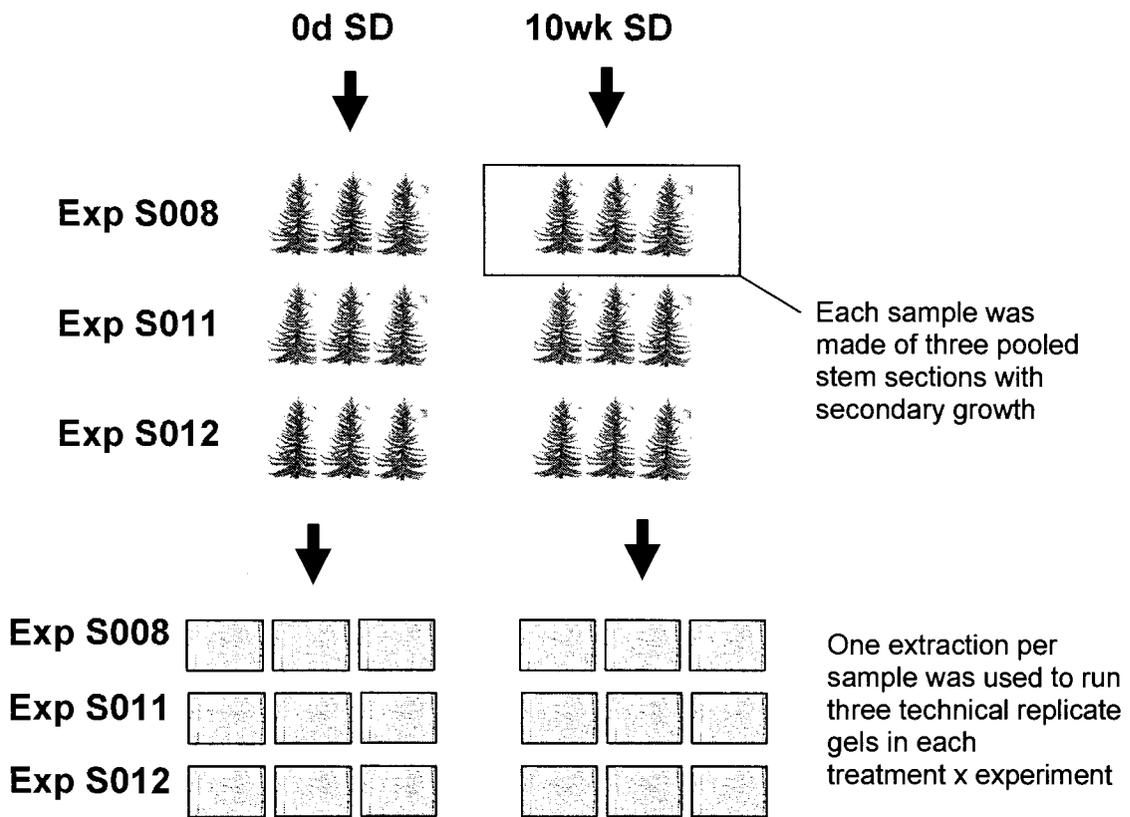
Figure 2.1 Experimental design. Three stem sections with secondary growth were pooled in each one of two experimental conditions: 0d SD or 10wk SD. The experiment was replicated three times to generate biological replicates (Exp S008, S011, S012). Each sample was used for a protein extraction, and each protein extract was run in three gels (technical replicates).

27

2.2 Protein extraction, precipitation and quantitation

Stems representing the previous year's growth were ground in a mortar to a fine powder using liquid nitrogen. Eighty milligrams of ground tissue were transferred into liquid nitrogen pre-cooled 1.5 mL microfuge tubes, and placed back in liquid nitrogen. To each tube 500 µL of extraction buffer were added, containing 100 mM of (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.4, 5 mM EDTA, 10 mM DTT, 10 % glycerol, 7.5 % PVPP, 0.3 % DIECA and a protease inhibitor cocktail (Bioshop Canada, Burlington, ON, Canada) containing 1 mM 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF), 0.3 µM aprotinin, 10 µM bestatin, 10 µM E64 and 100 µM leupeptin. Mixture was vortexed to homogenize and placed on ice. Metal beads were added to each tube and the tissue was further homogenized using a Mixer Mill MM 301 (Retsch, Haan, Germany) at 30Hz for 2 minutes. Tubes were placed back on ice and each sample was sonicated four times for 10 seconds with 2 seconds intervals at 40 % amplitude in an Ultrasonic Dismembrator 500 (Fisher Scientific, Pittsburgh, PA, USA). Homogenates were centrifuged in a microcentrifuge at 19100 g at 4°C for 30 minutes and the supernatant containing the protein was transferred to a new tube. Centifugation was repeated one or two times as required to remove particulate. To further purify protein extracts, four parts ice cold acetone were added per part supernatant (e.g. 400 µL of acetone per 100 µL of protein sample). The mixture was vortexed for 5 seconds and proteins precipitated at -20 °C for 60 minutes. Tubes were centrifuged at 19100 g at 4 °C for 15 minutes and the supernatant was discarded without disturbing the pellet. Pellets were air dried for approximately 15 minutes and resuspended using extraction buffer without DIECA or PVPP and with a DTT concentration of 0.1 mM. Protein extracts were stored at -20 °C until quantification.

Protein samples were diluted 1:20 and protein concentration quantified using the BCA™ protein assay kit (Pierce, Rockford, IL, USA) with the following minor modifications: The standard curve was prepared using extraction buffer without DIECA or PVPP and with a DTT concentration of 0.1 mM. Fifty microliters of standard or sample were mixed with 1 mL of BCA working reagent keeping a 1:20 ratio.

2.3 Proteomics analyses

2.3.1 One dimensional gel electrophoresis

Initial qualitative analysis of protein changes was performed using one dimensional gels with only one biological replicate comprising pooled material of the experiment S002 over a time course (as described in the plant material and experimental design section).

A BIO-RAD Mini-PROTEAN® II Cell casting mini-gel system was used for one dimensional gel electrophoresis, with a 12 % separating gel (0.375 M Tris-HCl pH 8.8 buffer, 0.1 % SDS, 12 % acrylamide solution, 0.05 % (w/v) ammonium persulfate and 0.05 % TEMED), and 3.9 % stacking gel (0.125 M Tris-HCl pH 6.8 buffer, 0.1 % SDS, 3.9 % acrylamide solution, 0.05 % (w/v) ammonium persulfate and 0.1 % TEMED). Fifteen micrograms of protein were diluted to 7.5 µL with extraction buffer without DIECA or PVPP and with a DTT concentration of 0.1 mM, and 7.5 µL of sample buffer added (62.5 mM Tris-HCl pH 6.8, 10 % glycerol, 2 % (w/v) SDS, 5 % β-mercaptoethanol, and 0.5 % (w/v) bromophenol blue). Samples were vortexed and centrifuged for a few seconds, and denatured for 5 minutes at 95 °C. The entire sample was added to the preformed wells and electrophoresed at 200 V. Molecular mass markers were included to facilitate molecular mass estimations (PageRuler™ prestained protein ladder, Fermentas Canada, Burlington, ON, Canada).

Gels were stained in a solution of 0.1 % (w/v) Coomassie Brilliant blue R-250, 17 % (v/v) acetic acid, and 42 % (v/v) methanol, and destained with a 30 % (v/v) methanol and 10 % (v/v) acetic acid solution. Gels were stored in 5 % (v/v) methanol / 7 % (v/v) acetic acid.

To dry the gels they were rinsed with MilliQ water and placed on top of Whatman paper and covered tightly with Saran Wrap. The gel mount was taken to the Gel Dryer Model 543 (BIO-RAD, Hercules, CA, USA), and let to dry for 3 hours using a gradient cycle to reach 80 °C.

2.3.2 Two-dimensional IEF / SDS-PAGE

For differential expression analyses in two dimensions three biological replicates (each one from pooled material as described above) with the codes S008, S011 and S012 were used for collection of tissue at 0d SD and 10wk SD To increase statistical significance and reduce variability we run three technical replicate (pseudoreplicates) two dimensional gels for each one of the biological replicate per treatment protein extractions giving a total of 18 gels.

One hundred micrograms of total protein was diluted to 50 µL using extraction buffer without DIECA or PVPP and with a DTT concentration of 0.1 mM, and mixed with 2 % IPG buffer (1 µL). This solution was added to 410 µL urea rehydration buffer (8 M urea, 2 % (w/v) CHAPS, 1 % IPG buffer 1 % (w/v) bromophenol blue, and 0.28 % (w/v) DTT). The mixture was vortexed briefly, then centrifuged and placed on ice, before being added to the Immobiline DryStrip reswelling tray (GE Healthcare, Chalfont St. Giles, UK). Immobiline DryStrip pH 3-10 NL 24 cm IPG strips (GE Healthcare, Chalfont St. Giles, UK) were placed on top of the solution for overnight passive rehydration, and covered with 3 mL oil to avoid dehydration.

Rehydrated strips were isoelectrofocused in an Ettan™ IPGphor II™ Isoelectric focusing unit (GE Healthcare, Chalfont St. Giles, UK) at 500 V for 1 hour, followed by 1000 V for 1 hour, then 8000 V until reaching 60000 VHs. Paper wicks where used to absorb excess water and strips were fully covered with oil to avoid dehydration and clogging. After electrofocussing, strips were placed in strip holders at -80 °C until used in the second dimension SDS-PAGE.

Proteins were separated in the second dimension using at 12 % separating gel (0.375 M Tris-HCl, pH 8.8 buffer, 0.1 % SDS, 12 % acrylamide solution, 0.05 % (w/v) ammonium persulfate and 0.05 % TEMED). Bottom plates were treated with bind silane solution (76 % ethanol, 2 % (v/v) acetic acid, 0.1 % (v/v) bind silane) before casting the gel to facilitate gel adhesion.

Prior to SDS-PAGE, strips were pre-equilibrated first in urea buffer with DDT (6M urea, 75 mM Tris-HCl pH 8.8, 29.3 % glycerol (v/v), 2 % SDS (w/v), 0.002 % bromophenol blue, and 0.5 % (w/v) DDT) for 15 minutes, then in urea buffer with iodoacetamide (6 M urea, 75 mM Tris-HCl pH 8.8, 29.3 % glycerol (v/v), 2 % SDS (w/v) 0.002 % bromophenol blue, and 4.5 % (w/v) iodoacetamide) for 15 minutes prior to electrophoresis. Equilibrated strips were laid on the surface of the acrylamide gel, and sealed with 1 % (w/v) low melting point agarose in 1x running buffer (25 mM Tris base, 192 mM glycine, 0.1 % SDS (w/v) and 0.002 % bromophenol blue). A molecular weight marker was also included (Unstained Protein Molecular Weight Marker, Fermentas, Burlington, ON, Canada).

Gels were placed on the Ettan™ DALT*six* long vertical electrophoresis system (GE Healthcare, Chalfont St. Giles, UK), and the chamber was filled with 1x running buffer (described above). Proteins were separated in the second dimension at 400 mA, 500 V and 10 w per gel. Gels were stained with Deep Purple™ Total Protein Stain (GE-Healthcare, Chalfont St. Giles, UK) using the manufacturer specifications. Gels were imaged using a Fujifilm Image Reader FLA-5000 with a LPG filter at a 532 nm wavelength at 600 V. Images were adjusted to reduce background using standardized conditions for every gel using a combination of L-Process (Science Lab 2003, V2.2 from Fujifilm) and Adobe® photoshop.

2.3.3 Differential expression analyses

Gels were imported into Image Master™ 2D Platinum (GE-Healthcare, Chalfont St. Giles, UK) and spots detected using a smoothness of 10, minimum area of 100 and a saliency of 2000. Spots across gels were matched using 10 landmarks per gel, corresponding to well defined proteins present across all 18 gels. A synthetic (artificial composite) gel containing all matched and unmatched spots was created to compare not only differentially expressed spots but also the presence/absence spots. Four hundred and ninety groups of corresponding spots across all the gels were confirmed by eye after automated matching to correct for missing or mis-matched spots. Volume reports were

generated from each spot group (same spot across all gels), were the volume corresponds to the pixel volume of the upper 75 % of the 3-dimensional image of the spot.

To normalize the volume values for spots across gels for subsequent quantitative analysis, the volume of each spot was divided by the standard deviation of all spot volume values in each corresponding gel. Reproducibility within biological replicates and between technical replicates was assessed by calculating the Pearson correlation coefficient (r^2).

To detect proteins that showed significantly different abundance between 0d SD and 10wk SD treatments, spot values were averaged among pseudoreplicates (technical replicates) and the values of the three biological replicates from 0d SD were compared to the values of the three biological replicates of 10wk SD using Significance Analysis of Microarrays (SAM, Tusher et al. 2001). For the SAM analysis of the 490 groups of spots, biological replicates were designated as blocks and a two class unpaired, unlogged, t-statistic with 100 permutations was performed to find differentially expressed proteins. In the resulting plot the False Discovery Rate (FDR) was adjusted to 5.2 %, and a 1.5 fold change cutoff was applied. A more permissive analysis was also performed with a 12.5 % FDR and no fold change.

2.3.4 Spot picking, digestion and mass spectrometry

Proteins were excised from gels using an Ettan^{TM} automated spot picker (GE healthcare, Chalfont St. Giles, UK). Protein-containing gel plugs were dispensed into 96 well plates containing 150 µL of MilliQ water. The samples were stored at -80 °C until digested.

To digest spots, the water in the wells was replaced with 50 µL of 100 % acetonitrile for dehydration. After 10 minutes, the samples were centrifuged briefly and the liquid was removed. Fifty microliters of 0.1 M NH_4HCO_3 and 50 µL of 100 % acetonitrile were added to each well, and plates vortexed for 10 minutes to de-stain the spots. Plates were then centrifuged and the liquid removed. A second dehydration cycle was performed, then samples dried in a SPD SpeedVac® (Thermo-Savant) concentrator

at room temperature for 10 minutes to remove residual acetonitrile. Proteins were reduced by incubating gel plugs in 30 µL of 10 mM DTT/0.1 M NH_4HCO_3 at 56 °C for 30 minutes, then plates were centrifuged briefly and liquid removed. Fifty microlitres of 100 % acetonitrile was added to each well, and plates incubated 5 minutes before centrifuging briefly and removing the liquid. Proteins were alkylated by incubating gel plugs in 30 µL of 55 mM iodoacetamide/0.1 M NH_4HCO_3 for 20 minutes in the dark. The liquid was removed and 150 µL of 0.1 M NH_4HCO_3 were added and the samples vortexed gently for 15 seconds every 3 minutes over 15 minutes. After centrifuging the samples and removing the liquid, two consecutive rounds of dehydration in NH_4HCO_3 / acetonitrile, including the speed vacuum step, were performed as described above. To digest the proteins 1 µg/µL of reconstituted Trypsin gold (mass spectrometry grade, Promega, Madison, WI, USA), was mixed with 40 mM NH_4HCO_3/10 % acetonitrile, to a concentration of trypsin of 0.02 µg/µL. Twenty microliters of this mixture were added to the wells with the gel plugs and the samples were incubated for 1 hour at room temperature. Twenty microlitres of the same trypsin solution was then added to each well, and the samples incubated at 37 °C for 16 hours. To stop the digestion, 15 µL of 0.4 % formic acid was added to each sample; plates were covered and incubated for 30 minutes at room temperature. The resulting liquid (but not the gel plug) from each well was transferred to a mass spectrometry plate (pre-cooled on ice). Twenty five microliters of 50 % acetonitrile/0.1 % formic acid was added to the plate with the gel pugs and the plate was covered again and after 30 minutes the liquid was transferred to the respective well in the mass spectrometry plate. The speed vacuum was used to concentrate the digested protein sample to approximately 45 µL. The samples were placed at -80 °C until the mass spectrometry run was performed.

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) was done using the Agilent 1100 series LC/MSD trap (Agilent Technologies, Santa Clara, CA, USA), using an AutoMS of 2, 622-922-1522 peak exclusion, a number of three precursor ions and a threshold of absorbance of 10000. Reliability of results was assessed by running BSA samples and confirming high score identities using MASCOT (Perkins *et al.* 1999). Each run lasted approximately 75 minutes and compounds were obtained by doing an

automatic search of compounds followed by compound deconvolution over the whole range of peptide peaks obtained during 75 minutes of run.

2.4 Annotation analyses

2.4.1 Peptide annotation and functional classification

To annotate peptide profiles obtained from LC-MS/MS, output sequences were compared one by one to the NCBI database using the Mascot algorithm (Perkins et al., 1999) with the following parameters: virdiplantae as taxonomical category, trypsin as digestion enzyme, up to two missed cleavages, carbamidomethyl as fixed modification, peptide tolerance of 2 Da, MS/MS tolerance of 1 Da and an ESI trap as instrument. The rest of the parameters were not modified. Since most of the sequence information for spruce species in NCBI is as ESTs, after the first search, all the peptides were re-submitted changing the database to EST_viridiplantae to increase the likelihood of finding a significant match. When the significant match was an EST that did not have informative annotation, the translated EST sequence was used for BLASTp (Altschul *et al.* 1990) and Interpro analysis (http://www.ebi.ac.uk/interpro/). Hits were considered appropriate annotation if the similarity between the EST and its resulting hit corresponded to the correct ORF of the original peptides.

Peptides that gave meaningful annotation were also classified into functional categories. The top blast hit obtained from the mascot hit for each profile was used to generate a list of Genbank accession identifiers which could be searched against Genbank. The fasta file generated was then loaded into Blast2GO (Conesa *et al.* 2005, Conesa and Götz 2007; Götz *et al.* 2008), allowing automatic blasting, mapping and annotation of each sequence to Gene Ontology (GO) categories (The Gene Ontology Consortium 2000). At the end of the standard annotation process, and increase in GO term annotation was produced by using two additional tools from Blast2GO. Automatic annotation of each sequence with the Interpro tool, which allows to add domain-related GO annotations; and automatic inference of biological function using Annex, which uses

an algorithm that allows to infer biological process and cellular compartment from molecular function (Conesa and Götz 2007; Götz *et al.* 2008). Categorization analyses were performed using level 2 of biological processes to give a broad classification of general processes. This higher order categorization was complemented by also obtaining GO biological process annotations at a more specific level (level 3) along with details of inferior levels (biological process, molecular function and cellular compartment), as well as Interpro annotation of domains (Appendix 6.3). All of these analyses were run under the program default conditions.

Using the same interface, map graphs and frequency tables were generated for the different levels of biological process determined by Gene Ontology. Blast2GO was also used to do an automatic mapping of the hit enzymes to metabolic pathways according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) classification (Ogata *et al.* 1999; Kanehisa *et al.* 2008). To see if outstanding categories were significant, the functional categories obtained were compared to a random sample of 5000 contigs from the spruce gene catalog (GCAT, https://genome.ulaval.ca/arborea/gcat/login) run under the same conditions as our sample in Blast2GO. The frequencies of functional categories between our sample and the random sample of 5000 sequences from GCAT were compared using chi-square analysis through a contingency table of 2*17 (samples*functional categories) using excel. To find out which functional categories were contributing to significant differences, I performed heterogeneity tests (Zar 1999), by generating 2*2 contingency tables to compare each one of the functional categories to the pooled data of all the other categories

2.4.2 Chitinase phylogenetic analysis

Since several chitinases were identified amongst the differentially expressed proteins, I conducted further phylogenetic analysis of spruce chitinases, including the chitinases identified in this study, to gain insight into relationships between these spruce chitinases and characterized chitinases from other species. These relationships, in turn, were hoped to reveal clues about possible function of the chitinases.

The differentially expressed chitinases from this study plus chitinases from a previously published phylogeny on plant chitinases that includes rice, Arabidopsis and other plant chitinases (Xu *et al.* 2007) were used to search chitinase contigs in the white spruce ESTs database (GCAT; https://genome.ulaval.ca/arborea/gcat/login) using tBLASTn. The redundant hits were filtered and the non-redundant sequences were used for BLASTx queries to filter out non-chitinase hits. ESTs representing putative white spruce chitinases were therefore identified from the database. The white spruce ESTs are also currently deposited in the Genbank database at the NCBI (http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucest).

Opening reading frames were deduced using a combination of ORF finder from VectorNTI® (Invitrogen, Carlsbad, CA, USA), Genescan (http://genes.mit.edu/GENSCAN.html), global alignments to known chitinases sequences using AlignX from VectorNTI®, and BLASTx. Because distinct EST contigs may represent different alleles of the same locus, contigs sharing greater than 95 % nucleotide similarity were omitted from the analysis.

To confirm sequences of the spruce chitinases the respective clones from the GCAT libraries were grown in sterile LB-Agar with ampicillin (50 µg/mL) over night at 37 °C. Individual colonies were picked and grown in 5 mL of sterile LB with ampicillin (50 µg/mL) over night at 37 °C with continuous shaking. Grown bacteria were precipitated at 6800 g for three minutes and the pellets were used for plasmid extraction using the QIAprep Spin Miniprep Kit (QIAGEN, Venlo, Netherlands). Plasmids were quantified using a nanodrop ND-1000 spectrophotometer (Nanodrop technologies Inc., Wilmington DE, USA).

Sequencing was performed using the BigDye terminator v3.1 cycle sequencing kit (Applied Biosystems, Foster city, CA, USA). Each sequencing reaction included 2 µL of BigDye premix, 3 µL of 5X sequencing buffer, 50 to 400 ng of plasmid template (1 µL), 1µL of 5 µM primer, and baxter water to a final volume of 20 µL. The extension reaction protocol had 25 cycles at: 96 °C for 30 seconds, 50 °C for 15 seconds and 60 °C for 1 minute. After the cycles were finished the reaction was cooled to 4 °C. The reaction was precipitated by adding 2 µL of 1.5 M NaOAc / 250 mM EDTA to a 1.5 mL microfuge tube and then adding the sequencing reaction for resuspension. Then 80 µL of EtOH (100

%) were added, and the mixture was vortexed and left on ice for 15 minutes. The samples were then centrifuged for 15 minutes at 4 °C (19100 g) and the supernatant was discarded. To the pellet, 500 µL of 70 % EtOH were added and the mixture was vortexed briefly. Samples were centrifuged again for 5 minutes (19100 g) and the supernatant was discarded again. Pellets were dried in a Savant Speed vac concentrator SVC-100H, (Savant Instruments, Inc. Farmingdale, NY, USA) for 10 minutes at room temperature. Samples were sequenced in a 3730 DNA analyzer (Applied Biosystems, Foster city, CA, USA).

Deduced amino acid sequences were obtained using a direct translation of the nucleotide ORFs using the translation tool in VectorNTI® (Invitrogen, Carlsbad, CA, USA). Phylogenies were built for the spruce chitinase-like sequences along with previously published chitinases (Xu et al. 2007) and selected functionally characterized chitinases from other plant species using MEGA4.0 (Tamura et al. 2007). The sequences were aligned using two different alignment algorithms – CLUSTAL and MUSCLE - with their default parameters. The CLUSTAL program used (Thompson et al. 1994) is integrated in the MEGA platform, while the MUSCLE analysis (Edgar 2004) was carried out using the EBI server (http://www.ebi.ac.uk/Tools/muscle/index.html). To build the phylogenies, four different phylogenetic algorithms were used (neighbor joining [NJ], minimum evolution [ME], maximum parsimony [MP] and unweighted pair group method with arithmetic mean [UPGMA]). For neighbor joining, the parameters were as follows: amino acids for data type, a phylogeny reconstruction analysis, NJ method, 1000 bootstrap replicates, pairwise deletion for missing data and gaps, Poisson correction for the model, include all substitutions, homogeneous pattern among lineages and uniform rates among sites. For minimum evolution the parameters were the same adding the search option of CNI (level = 1) with initial tree = NJ Max Trees = 1. For maximum parsimony the parameters were the same adding the search option of CNI (level = 1) with initial tree with random addition (10 reps), and using all sites for gaps and missing data. For UPGMA the parameters were the same and a pairwise deletion was chosen for gaps and missing data. The different algorithms at the alignment and phylogenetic levels were used to test for consistency of the clusters.

2.4.3 Chitinase *in silico* analysis

General properties of the three putative chitinase proteins identified by two dimensional gel electrophoresis were obtained using VectorNTI® (Invitrogen, Carlsbad, CA, USA), e.g. MW, pI and sequence length. Several strategies were used to identify functional domains of the three chitinases. SignalP (http://www.cbs.dtu.dk/services/SignalP/) was used to find putative signal peptides; PROSITE (http://www.expasy.ch/prosite/) was used to find chitin binding signals and putative asparagine glycosylation sites, YingOYang (http://www.cbs.dtu.dk/services/YinOYang/) was used to find n-acetylglucosamine glycosylation sites, Interpro (http://www.ebi.ac.uk/interpro/) was used to find the catalytic domains, and WoLF PSORT (http://wolfpsort.org/) was used to search for putative localization signals. Additionally, the chitinase sequences were aligned with AlignX (VectorNTI®, Invitrogen, Carlsbad, CA, USA) to a tobacco Class I chitinase (E.C. 3.2.1.14; NCBI protein accession number CAA34812.1), which was previously characterized as having a carboxy-terminal vacuolar signal (Neuhaus *et al.* 1991). Catalytic and substrate binding residues were found by using literature sources.

I also used the white spruce chitinase-like protein linear sequences along with two rye (*Secale cereale*) chitinase antifreeze proteins (GenBank accession numbers: AAG53609, AAG53610) and two chitinase and chitinase-like vegetative storage proteins from alfalfa (*Medicago sativa*) and banana (*Musa* sp.) respectively (GenBank accession numbers: AAN10048, AAS12600), to build 3-dimensional models by using homology modeling provided by Swiss model (http://swissmodel.expasy.org/repository/). Each one of the sequences was blasted against the GenBank NR database using BLASTp. For each analysis, nine sequences with different levels of identity (ranging from 60 % to 90 %), including one sequence with a predefined structure from the Protein Data Bank (PDB) (http://www.rcsb.org/pdb/home/home.do), were aligned with the query sequence using the T-Coffee (http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi) algorithm (Notredame *et al.* 2000). The alignments were used as input for protein homology modeling using the Swiss Model program (Arnold *et al.* 2006). For calculation of the 3D

structures only the core region corresponding to the catalytic region to the carboxylic end was taken. The upstream region from the catalytic domain that included signal peptide and chitin binding domain in some of the proteins was trimmed before performing the alignment in T-Coffee. As structural template for modeling each protein we used the top BLASTp hit having a structure, and for each case this structure was different with the exception of the spruce chitinases: Pg_GQ03602.B7_D17:5-:Contig3:1-897:+ and Pg_GQ03602.B7_D17:53:Contig1:1-1101:+. These parameters allow for better comparisons since core regions can be modeled more accurately because of their conservation level, and not all proteins are done on the same template, which would bias the similarities.

Graphics for the models were obtained using Raswin-Rasmol (http://rasmol.org, Sayle and Milner-White, 1995), and structural comparison was performed by calculation of the Root Mean Square Distance, (http://en.wikipedia.org/wiki/Root_mean_square_deviation_(bioinformatics)) using the pairwise structure comparison tool (SSAP) from the CATH database for protein structure classification (http://www.cathdb.info/). The output file (.pdb) was used as input for AFPredictor (Doxey *et al.* 2006), a Perl script which predicts ice binding surfaces. Output surface structures with Ordered Surface Carbons (OSCs) were visualized using 3D-Viewer from VectorNTI® (Invitrogen, Carlsbad, CA, USA). Surfaces were predicted using the algorithm proposed by Vashney *et al.* (1994), based on the surface created by a spherical probe as it rolls around the atoms of the modeled structure.

3.0 RESULTS

3.1 Protein changes in one-dimensional gels

To determine if visible changes in protein abundance could be detected during dormancy acquisition, woody stems of one year old seedlings subjected to a time course of short days followed by long days were evaluated under one-dimensional gel electrophoresis.

The time course showed several qualitative changes in the relative abundance of proteins of different sizes that appear or disappear through the time course (arrows in Figure 3.1). The presumed large subunit of the most abundant plant protein, Rubisco, appeared at the expected 55 kDa as a thick band. Several bands appeared in the protein profiles over the course of the short day treatment, and then disappeared or lowered their content after changing to long days, when trees were returned to optimal growth conditions following 3 months of cold treatment which helps to fulfill the chilling requirement to release dormancy. This included a band at approximately 30 kDa (arrow and an asterisk in Figure 3.1), which appeared after 25 SD and had a peak at 12wk of SD. These proteins represented potential candidate vegetative storage proteins, since it is predicted that these proteins will accumulate over the course of dormancy acquisition, and be broken down to their component amino acids during the initial phase of the subsequent growing season. The one dimensional gel electrophoresis analyses thus supported the idea that a proteomics approach may reveal candidate vegetative storage proteins.

Based on the protein profiles of each of these time points, it was decided to use samples representing 0 SD and 10wk SD treatments for two dimensional gel electrophoresis analysis.
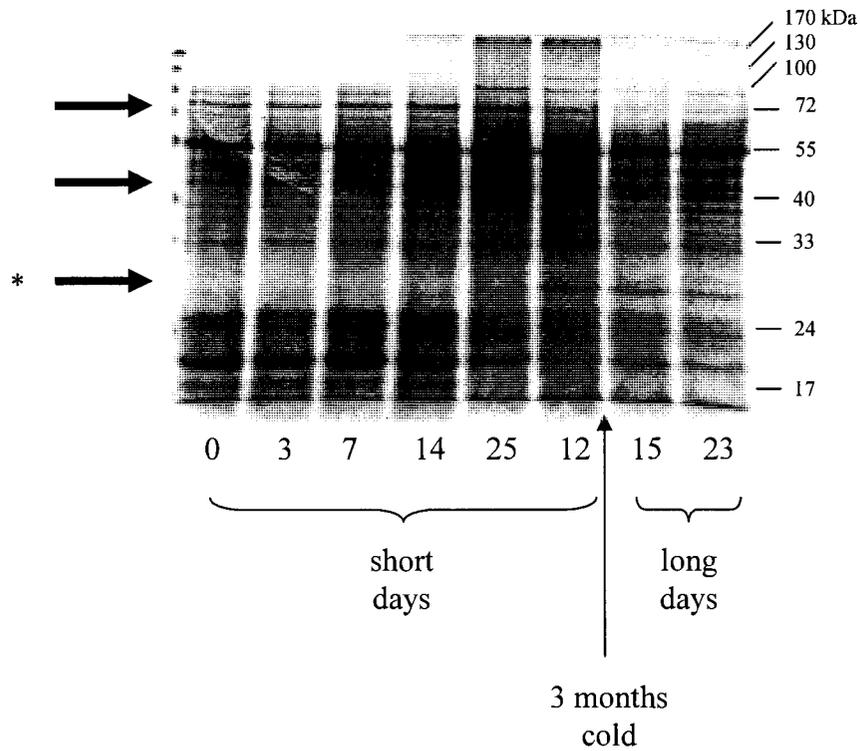
Figure 3.1 Coomasie blue stained SDS-PAGE stem protein changes in a time course of plants under short days (8 hours light, 16 hours dark) and then transferred to long days (16 hours light, 8 hours dark); lane 1 – after 0d SD, lane 2 – after 3d SD, lane 3 – after 7d SD, lane 4 – after 14d SD, lane 5 – after 25d SD, lane 6 – after 12wk SD, lane 7 – 15 days after switching to long days, lane 8 – 23 days after switching to long days. And arrow below indicates the application of a cold treatment for three months before putting the plants back into long days. Arrows on the left depict the appearance and disappearance of bands throughout the time course. An asterisk indicates a band appearing after 25d SD that has a peak after 12wk of SD. Numbers on the right indicate masses of the molecular weight standards that were loaded with the gel but that are not shown.

3.2 Differential protein expression in two-dimensional gels

After testing three techniques of staining (Coomassie Blue, Deep Purple and Silver staining), Deep Purple and Silver staining did not show large differences in the number of spots, while the sensitivity of Coomassie Blue was really low. However, since there was an apparent interference of the silver stained proteins with mass-spectrometry possibly because of the use of formaldehyde (Richert *et al.* 2004), we decided to stain all further gels using Deep Purple. Although sensitivity for both techniques may be in the same range (approximately 10 ng), deep purple also presented less background (result not shown) and its residual components have no interference with mass spectrometry. This allows using Deep Purple for both analytical and preparative gels.

Eighteen gels corresponding to two treatments (0d SD and 10wk SD), three biological replicates per treatment (pooled woody stems from three independently executed experiments) and three technical replicates (equivalent to pseudoreplicates) per biological replicate per treatment, were run to assess differential protein expression.

To try to achieve the maximum reproducibility in the protein gels, we used immobilized pH gradient strips for the isoelectrofocusing step, which ensures correct charge distribution; six gels were run at the same time (three technical replicates for two treatments) in a vertical electrophoresis system guaranteeing homogeneous temperature conditions; and the six gels were stained at the same time with Deep Purple™ (2-D Electrophoresis principles and methods handbook, GE Healthcare, 2004).

To test reproducibility among gels we used the Pearson coefficient of correlation. Gels were named according to the experiment number, treatment, and technical replicate. For example the technical replicate gel one, from the biological replicate (experiment S008), corresponding to 0d SD of treatment was named S8-0SD-1. The first part refers to the biological replicate, the second part to the treatment for short days, and the third part to the technical replicate (Table 3.1).

Reproducibility among the spots between gels within treatments was generally high. As would be expected, higher correlation coefficients (r^2) were found among technical replicates (Table 3.1) than among biological replicates. The correlation coefficients among technical replicates ranged from 1 to 0.630, whereas the correlation

coefficients among biological replicates ranged from 0.681 to 0.384. Differences in pseudoreplicates were generally due to staining, while differences in biological replicates of the same treatment could be due to factors like replication of the experiments in different dates and differences in protein expression inherent to each tree. As expected because of differences in expression, correlation coefficients were the lowest when comparing between samples of the two treatments; the average correlation in this case was 0.304.

Automated matching of protein spots over the 18 gels was performed using landmarks (common and well defined proteins present in all the gels), and then confirmed manually by checking each group of spots to correct for mismatches. We found a total of 490 groups of spots across the 18 gels which were quantitatively analyzed using the normalized absolute volume value for each spot. Figure 3.2 shows representative gels of the 2 conditions, while the same pair of gels outlining the differentially expressed proteins that could be annotated are shown in Appendix 7.1.

Statistically significant differences in spot intensities between 0d SD and 10wk SD samples was determined using the Significance Analysis of Microarrays package in Excel (Tusher et al, 2001; http://www-stat.stanford.edu/~tibs/SAM/). The values of the matched spots were averaged over the technical replicates for each of the three biological replicates per treatment to account for technical gel to gel variability. The three values of true biological replicates corresponding to 0d SD were then compared to three values of 10wk SD samples for each of the 490 spots, i.e. n = 3 for these analyses. Because each of the three independent experiments was executed in a different growth chamber, each biological replicate of each condition was designated as a block to account for this systematic source of variation. In cases where a spot was not detected (i.e. was below detectable limits) for a given biological replicate, the value was designated as 0.000001, since it could not be assumed that these spots were entirely absent.

Using SAM, a total of 216 statistically significant differentially expressed proteins were identified, using a false discovery rate (FDR) of 5.2 % and a fold change cut-off filter of 1.5 (Figure 3.3). Under these parameters a median number of 11.23 false positives could be found. Of these 216 significantly differentially expressed proteins, 203 were upregulated after 10wk SD, while only 13 showed higher expression on 0d SD. A

second more permissive analysis was also carried out using an FDR of 12.5 % and no fold change cut-off filter, which revealed 245 and 65 proteins upregulated after 10wk SD and 0d SD respectively. Since this second analysis in more permissive the median number false positives increased to 38.96.

3.3 Mass spectrometry analysis

A total of 161 spots were analyzed by LC-MS/MS; most of these analyzed peptides showed significant differential expression at an FDR of 5.2 %, while a small number of peptides were included that showed significant differential expression at an FDR of 12.5 %. LC-MS/MS on trypsin digested samples showed a good intensity signal (10^6 or 10^7) which differs from the background signal (10^4) for the peaks corresponding to the peptide profiles (an example is given in Figure 3.4). Peptide profiles were annotated using MASCOT (Perkins et al., 1999) to query first against virtual profiles from NCBI protein sequences, and then against the EST database of plants (dbEST plants). Queries against these databases were performed first with automatic charge assignment (deconvoluted), and then when no charge was assigned to the sequences (undeconvoluted), so that MASCOT could look for all possibilities of charges in each peak. As such, four different queries were made for each LC-MS/MS profile. The level of confidence of annotation is higher when hits to the same protein are found in more than one of the four query analysis, and when the probability- based Mowse score (Papin *et al.* 1993) from MASCOT is higher (Table 3.2). The Mowse score from MASCOT is based on the comparison of the experimental peptide profiles to the calculated peptide masses in a database. Since Mowse values for MASCOT are probabilities turned into scores (http://www.matrixscience.com/help/scoring_help.html) a significance level of 0.05 is assigned to assess if a match is valid. Under these criteria all the proteins with annotation in Table 3.2 have significant scores over the threshold, meaning that the probability of the annotation to be a random match is less than 0.05.

Results of this analysis are shown in Table 3.2. This table also summarizes for each protein the fold change difference in expression between 0d SD and 10wk SD

treatments, and the FDR level at which the protein was found to be significantly differentially expressed (q value below 5.2 % or 12.5 %), which corresponds to the minimum FDR at which the comparison for each protein is significant. A total of 110 of the analyzed peptides could be annotated by sequence similarity to NCBI accessions with meaningful annotation (Mowse probability based score) in one of the four analyses (NCBI protein or plant dbEST, convoluted or deconvoluted charge assignment), while another 12 of the analyzed peptides were annotated as hypothetical proteins and 39 gave no meaningful similarities when queried against NCBI databases. Many of the analyzed peptides gave hits to the same protein accession (e.g. ATPase beta subunit [*Cycas revoluta*]) giving a level of redundancy of close to 28 %. Of the annotated proteins, approximately one quarter were mainly annotated by querying against the plant EST database, which contains over 500000 spruce ESTs as well as several hundred thousand ESTs from other conifer species. Most of proteins that were mainly annotated by using the EST database also had good hits in the NCBI protein database (Table 3.2).

Table 3.1 Correlation coefficients (r^2) of all the protein spots between paired gels. Names of gels are divided in 3 parts. First part refers to the biological replicate corresponding to experiments S008 (S8), S011 (S11) and S012 (S12); the second part corresponds to the treatment for 0d SD (0SD) or 10wk SD (10SDW); and the third part to the technical replicate (1, 2 or 3). Correlation coefficients within treatment are shaded green and yellow while coefficients between treatments are in pink.

pH 3-10

A

166.0 kDa
66.2
45.0
35.0
25.0
18.4
14.4

pH 3-10

B

166.0 kDa
66.2
45.0
35.0
25.0
18.4
14.4

Figure 3.2 2D-SDS PAGE for samples under 0d SD (A) and 10wk SD (B). The two gels correspond to the second technical replicate from experiment S011. An increase in the number of detected proteins is evident after 10wk SD. Numbers on the right indicate masses of the molecular weight standards that were loaded with the gel but that are not shown.
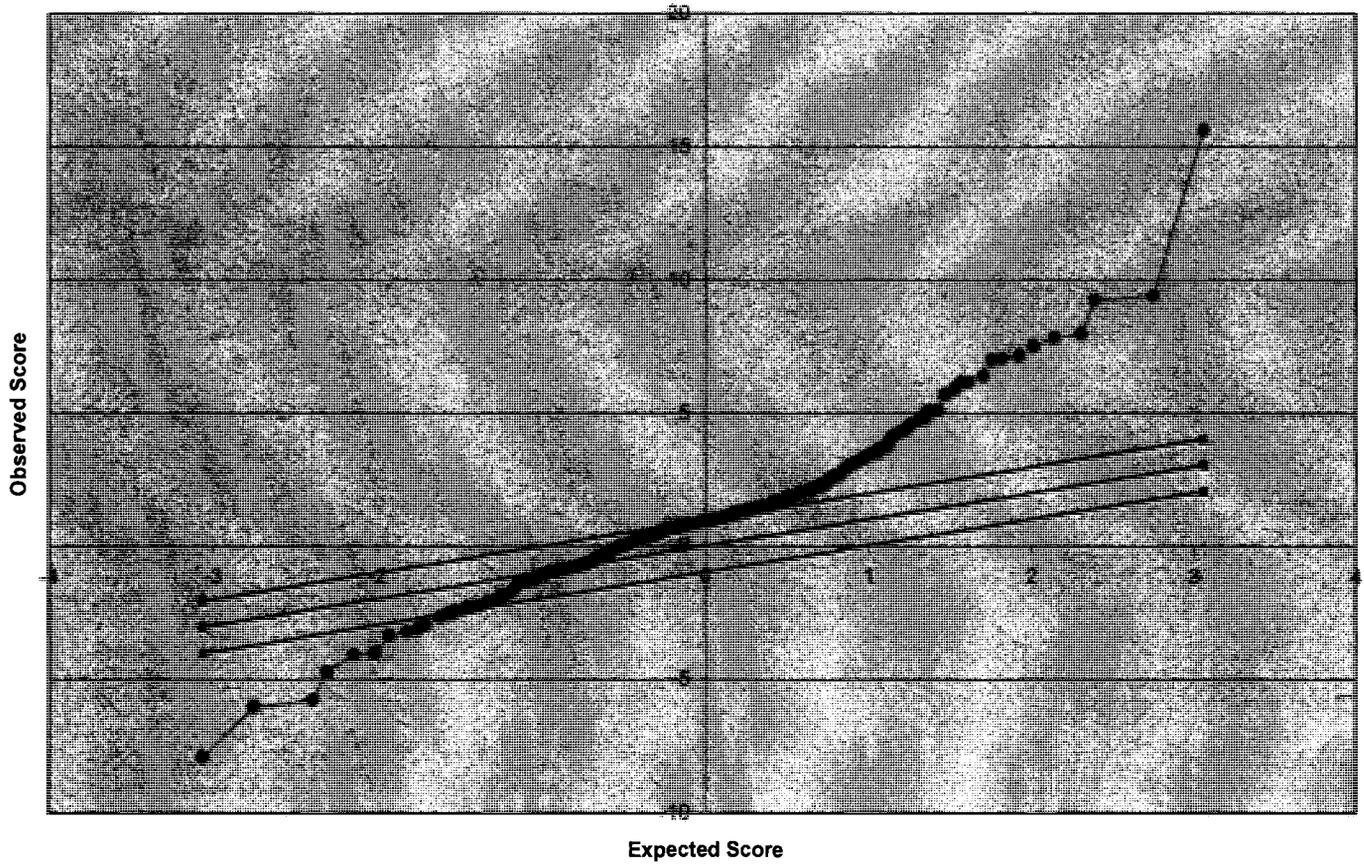
Figure 3.3 Significant Analysis of Microarrays (SAM) plot. Proteins overexpressed at 0d SD are in green while those overexpressed upon 10wk SD are in red. A total of 216 proteins were found as differentially expressed. The median number of false positive was 11.23 and the false discovery rate 5.2 %. The two external solid lines indicate the limits for a 1.5-fold change. The solid line in-between them corresponds to the region where the observed score equals the expected score.
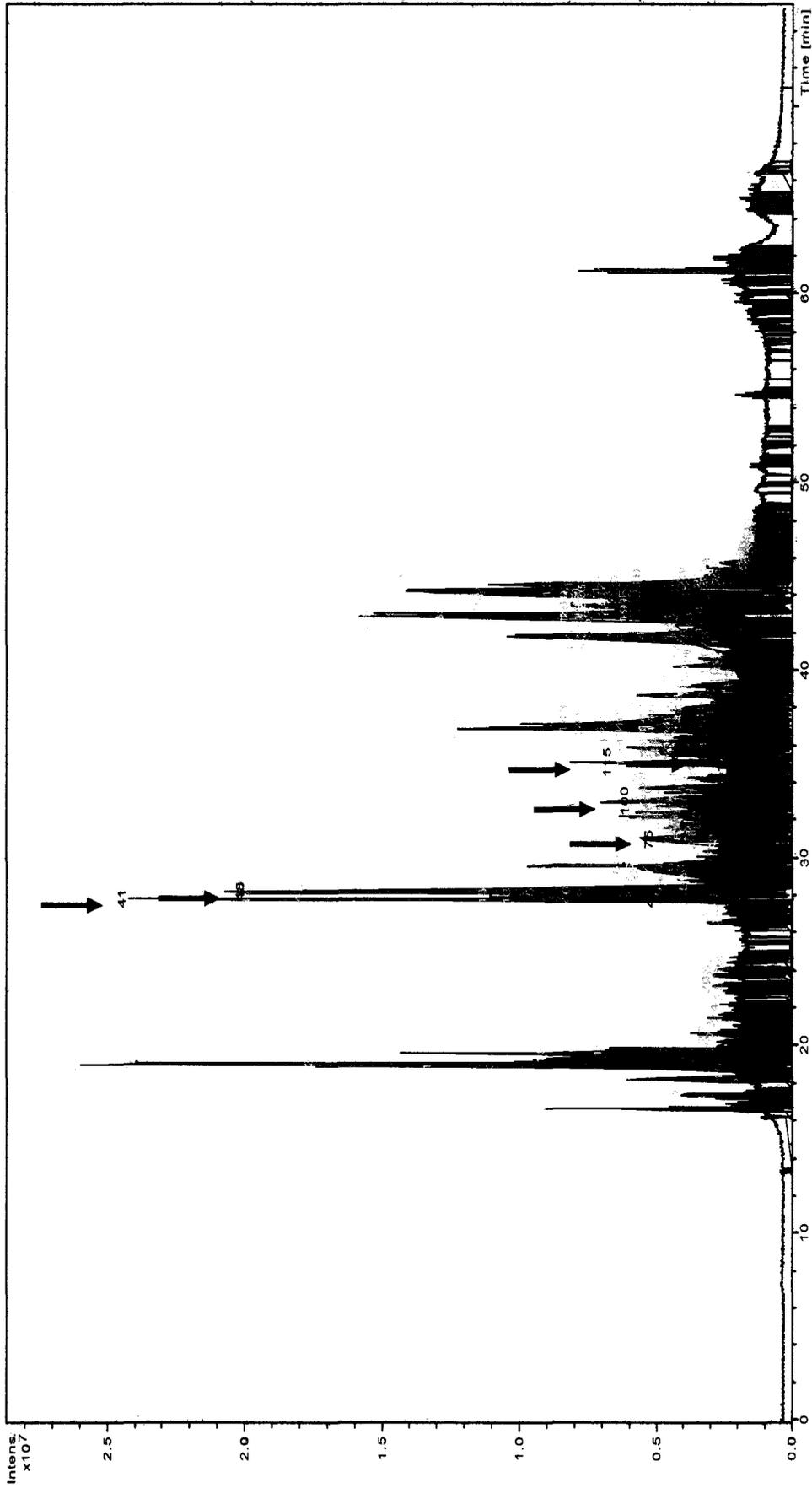
Figure 3.4 LC-MS/MS peptide profile of a chitinase-like protein (spot 477 in Table 3.2). This protein presented one of the highest MASCOT scores. Green and red peaks correspond to the tandem rounds of mass spectrometry. Arrows indicate hit peptide peaks in Mascot.

Table 3.2 Differentially expressed proteins with annotations and levels of significance. The score (d) represents the modified t-statistic for SAM. Q-value (%) is the minimum FDR at which the gene is significant. Asterisks at the query columns correspond to significance of mascot mowse scores over the 4 analyses performed: (*) score over threshold to 59, (**) score from 60 to 99, (***) score from 100 to 199, (****) score over 200. Query columns against the NCBI protein database (Prot) or plant EST database (Ests), correspond to: (d) deconvoluted analysis, (u) undeconvoluted analysis. A fold change value preceded by an asterisk corresponds to upregulation in 0d SD while absence of the asterisk corresponds to protein upregulated after 10wk SD.

	SAM analysis							Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
470	1.523	1.320E+11	3.182	55	BAD02824.1	putative class I chitinase [Taxodium distichum]	3E-117	*			
729	5.686	1.018E+11	0.000	no hit							
462	5.029	7.915E+10	0.000	112	AAP03088.1	class Ia chitinase [Galega orientalis]	2E-89		**		***
451	1.988	7.276E+10	1.622	204	AAL77589.1	chloroplast ribose-5-phosphate isomerase [Spinacia oleracea]	1E-25			****	***
333	7.009	7.075E+10	0.000	48	BAF81517.1	putative lactoylglutathione lyase [Brassica rapa]	2E-129	*			
455	1.902	7.007E+10	1.622	75	BAD07864.1	putative secretory carrier membrane protein [Oryza sativa Japonica group]	1E-15				**

	SAM analysis							Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
463	4.125	6.445E+10	0.000	50	BAD44910.1	zinc knuckle domain containing protein-like [Oryza sativa Japonica group]	1E-103	*			
362	2.100	5.940E+10	1.622	no hit							
473	1.818	5.846E+10	2.096	335	AAT09427.1	class II chitinase [Picea abies]	2E-106	****	****	****	****
695	2.345	4.932E+10	1.284	54	ABH02600.1	ATP synthase beta subunit [Diarrhena obovata]	0		*		
184	3.468	4.744E+10	0.000	no hit							
779	3.390	4.702E+10	0.000	72	AAT38562.1	chloroplast serine acetyltransferase [Thlaspi goesingense]	2E-33				**
468	1.326	4.602E+10	4.166	314	AAT09427.1	class II chitinase [Picea abies]	3E-108	***	****	***	****
378	3.748	3.808E+10	0.000	56	CAL56731.1	DNA repair protein-related (ISS) [Ostreococcus tauri]	0		*		
928	7.980	3.141E+10	0.000	48	NP_195035.2	APM1 (Aminopeptidase M1) [Arabidopsis thaliana]	1E-89	*			
273	1.496	3.025E+10	3.182	no hit							

	SAM analysis							Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
641	3.073	3.012E+10	0.254	no hit							***
459	1.725	2.801E+10	2.096	140	AAP03088.1	class Ia chitinase [Galega orientalis]	2E-89	***	***	***	***
634	2.470	2.664E+10	0.889	52	AAT09427.1	class II chitinase [Picea abies]	2E-106	*	*		
854	1.664	2.522E+10	2.284	no hit		alcohol dehydrogenase					
549	5.125	2.417E+10	0.000	50	AAC49545.1		0	*			
241	1.218	2.262E+10	4.166	62	ABH02600.1	ATP synthase beta subunit [Diarrhena obovata]	0		**		
546	4.395	2.012E+10	0.000	no hit							
96	3.244	1.993E+10	0.254	no hit							
240	2.285	1.990E+10	1.543	56	AAL27805.1	ATPase beta subunit [Cycas revoluta]	0		*		
745	2.102	1.957E+10	1.622	63	CAN71106.1	hypothetical protein [Vitis vinifera]	0		**		
97	3.379	1.945E+10	0.000	52	BAD16023.1	hypothetical protein [Oryza sativa Japonica Group]	6E-93	*	*		
477	1.092	1.891E+10	5.200	270	AAT09427.1	class II chitinase [Picea abies]	2E-106	***	****	***	****

	SAM analysis					Query					
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
373	4.585	1.876E+10	0.000	56	ABH02600.1	ATP synthase beta subunit [Diarrhena obovata]	0		*		
853	2.878	1.857E+10	0.254	53	EDQ64560.1	predicted protein [Physcomitrella patens subsp. patens]	0	*			
367	2.037	1.830E+10	1.622	no hit							
246	1.542	1.821E+10	2.284	no hit							
936	1.559	1.728E+10	2.284	no hit							
706	1.914	1.721E+10	1.622	no hit							
262	1.292	1.658E+10	4.166	no hit							
673	1.924	1.583E+10	1.622	63	AAF08180.1	NADH dehydrogenase subunit F [Cercidiphyllum japonicum]	0		**		
590	4.331	1.564E+10	0.000	59	AAD25591.1	Mutator-like transposase [Arabidopsis thaliana]	8E-175	*			
391	1.792	1.478E+10	2.096	52	CAL54485.1	putative translational activator (ISS) [Ostreococcus tauri]	0	*			

	SAM analysis								Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)	
604	1.383	1.384E+10	3.182	50	AAF76340.1	NADH dehydrogenase subunit F [Penaea mucronata]	0		*			
145	0.972	1.352E+10	6.728	no hit								
573	3.444	1.309E+10	0.000	no hit								
264	1.256	1.136E+10	4.166	no hit								
614	4.801	1.124E+10	0.000	79	CAA10904.1	asparaginyl-tRNA synthetase [Arabidopsis thaliana]	2E-82	**				
100	1.842	1.099E+10	2.096	no hit								
307	1.166	1.089E+10	5.200	no hit								
336	1.146	1.063E+10	5.200	53	EDQ48928.1	predicted protein [Physcomitrella patens subsp. patens]	3E-126		*			
149	1.655	9.370E+09	2.284	no hit								
637	3.686	9.239E+09	0.000	no hit								
339	1.954	9.232E+09	1.622	no hit							**	
585	1.168	8.121E+09	5.200	no hit								

	SAM analysis							Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
99	2.221	8.041E+09	1.543	53	NP_190566.1	phototropic-responsive protein, putative [Arabidopsis thaliana]	0	*	*		
625	1.245	7.909E+09	4.166	no hit							
694	1.253	7.765E+09	4.166	no hit							
567	1.882	6.915E+09	1.622	no hit							
598	1.322	6.713E+09	4.166	56	NP_179846.2	kinesin motor protein-related [Arabidopsis thaliana]	0		*		
632	1.247	6.044E+09	4.166	no hit							
327	1.190	5.785E+09	4.166	no hit							
322	1.014	5.602E+09	6.728	no hit							
583	1.661	5.455E+09	2.284	no hit							
782	0.937	4.886E+09	6.728	50	ABH02600.1	ATP synthase beta subunit [Diarrhena obovata]	0		*		
219	0.940	4.825E+09	6.728	no hit							
571	1.441	4.726E+09	3.182	no hit							
452	0.957	4.088E+09	6.728	no hit							

	SAM analysis								Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
101	7.504	8.603E+01	0.000	no hit							
647	9.246	7.755E+01	0.000	no hit							
600	6.188	6.135E+01	0.000	74	CAN63377.1	hypothetical protein [Vitis vinifera]	0		**		
818	3.406	5.159E+01	0.000	54	ABN09053.1	Cyclin-like F-box; Galactose oxidase, central [Medicago truncatula]	6E-15	*			
551	1.877	3.883E+01	1.622	no hit							
457	4.522	3.288E+01	0.000	54	BAD29556.1	hydroxyproline-rich glycoprotein family protein-like [Oryza sativa Japonica group]	2E-42	*			
778	2.168	3.250E+01	1.543	370	BAA31157.1	leghemoglobin [Pisum sativum]	0.076			****	****
481	2.970	2.548E+01	0.254	no hit							
387	0.929	2.378E+01	6.728	no hit							
372	5.730	2.279E+01	0.000	122	AAA21277.1	2-phospho-D-glycerate hydrolase	0	**	***	***	***
272	5.865	2.237E+01	0.000	no hit							

	SAM analysis							Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
371	6.197	1.844E+01	0.000	57	AAD52863.1	maturase-like protein [Adesmia volckmannii]	0	*			
317	1.047	1.838E+01	5.200	no hit							
701	9.416	1.764E+01	0.000	52	CAO43835.1	unnamed protein product [Vitis vinifera]	0	*			
518	1.959	1.699E+01	1.622	no hit							
181	4.770	1.610E+01	0.000	54	CAL44986.1	putative Na/H antiporter [Cymodocea nodosa]	0		*		
223	3.182	1.412E+01	0.254	58	CAA29003.1	P700 chlorophyll a-apoproteins 84 KD protein [Pisum sativum]	0		*		
259	2.788	1.368E+01	0.494	no hit							
98	4.116	1.329E+01	0.000	49	NP_001031936.1	hydrolase, hydrolyzing O-glycosyl compounds [Arabidopsis thaliana]	2E-148	*			
340	1.302	1.310E+01	4.166	52	ABA97297.1	hypothetical protein LOC_Os12g15610 [Oryza sativa (japonica cultivar-group)]	3E-23	*			
606	3.310	1.230E+01	0.254	114	AAT09427.1	class II chitinase [Picea abies]	6E-106	***	**	***	**

	SAM analysis							Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
82	1.628	1.169E+01	2.284	50	CAN71106.1	hypothetical protein [Vitis vinifera]	0		*		
106	2.504	1.132E+01	0.889	75	AAT38562.1	chloroplast serine acetyltransferase [Thlaspi goesingense]	1E-33			**	**
880	3.283	1.103E+01	0.254	no hit							
758	5.989	1.094E+01	0.000	no hit							
467	7.172	1.087E+01	0.000	206	BAD02824.1	putative class I chitinase [Taxodium distichum]	3E-117	***	****	***	****
319	1.811	1.013E+01	2.096	no hit							
603	2.643	1.008E+01	0.889	59	ABH02600.1	ATP synthase beta subunit [Diarrhena obovata]	0		*		
744	0.927	9.805E+00	6.728	no hit							
756	1.526	9.688E+00	3.182	54	XP_001786655.1	predicted protein [Physcomitrella patens subsp. patens]	8E-107	*			
291	2.261	9.126E+00	1.543	no hit							
754	2.007	8.888E+00	1.622	no hit							
810	1.847	8.647E+00	2.096	no hit							
626	4.284	8.597E+00	0.000	no hit							

	SAM analysis							Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
607	2.344	8.530E+00	1.284	no hit							
430	3.135	8.314E+00	0.254	56	NP_191562.1	SPL12 (SQUAMOSA PROMOTER-BINDING PROTEIN-LIKE 12); transcription factor [Arabidopsis thaliana]	0	*			
609	1.125	8.124E+00	5.200	51	BAC53939.1	MCM protein-like protein [Nicotiana tabacum]	0	*			
838	2.703	7.933E+00	0.889	54	NP_568662.1	O-acetyltransferase-related [Arabidopsis thaliana]	0		*		
	2.703	7.933E+00	0.889	54	AAR20302.1	maturase[Euonymus fortunei]	0		*		
858	1.200	7.857E+00	4.166	no hit							
503	2.252	7.837E+00	1.543	no hit							
574	0.948	7.805E+00	6.728	56	BAA22288.1	polyprotein [Oryza australiensis]	0		*		
178	1.561	7.752E+00	2.284	no hit							
597	2.064	7.574E+00	1.622	52	CAL44986.1	putative Na/H antiporter [Cymodocea nodosa]	0	*			

59

	SAM analysis						Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
597	2.064	7.574E+00	1.622	52	CAL55258.1	putative chloroplast 1-hydroxy-2-methyl-2-(ISS) [Ostreococcus tauri]	0		*		
773	1.607	6.813E+00	2.284	no hit							
798	2.144	6.709E+00	1.622	52	NP_189551.1	glycine-rich protein [Arabidopsis thaliana]	3E-13		*		
491	6.415	6.496E+00	0.000	203	BAD90814.1	thaumatin-like protein [Cryptomeria japonica]	2E-92	*		****	***
813	3.501	6.358E+00	0.000	no hit							
105	7.063	6.147E+00	0.000	61	ABH02600.1	ATP synthase beta subunit [Diarrhena obovata]	0	*	**		
187	1.207	6.127E+00	4.166	54	CAC19876.1	acetyl-CoA carboxylase [Brassica napus]	0	*			
834	1.237	6.008E+00	4.166	no hit							
310	3.506	5.825E+00	0.000	no hit							
158	1.377	5.812E+00	3.182	no hit							
749	1.792	5.716E+00	2.096	no hit							

	SAM analysis								Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)	
693	2.163	5.445E+00	1.543	56	EAZ32923.1	hypothetical protein OsJ_016406 [Oryza sativa (japonica cultivar-group)]	1E-173	*				
297	1.610	5.382E+00	2.284	49	BAA19156.1	HMG-1 [Canavalia gladiata]	8E-60	*				
859	0.952	5.279E+00	6.728	no hit								
320	4.005	5.260E+00	0.000	51	CAN63377.1	Hypothetical protein [Vitis vinifera]	0		*			
412	1.507	5.057E+00	3.182	55	ABH02600.1	ATP synthase beta subunit [Diarrhena obovata]	0	*				
425	1.403	4.892E+00	3.182	57	BAA94510.1	protein kinase 2 [Populus nigra]	0	*				
217	1.687	4.803E+00	2.284	no hit								
588	2.052	4.625E+00	1.622	51	ABU75152.1	AtpB [Brucea javanica]	0		*			
363	2.543	4.573E+00	0.889	no hit								
688	2.205	4.478E+00	1.543	55	BAC16059.1	putative cis-zeatin O-glucosyltransferase [Oryza Japonica group]	0	*	*			
361	2.671	4.378E+00	0.889	58	ABH02600.1	ATP synthase beta subunit [Diarrhena obovata]	0		*			

61

	SAM analysis							Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
712	1.274	4.359E+00	4.166	no hit							
615	1.366	4.282E+00	3.182	no hit							
124	7.829	4.152E+00	0.000	no hit							
800	2.404	4.149E+00	1.284	73	AAC49545.1	alcohol dehydrogenase	0	**			
836	1.527	3.973E+00	3.182	62	EDQ48531.1	predicted protein [Physcomitrella patens subsp. patens]	8E-107	**			
564	2.024	3.958E+00	1.622	no hit							
326	1.053	3.936E+00	5.200	no hit							
777	1.126	3.897E+00	5.200	no hit							
520	1.853	3.879E+00	2.096	62	ABH02600.1	ATP synthase beta subunit [Diarrhena obovata]	0		**		
376	15.613	3.877E+00	0.000	330	CAB96173.1	enolase [Spinacia oleracea]	0	***	****		****
270	0.935	3.846E+00	6.728	no hit							
229	3.080	3.789E+00	0.254	no hit							
628	1.356	3.784E+00	3.182	50	XP_001782328.1	predicted protein [Physcomitrella patens subsp. patens]	3E-114		*		

| | SAM analysis | | | | | | | Query | | | |
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
350	3.094	3.783E+00	0.254	69	NP_177466.1	nascent polypeptide-associated complex (NAC) domain-containing protein [Arabidopsis thaliana]	6E-56			**	
156	2.304	3.721E+00	1.284	49	CAL44986.1	putative Na/H antiporter [Cymodocea nodosa]	0		*		
269	3.025	3.649E+00	0.254	no hit							
283	1.643	3.643E+00	2.284	no hit							
139	1.553	3.613E+00	2.284	no hit							
530	1.014	3.567E+00	6.728	no hit							
410	3.219	3.541E+00	0.254	no hit							
426	2.054	3.534E+00	1.622	no hit							
298	3.280	3.525E+00	0.254	52	ABA46972.1	cysteine protease Mir1 [Zea diploperennis]	1E-117		*		
417	4.216	3.479E+00	0.000	68	BAD69015.1	putative auxin-independent growth promoter [Oryza sativa (japonica cultivar-group)]	1E-114			**	
843	1.892	3.406E+00	1.622	60	AAL27805.1	ATPase beta subunit [Cycas revoluta]	0		**		

63

	SAM analysis							Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
544	1.442	3.386E+00	3.182	no hit							
796	2.041	3.361E+00	1.622	75	ZP_02965050.1	NHL repeat containing protein [bacterium Ellin514]	1E-26	**			
696	2.173	3.316E+00	1.543	59	BAD45437.1	speckle-type POZ protein-like [Oryza sativa (japonica cultivar-group)]	5E-12		*		
867	1.638	3.310E+00	2.284	no hit							
265	1.008	3.282E+00	6.728	54	AAL27805.1	ATPase beta subunit [Cycas revoluta]	0		*		
561	2.413	3.243E+00	1.284	149	ABO69625.1	synaptonemal complex protein ZYP1 [Brassica oleracea]	8E-05			***	***
328	5.097	3.237E+00	0.000	54	ABD32884.1	SH2 motif; Alcohol dehydrogenase, zinc-containing; Resolvase, Rnase H-like fold; Nucleic acid binding, OB-fold, subgroup [Medicago truncatula]	0		*		
172	1.341	3.118E+00	3.182	no hit							
650	4.721	3.103E+00	0.000	no hit							

64

SAM analysis								Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
584	2.919	3.040E+00	0.254	no hit							
505	1.798	2.983E+00	2.096	no hit							
562	2.198	2.915E+00	1.543	56	ABA97297.1	hypothetical protein LOC_Os12g15610 [Oryza sativa (japonica cultivar-group)]	3E-23	*			
166	1.493	2.819E+00	3.182	51	BAF79940.1	receptor-like kinase [Marchantia polymorpha]	0		*		
523	1.412	2.770E+00	3.182	no hit							
516	3.094	2.724E+00	0.254	116	AAM77651.1	cpl0-like protein [Gossypium hirsutum]	9E-90	***	***	***	**
542	1.927	2.679E+00	1.622	no hit							
675	1.391	2.676E+00	3.182	97	CAH58634.1	thioredoxin-dependent peroxidase [Plantago major]	6E-56	**	**	**	**
286	1.313	2.668E+00	4.166	no hit							
822	1.291	2.648E+00	4.166	no hit							
472	1.597	2.604E+00	2.284	66	XP_001698933.1	hypothetical protein CHLREDRAFT_193403 [Chlamydomonas reinhardtii]	0	**			

	SAM analysis								Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
559	2.364	2.601E+00	1.284	53	AAP73860.1	putative DNA gyrase subunit [Oryza sativa (japonica cultivar-group)]	0	*			
382	0.932	2.577E+00	6.728	no hit							
109	2.088	2.536E+00	1.622	no hit							
874	3.687	2.415E+00	0.000	no hit							
169	1.360	2.361E+00	3.182	54	AAP84375.1	mutant FCA-D1 [Triticum aestivum]	5E-95	*			
227	1.154	2.351E+00	5.200	no hit							
108	2.318	2.334E+00	1.284	78	BAA92724.1	putative chaperonin 60 beta precursor [Oryza sativa Japonica group]	1E-57	**		**	**
329	1.250	2.327E+00	4.166	no hit							
629	2.023	2.322E+00	1.622	no hit							
575	1.201	2.314E+00	4.166	no hit							
277	4.300	2.296E+00	0.000	no hit							
260	1.535	2.292E+00	2.284	no hit							

	SAM analysis							Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
314	1.021	2.208E+00	6.728	68	CAL57543.1	Glucose-repressible alcohol dehydrogenase transcriptional effector CCR4 and related proteins (ISS) [Ostreococcus tauri]	0		**		
71	1.712	2.202E+00	2.096	57	EAZ11596.1	hypothetical protein OsJ_001421 [Oryza sativa (japonica cultivar-group)]	5E-150		*		
612	1.486	2.120E+00	3.182	no hit							
788	1.611	2.068E+00	2.284	no hit							
285	4.182	2.052E+00	0.000	166	AAQ22131.1	quinone oxidoreductase x [Fragaria ananassa]	6E-75	***	***	***	***
529	1.437	2.045E+00	3.182	124	AAA21277.1	2-phospho-D-glycerate hydrolase	0	***	***	***	***
740	2.044	2.022E+00	1.622	no hit							
170	2.671	2.019E+00	0.889	no hit							
414	3.467	1.988E+00	0.000	no hit							
111	3.679	1.973E+00	0.000	no hit							
400	2.287	1.971E+00	1.543	187	AAA21277.1	2-phospho-D-glycerate hydrolase	0	***	***	***	***

	SAM analysis							Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
429	0.982	1.967E+00	6.728	no hit							
484	1.441	1.926E+00	3.182	no hit							
532	1.184	1.919E+00	4.166	no hit							
90	5.097	1.900E+00	0.000	497	Q43116	Protein disulfide-isomerase precursor (PDI)	3E-96	****	****	****	****
734	1.085	1.880E+00	5.200	81	AAL58654.1	ribulose-1,5-bisphosphate carboxylase/oxygenase [Hexalectris revoluta]	0		**		
365	1.259	1.862E+00	4.166	no hit							
680	1.625	1.861E+00	2.284	no hit							
144	1.320	1.858E+00	4.166	50	CAN63377.1	hypothetical protein [Vitis vinifera]	0		*		
724	2.879	1.850E+00	0.254	no hit							
883	1.089	1.840E+00	5.200	no hit							
784	2.821	1.825E+00	0.494	no hit							
117	1.562	1.789E+00	2.284	no hit							

SAM analysis					top BLASTp hit	BLAST annotation	E-value	Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score				Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
263	2.197	1.789E+00	1.543	423	ABK76304.1	chloroplast sedoheptulose-1,7-bisphosphatase [Morus alba var.multicaulis]	3E-165	****	****	****	****
347	1.581	1.769E+00	2.284	57	NP_199328.1	RTL3 (RNASE THREE-LIKE PROTEIN 3); double-stranded RNA binding / ribonuclease III [Arabidopsis thaliana]	0		*		
868	1.170	1.744E+00	5.200	no hit							
538	2.517	1.733E+00	0.889	52	ABH02600.1	ATP synthase beta subunit [Diarrhena obovata]	0		*		
486	1.457	1.730E+00	3.182	193	ABG73467.1	6-phosphogluconolactonase [Oryza brachyantha]	2E-84	***	***	***	***
154	1.600	1.700E+00	2.284	no hit							
208	2.172	1.692E+00	1.543	no hit							
136	1.148	1.660E+00	5.200	209	P17614	ATP synthase subunit beta, mitochondrial precursor	0	***	****	***	****

	SAM analysis							Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
668	2.090	1.656E+00	1.622	50	CAL50172.1	A Chain A, Crystal Structure Of Peroxisomal Acyl-Coa Oxidase-Ii (ISS) [Ostreococcus tauri]	0				
212	1.046	1.655E+00	5.200	no hit					*		
586	0.981	1.635E+00	6.728	52	ABL97963.1	hydrogen-transporting ATP synthase [Brassica rapa]	5E-71	*			
464	1.043	1.614E+00	5.200	no hit							
411	3.704	1.588E+00	0.000	50	BAD44910.1	zinc knuckle domain containing protein-like [Oryza sativa (japonica cultivar-group)]	1E-104		*		
579	1.216	1.576E+00	4.166	no hit							
364	1.894	1.574E+00	1.622	no hit							
346	1.362	1.567E+00	3.182	no hit							
209	1.679	1.563E+00	2.284	no hit							
847	1.281	1.545E+00	4.166	no hit							
447	1.033	1.537E+00	5.200	no hit							
73	1.408	1.529E+00	3.182	no hit							
408	1.446	1.511E+00	3.182	no hit							

SAM analysis								Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
409	1.747	1.478E+00	2.434	no hit							
742	1.903	1.464E+00	1.886	no hit							
235	1.726	1.440E+00	2.434	no hit							
140	1.479	1.438E+00	3.352	no hit							
931	1.091	1.416E+00	6.081	52	CAL44986.1	putative Na/H antiporter [Cymodocea nodosa]	0		*		
686	2.711	1.398E+00	0.867	70	NP_180685.1	ATCDT1A/CDT1A (ARABIDOPSIS HOMOLOG OF YEAST CDT1 A); cyclin-dependent protein kinase/ protein binding [Arabidopsis thaliana]	9E-31	**	*		
768	0.964	1.419E+00	8.095	no hit							
592	0.964	1.387E+00	8.095	no hit							
379	1.043	1.381E+00	6.081	no hit							
433	1.306	1.375E+00	4.787	no hit							
489	1.149	1.359E+00	6.081	no hit							
531	2.636	1.338E+00	0.867	52	Q9GI85	Maturase K (Intron maturase)	0		*		

SAM analysis								Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
723	0.989	1.308E+00	8.095	54	ABJ55956.1	NBS-containing resistance-like protein [Prunus avium]	4E-85		*		
115	0.928	1.292E+00	8.095	no hit							
613	1.299	1.257E+00	4.787	no hit							
278	1.478	1.248E+00	3.352	no hit							
576	1.314	1.242E+00	4.787	no hit							
385	1.344	1.191E+00	3.352	no hit							
207	-1.125	*8.685E-01	10.759	no hit							
454	-0.995	*8.569E-01	12.314	no hit							
113	-1.302	*8.146E-01	9.635	no hit							
352	-0.967	*7.721E-01	12.314	no hit							
869	-0.915	*7.700E-01	12.314	no hit							
483	-1.152	*7.668E-01	10.759	no hit							
837	-1.585	*7.143E-01	8.490	no hit							
318	-1.355	*7.058E-01	9.635	no hit							
487	-1.860	*6.961E-01	8.320	no hit							
296	-2.379	*6.840E-01	8.213	no hit							

	SAM analysis							Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
289	-1.110	*6.829E-01	10.759	no hit							
651	-2.265	*6.793E-01	8.213	no hit							
211	-1.028	*6.680E-01	10.759	no hit							
399	-2.431	*6.341E-01	5.200	no hit							
384	-0.921	*6.298E-01	12.314	no hit							
450	-1.835	*6.253E-01	8.320	no hit							
815	-2.975	*6.223E-01	2.284	no hit							
498	-0.959	*6.020E-01	12.314	no hit							
930	-1.683	*5.834E-01	8.361	77	CAA83548.1	PsHSC71.0 [Pisum sativum]	4E-30	*	**		**
438	-1.162	*5.791E-01	10.759	no hit							
702	-1.964	*5.659E-01	8.320	no hit							
833	-3.361	*5.551E-01	1.622	84	CAA34161.1	ribulose-1,5-carboxylase/oxygenase [Larix laricina]	3E-74				**
807	-0.950	*5.521E-01	12.314	no hit							
221	-1.374	*5.359E-01	9.635	no hit							
453	-2.111	*5.352E-01	7.148	no hit							
197	-1.601	*5.349E-01	8.490	no hit							

| | SAM analysis | | | | | | | Query | | | |
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
316	-2.541	*5.336E-01	5.200	no hit							
143	-1.320	*5.271E-01	9.635	52	ABA98741.2	retrotransposon protein, putative, Ty3-gypsy subclass [Oryza sativa (japonica cultivar-group)]	0		*		
152	-0.912	*5.260E-01	12.314	67	P17614	ATP synthase subunit beta, mitochondrial precursor	0	*	**		
510	-0.845	*5.206E-01	12.609	57	CAN66453.1	hypothetical protein [Vitis vinifera]	0		*		
556	-0.990	*5.151E-01	12.314	no hit							
658	-0.856	*5.111E-01	12.609	53	NP_179846.2	kinesin motor protein-related [Arabidopsis thaliana]	0	*			
275	-3.215	*5.100E-01	1.622	no hit							
557	-2.042	*4.999E-01	7.148	no hit							
388	-1.008	*4.833E-01	12.314	no hit							
851	-1.371	*4.591E-01	9.635	no hit							

74

| | SAM analysis | | | | | | Query | | | | |
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
954	-1.323	*4.477E-01	9.635	54	AAL27805.1	ATPase beta subunit [Cycas revoluta]	0		*		
168	-0.953	*4.360E-01	12.314	no hit							
515	-0.870	*4.315E-01	12.561	no hit							
405	-1.401	*4.307E-01	9.635	no hit							
619	-1.390	*4.241E-01	9.635	no hit							
705	-3.107	*4.088E-01	2.096	no hit							
852	-2.546	*4.057E-01	5.200	no hit							
938	-0.955	*4.056E-01	12.314	no hit							
639	-2.118	*3.900E-01	7.148	no hit							
127	-2.185	*3.857E-01	7.020	no hit							
605	-4.069	*3.838E-01	0.325	55	NP_680770.1	FAT domain-containing protein / phosphatidylinositol 3- and 4-kinase family protein [Arabidopsis thaliana]	0		*		
568	-7.917	*3.691E-01	0.000	no hit							
709	-2.232	*3.086E-01	7.020	90	AAL33589.1	methionine synthase [Zea mays]	2E-43			**	**

	SAM analysis							Query			
Spot ID	Score(d)	Fold Change	q-value(%)	MASCOT score	top BLASTp hit	BLAST annotation	E-value	Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
540	-0.833	*2.716E-01	12.609	no hit							
877	-6.026	*2.715E-01	0.000	no hit							
390	-1.801	*2.645E-01	8.320	no hit							
570	-2.330	*2.566E-01	7.020	no hit							
396	-2.354	*2.540E-01	7.020	no hit							
935	-4.714	*1.944E-01	0.325	140	AAR24912.1	fructokinase 3 [Solanum lycopersicum]	1E-103	**	***	**	***
358	-0.947	*1.849E-01	12.314	no hit							
794	-1.834	*1.838E-01	8.320	no hit							
959	-2.659	*1.744E-01	5.200	66	AAL27805.1	ATPase beta subunit [Cycas revoluta]	0		**		
593	-4.029	*8.312E-02	0.325	no hit							
539	-2.261	*5.208E-02	7.020	no hit							
953	-1.115	*1.542E-02	10.759	no hit							
961	-1.401	*4.845E-03	9.635	57	NP_201224.1	unknown protein [Arabidopsis thaliana]	0		*		
945	-0.892	*4.660E-04	12.561	no hit							

	SAM analysis			MASCOT score	top BLASTp hit	BLAST annotation	E-value	Query			
Spot ID	Score(d)	Fold Change	q-value(%)					Prot. (d)	Prot. (u)	Ests (d)	Ests (u)
955	-5.765	*1.551E-11	0.000	no hit							

3.4 Annotation analysis

3.4.1 Annotation to functional categories

Automated annotation performed using Blast2GO was used to create a combined graph to map the GO terms of the different annotated proteins according to their parent-children relationships and the abundance in each respective category (results not shown). The combined graph was then used to obtain frequencies of proteins per biological processes of second level (Table 3.4) and the respective relative frequencies (Figure 3.5), which gives resolution of broad functional groups. This bar graph provides a general overview of the biological processes represented by the annotated proteins and its comparison to the categories of a random 5000 sequences sample from GCAT (the comparison of these two analyses will be described below). Figure 3.5 reflects the diversity of categories populated by different members of each group. In other words, if there was more than one hit under the same annotation the category did not increase in size since the hit was redundant.

As seen in Table 3.4, metabolic processes (47 proteins) and cellular processes (44 proteins) were the two GO terms with the most members. The metabolic processes category covers reactions of anabolism and catabolism to transform chemicals. In general metabolic processes have to do with small molecules, i.e. metabolites, but also include processes related to DNA repair and replication as well as protein synthesis and degradation. Cellular processes cover a wide array of actions carried out in a cell or between cells, such as cellular communication. The next most represented category was response to stimulus (14 proteins), a term meaning that cells were reacting to an external signal in terms of characteristics like secretion, enzyme production and gene expression. The next two most abundant categories had to do with localization; these proteins involve transportation of substances to a fixed location. The remaining terms showed relationships with interaction with biotic stimulus (multi-organism process), regulation of the rate and quality of production of substances (biological regulation), and a variety of

development processes (reproductive process, developmental process, reproduction, multicellular organismal process).

To gain more insight into specific processes represented by these general categories (level 2), I classified the proteins with greater specificity using level 3 of biological process annotation. Appendix 7.2 shows that many of the annotated proteins are implicated in general metabolic processes including cellular, primary and macromolecule metabolism, reflecting a shift in metabolic pathways taking place in the cell. Biosynthetic processes that take place included the activation of proteins involved in phototosynthesis and carbon utilization (sedoheptulose 1-7 biphosphatase, chlorophyll a-apoprotein, RuBisco, ribose 5-phosphate isomerase), energy production (ATP synthase and NADH dehydrogenase) and synthesis of proteins involved in diverse processes (e.g. acetyl transferases, amino acid metabolism proteins).

Transport was another prominent category, containing proteins involved in the transport of ions like sodium and protons (ATP synthase, Na/H antiporter), electron transport (Chlorophyll apoprotein, NADH dehydrogenase) and membrane related transport (secretory carrier membrane protein). The previous categories outlined the importance of synthesis, transport and generation of energy, but at the same time enzymes that can be involved in interconversion processes were also found (6-phosphogluconolactonase, ribose 5-P isomerase, fructokinase, enolase, chitinases).

In addition to carbon metabolism, some proteins were related to the term of nitrogen compound metabolic process. Among this there were components of signal transduction (receptor-like kinase and a protein kinase) and enzymes related to amino acid metabolism (methionine synthase, asparaginyl t-RNA synthase, serine acetyltransferase).

Several proteins were classified into more than one GO term. In the general stress response we found genes involved in redox reactions (lactoglutathione lyase, thioredoxin-dependent peroxidase, quinone oxidoreductase) plus genes involved in energy harvest and production (Rubisco and ATP synthase) and a protective protein (heat shock protein). Many of the abiotic response proteins were common with the general stress response proteins category (lactoglutathione lyase, quinine oxidoreductase, Rubisco), but additionally an enolase and a phototropic responsive protein were also found in this term.

Responses to biotic stimulus had the same members as the response to other organism term. In these categories I found chitinases, heat shock proteins, chaperonins and a thaumatin like protein. Finally a chaperonin and a chitinase were also annotated to defense response, immune response and death. The mixed categorization in the responses to stimulus and stress showed that many proteins act in different pathways and under different influences.

A response at the cellular level could be characterized by the group of proteins belonging to cellular component organization and biogenesis. In this group I found proteins related to chromatin and nucleic acid binding (the high mobility group protein HMG-1 and a protein with multiple domains - ABD32884.1), and also potential cell wall remodeling enzymes (chitinases).

3.4.2 Enzyme annotation

Blast2GO automated mapping was performed on the proteins that could be annotated to metabolic or regulatory pathways using the KEGG classification. Twenty four non-redundant hits were annotated to one or more pathways (Table 3.3). From the pathways we can see that the induction of dormancy implicates changes related to amino acid synthesis and metabolism, sugar metabolism, carbon fixation and utilization, lipid metabolism and the energy status of the cell. In Table 3.3 the sequence description of the hits might not coincide exactly with the blast hits of Table 3.2 for the same sequence identifier. The reason for this is that blast2GO uses a language algorithm to annotate the most likely function by checking not only the top hits, but the 20 top hits (Conesa and Götz 2007).

Pathways represented by at least one sequence included oxidative phosphorylation, carbon fixation, glycolysis or gluconeogenesis, pyruvate metabolism, alanine and aspartate metabolism, and the pentose phosphate pathway. Some of these pathways can be represented by three enzymes while others may be represented by one or two enzymes. As seen in Table 3.3 several enzymes belong to more than one pathway.

3.4.3 Comparison of functional categories to a random sample

To establish the relevance of the major functional categories arising in these analyses, I subjected a random sample of 5000 sequences from the spruce database (GCAT v2.0) to the same annotation process as the set of proteins obtained from the two-dimensional gel electrophoresis study.

The first step in Blast2GO to get the GO-associated terms is to perform BLAST in all the sequences. While from our sample all the sequences submitted to Blast2GO had a BLASTp hit, only 2615 cDNAs from the 5000 random sequences sample yielded similarity to a sequence with meaningful annotation. After performing the mapping which associates GO terms to the BLAST hits, I then performed automated annotation of all the sequences using the Interpro and Annex tools from Blast2GO. After this analysis, 75 % of our sample sequences could be annotated to at least one GO term, while 71 % could be annotated in the 5000 sequences random sample.

To establish if there was a significant difference between the relative frequencies of sequences in each of the functional categories corresponding to biological process (level 2) of our sample and the respective categories in the random sample (Figure 3.5), a table was created with the frequencies found in each category (Table 3.4). The chi-square test of the contingency table showed that the relative frequencies for both sets of data (our sample and the random sample) are not the same for each functional category ($P =$ 0.00215). Since there was a general significant difference we performed heterogeneity tests (Zar 1999) to see which terms contributed to this global difference.

Table 3.3 Automated pathway annotation using blast2GO. The annotation is based on KEGG metabolic pathways. The top BLASTp hit corresponds to the same hit in Table 3.2. Sequence description is given by a language algorithm which uses the top 20 BLAST hits (Conesa and Götz 2007).

Spot ID	top BLASTp hit	Sequence Description	KeggMaps - top pathway(s)
614	CAA10904.1	asparaginyl-trna synthetase	Alanine and aspartate metabolism; Aminoacyl-tRNA biosynthesis
98	NP_001031936.1	catalytic cation binding hydrolyzing o-glycosyl compounds	Alanine and aspartate metabolism; Pantothenate and CoA biosynthesis; D-Glutamine and D-glutamate metabolism; Arginine and proline metabolism; Tyrosine metabolism;beta-Alanine metabolism; Biosynthesis of siderophore group nonribosomal peptides; Phenylalanine metabolism; Urea cycle and metabolism of amino groups
459,	AAP03088.1	class I chitinase	Aminosugars metabolism
378	CAL56731.1	dna repair protein	Base excision repair
597	CAL55258.1	hydroxymethylbutenyl 4-diphosphate synthase	Biosynthesis of steroids
263	ABK76304.1	chloroplast sedoheptulose--bisphosphatase	Carbon fixation
451	AAL77589.1	ribose 5-phosphate isomerase	Carbon fixation; Pentose phosphate pathway
106	AAT38562.1	serine acetyltransferase	Cysteine metabolism; Sulfur metabolism
187	CAC19876.1	cytosolic acetyl-carboxylase	Fatty acid biosynthesis; Pyruvate metabolism; Tetracycline biosynthesis; Fatty acid biosynthesis; Propanoate metabolism
668	CAL50172.1	acyl-coenzyme a oxidasepalmitoyl	Fatty acid metabolism; alpha-Linolenic acid metabolism; Biosynthesis of unsaturated fatty acids
935	AAR24912.1	pfkb-type carbohydrate kinase family protein	Fructose and mannose metabolism; Starch and sucrose metabolism; Pentose phosphate pathway
400	AAA21277.1	enolase (2-phospho-d-glycerate hydroylase)	Glycolysis / Gluconeogenesis

Spot ID	top BLASTp hit	Sequence Description	KeggMaps - top pathway(s)
734	AAL58654.1	ribulose--bisphosphate carboxylase oxygenase	Glyoxylate and dicarboxylate metabolism; Carbon fixation
675	CAH58634.1	thioredoxin-dependent peroxidase	Methane metabolism; Phenylalanine metabolism; Phenylpropanoid biosynthesis
709	AAL33589.1	methionine synthase	Methionine metabolism; Methionine metabolism; One carbon pool by folate
586	ABL97963.1	atp synthase delta mitochondrial precursor	Oxidative phosphorylation; Photosynthesis
604	AAF76340.1	nadh dehydrogenase subunit f	Oxidative phosphorylation; Ubiquinone biosynthesis
673	AAF08180.1	nadh-plastoquinone oxidoreductase subunit 5	Oxidative phosphorylation; Ubiquinone biosynthesis
105	ABH02600.1	atp synthase beta subunit	Oxidative phosphorylation; Photosynthesis
486	ABG73467.1	6-phosphogluconolactonase-like protein	Pentose phosphate pathway
333	BAF81517.1	lactoylglutathione lyase	Pyruvate metabolism
688	BAC16059.1	cis-zeatin o-glucosyltransferase	Pyruvate metabolism; Alanine and aspartate metabolism; Butanoate metabolism; Valine, leucine and isoleucine biosynthesis; Glycolysis / Gluconeogenesis
800	AAC49545.1	alcohol dehydrogenase	Fatty acid metabolism; Metabolism of xenobiotics by cytochrome P450; Glycolysis / Gluconeogenesis; Tyrosine metabolism;1- and 2-Methylnaphthalene degradation
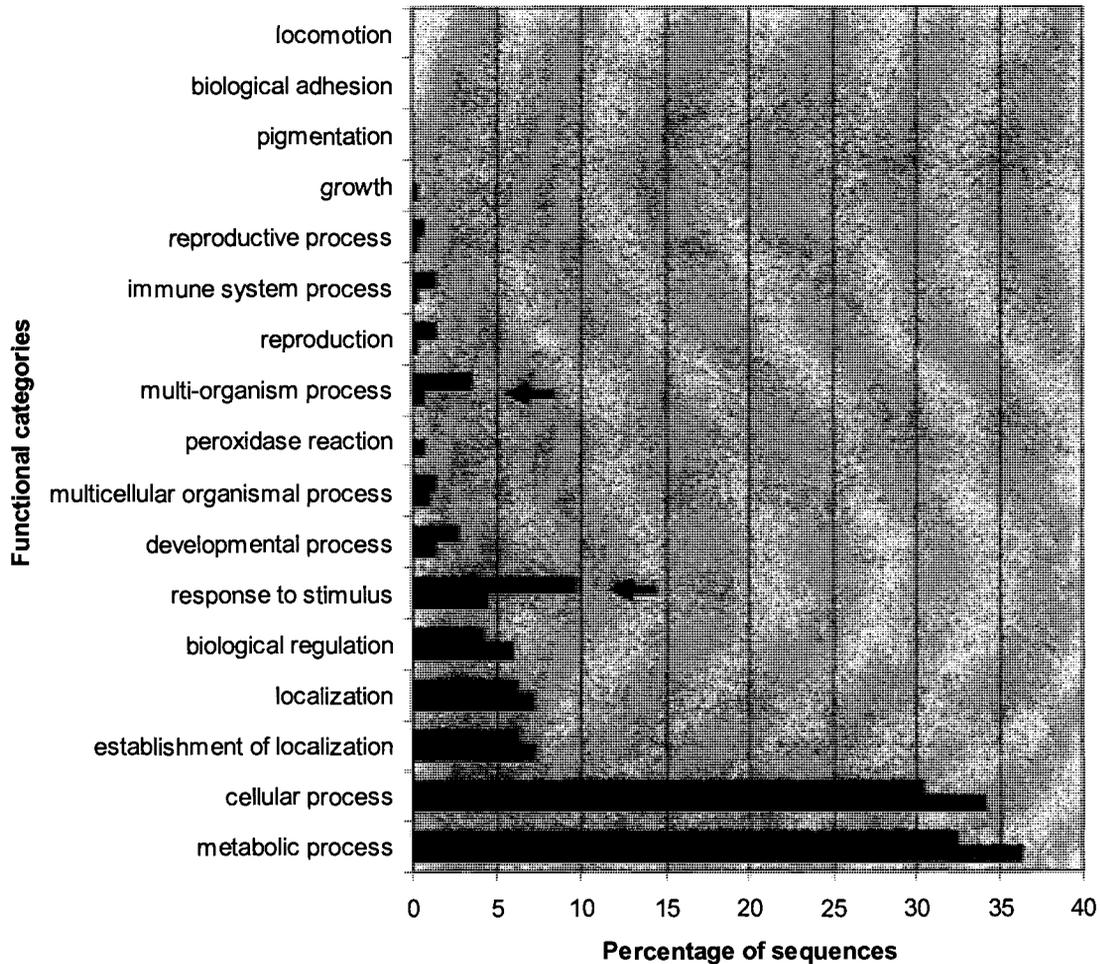
Figure 3.5 Bar graph of the relative frequencies of sequences in each functional category (biological process, level 2), from proteins annotated in this study (dark red), and a random sample of 5000 sequences from the GCAT database (light blue). The relative frequencies of the two sets of data are significantly different (P= 0.00215). The largest contribution to this significant difference comes from response to stimulus (P= 0.003575) and multi-organism process (P= 0.000178), indicated by arrows.

Table 3.4 Frequencies of the different biological processes (level 2) functional categories determined using Blast2GO. The number of sequences from the random sample of 5000 sequences from GCAT is compared to the number of sequences of this study. Heterogeneity tests (chi-square) were performed to establish which categories had significant differences between the samples. (P) is the probability value of the chi-square statistic. (**) significant at an α level of 0.001, (*) significant at an α level of 0.05.

Functional categories	Number of sequences (random sample)	Number of sequences (this study)	heterogeneity chi-square (P)
metabolic process	1018	47	0.346176
cellular process	956	44	0.356967
establishment of localization	202	9	0.652018
localization	202	9	0.652018
biological regulation	167	6	0.365155
response to stimulus	124	14	**0.003575
developmental process	39	4	0.179613
multicellular organismal process	26	2	0.583115
peroxidase reaction	18	0	0.333432
multi-organism process	18	5	**0.000178
reproduction	9	2	*0.041324
immune system process	9	2	*0.041324
reproductive process	8	1	0.388772
growth	8	0	0.519757
pigmentation	1	0	0.820175
biological adhesion	1	0	0.820175
locomotion	1	0	0.820175

The results of these analyses are shown in the third column of Table 3.4. As seen from these analyses, the observed values for categories corresponding to response to stimulus and multi-organism process deviate from the expected values and show that their relative frequency is quite different from that of the random sample (the probabilities in this case are below 0.001). The categories of reproduction and immune system response also show a significant difference at an α level of 0.05, but this result should be taken with caution since the number of hit sequences is low in our sample (2 hit sequence in each category, Table 3.4), and generally a count number below 5 is not as reliable (Richard Moses, personal communication).

3.5 Chitinase *in-silico* characterization

Nine spots annotated as chitinases (chitinase-like proteins, since function was not established) had a minimum of 10-fold change and showed greater abundance after 10wk SD (Table 3.2). The large fold changes in expression levels for these chitinase-like proteins were because the spots were below detectable limits at 0d SD, leading to a very small denominator. The sum of the normalized volume values of these protein spots accounted for almost 4 % of all detected signal for protein volumes in the second treatment, but only 0.19 % in 0d SD. The nine spots annotated as chitinase-like proteins corresponded to three non-redundant contig sequences in GCAT; multiple hits to the same contig sequence by peptides representing different spots could have been caused by isomers or post-translational modifications. The GCAT contig identifiers of these chitinase-like proteins: Pg_GQ0134.B7.1_C02:53:Contig1:1-1185:+, Pg_GQ03602.B7_D17:53:Contig1:1-1101:+, Pg_GQ03602.B7_D17:5-:Contig3:1-897:+ correspond to spot identifiers 459, 473 and 477 respectively in Table 3.2 and the respective sequences are included in Appendix 7.3.

Of the annotated proteins, these chitinase-like proteins represented the best vegetative storage protein candidates. As such, these proteins were characterized in more detail.

3.5.1 Phylogenetic analysis

To better infer potential roles of the chitinase-like proteins during dormancy acquisition in spruce, a phylogenetic analysis was carried out on spruce chitinase-like sequences obtained from GCAT (materials and methods), along with chitinases and chitinase-like sequences from Arabidopsis, rice and other plant species (Figure 3.6). Most non-spruce sequences used to construct the phylogenetic tree are those used in a phylogenetic study of plant chitinases by Xu *et al.* (2007). Additional functionally characterized chitinases from different plants were added in order to facilitate inference of putative function. The GCAT chitinase-like sequences were obtained by using all the non-spruce sequences mentioned above to perform a batch tBLASTn against the GCAT database. The search yielded 4256 redundant contig hits with an e-value below E-04; these sequences were filtered for redundancy and almost 600 non-redundant hits were then used for BLASTx queries against the NCBInr database finding that almost 100 hits were strongly annotated as chitinase. From these, a multiple pairwise nucleotide alignment was used to filter out sequences with more than 95 % identity, which have a high likelihood of representing multiple alleles of the same locus, resulting in 31 non-redundant chitinase-like sequences. Phylogenies were constructed with deduced amino acid sequences in Mega 4.0 using permutations of two alignment algorithms (CLUSTAL and MUSCLE) and four phylogenetic algorithms (NJ, MP, ME, UPGMA). The neighbour-joining tree derived from the CLUSTAL alignment is shown in Figure 3.6. This tree with 1000 bootstrap replicates yielded 6 well supported clusters with high bootstrap scores (Figure 3.6). Four of the 6 clusters were supported by 100 % bootstrap values; cluster I was supported by a 99 % score and cluster IV by a 93 %. Although the topologies of all trees are not identical, the clusters formed in each one of the eight analyses contain the same members and there is a clear separation in the tree between two superfamilies: the 19-glycosyl hydrolases (clusters I, IV and VI) and the 18 glycosyl hydrolases (clusters II, III and V) (Appendix 7.4). The glycosyl hydrolases are a group of proteins that catalyze the cleavage of *O*-glycosyl linkages. There is over 80 families of these enzymes (Henrissat and Davies 2000), and families 18 and 19 are distinguished by

their difference in their catalysis mechanism. Family 18 performs substrate assisted catalysis, while family 19 uses acid catalysis.

The number of clusters and the cluster structure (members of each group) is the same as that which has been previously published (Xu *et al.*, 2007). Cluster I contains members of traditional chitinase classes I, II and VI; cluster II forms a new uncharacterized group; clusters III, IV and V correspond to the same respective traditional classes; finally cluster VI corresponds to the traditional class VII chitinase family. The traditional chitinase classes were classified according to function and primary structure (Hamel *et al.* 1997; reviewed in Kasprzewska 2003). As in the study by Xu *et al.* (1997), here we try to make a correspondence of such traditional classes to the sequence-based clusters in the tree.

Chitinase-like sequences from GCAT were distributed along 5 of the 6 clusters formed: 13 chitinases in cluster IV, 9 chitinases in cluster I, 2 chitinases in cluster VI, 4 chitinases in cluster 5 and 3 chitinases in cluster II. Cluster III was the only cluster that did not include spruce chitinases. Additional previously characterized chitinases from white spruce (AAA85364) and Norway spruce (AAQ17050), which are involved in embryogenesis (Dong and Dunstan 1997; Wiweger *et al.* 2003) are also members of cluster IV.

The three non-redundant chitinase-like sequences found to be differentially up-regulated after 10wk SD all belong to cluster I in the phylogenetic tree, which also includes chitinases involved in response to pathogens in *Solanum tuberosum* (potato) (Buchter *et al.* 1997) and *Capsicum annum* (pepper) (Hong *et al.* 2000), and chitinases that respond to cold in *Secale cereale* (rye) (Antikainen *et al.* 1996; Griffith *et al.* 1997; Hon *et al.* 1995; Pihakaski-Maunsbach *et al.* 2001; Yeh *et al.* 2000; Yu and Griffith 1999).

### 3.5.2 General characteristics and domains of the chitinase proteins

In silico analyses were conducted on the deduced amino acid sequences of the three chitinase-like sequences arising from the proteomics study to further characterize

these proteins. The main properties of the putative chitinase translation products are summarized in Table 3.5. The proteins range from 262 to 301 amino acids and their molecular weights are close to 30 kDa which is very similar to the MW of the band that appeared after 25d SD in the one dimensional SDS-PAGE analysis (Figure 3.1). Their isoelectric point is close to neutral but PgGQ0134.B7.1_C02:53:Contig1:1-1185:+ has a pI more in the alkaline side.

The three spruce chitinase-like sequences were aligned to a tobacco traditional Class I chitinase sequence (CAA34812) using the CLUSTAL algorithm (Figure 3.7). The reason for this was that the tobacco sequence was previously characterized as having a carboxy-terminal vacuolar signal (Neuhaus *et al.* 1991). This signal is considered to be necessary and sufficient to target the chitinase to a vacuole, which has been shown to be a characteristic of many woody plant VSPs (Stepien *et al.* 1994). Additionally, the tobacco chitinase contains all the possible domains present in traditional class I chitinases, which makes comparison of domains in the spruce chitinase-like sequences much easier.

The alignment was manually edited in the signal peptide region to match the beginning of the sequences (Figure 3.7). The two most similar sequences from the alignment were Pg_GQ03602.B7_D17:5-:Contig3:1-897:+ and Pg_GQ03602.B7_D17:53:Contig1:1- with an identity score of 95 % and a score of 97 % for conserved amino acids (amino acids that are not equal but share physio-chemical properties) (Table 3.6). Both of these sequences have similar levels of identity and conservation to Pg_GQ0134.B7.1_C02:53:Contig1:1-1185:+ (close to 70 % and 80 % respectively). As would be expected, the levels of identity and similarity of the three spruce sequences with the tobacco Class I chitinase were always lower than between spruce chitinase-like sequences (Table 3.6).

Domains in the three chitinase-like sequences were identified using SignalP, PROSITE and Interpro, and are illustrated in Figure 3.7. SignalP (http://www.cbs.dtu.dk/services/SignalP/) was used to identify putative signal peptides in the chitinase sequences. The signal peptides that were identified, which suggest processing via the endoplasmic reticulum, cover the same region in the three spruce chitinases-like sequences but differ in the tobacco sequence. The signal peptides detected showed high confidence for the three spruce sequences (Appendix 7.5).
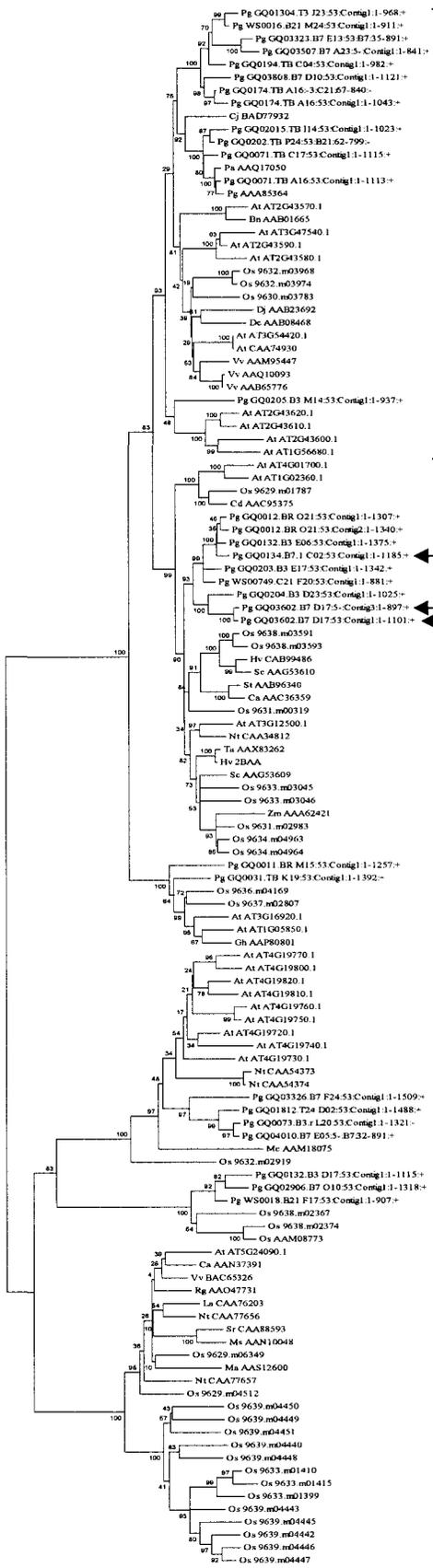
Figure 3.6. Phylogeny of plant chitinases using NJ and 1000 bootstrap replicates. Thirty one non-redundant chitinase-like sequences from GCAT were aligned to 90 previously characterized chitinase and chitinase-like sequences form diverse plant species using CLUSTAL (Thompson *et al.* 1994) and a phylogenetic tree was constructed using the NJ algorithm included in the MEGA 4.0 package (Tamura *et al.* 2007). Arrows show the position of the three non-redundant chitinase-like sequences from white spruce.

Table 3.5 General analysis of the 3 non-redundant chitinase-like sequences up-regulated after 10wk SD.

Analysis	PgGQ0134.B7.1_C02: 53:Contig1:1-1185:+	PgGQ03602.B7_D17:5 -:Contig3:1-897:+	PgGQ03602.B7_D17:5 3:Contig1:1-1101:+
Length	301 aa	262 aa	262 aa
Molecular Weight	32301.50 Da	28758.86 Da	28609.65 Da
Isoelectric Point	8.03	5.97	6.36
Charge at pH 7	1.94	-2.01	-1.02

Interestingly, a PROSITE analysis (Bairoch 1991, http://www.expasy.ch/prosite/) for chitin binding domains revealed that the chitin binding domain essential for substrate binding in chitinase proteins is not present in any of the three white spruce sequences but is represented in the tobacco sequence by a cysteine rich region (Figure 3.7). The chitinase catalytic domain was annotated using the pFAM algorithm (Sonhammer *et al.* 1997) in Interpro (Apweiler *et al.* 2001, http://www.ebi.ac.uk/interpro/) (Appendix 7.6). This domain is present in all sequences, but there is a 14 amino acid deletion in two of the spruce sequences which has been previously reported for some members of the traditional Class II chitinases (Hamel *et al.* 1997). An additional carboxy-terminal extension with no real consensus sequences (Neuhaus *et al.* 1991) is present in the tobacco sequence. This short sequence of 7 amino acids is said to be necessary and sufficient for vacuolar targeting (Neuhaus *et al.* 1991). The spruce chitinase-like sequence PgGQ0134.B7.1_C02:53:Contig1:1-1185:+ also appears to have an extension, but does not exhibit conserved amino acids in this region when compared to the tobacco sequence. However, this region seems to have many hydrophobic amino acids (underlined in Figure 3.7). WoLF PSORT analysis (Horton *et al.* 2007, http://wolfpsort.org/) showed at least 1 association of each one of the spruce chitinase-like sequences with other proteins that may be directed to the vacuole, but a direct vacuolar signal was not found (Appendix 7.7).

Since glycosylation is one of many modifications that proteins may experience, and addition of sugars can change protein isoforms affecting their pI and MW, we also investigated these characteristics because of the number of redundant hits found to chitinases. I was not able to find any potential glycosylation sites on asparagines by the prediction performed by PROSITE, but a number of N-acetylglucosamine potential glycosylation sites are found when using YingOYang (Gupta 2001, http://www.cbs.dtu.dk/services/YinOYang/) (Appendix 7.8). For the YingOYang prediction, a hit was considered strong when the calculated value was over the two possible thresholds with a good separation from the upper threshold, intermediate when it was still over to the two thresholds but closer to the upper threshold and low when it was only over the lower threshold. The predictions (Appendix 7.8) show one site with a strong probability of being glycosylated, two sites with an intermediate probability and

seven sites with a low probability for the spruce chitinase-like sequence: PgGQ0134.B7.1_C02:53:Contig1:1-1185:+; only one site with an intermediate probability and one with a low probability of being glycosylated for PgGQ03602.B7_D17:5-:Contig3:1-897:+; and two sites with intermediate probability and two with a low probability of glycosylation for PgGQ03602.B7_D17:53:Contig1:1-1101:+.

3.5.3 Tertiary structure characteristics

To identify possible structural and potential functional similarity we performed homology modeling of the three non-redundant chitinase-like sequences (Pg_GQ0134.B7.1_C02:53:Contig1:1-1185:+, Pg_GQ03602.B7_D17:5-:Contig3:1-897:+ and Pg_GQ03602.B7_D17:53:Contig1:1-1101:+) along with two rye chitinases having antifreeze properties (AAG53609, AAG53610), and a chitinase from alfalfa and chitinase-like protein from banana that have been classified as having a storage function (AAN10048, AAS12600).

The models obtained using swiss-model (Arnold et al. 2006) and displayed in Raswin/Rasmol (Sayle and Milner-White 1995) are shown in Figure 3.8. The chitinases from cluster I (spruce and rye) show a catalytic cleft (arrows in Figure 3.8) similar to the one reported for a barley endochitinase, which is also displayed in the same figure as the comparison template (Holm and Sander 1994). This cleft contains a buried catalytic glutamic acid. The external extended loop seen in Pg_GQ0134.B7.1_C02:53:Contig1:1-1185:+, AGG53609 and 2BAA (boxes in Figure 3.8) represents a region where a conserved tryptophan binds the substrate and that is absent in some class II chitinases, like the two Pg_GQ03602.B7_D17 sequences (Figure 3.7) but not in all since the barley endochitinase is classified as a class II chitinase.

As seen from the figure all the proteins belonging to cluster I (spruce chitinases and rye chitinases) showed very similar structures with a high content of α-helices, while the two chitinases with characterized storage properties from *Medicago sativa* and *Musa sp.* belonging to cluster III have a different shape with an (α/β) eight barrel fold.

Table 3.6 Percentage scores for identity (first value in each cell) and conserved aminoacids (second value in each cell) for the pairwise alignments of Figure 3.7.

	PgGQ03602.B7_ D17:5:Contig3:1 -897:+	PgGQ03602.B7_ D17:53:Contig1: 1-1101:+	PgGQ0134.B7.1_ C02:53:Contig1: 1-1185:+	NtCAA34812
PgGQ03602.B7_ D17:5:Contig3:1 -897:+		95 %/97 %	68 %/80 %	57 %/73 %
PgGQ03602.B7_ D17:53:Contig1: 1-1101:+			68 %/79 %	58 %/72 %
PgGQ0134.B7.1_ C02:53:Contig1: 1-1185:+				58 %/75 %
NtCAA34812				

```
                            1                                      40
PgGQ03602.B7_D17:5     (1) M■■■■■■L■■■■■■■■S■■--------
PgGQ03602.B7_D17:53    (1) M■■■■■■L■■■■■■■■S■■--------
PgGQ0134.B7.1_C02:53   (1) MK■RFSL■V■ALV-■AT■NFY--■S■E■C■QA■L
NtCAA34812             (1) MRLCKFTA■SS■LFSL■LSASAEQCGSQAG■CP■C

                           41                                     80
PgGQ03602.B7_D17:5    (33) --------------------------------■
PgGQ03602.B7_D17:53   (33) --------------------------------■
PgGQ0134.B7.1_C02:53  (38) ■G------------------------VQG■
NtCAA34812            (41) KFGWCGNTNDYCGPGNCQSQCPGGPTPTPPTPP■GGD■

                           81                                    120
PgGQ03602.B7_D17:5    (34) ■F■I■D■F■LKHRND■C■KGFY■Y■FI■AA■FP
PgGQ03602.B7_D17:53   (34) ■F■I■D■F■R■LKHRND■C■KGFY■Y■FI■AA■FP
PgGQ0134.B7.1_C02:53  (46) ■I■F■LLKHRND■VCK■KGFY■YA■FIAAADAFP
NtCAA34812            (81) ■I■S■FDQMLKHRNDN■CQ■KGFY■Y■FINAARSFP

                           121                                   160
PgGQ03602.B7_D17:5    (74) ■FGN■G■E■KRE■AAFF■QTSHE┃TTGGW■TAPDGPYAW
PgGQ03602.B7_D17:53   (74) ■FGN■G■E■KRE■AAFF■QTSHE┃TTGGW■TAPDGPYAW
PgGQ0134.B7.1_C02:53  (86) GFG■TGD■QKRE■AAFF■QTSHE■TTGGWPTAPDGPYAW
NtCAA34812           (121) GFG■SGDT■A■KRE■AAFF■QTSHE┃TTGGWATAPDGPYAW

                           161                                   200
PgGQ03602.B7_D17:5   (114) GYC■■EQD■D■---------┃----E■Y■GRGP■Q■
PgGQ03602.B7_D17:53  (114) GYC■■EQD■D■---------┃----E■Y■GRGP■Q■
PgGQ0134.B7.1_C02:53 (126) GYC■EQ■N■G■QA■SE■S■KRY■GRGP■Q■W
NtCAA34812           (161) GYC■LREQ■G-■TP■GQ■P■RKY■GRGP■Q■H

                           201                                   240
PgGQ03602.B7_D17:5   (140) K┃NYEA■G■A■G■D■NNPN■VA■■SFK■A■WFWMN
PgGQ03602.B7_D17:53  (140) KSNYEA■G■A■G■D■NNPG■VA■■SFK■A■WFWM■T
PgGQ0134.B7.1_C02:53 (166) ■NY■G■A■G■DG■NNP■VANDA■SFK■A■WFWM■A
NtCAA34812           (200) ■NY■CG■A■GVD■NNP■VA■D■V■SFK■A■WFWM■

                           241                                   280
PgGQ03602.B7_D17:5   (180) QPPKPSSHIV■G■W■PS■D■AA■R■G■GV■TNIINGG
PgGQ03602.B7_D17:53  (180) Q■PKPSPHTV■G■W■PS■D■AA■R■G■GV■TNIINGG
PgGQ0134.B7.1_C02:53 (206) Q■PKPS■HNV■AGGWGPSS■D■AA■RAAG■GV■TNIINGG
NtCAA34812           (240) Q■PKPS■HDV■G■WQPS■GDRAANR■G■GV■TNIINGG

                           281                                   320
PgGQ03602.B7_D17:5   (220) ■ECG■G■DSN■DRIGFY■YC■ILGAS■G■N■DC■Q■
PgGQ03602.B7_D17:53  (220) ■ECG■G■DSN■DRIGFY■AYC■ILGAS■G■N■DC■Q■
PgGQ0134.B7.1_C02:53 (246) ■ECG■G■DS■DRIGFY■YC■ILG■S■G■N■DC■Q■P
NtCAA34812           (280) ■ECG■G■DS■VQDRIGFY■YCSILG■SPGDN■DCGNQ

                           321              336
PgGQ03602.B7_D17:5   (260) FG■-------------
PgGQ03602.B7_D17:53  (260) FG■-------------
PgGQ0134.B7.1_C02:53 (286) FGF■AQSQPRLIKTVV
NtCAA34812           (320) FGN■LLVDTM------
```

96

Figure 3.7 Alignment of upregulated chitinase-like sequences to a tobacco class I chitinase. The spruce chitinase-like sequences were aligned to a tobacco sequence, since this sequence has all possible features of a traditional class I chitinase, plus a carboxy terminal extension which has been linked to vacuolar localization (Neuhaus *et al.* 1991). Identical amino acids (yellow background), conservative (blue background), block of similar (green background), weekly similar (green foreground). Block lines: blue (signal peptide, upper bar for spruce chitinase-like sequences and lower for tobacco), red (chitin binding domain), black (catalytic domain), green (carboxy terminal extension). Two rectangular boxes enclose sites for: a conserved glutamic acid (E) which is said to be the catalytic amino acid; a tryptophan (W) involved in substrate binding; and a variable tyrosine (Y) – serine (S) – asparagine (N) involved in substrate binding. Strongly hydrophobic amino acids in the CTE are underlined. Sequence names of GCAT chitinase-like sequences were trimmed.

To support these qualitative similarities and differences the structures of all these chitinases (with the exception of the barley chitinase which was only used to illustrate a model chitinase) were aligned and the Root Mean Square Distance (RMSD) was calculated for every pairwise structure (Table 3.7).

The RMSD scores showed that in general all the evaluated proteins from cluster I (spruce and rye proteins) are very close structurally, with distances of no more than 1.37 Å and as small as 0.03 between two of the spruce chitinases. The distances between proteins of cluster I and proteins of cluster III (alfalfa and banana) were higher than 13.89 and as high as 17.34, while the 2 proteins of cluster III had a distance of 1.27 between them. Pg_GQ0134.B7.1_C02:53:Contig1:1-1185:+, was apparently more closely related at the structural level to the two rye antifreeze proteins (1.08 and 0.54 Å) than to the two other spruce chitinases (1.31 Å to each). This fact is interesting since in the phylogenetic tree (Figure 3.7) all the spruce chitinases are more closely related to each other than to other chitinases in cluster I. Spruce chitinases Pg_GQ03602.B7_D17:5-:Contig3:1-897:+ and Pg_GQ03602.B7_D17:53:Contig1:1-1101:+ are also the most similar pair of proteins (0.03 Å), and have also high similarity to the rye antifreeze proteins (ranging from 0.86 to 1.37 Å). However, precaution has to be taken in the assumptions of the RMSD values since the process of homology modeling using just one template structure per model can result in very similar structures. This is a result of the energy minimization process not moving the models far apart from the template (Warren Gallin, personal communication).

The proteins structures were used for prediction of ice binding surfaces using AFPredictor (Doxey *et al.* 2006). As expected, the chitinase proteins with storage function from cluster III showed no evidence of ordered surface carbons (OSCs) that could potentially bind to an ice surface (Table 3.8). The three spruce chitinase models showed potential OSCs (Table 3.8 and Figure 3.9) and consequently also values for fraction of surface area (FSA) covered by the OSC arrangement and total surface area (TSA) in Ångströms (Table 3.8). Surprisingly, one of the two rye antifreeze proteins did not predict OSCs. All of the proteins predicting ice binding surfaces had four OSCs with the exception of Pg_GQ0134.B7.1_C02:53:Contig1:1-1185:+ which predicted a number of six. The proteins where OSCs were observed (Figure 3.9) show spatially regular surface carbons in relatively planar surfaces, with undulations between them.
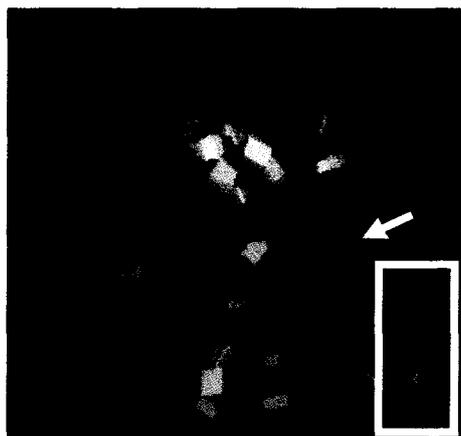
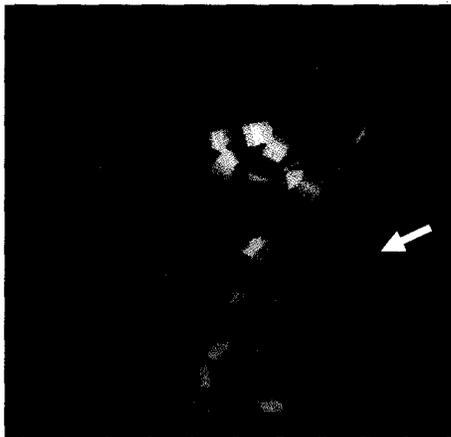Pg_GQ0134.B7.1_C02:53:Contig1:1-1185:+
Spruce chitinase-like

Pg_GQ03602.B7_D17:5-:Contig3:1-897:+
Spruce chitinase-like

Pg_GQ03602.B7_D17:53:Contig1:1-1101:+
Spruce chitinase-like

AAG53609
Rye chitinase (antifreeze protein)

AAG53610
Rye chitinase (antifreeze protein)

AAN10048
Alfalfa chitinase (storage protein)

AAS12600
Banana chitinase-like (storage protein)
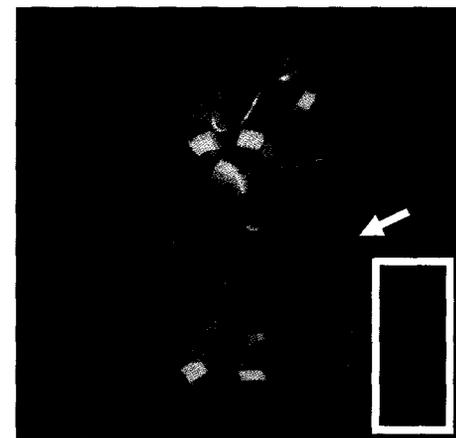
2BAA
Barley chitinase (catalytic protein)

Figure 3.8. Three dimensional structures of chitinases. Models of the three non-redundant spruce chitinases-like sequences (Pg_GQ0134.B7.1_C02:53:Contig1:1-1185:+, Pg_GQ03602.B7_D17:5-:Contig3:1-897:+, Pg_GQ03602.B7_D17:53:Contig1:1-1101:+), two chitinases from rye with antifreeze properties (AAG53609, AAG53610), 2 chitinases with storage properties from alfalfa (AAN10048) and Banana (AAS12600), and a crystal structure model of a chitinase from barley (2BAA). Arrows indicate the catalytic cleft, and boxes outline the outside loop with a conserved tryptophan (W) for substrate binding. The structures correspond to the ribbons graphics for tertiary structures from Raswin/Rasmol (Sayle and Milner-White 1995).

Table 3.7 Root Mean Square (in **Ångströms**) distances of aligned 3D structures. The distances were calculated by performing all pairwise structural alignments using the pairwise structure comparison tool (SSAP) from the CATH database for protein structure classification (Orengo *et al.* 1997, **http://www.cathdb.info/**).

	Pg_GQ013 4.B7.1_C02 :53:Contig 1:1-1185:+	Pg_GQ036 02.B7_D17: 5- :Contig3:1-897:+	Pg_GQ036 02.B7_D17: 53:Contig1 :1-1101:+	AAG53609	AAG53610	AAN10048	AAS12600
Pg_GQ013 4.B7.1_C02 :53:Contig 1:1-1185:+		1.31	1.31	1.08	0.54	17.14	17.1
Pg_GQ036 02.B7_D17: 5- :Contig3:1-897:+			0.03	0.88	1.37	17.34	16.96
Pg_GQ036 02.B7_D17: 53:Contig1 :1-1101:+				0.86	1.37	13.89	15.86
AAG53609					0.92	14.44	17.17
AAG53610						15.76	16.23
AAN10048							1.27
AAS12600							

Table 3.8 Prediction of Ordered Surface Carbons (OSCs), Total Surface Area (TSA) and Fraction of Surface Area (FSA), for the modeled protein structures.

PROTEIN	OSCs	TSA	FSA
Pg_GQ0134.B7.1_C02:53:Contig1:1-1185:+	6	214.532	0.020
Pg_GQ03602.B7_D17:5-:Contig3:1-897:+	4	104.518	0.010
Pg_GQ03602.B7_D17:53:Contig1:1-1101:+	4	104.573	0.010
AAG53609	0	0.000	0.000
AAG53610	4	152.838	0.015
AAN10048	0	0.000	0.000
AAS12600	0	0.000	0.000
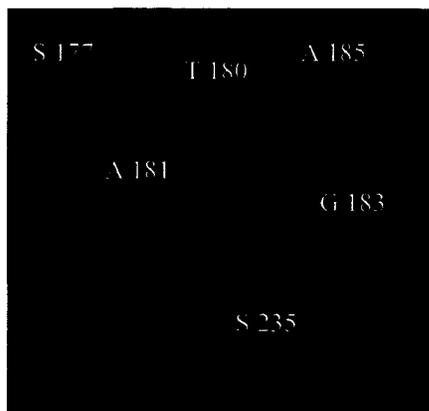
Pg_GQ0134.B7.1_C02:53:Contig1:1-1185:+
Spruce chitinase-like
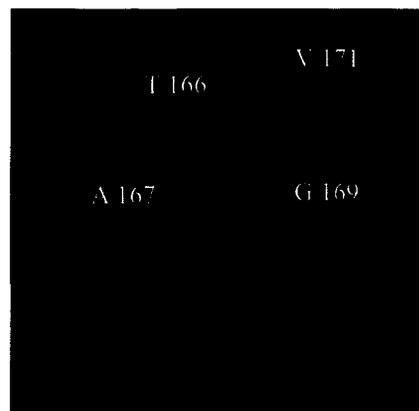
Pg_GQ03602.B7_D17:5-:Contig3:1-897:+
Spruce chitinase-like
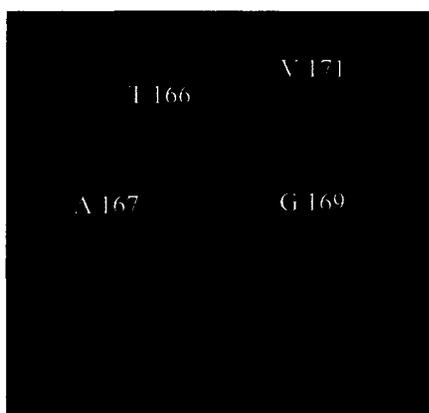
Pg_GQ03602.B7_D17:53:Contig1:1-1101:+
Spruce chitinase-like
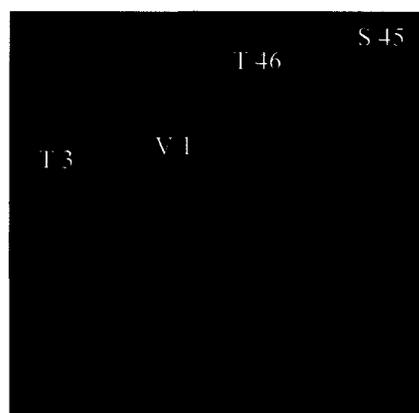
AAG53610
Rye chitinase (antifreeze protein)

Figure 3.9 Surface structure of proteins with predicted ice binding surfaces. The amino acids that were found to have OSCs are shown in each image in red accompanied by the on letter code for the amino acid and its position in the trimmed protein linear chain (from the catalytic domain downstream. Appendix 7.9).

4.0 DISCUSSION

4.1 Differential protein expression in 2D gels

The absolute volume value for each spot normalized to the standard deviations of all volume values per gel were used to do quantitative comparisons using Significance Analysis of Microarrays (SAM). Volume was chosen over other possible quantitative measures offered by the software (Image Master™ 2D Platinum) such as intensity since it is less biased by artifacts created by staining or capturing. Additionally since it used the upper 75% of the 3D image of the spot it eliminated the background that can surround the spot (Image Master™ 2D Platinum Software, user manual for version 5.0, SIB, 2003). SAM analysis was chosen because although student t-tests are often used for proteome 2D gel electrophoresis analysis (Lippert *et al.* 2005, 2007; Li *et al.* 2008; Romero-Puertas *et al.* 2008; Wang *et al.* 2008), the probability of obtaining false positives increases with the number of analyzed pairs of proteins (Type I error) which for microarray analysis and proteomic analysis tends to be high (Tusher *et al.* 2001). SAM uses a modified t-test to assign a score of differential expression for each spot based on a standard deviation of repeated measurements for the values of the analyzed gene (Tusher *et al.* 2001). The False Discovery Rate (FDR) which is analogous to the α value for t-tests is calculated globally, but a specific FDR is obtained for each gene (similar to a p value) allowing the assessment of the level of significance for each gene.

Statistical analysis with SAM found that most of the differentially expressed proteins were accumulated to greater levels in the second condition (10wk SD); 89% and 79% of the differentially expressed proteins using FDRs of 5.2 % and 12.5 % respectively, were up-regulated after 10wk SD treatment. This is in contrast with a published transcriptomics study of cambial zone tissue during dormancy in poplar, where a larger number of transcripts are found in active cambium when compared to dormant tissues (Schrader *et al.* 2004). However, the same study highlighted the fact that dormant cambium generally can also have a high number of transcripts due to a higher redundancy of gene products like VSPs. The cessation of growth during dormancy does not mean a

cessation in the metabolic apparatus of the plant; in fact Schrader *et al.* (2004) demonstrated that even some cell cycle genes show no transcript down-regulation upon dormancy acquisition, and may be involved in keeping a basal metabolism for rapid reactivation upon spring regrowth. The activation of genes involved in growth and hormone control, responsiveness to stress and genes involved in changes of carbon fixation and utilization are a common feature during the transition to dormancy as supported by studies in proteomics, transcriptomics, physiology and metabolomics (Druart *et al.* 2007; Jeknic and Chen, 1999; Morin *et al.* 2007; Park *et al.* 2008; Ruttink *et al.* 2007; Schrader *et al.* 2004). Therefore an upregulation of relatively abundant gene products, such as what is generally detected using proteomics analysis, is not unreasonable.

4.2 Mass spectrometry analysis

Although we received good relative abundance signals (10^6 or 10^7) for the peptide profile outputs from all the analyzed samples, the overall success rate of protein identification from the peptide profiles was less than 50%. Several factors likely contributed to this relatively low success rate. (1) Not all currently available spruce ESTs had been deposited into NCBI's dbEST at the time that the MASCOT analyses were conducted, including those ESTs obtained from three dormancy-related libraries. It has been reported that protein identification is greatly increased when using additional ESTs libraries corresponding to the same species and tissues of study (Lippert *et al.* 2005, 2007). At the time of analysis, approximately 450,000 *Picea* spp. ESTs were available in dbEST; currently, there are approximately 530,000 *Picea* spp. ESTs in dbEST. (2) EST data is fragmentary and represented sequences may not coincide with over represented peptides from mass spectrometry analysis. (3) Losses by partial digestion with trypsin or low recovery from the gel plugs caused by competitive binding and ionization introduced during the two dimension separation (Ross *et al.* 2002) can cause additional failure for identification. Although some information can be obtained from mass spectrometric analyses to get more accurate identifications for our kind of data, additional techniques

like using different enzymes or remeasuring the spectrum with different parameters (Lahm and Langen 2000), could be used. However doing multiple analyses for the large number of proteins that we had can be difficult due to the lack of automation.

4.3 Functional categorization of the differentially expressed proteins

Several classification and annotation bioinformatic tools were used to infer function of the 122 significantly differentially expressed proteins that I was able to annotate using MASCOT. Most of these 122 annotated proteins were upregulated after 10wk SD. As such, the changes discussed here represent proteins that showed increased abundance after 10wk SD unless stated otherwise.

4.3.1 Cell cycle

Modulation of the cell cycle has been demonstrated to be a major process implicated in dormancy acquisition (Schrader *et al.* 2004; Druart *et al.* 2007; Ruttink *et al.* 2008). We identified proteins that corresponded to genes associated with the cell cycle. Similarities were found to a cdc 10 dependent transcript (*AtCDT1a*) from *A. thaliana*, a minichromosome maintenance protein (*MCM*-like protein), and a putative DNA gyrase subunit with domains of topoisomerase. It is interesting that the first two proteins (*CDT1* and *MCM*-like protein) are part of the origin recognition complex (ORC) that binds DNA for replication forming the pre-replication complex during G1 (Nishitani and Lygerou 2002). These proteins have been shown to increase their expression during the G1 phase in different organisms before replication happens in the S phase (Nishitani and Lygerou 2002). It has been shown that during autumn, the cambial cells of balsam fir (*Abies balsamea*) are arrested in G1 phase (Mellerowicz *et al.* 1989), and this would mean that proteins expressed during this time could show differential expression, since these proteins are not expressed during mitotic activity. In fact the *AtCDT* protein from *Arabidopsis thaliana* should be degraded by targeting of cyclin dependent kinases at the end of DNA replication to avoid multiple replication cycles before mitosis (Castellano *et*

al, 2004; Raynaud *et al.* 2005); this *Arabidopsis* protein has also been shown to be involved in plastid division (Raynaud *et al.* 2005). CDT proteins have a central role in licensing chromatin for replication, in fact, CDT along with Cdc6/18 are in charge of helping to load the MCM protein which is itself a helicase (Nishitani and Lygerou 2002).

4.3.2 Regulatory proteins

Several proteins with putative regulatory functions were identified. A protein similar to a high mobility group protein (*HMG-1*) from *C. gladiata* was detected. This protein was inferred to be implicated in the regulation of a gene encoding a seed storage protein (Yamamoto and Minamikawa 1998). High mobility group proteins are involved in regulation of transcription and in *Arabidopsis* the analysis of this group of proteins showed that the proteins associate to chromatin in interphase cells but not in mitotic cells (Launholt *et al.* 2006). Two high mobility group proteins were also reported after four weeks of short days towards bud formation and dormancy transition in poplar (Ruttink *et al.* 2007). Since the high mobility group proteins seem to have a structural and regulatory function, their activation could be involved in changes to reprogram the transcriptome. Likewise I saw the accumulation of a protein similar to a carbon catabolite repressor (CCR4). In *S. cerevisae,* this protein is involved in the transcriptional and structural chromatin control of alcohol dehydrogenase (*ADHII)* and has been inferred to be involved in the transcription of other genes (Denis and Malvar 1990). The massive changes in the cambial transcriptome of poplar through dormancy acquisition have been said to be due in part to extensive chromatin remodeling (Druart *et al.* 2007), and the presence of these controllers of transcription and chromatin could be related to the same modulations.

A nascent polypeptide associated complex (NAC) domain-containing protein was also among the upregulated transcriptional regulators. In *Arabidopsis thaliana* roots, two NAC domain containing proteins were upregulated upon salt stress (Yuanqing *et al.* 2007) which is congruent with a stress response regulation in our case. The NAC has been shown to associate with the nascent regions of proteins that emerge from the

ribosome, and to exert a protective function to avoid premature interaction with other proteins in the cytosol (Wang *et al.* 1995). NAC also ensures co-translational targeting to the cytosolic space for polypeptides with out a signal peptide (Möller *et al.* 1998).

An SPL12 (SQUAMOSA promoter-binding like protein) had an 8.3-fold change after 10wk SD. Two SQUAMOSA promoter binding protein-like genes are also strongly regulated upon short day induction in poplar buds (Ruttink *et al.* 2007). The SQUAMOSA binding proteins were first discovered as transcriptional regulators of early floral development in snapdragon (*Antirrhinum majus*), and their activation precedes that of the SQUAMOSA gene which is involved in flower development (Klein *et al.* 1996). However, genes with homologous domains and similar binding properties as SPL12 have been also found to respond to micronutrient and oxygen status in *Chlamydomonas* (Kropat *et al.* 2005). It seems interesting that recently more attention has been given to floral-development related genes which also seem to be related to growth cessation (Bohlenius *et al.* 2006) due in part to their control by photoperiod. Another upregulated protein also related to flowering was an FCA protein. The FCA gene is involved in flowering control, but *fca* mutants showed that the flowering control of FCA is constitutive and autonomous, since the mutation can be suppressed by vernalization or incandescent light (Martinez-Zapater and Somerville 1990). When active, FCA represses FLC (Flowering Locus C), which acts as a repressor of floral induction (reviewed in Eckardt *et al.* 2002). Interestingly FCA has been now characterized as an ABA receptor, and the interaction of ABA with FCA disrupts this gene's interaction with the 3' processing factor FY, which is necessary for FLC repression (Razem *et al.* 2006). Since ABA has been shown to act downstream of short photoperiod during transition to dormancy (Druart *et al.* 2007; Ruttink *et al.* 2007), the potential control of FCA by ABA seems interesting.

We also found a protein similar to NBS resistance-like protein from sweet cherry (*Prunnus avium*) that presented a 1.3-fold change; the NBS containing proteins have been related to pathogen recognition and signal transduction. An NBS-LRR protein was identified in Norway spruce upon pathogen infection (Jøhnk *et al.* 2005), and an NBS disease resistance protein was found to increase transcript expression in late winter in eastern cottonwood (Park *et al.* 2008). The expression of all these stress-response

proteins shows a preparation of the plant for abiotic and biotic stresses over winter when the anabolic metabolism is depressed.

Finally, a speckle-type POZ protein was also upregulated upon short day induction. In Arabidopsis, these proteins associate with cullin proteins to build functional E3 ligase complexes for substrate protein ubiquitination (Weber *et al.* 2005). Ubiquitin tagging is essential for protein degradation via the proteosome. Apparently the POZ proteins are involved in substrate recruitment to the ligase complex and the large number of homo and heterodimers that they can form allows for diverse substrate specificity (Weber *et al.* 2005). In human cells a cullin-POZ complex is responsible for ubiquitination of phosphatidylinositol phosphatase kinases (Bunce *et al.* 2008), which is interesting in this case since we found a phosphatidylinositol 3-4 kinase family protein downregulated after 10wk SD. In eukaryotic cells, phosphatidylinositols and its phosphorilated derivatives, the phosphoinositides, have been shown to be involved in diverse processes including growth, differentiation, cytoskeletal arrangement and membrane trafficking; and phospatidylinositol kinases and phosphatase kinases are involved in the synthesis of phosphoinositides (reviewed in Odorizzi *et al.* 2000). Additional similarities to a protein kinase 2 from black poplar (*Populus nigra*) and a receptor like kinase of liverwort (*Marchantia poymorpha*) were also found among our regulated proteins, but these classifications are too broad to be related to a specific process.

4.3.3 Response to stress

GO categories of responses to different stimuli (stress, biotic, abiotic, defense, immune and death) had many common responses with cold acclimation. It has been shown that not only low temperature but short photoperiod alone induces the cold acclimation in temperate trees like hybrid aspen (Welling *et al.* 2002). Therefore the upregulation of some of these proteins would be a consequent response in terms of the transition to winter and also of the stress response.

Chaperonins were classified in this group of responses to stress and along with the protein disulfide isomerase (PDI), they may be involved in avoidance of protein aggregation and cold denaturation of tertiary structure, correct protein folding and structure stabilization (Boston *et al.* 1996). While chaperones are especially important in preventing aggregation and promoting renaturation of aggregated proteins upon stress, PDI is classified as a foldase and acts as a catalyst of disulfide bonds in nascent polypeptides. Chaperones on the other hand do not always act as catalysts, instead binding to proteins that are under an unstable, non-native structural state (reviewed in Boston *et al.* 1996). Interestingly, although PDI is known as a thiol-oxidoreductase that catalyzes disulfide bond formation in the ER, in soybean some PDI family members (*GmPDIL-1 and GMPDIL-2*) may act as chaperones of seed storage proteins (Kamauchi *et al.* 2008). When in concentration excess, PDI can stop lysozyme protein aggregation, acting therefore as a chaperone (Puig and Gilbert 1994). Interestingly, the heat shock protein (HSC71.0) that belongs to the HSP70 group (another group of chaperones) is down regulated (spot 930 on Table 3.2). Heat shock proteins from this group are amongst the most conserved proteins in plants and are fundamental to many cellular processes through folding of nascent polypeptides (Boston *et al.* 1996). However HSC71.0 is not responsive to heat stress and seems to be constitutively expressed in pea leaves (DeRocher and Vierling 1995), so the protein does not have to be necessarily upregulated upon stress.

In addition to these protective proteins, I also found other proteins that respond to abiotic or general stresses. For example, several genes associated with abiotic stress responses are upregulated in early winter in eastern cottonwood (Park *et al.* 2008), and in European aspen, cambial zone cells show upregulation of oxidative stress response genes as cold hardiness develops in autumn (Druart *et al.* 2007). A thioredoxin-dependent peroxidase had almost a 3 fold change after 10wk SD in our experiment. This type of enzyme with similarity to plant 2Cys-peroxiredoxins was isolated in *C. reinhardtii*, and demonstrated to act as an antioxidant agent against reactive oxygen species (ROS) and to protect DNA against ROS induced degradation (Goyer *et al.* 2002). The activation of antioxidant genes may help plants to cope with ROS that result from several stresses during winter including cold or drought which generally cause an upregulation of reactive

oxygen scavenging enzymes. The upregulation of enzymes can also occur as a consequence of the over-excitation of the photosystem II (reviewed in Apel and Hirt, 2004). In general, the photo-inhibitory effect is a result of high excess of incident light with respect to the processing capacity of the photosystem, but a reduction of the photosynthetic antenna during cold acclimation can have similar effects (reviewed in Hüner et al. 1998). Although ROS have been historically classified as being detrimental for cell, and said to be a result of biotic and abiotic stresses, a new role in signaling and general metabolism has emerged (Apel and Hirt 2004; Foyer and Noctor 2005); for example in Arabidopsis, hydrogen peroxide activates the expression of several MAP kinases resulting in tolerance to freezing and cold (reviewed in Apel and Hirt 2004). A second role for the upregulated thioredoxin-dependent peroxidase could be in lignification. Peroxidases catalyze the formation of the monolignols radicals which allows subsequent polymerization of lignin. The presence of peroxidases associated with the developmental process of lignification has been previously characterized for Norway spruce (Koutaniemi et al. 2007), and increases in peroxidase activity have also been seen in autumn and winter in Scots pine, Norway spruce and silver birch (Marjamaa et al. 2003). Likewise, there is significant lignification that takes place in woody stems during the final stages of growth before the onset of dormancy (Janice Cooke, personal comment).

Additional toxic products can also arise from primary metabolism. The upregulation of lactoylglutathione lyase (glyoxylase I) suggests an additional detoxification mechanism. Methylglyoxal is one of these cytotoxic byproducts and when accumulated in cells, it can cause problems like increased protein degradation through amino acid modification and impaired proliferation. Glyoxylase along with reduced glutathione can transform this toxic product into lactate. Methylglyoxal was shown to be overproduced in shoots and roots of two rice varieties upon cold, drought and cold stress, and was defined as a general abiotic response product (Yadav et al. 2005). Furthermore, in the same study, tobacco plants transformed with antisense glyoxylase I, showed more than 100% methyglyoxal production than wild type plants or glyoxylase overexpressors. This shows that this protein may be acting in a similar way in dormancy acquisition

associated with overwintering, where there is both cold acclimation and acquisition of desiccation tolerance.

A protein with similarity to a strawberry (*Fragaria* x *ananassa*) quinone oxidoreductase seems to be part of a genereral abiotic response too. Some of its lower GO terms show responsiveness to cold (Appendix 7.2). In general quinone oxidoreducateses are associated with the reduction of quinones. In strawberry, the expression of this protein is associated with ripening (Raab *et al.* 2006). However, oxidoreductases are also classified as being involved in oxidative phosphorylation (Table 3.3), and as part of the general electron transport chain, its presence could be related to adjustments in energy status.

On the side of responses to biotic stimulus, I found a protein with more than 6 fold change having similarity to a thaumatin-like protein from *Cryptomeria japonica*. Thaumatin-like proteins are classified as pathogenesis-related (PR) proteins that are considered to be activated as a general defense response in plants. Thaumatin-like proteins make up the PR-5 family (reviewed in Anzlovar and Dermastia 2003). Thaumatin-like proteins from *C. japonica* are apparently activated in response to ROS (Futamura *et al.* 2005). Other thaumatins can also act as osmotins, responding to osmotic stress, water stress and cold (reviewed in Anzlovar and Dermastia 2003). In fact, a thaumatin/osmotin-like protein exhibited a 63-fold change increase in dormant cambium of poplar compared to actively growing cambium (Schrader *et al.* 2004). In this study the expression of the protein was speculated to be a response for survival at low temperatures and therefore the osmotin could be acting as a cryoprotectant or antifreeze protein. In winter rye, a characterization of polypetides with antifreeze activity showed that some of them had similarities to thaumatin-like proteins (Hon *et al.* 1995; Griffith *et al.* 1997).

Chitinase-like sequences, which were identified multiple times in this study, are also associated with stress responses. Nine spots in the gel were identified as chitinase-like sequences and each one exhibited a greater than 10-fold change after 10wk SD treatment. Chitinases catalyze the cleavage of β-1,4- glycosidic bonds in polymers of N-acetylglucosamine, which are integral part of chitin, a molecule that is common of fungal cell walls, bacteria and insects. Additional substrate include lipochitooligosachrides (Nod factors) produced by symbiotic bacteria, bacterial peptidoglycans and arabinogalactan

proteins (AGPs) from plants (reviewed in Kasprzewska 2003). Therefore chitinases have been often associated with plant defense (reviewed in Kasprzewska 2003) but more recently, chitinase-like proteins have been also shown to respond to a diverse array of influences and processes including embryogenesis, chilling and frost resistance, programmed cell death, nodulation, mycorrhizal associations, and development and growth processes (reviewed in Kasprzewska 2003). Chitinase proteins have also been characterized as vegetative storage proteins in alfalfa (Avice et al., 2003) and a chitinase-like protein also has a storage role in banana (Peumans *et al.* 2002). In spruce species, chitinases have been found to respond to some of these influences. In Norway spruce, expression of a Class IV chitinase has been related to signaling for transition from proembryogenic masses to somatic embryos (Wiweger *et al.* 2003) and a class IV chitinase also seems to be related to embryogenesis events in white spruce (Dong and Dunstan 1997). Class IV and class II chitinases have been shown to respond to pathogen attack in Norway spruce (Hietala *et al.* 2004; Jøhnk *et al.* 2005), in conjunction with peroxidases (Nagy *et al.* 2004). In Norway spruce, two chitinases cause a decrease in the elicitor levels from ectomycorrhizal fungi to adjust their symbiotic relationships (Salzer *et al.* 1997). From these studies it is apparent that the general functions of chitinases also apply to the *Picea* genus. A more in-depth analysis of the role of our expressed chitinase-like sequences is given in Section 4.4 below since these proteins represented the most apparent changes in the proteome of secondary stems between 0d SD and 10wk SD treatment.

4.3.4 Primary metabolism and energy status

Several proteins were identified that encode enzymes involved in primary metabolism of carbon compounds, including enzymes involved in carbon fixation, the pentose phosphate pathway (PPP), sugar metabolism, glycolysis and gluconeogenesis (Table 3.3). We identified three enzymes related to carbon fixation (sedoheptulose-bisphosphatase, ribose-5 phosphate isomerase and Rubisco), for this latter enzyme, I found the large subunit upregulated and the small subunit downregulated. We also saw a

protein from the photosystem I (P700 chlorophyll a-apoprotein 84 KD protein), and multiple spots that annotated as ATP synthase, required for proton gradient generation of ATP. After an initial depression of the photosynthetic apparatus upon cold acclimation, a need for increased sucrose synthesis is required and therefore fixation of carbon increases again to maintain a balance (reviewed in Stitt and Hurry 2002). In the transition to dormancy, plants start cold acclimating and one important change is the accumulation of sugars like sucrose that along with other sugars may serve as osmoprotectants, to protect specific cellular compartment or as regulators of volume during cycles of dehydration and rehydration (reviewed in Stitt and Hurry 2002). Metabolite data from dormant cambial zone tissue in poplar showed increased levels of sucrose along with other sugars (Druart *et al.* 2002) and in oaks increasing levels of soluble carbohydrates are evident from October to January (Morin *et al.* 2007). We only found one enzyme (PFKb-type carbohydrate kinase, a fructokinase) which could be involved in sucrose or starch metabolism, but this enzyme is downregulated at 10wk SD (Table 3.2). While sucrose synthesis is important, it seems like starch breakdown also plays an important role in the transition to dormancy to produce carbon sources for different metabolic processes (Druart *et al.* 2007; Schrader *et al.* 2004). The need for starch breakdown seems to come from an apparent decline in photosynthesis going in to autumn (Keskitalo *et al.* 2005). This seems to argue against the up-regulation of the carbon fixation apparatus discussed before. However, the study by Keskitalo *et al.* (2005) was performed in leaves where a decline of the whole photosynthetic machinery is concomitant with the process of senescence; while in stems, the photosynthetic apparatus goes through a series of cycles to adjust to new conditions..

An additional interesting fact is the upregulation of the large Rubisco subunit (Table 3.2). This protein was not only classified under the biosynthetic category but also as responsive to abiotic stress according to the GO terms (Appendix 7.2). Rubisco has been considered to be an unconventional vegetative storage protein (reviewed in Cooke and Weih, 2005) because it is synthesized in amounts far in excess over that needed for enzymatic activity, which is thought to serve an additional function than just carbon fixation. Additionally Rubisco breaks down during senescence, providing carbon and nitrogen sources for future growth.

Along with these changes, I found multiple hits to enolase which could be involved in glycolysis or gluconeogenesis. In the study by Druart *et al.* (2007), the increased abundance of transcripts corresponding to enzymes from glycolysis was said to aid in the increased demand for energy and carbon skeletons during the transition to dormancy. It is interesting that enolase is also activated by light stimulus (lower level GO terms in Appendix 7.2), which could indicate activation via a short day photoperiod. Enolase could also be involved along with alcohol dehydrogenase in the anaerobic degradation of sucrose. ADH has been shown to be upregulated in late winter and was speculated to respond to possible anoxic conditions due to ice-encasement (Park *et al.* 2008).

The presence of enzymes related to the pyruvate and fatty acid metabolism (acetyl-CoA carboxylase, acyl-coenzyme A oxidase II) could be related to the need to restructure the membrane lipid composition for overwintering and to synthesize new vacuoles for storage of proteins, starch and lipids. New vacuolar membranes are needed once the main vacuole disintegrates and multiple vacuoles appear accompanied by abundant lipid droplets (Farrar and Evert 1997). Many of these multiple vacuoles that appear in the fusiform and ray cells of the cambium of *Robinia pseudoacacia* contain electron dense molecules that give positive reaction for proteins with protein binding dyes; protein bodies of the same kind also appear in poplar (Sauter and van Cleve 1994). These vacuoles have been shown to accumulate vegetative storage proteins that function as nitrogen and carbon reserves over winter in different perennial species (Clausen and Apel 1991; Sauter *et al.* 1989; reviewed in Stepien *et al.* 1994). Besides the formation of vacuolar bodies (oleosomes) that appear overwinter in *R. pseudoacacia*, but also in poplar ray cells, an additional need for fatty acid biosynthesis could come from the increase in the amount of endoplasmic reticulum (Sauter and van Cleve 1994). Indeed, sometimes this lipid accumulation is accompanied by an increase in free glycerol, which could be linked to cell freezing protection.

Two enzymes from the pentose phosphate pathway (ribose 5-P isomerase and phophogluconolactonase) were also up-regulated after 10wk SD. In the cambial meristems of poplar the pentose phosphate pathway enzymes have been related to the generation of reducing power for different metabolic processes, but this occurs in the

early phase of cambial reactivation and not in the actual transition to dormancy (Druart *et al.* 2007). However, in the study of Park *et al.* (2008), genes from the pentose phosphate pathway are upregulated during early winter, and their presence was related to the possibility that the pathway provides the reducing power necessary for lipid desaturation so the membrane can maintain fluidity during winter. A failure to maintain this condition results in structural disruption that affects chlorophyll losses in the chloroplast and proton trafficking in the mitochondria among other membrane related processes (reviewed in Stitt and Hurry 2002). So, an activation of this pathway could provide both reducing power for basal processes and maintenance of membrane properties.

Another change that can be detected is a shift in energy status. Multiple hits to ATP synthase complex enzymes and to NADH dehydrogenase complex indicated probable activity in the mitochondria thorough oxidative phosphorylation (Table 3.3), although additional functions of ATP synthases are also coupled to the chloroplast membranes (as mentioned before) and to the plasma membrane for nutrient and sugar mobilization (Alves *et al.* 2007). The increased demand of energy can be a consequence of the necessary investment required for the transition to dormancy. New cryoprotectants, stress response proteins and synthesis of soluble sugars, or even as mentioned before, the adjustments in carbon fixation resulting from photosynthetic changes, are needed to cope with changing environmental conditions. Several studies showed an increase in the machinery of this pathway upon stress or upon increases in metabolizable sugars (e.g. González *et al.* 2007). However, the same study showed that increases in oxidative phosphorylation have been seen upon dormancy release due to the needs of cell division and growth until the full photosynthetic apparatus is reestablished. ATP synthase and NADH dehydrogenase are also classified in the category of transport since its coupling to hydrogen and electron transport allow for the energy processes to take place.

Some of the proteins from Table 3.3 were classified by KEGG as being involved in amino acid metabolism or protein synthesis (asparaginyl-trna synthetase, serine acetyltransferase, methionine synthase); and some of this were also part of the nitrogen compound metabolic process category. The amino acid pool is essential for the protein synthesis that takes place in the transition to dormancy due to the metabolic shift that accompanies the expression of genes to withstand the winter conditions. Free amino acids

including serine, aspartate and glutamate can be detected in autumn and winter in the cambial zone of aspen (Druart *et al.* 2007) and similar changes are detected in *A. thaliana* as a result of cold treatments (Cook *et al.* 2004). Amino acids like glutamine/glutamate, asparagine/aspartate are important for nitrogen recycling in conifers (reviewed in Canton *et al.* 2005) and as a starting point for the synthesis of other essential amino acids (reviewed in Canovas *et al.* 2007). The enzymes of amino acid metabolism found in our study map to eight possible amino acid related pathways, but most of them only have one or two hit enzymes, so it is difficult to make inferences on their possible implications.

4.4 Characterization of spruce chitinase-like sequences

Multiple chitinase-like sequences were identified in this study, and were amongst the proteins showing the greatest fold-change in abundance between 0d SD and 10wk SD, demonstrating that these proteins likely play an important role or roles in the transition to dormancy. One of the objectives of this study was to identify proteins that could potentially serve as VSPs; that is proteins that are accumulated when resources are available and broken down during a subsequent period when resources are limiting. The accumulation pattern, structural characteristics and the fact that no other protein that was annotated in this study represented a logical candidate, led me to speculate that the chitinase-like sequences found could function as vegetative storage proteins (VSPs). These constitute a group of diverse proteins recruited from different evolutionary origins that can be characterized by a role of storage rather than an enzymatic function (Coleman 2004), but that in recent studies have been shown to sometimes retain an enzymatic function. Chitinases have been shown to function as VSPs in alfalfa (Meuriot et al., 2004) and banana (Peumans *et al.* 2002). Consequently, I decided to focus on further characterization of the three chitinase-like sequences identified in this study.

In this study, chitinase-like sequences were the most distinct group of up-regulated proteins after 10wk SD; in fact the large fold changes observed in seven of the nine chitinase-like proteins identified by LC-MS/MS (Table 3.2) indicated that these proteins showed an absence-presence expression pattern between the active and dormant

stems, and that this pattern was correlated with the shift to short day conditions. Although seasonal VSPs from trees accumulate in response to the transition to winter (reviewed in Stepien *et al.* 1994), photoperiod seems to be sufficient to activate accumulation of these proteins (Coleman *et al.* 1992; Zhu and Coleman, 2001a, 2001b).

The nine spots in 2D gels with similarity to chitinases constituted an average of around 4% of the total normalized volume of all protein spots in the gels. In a classic definition of storage protein, the more traditional seed reserve proteins, which are localized in storage vacuoles, constitute at least 5% of the total protein, being stored for nutritional needs in a later stage they usually lack any other metabolic or structural role (Staswick, 1994); these characteristics are similar to most of the patterns followed by VSPs. Accumulation for characterized VSPs however vary from 7 to 15% in the transition to winter for interior spruce shoots (Roberts *et al.* 1990), 30% in the bark tissue of *Salix microstachya* over winter (Wetzel and Greenwood 1991), and as much as 62% after a 24 short day induction in bark of poplar (Coleman *et al.* 1991). Although these chitinase-like proteins do not account for such high percentages, they represented the most significant protein accumulation observed in my study. Additionally the 10wk SD time point only represents transition to dormancy, rather than fully dormant trees, so further accumulation could be possible.

4.4.1 Chitinases phylogeny

To further infer possible functions of these chitinase-like sequences in spruce during dormancy acquisition, I constructed a phylogenetic tree with chitinases and chitinase-like sequences from other species and all available chitinase-like sequences from the GCAT spruce database.

To confirm correct clustering of the sequences, in addition to a bootstrap analysis with 1000 replicates, we used two different alignment algorithms (Clustal and MUSCLE), and four phylogenetic algorithms (NJ, MP, ME, UPGMA). While Clustal (Thompson *et al.* 1994) is widely used because of its speed and consistent clustering mainly when aligned sequences have medium to high levels of similarity, MUSCLE

118

(Edgar 2004) uses a similar progressive alignment as CLUSTAL, but the distance between pairwise sequences is calculated using k-mers (similar subsequences), resulting in an improved alignment of blocks of high similarity. MUSCLE was used because it achieves accuracy comparable to other alignment software like T-Coffee or MAFFT, and seems to be faster (Edgar 2004). The phylogenetic algorithms were chosen because they are the four possible analysis to be performed using MEGA4 (Kumar *et al.* 2008; Tamura *et al.* 2007). Each of the algorithms clusters sequences using a slightly different principle covering either distance or character based methods. Although UPGMA is considered to not be as reliable because a constant rate of evolution is assumed and the distance to the ancestral nodes for any pair of sequences is apparently equal (Kumar *et al.* 2004), it was also tested since the objective was to determine if these methods produced robust clusters.

All of the analyses showed consistent clustering from the evaluated sequences (Appendix 7.4), demonstrating that the clusters are reliable. All the main clusters from the CLUSTAL/NJ analysis had bootstrap values over 93% and most of them were supported by a 100% score (Figure 3.6), giving additional support to the correct clustering. The rest of the analysis done either with CLUSTAL or MUSCLE in combination with the four phylogenetic algorithms yielded clusters with close to 90% or over 90% support. The exception may be the analyses of MP where clusters I and IV still grouped together the same chitinase members, but the clusters had lower bootstrap support (Appendix 7.4). This could have something to do with the MP algorithm, which usually is able to find more than one optimal tree, and therefore this could decrease the support for one type of clustering. The clustering with MP can be further complicated since there is no defined outgroup. Attempts to find an enzymatic outgroup (other glycosyl-hydrolases), or a chitinase from another taxonomic group (algae, fungi) were performed (not shown). However since the diversity of glycosyl hydrolases is so huge (Henrissat and Davies 2000) it was difficult to select an outgroup using this approach. Algal chitinases got incorporated into clusters, and although fungal chitinases were positioned as outgroups for the two big families of 19-glycosyl hydrolases and 18-glycosyl-hydrolases, a clade analysis would have required separating the clusters from each family. This ancestry for the two glycosyl hydrolase families is in agreement with

the divergence of the chitinases, which not only precedes the division between angiosperms and gymnosperms, but also the division of fungi and plants (Hamel *et al.* 1997).

Since our objective in performing this phylogenetic analysis was to make inferences about protein function based on relationships to functionally characterized chitinases from other species, the most important factor was to have consistency in our clusters. The fact that the tree topologies resemble the topology of the previously published chitinase tree (Xu *et al.* 2007), even with the addition of sequences from other plants and the white spruce putative chitinases, shows that the clusters are robust.

Traditionally, chitinases have been grouped in seven classes according to structural and functional characteristics (reviewed Kasprzewska 2003), and though initially plant chitinases were classified as having a central role against pathogen attack, more recent research shows that some chitinases are involved in responses to other stresses as well as growth and developmental processes (reviewed Kasprzewska 2003). In the recent study of Xu *et al.* (2007), most of the proteins from each of the traditional classes grouped together into corresponding clusters in their phylogenetic analysis, with the exception of traditional classes I, II and VI which were clustered together and an additional group designated as cluster II in the tree. Since my tree conserved the tree topology, my groups are consistent with this idea. The fact that traditional classes I, II and VI are clustered in one group in my tree is consistent with their close functional relation. The chitinases from these classes along with Class IV and VII chitinases belong to the PR-3 class of proteins. Classes IV and VII which in my tree correspond to clusters IV and VI are still in close proximity to the members of its PR subfamily in the tree (Figure 3.6). In fact all the upper part of the tree that includes all the aforementioned families corresponds to the 19-glycosyl hydrolase family of proteins while the lower part having clusters II, III and V, correspond to the 18-glycosyl hydrolases. This shows a congruent evolutionary-functional trend since the mechanisms of catalysis of these two enzyme super families are divergent (reviewed in Kasprzewska 2003), and is in agreement with my attempts to root the chitinases with fungal chitinases (discussed above).

Although class IV chitinases seem to be structurally more similar to class I chitinases than class II members, since they share a signal peptide, a chitin binding domain or cysteine-rich domain (CRD) and a catalytic domain (Hamel *et al.* 1997), while class II sequences lack the CRD, the overall sequence similarity of class IV members to class I and class II members is only around 50% (Beintema 1994; Hamel *et al.* 1997) while members of class I and II can share over 60% similarity. Evidently the similarity between class I and IV is also lower in this study, since the clusters were clearly separated. The presence of class I and II members in the same cluster and of a separated class IV as in my tree has also been reported before (Hamel *et al.* 1997), and has been attributed to amino acid changing substitutions which in consequence alter the primary structure of the members of this class.

In the lower part of the tree corresponding to the 18-glycosyl-hydrolases we found four white spruce putative chitinases in cluster V (traditional class V), three members in the newly formed cluster II, which still seems closely related to cluster V, and no white spruce chitinases in cluster III (traditional class III chitinases). As mentioned before, the traditional class III and V chitinases belong to different PR families, and a distinct glycosylase family. These groups seem to have some relationship with exochitinases from prokaryotes (reviewed in Beintema 1994), which shows that they could have diverged even before the divergence of prokariotes and eukaryotes, or at least from a different ancestral sequence (Hamel *et al.* 1997). Structurally, class III chitinases are very different from other plant chitinases, possessing a (β/α) eight barrel topology which they share with other proteins that do not posses chitinase activity (Hamel *et al.* 1997; van Schelting *et al.* 1996). The fact that some of the enzymes from this glycosyl hydrolase group, like concanavalin, do not have enzymatic activity has to do with changes in the catalytic amino acids and in the external loops involved in substrate binding. However the absence of traditional class III chitinases from white spruce does not mean absence of these sequences in the genome, since the chitinase survey was made based on specific EST libraries.

4.4.2 Chitinase domains

In addition to the expression patterns of our nine chitinase-like spots in 2D-gels, the three non-redundant chitinase-like sequences characterized from these spots presented a molecular weight (MW) of approximately 30 kDa and an isolectric point (pI) in the range of 6 to 8 (Table 3.5). These MW corresponds to the most prominent band observed in the one-dimensional SDS-PAGE time course analysis (Figure 3.1). This 30 kD band seemed to appear after 25d SD and stay until the end of the time course, which is also consistent with the expression patterns of other VSPs like BSP from poplar, which is 32 kDa and starts accumulating after 10 days of short photoperiod (Coleman *et al.* 1991). These patterns were also seen in the 30 and 27 kDa VSPs from interior spruce which are not detected at the beginning of autumn, but accumulate 20 days after (Roberts *et al.* 1991). Many VSPs are close to a MW of 30 kDa, but the range of classified VSPs varies from larger proteins like the 94 kDa lipoxygenase from soybean (Tranberger *et al.* 1991), to smaller proteins of around 16 kDa found in peach (Gomez and Faurobert 2002) and lectins of black mulberry (Van Damme *et al.* 2002).

The fact that we found that the nine proteins spots in the gels corresponded to three non-redundant chitinase-like sequences could be due to potential post-translational modifications. For example many of the VSPs from diverse plants like soybean and poplar have been described as being glycoproteins, and this modification is thought to provide thermostability during winter (reviewed in Stepien *et al.* 1994). From my PROSITE analysis I found that the spruce chitinase-like sequences did not have potential asparagines to be glycosylated. I also used YingOYang to predict O-N-acetylglucosamine glycosylation sites. From the results we can see that PgGQ0134.B7.1_C02:53:Contig1:1-1185:+ appears to be the only chitinase to have a considerable number of putative sites for glycosylation. The other two chitinase-like sequences from spruce also have putative sites but they are fewer and predicted with less confidence (Appendix 7.8). Although these predictions are based on neural networks, they are only probabilities of occurrences and would have to be tested to be confirmed.

The three non-redundant chitinase-like sequences corresponding to the upregulated proteins identified in this study were grouped in cluster I, and therefore they

122

were aligned to a tobacco class I chitinase (CAA34812) that was first characterized as having a short c-terminal vacuolar localization signal at the carboxylic terminal region (Neuhaus *et al.* 1991). The sequence was selected for the alignment because it belonged to a group of previously characterized class I chitinases having all the potential domains characterized for chitinases (Hamel *et al.* 1997). Thus, the alignment combined with additional data mining tools should reveal the basic primary features of putative functionality. As expected, the chitinase-like sequences from spruce exhibited the highest sequence similarity to each other rather than to the tobacco chitinase. The two chitinase-like sequences designated with the GCAT cluster identifier GQ03602.B7_D17 were closely related at the amino acid level (Table 3.6), but using a 95% nucleotide cutoff filter we distinguish them as being different chitinases. Interestingly the identity of these two sequences to the other spruce chitinase-like sequence is only close to 70%, which could mean the presence of either many substitutions or some structural differences. In fact many evident sequence changes can be easily detected by looking at the alignment form Figure 3.7. First the region corresponding to the signal peptide has many substitutions between PgGQ0134.B7.1_C02:53:Contig1:1- and the other two spruce chitinase-like sequences; these substitutions should not affect the functionality of the signal peptides which largely depend of the biochemical properties of the amino acids, so different amino acids with similar properties can serve the same function. Second, there is a longer extension between the signal peptide of PgGQ0134.B7.1_C02:53:Contig1:1-1185 and its catalytic domain as compared to the connecting region of the other two spruce chitinase-like sequences (Figure 3.7); this region serves no apparent function but may be residual from the evolutionary deletion of the chitin binding domain. Third, a 14 codon deletion is evident in the catalytic region for the GQ03602.B7_D17 sequences; this deletion has been reported for sequences of tomato (*Lycopersicon esculentum*), tobacco (*Nicotiana tabacum*) and garden petunia (*Petunia hybrida*) classified as traditional class II chitinases (Hamel *et al.* 1997), but not all class II chitinases have this deletion. Finally there is carboxy-terminal extension present in PgGQ0134.B7.1_C02:53:Contig1:1-1185 and lacking in the other two spruce sequence. Besides these differences accounting for the low similarity there are substitutions also in the more conserved catalytic domain, although here the conservation of the three sequences is higher.

Hamel *et al.* (1997), classified class II chitinases as sequences lacking a cysteine-rich domain (CRD) that corresponds to the chitin binding domain (CBD), and a carboxy-terminal extension (CTE) with or without the 14 codon deletion. The three white spruce sequences lack the CRD but PgGQ0134.B7.1_C02:53:Contig1:1-1185:+ does apparently have CTE (Figure 3.7) which is a feature of only class I chitinases (reviewed in Beintema 1994; Hamel *et al.* 1997).

Searching for a possible VSP function, I analyzed the main domains in these proteins. VSPs themselves have been found in special protein bodies (vacuoles) in the woody tissues of perennials (reviewed in Stepien *et al.* 1994). In the ray cells of poplar, immunogold labeled protein antibodies were localized to vacuolar bodies (Sauter *et al.* 1989). These vacuoles lost most of the electron dense material over spring, showing the seasonal accumulation of the protein bodies. These results show that at least some VSPs accumulate in vacuoles, and therefore some signal for vacuolar localization may be indicative of VSP function.

The first requirement for vacuolar localization is to follow the endoplasmic reticulum (ER) path which can guide proteins to an extracellular destination or towards specialized vacuoles (Chrispeels and Raikhel 1992). I used Signal P to assess the probability that the three chitinase-like sequences were targeted according to this pathway. The three sequences had a highly reliable annotation for a signal peptide in the N-terminal region of the proteins (Appendix 7.5). It has been shown that not all chitinases have a strong signal peptide (Xu *et al.* 2007). However, since VSPs tend to be localized in vacuoles the presence of a signal peptide in the chitinase-like sequences found in my study shows that they could be putatively directed through the endoplasmic reticulum to vacuoles.

For many class I chitinases, the signal peptide is followed by a cysteine-rich region referred to as the chitin binding domain (Hamel *et al.* 1997) or hevein domain (Beintema 1994). In tobacco when the domain was deleted from a class I chitinase the affinity for the substrate was decreased and catalytic activity was lowered by three-fold when compared to the protein that had the chitin binding domain (Iseli *et al.* 1993). Also in tobacco, the specific activities of class I chitinases were 10 to 15 fold higher than those of class II chitinases, and these latter proteins fail to inhibit fungal growth as opposed to

their class I counterparts which posses the domain (Sela-Buurlage *et al.* 1993). This evidence demonstrates that although the absence of a chitin binding domain does not completely suppress enzymatic activity, it does reduce it; in the strict sense a VSP is a protein that has lost enzymatic activity and therefore the white spruce chitinase products would likely have reduced enzymatic activity due to the lack of this domain. Interestingly, there seems to be a trend towards the deletion of domains supported by phylogenetic studies and by the presence of repetitive sequences on the ends of these regions (Hamel *et al.* 1997).

Following the chitin binding domain, the next conserved region that was found in all the aligned proteins was the catalytic domain (Figure 3.7). This section, which spans most of the protein, is responsible for the catalysis of the glycosidic bond in polymers of N-acetylglucosamine. The domain forms a catalytic cleft which has been previously observed and characterized as having the necessary amino acids for substrate binding and cleavage of glycosidic bonds (Hart *et al.* 1995; Holm and Sander 1994). This catalytic cleft is evident in the homology models generated from the proteins of cluster I when compared to the crystallized barley chitinase (Figure 3.8).

Other interesting features of the aligned proteins (Figure 3.7) are their catalytic and substrate binding amino acids. Holm and Sander (1994) made a comparison of lysozyme structures across species to find the residues involved in catalysis. An invariant glutamic acid was shown to be the main catalytic residue; this residue is conserved in all the aligned sequences (first box in Figure 3.7 alignment). A second residue thought to be important for substrate binding, a tryptophan (Hart *et al.* 1995), is present only in the PgGQ0134.B7.1_C02:53:Contig1:1-1185:+ and in the tobacco sequence. The reason for this is that in the other two spruce sequences, this region corresponds to the 14 amino acid deletion typical of some plant chitinases (Hamel *et al.* 1997). In the model of PgGQ0134.B7.1_C02:53:Contig1:1-1185:+ this region corresponds to an outside loop (box in model of Figure 3.8) which is at the entrance of the catalytic cleft. The tryptophan of this loop is apparently important for substrate binding and is present in the barley class II endochitinase (2BAA) as Trp103, in the homologous loop region of the crystal structure (Figure 3.8).

A final important residue involved in substrate binding is the tyrosine located in position 202 of the tobacco sequence (Figure 3.7). The Tyr 202 is also present in the PgGQ0134.B7.1_C02:53:Contig1:1-1185:+ sequence but is absent from the other two spruce chitinase-like sequence where it is replaced by either serine or asparagine. This pattern is congruent to what has been found in other class II chitinases (Hamel *et al.* 1997), and it has been speculated to be an additional reason for the decreased activity that proteins with mutations in that residue have, when compared to proteins with the Tyrosine (Verburg *et al.* 1993).

The last feature is the presence of a carboxy-terminal extension (CTE), typical of class I chitinases which are targeted to the vacuole (Hamel *et al.* 1997). This feature is present in the tobacco sequence and also in PgGQ0134.B7.1_C02:53:Contig1:1-1185:+. Wood tobacco (*Nicotiana sylvestris*) plants transformed with a chitinase carrying the carboxy-terminal amino acids, or with a chimeric cucumber chitinase with the carboxy-terminal amino acids from the tobacco sequence, showed that the carboxy extension was necessary to obtain increased enzymatic activity in vacuoles (Neuhaus *et al.* 1991). Plants expressing the enzyme variants without the extension exhibited enzyme targetting to the intercellular spaces. The CTE is therefore a localization signal that seems necessary and sufficient for vacuole targeting. However, there is no apparent similarity in carboxy-terminal vacuolar localization signals between the tobacco protein and other proteins such as the barley lectin or the *Ipomoea* sporamin (reviewed in Chrispeels and Raikhel, 1992; Neuhaus *et al.* 1991), but the extensions seem to be rich in hydrophobic amino acids. One of our white spruce chitinase-like sequences shows a carboxy-terminal extension (Figure 3.7), which could be a vacuolar localization signal; seven of the 13 amino acids are hydrophobic in the spruce sequence, while four out of seven have this characteristic in the tobacco sequence. As mentioned before, plant VSPs are usually stored in vacuoles (Stepien *et al.* 1994). In seed tissues up to 50% of synthesized proteins accumulate in vacuoles as storage proteins, and stress signals can trigger the accumulation of different kinds of proteins in these seed vacuoles, including glucanases and chitinases (Chrispeels and Raikhel 1992). The absence of CTE domains in most chitinases classes is thought to be related to the need for apoplastic chitinases that are a first line of defense against pathogens (Hamel *et al.* 1997). If there are sufficient amounts

of apoplastic chitinases to cover this apoplastic function, the preservation of a CTE to be targeted to the vacuole could argue for a different function.

It is interesting that although PgGQ0134.B7.1_C02:53:Contig1:1-1185:+ lacks the CRD and, therefore, should be classified as a class II chitinase according to Hamel *et al.* (1997), the rest of its features resemble the tobacco class I chitinase (Figure 3.7) and in general the features seem to be closer to a class I vacuolar chitinase. The catalytic domain has no major deletions and has conserved amino acids for susbstrate binding and catalysis (described above), and the presence of a CTE at the end of the sequence, resemble more the description of a class I chitinase that a class II.

4.4.3 Possible roles for white spruce chitinases

Members of cluster I (Figure 3.6) have been characterized as being responsive to different elicitors. While chitinase proteins like the *Solanum tuberosum* (AAB96340) or the *Capsicum annum* (AAC36359) seem to be responsive mainly to pathogens (Buchter *et al.* 1997; Hong *et al.* 2000), there are also a number of genes related to cold response in this cluster. For example, the transcript levels of the *Cynodon* chitinase (AAC95375) increased upon cold acclimation in crown tissues of two plant varieties (de los Reyes *et al.* 2001), and two well characterized rye chitinases (AAG53609 and AAG53610) have antifreeze activity (Antikainen *et al.* 1996; Griffith *et al.* 1997; Hon *et al.* 1995; Pihakaski-Maunsbach *et al.* 2001; Yeh *et al.* 2000; Yu and Griffith 1999). However, to my knowledge, there are no chitinases in cluster I which has been reported as having a vegetative storage function. Chitinases from banana (Peumans *et al.* 2002) and from alfalfa (Meuriot *et al.* 2004) that seem to behave as VSPs belong to cluster III (class III) in my phylogenetic analysis (Figure 3.6).

Since I characterized all the domains and catalytic units that may have a role in defense, I also searched for similarities between the three spruce chitinase-like sequences and chitinases having antifreeze activity from rye, as well as the two chitinases from cluster III which have a vegetative storage function. Since the actual function of a protein is closely related to the structural characteristics, and since there is little similarity of

cluster III chitinases (class III) to all the classes of the 19-glycosyl hydrolase chitinases (classes I, II, IV, VI and VII), I decided to examine the structural characteristics of these proteins by building 3D models using homology modeling. Since the actual function of a protein is performed after protein processing, we trimmed the signal peptides for all the sequences which are typically cut from the proteins in the ER (Chrispeels and Raikhel 1992). We also eliminated the remaining regions upstream from the catalytic domains, since these regions are difficult to resolve in homology modeling because of the high variability between sequences. For modeling, we used the top model result from the BLAST search for each protein, along with 9 additional proteins chosen with levels of similarity from 60 to close to 100%, thus guaranteeing that the models are based on true alignments of the most conserved regions, and that models are not biased towards the comparison to only one specific model from the protein data bank (PDB).

From Figure 3.8, it is evident that all the chitinases from cluster I have a very similar structure showing the catalytic cleft, thus resembling the structure of previously resolved chitinases (Holm and Sander 1994; Huet *et al.* 2006; Ubhayasekera *et al.* 2007) and the crystal structure of the barley chitinase (Figure 3.8). The structures present a high content of α-helices (approximately ten), which is typical of 19-glycosyl hydrolases like the barley endochitinases (Hart *et al.* 1995) and the papaya chitinase (Huet *et al.* 2006, 2008). At the same time the only major evident change in cluster I chitinases is an external loop present in PgGQ0134.B7.1_C02:53:Contig1:1-1185:+, AAG53609 and 2BAA (boxes Figure 3.8). As mentioned before, this loop at the entrance of the catalytic cleft has a conserved tryptophan (W) that seems to be involved in substrate binding.

Since class III chitinases appear to be derived from an ancestor that is different from chitinases of classes I, II and IV (Hamel *et al.* 1997), they should in consequence be quite different. The homology models obtained for the alfalfa and banana proteins show that the two chitinases from class/cluster III have a completely different tertiary structure compared to the cluster I chitinases from spruce and rye (Figure 3.8). Hamel *et al.* (1997) classified the chitinases according to their domains structure, and showed that class III chitinases lack a CRD, a CTE and have low similarity even in the catalytic region when compared to the other three classes. Their (β/α) eight barrel topology, which can be observed in Figure 3.8, is typical of other glycosyl hydrolases without lysozyme or

chitinase activity (Hamel *et al.* 1997). Because of its structural particularities and the way they are expressed they could be involved in processes of pathogen defense, morphogenesis, differentiation and embryogenesis (reviewed in Hamel *et al.* 1997); but they can also have a storage function, as is the case for the two modeled proteins, which would be consistent with a lack of chitinase activity.

To have a quantitative measure of the similarities between the 3D models we calculated the root mean square distance (RMSD) to assess the average distance in Ångströms between structures. While the distances between all the modeled members from cluster I are small (Table 3.7), the distances to the proteins of cluster III are 10 times higher. A comparison of two crystallographic models of the main chains of the pancreatic kallikrein (serine protease) showed that the average RMSD was 0.37Å (reviewed in Bajaj and Blundell, 1984), demonstrating that in fact the distances of our cluster 1 models, which are distinct proteins, are quite small. This is additional support for the phylogenetic separation (Figure 3.6) and all the differences in domains that these two groups have. The differences in structures between members of clusters I and III do not necessarily argue against a possible role in storage of the white spruce chitinase-like proteins of cluster I. As evidenced from the domain analysis, the potential decreased catalytic activity, lack of certain domains or directionality to the vacuoles could favor a storage function without ruling out other possible functions, but these hypotheses have to be tested. Additionally, through evolution the small changes in secondary structure or in the loop regions which are more prone to variation may result in the rise of new functionalities besides the main function of the protein (Bajaj and Blundell, 1984).

According to the RMSD values, PgGQ0134.B7.1_C02:53:Contig1:1-1185:+ seems to be more closely related structurally to the rye antifreeze chitinases than to the other two spruce proteins; although the differences are small, the phylogeny derived from primary sequence data groups all white spruce chitinase-like sequences of the cluster together and the rye sequences are in a different subgroup. It has been long known that conservation of the topological characteristics in the whole protein and mainly in the catalytic regions is associated with conservation of function, and that changes at the amino acid level are less restricted since different amino acid chains can fold into the

same tertiary structure (Bajaj and Blundell 1984). Therefore, tertiary structure is more closely related to function than the primary sequence characteristics.

Chitinases from cluster III (class III), have a very different structure (Figure 3.8) which is typical of the 18-glycosyl hydrolases (van Schelting *et al.* 1996). As mentioned before, some of the enzymes of this group like concanavalin B, which shares sequence similarity with chitinases, do not posses chitinolytic activity (Hennig *et al.* 1995). This lack of enzymatic activity may be due to the change in the catalytic amino acid and a difference in the size of the catalytic cleft and can result in new functions. The induction of the 32 kDa chitinase from alfalfa seems to follow this pattern. As a VSP, the protein accumulates in rye taproots upon treatment with methyl jasmonate which is known to change the nitrogen source-sink relationships (Meuriot *et al.* 2004). An increase in transcript abundance for the respective gene is also evident in the transition to winter, but there seems to be a repression of the accumulation when temperatures decrease, and therefore an additional function such as antifreeze activity can be ruled out (Dhont *et al.* 2006). This enzyme maintains chitinolytic activity, showing that the loss of enzymatic activity is not necessary to comply with a storage function.

Given the structural similarities between the spruce chitinase-like sequences and the chitinases with characterized antifreeze activity, we then turned to analyze a possible role of the white spruce chitinase-like sequences as antifreeze proteins. Antifreeze proteins inhibit the growth of extra-cellular ice and stop recrystallization, which can dehydrate and damage the cells during winter; they do this by binding ice surfaces through van der Waals interactions and hydrogen bonding (Griffith and Yaish 2004). Plant antifreeze proteins have been reported in ferns, gymnosperms and angisoperms, and they can be found almost in every plant structure (reviewed in Griffith and Yaish 2004).

Since there is no apparent consensus sequence for ice binding sequences of plant antifreeze proteins, (Atici and Nalbantoglu 2003; Griffith and Yaish 2004) it is more likely that the tertiary structure has typical binding features which depend on amino acid biochemical characteristics. However, antifreeze proteins are structurally diverse and the study of protein-ice interaction is extremely difficult because the protein can not be crystallized with ice (Doxey 2006; Jia and Davies 2002). Therefore, after modeling our proteins, we used AFPredictor (Doxey *et al.* 2006) to predict potential ice-binding

surfaces, which uses the premise that spatially regular surface atoms have a higher probability of binding a regular surface like ice. As shown in Figure 3.9, the ordered surface carbons (OSCs) of the models have spatial regularity and undulations between them, and they are found in a mostly coplanar surface (Doxey *et al.* 2006). Close contact with the ice is achieved by regular undulations in the ice dock with undulations in the protein surface (Jia and Davies 2002). The hydrophobic or slightly hydrophobic character of the amino acids shown in the predicted ice binding surface (Figure 3.9) agrees with models generated for fish and insect antifreeze proteins (AFPs) (Jia and Davies 2002).

One argument against these proteins being involved in ice binding is that only a small portion of the proteins have predicted OSCs; the fraction of surface area occupied is only 1-2% (Table 3.8). In general, a considerable surface of the protein should be involved in ice binding (Jia and Davies 2002). However, the antifreeze proteins in rye form oligomeric complexes even between different protein types like glucanases and chitinases (Yu and Griffith 1999), and therefore oligomerization could be a mechanism to increase ice binding surface.

Additionally, there were no OSCs predicted in one of the rye antifreeze protein (AAG53609); although this protein has proven antifreeze activity (Yeh *et al.* 2000). The rye proteins could have another or an additional mechanism than OSCs for ice binding (Andrew Doxey, personal communication). However, there does not appear to be a common domain in the two rye chitinases which is distinct to a tobacco chitinase which has no antifreeze activity (Yeh *et al.* 1997). In fact, in the chitinase from carrot, the ice binding surface appears to be mainly hydrophilic (reviewed in Griffith and Yaish 2004). In conclusion, although an ice binding function is possible for the spruce chitinase-like sequences because of their structural characteristics, this hypothesis has to be tested.

5.0 CONCLUSIONS

In this study we investigated the changes in the protein expression profiles between woody stems undergoing active growth and those that were subjected to 10 weeks of short day photoperiod. The experiments with 2D SDS-PAGE showed a big shift in protein abundance reflected in a large number of over-expressed proteins after 10 weeks of short days.

From the functional annotation analysis of proteins we could see up-regulation of stress responsive proteins involved in protein protection, oxidative stress, defense against pests and pathogens, and osmotic protection. A second group was characterized by proteins involved in changes in primary carbon metabolism and energy status representing major pathways such as photosynthesis, the pentose phosphate pathway, glycolysis and fatty acid metabolism, along with enzymes involved in oxidative phosphorylation and ATP synthesis. Finally, proteins were identified that are implicated in amino acid metabolism. All these changes that occur during the transition towards dormancy point to an adjustment of the plant to withstand a period where carbon and nitrogen nutrient resource acquisition is halted, and defenses against biotic and abiotic stresses have to be synthesized ahead to be ready in a period where anabolism may be reduced.

Multiple chitinases were identified in the two dimensional gel analysis; most of these proteins had a MW which was close to a major band of 30 kDa found in 1D PAGE which appeared after 25 days of short days and stayed until the end of the time course but in a lower proportion. The molecular weight and pattern of abundance change between the two conditions prompted me to further examine these proteins as potential VSPs.

The phylogenetic analysis showed that the characterized chitinases-like proteins were found in a cluster that has members of traditional classes I, II and VI, these members are responsive to pathogens and cold acclimation, while previously characterized VSP chitinases belong to the evolutionary divergent class III (cluster III in our tree).

Domain analysis showed that two of our chitinase-like sequences seem to belong to the traditional class II chitinases which lack a vacuolar localization signal and

therefore are likely to be secreted to the apoplast. One of our chitinases however, PgGQ0134.B7.1_C02:53:Contig1:1-1185:+, seemed to lack a chitin binding domain, like class II chitinases, but had a carboxy-terminal extension which could be a vacuolar targeting signal. This sequence seemed to have more features of a class I chitinase (some of which are targeted to the vacuole) than a class II, and therefore could be directed to a vacuole as is the case for the VSPs that have been found in other plants. The chitinase seems to have conservation in critical amino acids found in other chitinases for catalysis, and therefore it could also be performing a defense function like many other chitinases.

Models of the white spruce chitinase show overall structural similarity to antifreeze proteins from rye and quite different structures from the VSP chitinases from banana and alfalfa. A prediction for possible ice binding surfaces showed that the three spruce chitinases had a small subsection which could bind ice, but further investigation in this respect is needed.

To summarize, upon short day treatments in stems we could find shifts in stress, carbon and energy metabolism. The most evident change was the accumulation of a group of chitinases; in silico characterization of these chitinases revealed that they could potentially serve multiple roles such as storage, defense and antifreeze proteins, all of which seem useful for the plant going into winter.

To examine the likelihood of these characteristics several experiments can be conducted:

1. Specific antibodies to chitinases can be used to perform immunodetection in histological preparations to confirm if these proteins are directed to the vacuole and constitute abundant proteins.

2. Chitin binding and catalytic assays of chitinases can be performed by affinity isolation or purification of overexpressed protein, to test the activity of the chitinases on sugar substrates.

3. Antifreeze activity can be assessed by crystal morphology assays, using pure protein extracts.

6.0 REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215: 403-410.

Alves G, Decourteix M, Fleurat-Lessard P, Sakr S, Bonhomme M, Améglio T, Lacointe A, Julien JL, Petel G, Guilliot A. 2007. Spatial activity and expression of plasma membrane H^+ - ATPase in stem xylem of walnut during dormancy and growth resumption. Tree Physiology 27: 1471-1480.

Anonymous. 2004. 2-D Electrophoresis principles and methods handbook. GE-Healthcare. Handbook 80-6429-60AC.

Antikainen M, Griffith M, Zhang J, Hon WC, Yang DSC, Pihakaski-Maunsbach K. 1996. Immunolocalization of antifreeze proteins in winter rye leaves, crowns, and roots by tissue printing. Plant Physiology 110: 845-857.

Anzlovar S, Dermastia M. 2003. The comparative analysis of osmotins and osmotin-like PR-5 proteins. Plant Biology 5: 116-124.

Apel K, Hirt H. 2004. Reactive oxygen species: metabolism, oxidative stress and signal transduction. Annual Review of Plant Biology 55: 373-399.

Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MD, Durbin R, Falquet L, Fleischmann W, Gouzy J, Hermjakob H, Hulo N, Jonassen I, Kahn D, Kanapin A, Karavidopoulou Y, Lopez R, Marx B, Mulder NJ, Oinn TM, Pagni M, Servant F, Sigrist CJ, Zdobnov EM. 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acid Research 29: 37-40.

Arnold K, Bordoli L, Kopp J, Sschwede T. 2006. The Swiss model workspace: a web-based environment for protein structure and homology modeling. Bioinformatics 22: 195-201.

Arora R, Wisniewski ME, Scorza R. 1992. Cold Acclimation in genetically related (sibling) deciduous and evergreen peach (*Prunus persica* [L.] Batsch) I. Seasonal changes in cold hardiness and polypeptides of bark and xylem tissues. Plant Physiology 99: 1562-1568.

Atici O, Nalbantoğlu B. 2003. Antifreeze proteins in higher plants. Phytochemistry 64: 1187-1196.

Avice JC, Le Dily F, Goulas E, Noquet C, Meuriot F, Volenec JJ, Cunningham SM, Sors TG, Dhont C, Castonguay P, Nadeau P, Bélanger G, Chalifour FP, Ourry A. 2003.

Vegetative storage proteins in overwintering storage organs of forage legumes: roles and regulation. Canadian Journal of Botany 81: 1198-1212.

Bairoch A. 1991. PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acid Research 19 Suppl: 2241-5.

Bajaj M, Blundell T. 1984. Evolution and the tertiary structure of proteins. Annual Review of Biophysics and Bioengineering 13: 453-492.

Beintema JJ. 1994. Structural features of plant chitinases and chitin-binding proteins. FEBS Letters 350: 159-163.

Binnie SC, Grossnickle SC, Roberts DR. 1994. Fall acclimation patterns of interior spruce seedlings and their relationship to changes in vegetative storage proteins. Tree Physiology 14: 1107-1120.

Böhlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, Strauss SH, Nilsson O. 2006. *CO/FT* regulatory module controls timing of flowering and seasonal growth cessation in trees. Science 312: 1040-1043.

Boston RS, Viitanen PV, Vierling E. 1996. Molecular chaperones and protein folding in plants. Plant Molecular Biology 32: 191-222.

Buchter R, Stromberg A, Schmelzer E, Kombrink E. 1997. Primary structure and expression of acidic (class II) chitinase in potato. Plant Molecular Biology 35: 749-761.

Bunce MW, Boronenkov IV, Anderson RA. 2008. Coordinated activation of the nuclear ubiquitin ligase Cul3/SPOP by the generation of phosphatidylinositol 5-phosphate. Journal of Biological Chemistry 283: 8678-8686.

Burns RM., Honkala BH. 1990. Silvics of North America: 1. Conifers; 2. Hardwoods. Agriculture Handbook 654, pp 1-877. U.S. Department of Agriculture, Forest Service, Washington, DC.

Canovas, FM, Avila C, Canton FR, Cañas RA, de la Torre F. 2007. Ammonium assimilation and amino acid metabolism in conifers. Journal of Experimental Botany 58: 2307-2318.

Cantón FR, Suárez MF, Cánovas FM. 2005. Molecular aspects of nitrogen mobilization and recycling in trees. Photosynthesis Research 83: 265-278.

Castellano MdM, Boniotti MB, Caro E, Schnittger A, Gutierrez C. 2004. DNA replication licensing affects cell proliferation or endoreplication in a cell type-specific manner. Plant Cell 16: 2380-2393.

Chrispeels MJ, Raikhel NV. 1992. Short peptide domains target proteins to plant vacuoles. Cell 68: 613-616.

Clapham DH, Dormling I, Ekberg I, Eriksson G, Qamaruddin M, Vince-Prue D. 1998. Latitudinal cline of requirement for far-red light for the photoperiodic control of budset and extension growth in *Picea abies* (Norway spruce). Physiologia Plantarum 102: 71-78.

Clausen S, Apel K. 1991. Seasonal changes in the concentration of the major storage protein and its mRNA in xylem ray cells of poplar trees. Plant Molecular Biology 17: 669-678.

Coleman GD, Chen THH, Ernst SG, Fuchigami L. 1991. Photoperiod control of poplar bark storage protein accumulation. Plant Physiology 96: 686-692.

Coleman GD, Chen THH, Fuchigami LH. 1992. Complementary cDNA cloning of poplar bark storage protein and control of its expression by photoperiod. Plant Physiology 98: 687-693.

Coleman GD, Englert JM, Chen THH, Fuchigami LH. 1993. Physiological and environmental requirements for poplar (*Populus deltoides*) bark storage protein degradation. Plant Physiology 102: 53-59.

Coleman GD. 2004. Physiology and regulation of seasonal nitrogen cycling in woody plants. Journal of Crop Improvement 10: 237-259.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics Application Notes 21: 3674-3676.

Conesa A, Götz S. 2007. Blast2GO: a comprehensive suite for functional analysis in plant genomics. International Journal of Plant Genomics 2008: 0-39.

Cook D, Fowler S, Fiehn O, Thomashow MF. 2004. A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of *Arabidopsis*. Proceedings of the National Academy of Sciences USA 101: 15243-15248.

Cooke JEK, Weih M. 2005. Nitrogen storage and seasonal nitrogen cycling in *Populus*: bridging molecular physiology and ecophysiology. New Phytologist 167: 19-30.

Davis JM, Egelkrout EE, Coleman GD, Chen THH, Haissig BE, Riemenschneider DE, Gordon MP. 1993. A family of wound-induced genes in *Populus* shares common features with genes encoding vegetative storage proteins. Plant Molecular Biology 23: 135-143.

De los Reyes BG, Taliaferro CM, Anderson MP, Melcher U, McMaugh S. 2001. Induced expression of the class II chitinase gene during cold acclimation and dehydration of bermudagrass (*Cynodon* sp.). Theoretical and Applied Genetics 103: 297-306.

Denis CL, Malvar T. 1990. The *CCR4* gene from *Saccharomyces cerevisiae* is required for both nonfermentative and *spt*-mediated gene expression. Genetics 124: 283-291.

DeRocher A, Vierling E. 1995. Cytoplasmic HSP70 homologues of pea: differential expression in vegetative and embryonic organs. Plant Molecular Biology 27: 441-456.

DeWald DB, Mason HS, Mullet JE. 1992. The soybean vegetative storage proteins VSPα and VSPβ are acid phosphatases active on polyphosphates. Journal of Biological Chemistry 267: 15958-15964.

Dhont C, Castonguay Y, Avice JC, Chalifour FP. 2006. VSP accumulation and cold-inducible gene expression during autumn hardening and overwintering in alfalfa. Journal of Experimental Botany 57: 2325-2337.

Dong JZ, Dunstan DI. 1997. Endochitinase and β-1,3-glucanase genes are developmentally regulated during somatic embryogenesis in *Picea glauca*. Planta 201: 189-194.

Doxey AC, Yaish MW, Griffith M, MacConkey BJ. 2006. Ordered surface carbons distinguish antifreeze proteins and their ice-binding regions. Nature Biotechnology 24: 852-855.

Druart N, Johansson A, Baba K., Schrader J, Sjodin A, Bhalerao RR, Resman L, Trygg J, Moritz T., Bhalerao RP. 2007. Environmental and hormonal regulation of the activity-dormancy cycle in the cambial meristem involves stage-specific modulation of transcriptional and metabolic networks. Plant Journal 50: 557-573.

Eckardt NA. 2002. Alternative splicing and the control of flowering time. Plant Cell 14: 743-747.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acid Research 32: 1792-1797.

Eriksson ME, Moritz T. 2002. Daylength and spatial expression of gibberellin 20-oxidase isolated from hybrid aspen (*Populus tremula* L. * *P. tremuloides* Michx.). Planta 214: 920-930.

Espinosa-Ruiz A, Saxena S, Schmidt J, Mellerowicz E, Miskolczi P, Bako L, Bhalerao RP. 2004. Differential stage-specific regulation of cyclin-dependent kinases during cambial dormancy in hybrid aspen. Plant Journal 38: 603-615.

Farrar JJ, Evert RF. 1997. Seasonal changes in the ultrastructure of the vascular cambium of *Robinia pseudoacacia*. Trees 11: 191-202.

137

Fowler S, Thomashow MF. 2002. Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. Plant Cell 14: 1675-1690.

Foyer CH, Noctor G. 2005. Oxidant and antioxidant signaling in plants: a re-evaluation of the concept of oxidative stress in a physiological context. Plant Cell and Environment 28: 1056-1071.

Futamura N, Tani N, Tsumura Y, Nakajima M, Sakaguchi M, Shinohara K. 2005. Characterization of genes for novel thaumatin-like proteins in *Cryptomeria japonica*. Tree Physiology 26: 51-62.

Glerum C. 1985. Frost hardiness of coniferous seedlings: principles and applications. In Duryea ML. (ed) Proceedings: Evaluating seedling quality: principles, procedures and predictive abilities of major tests. Workshop held October 15-18, 1984. Forest Research Laboratory. Oregon State University , Corvallis. ISBN 0-87437-000-0.

Gomez L, Faurobert M. 2002. Contribution of vegetative storage proteins to seasonal nitrogen variations in the young shoots of peach trees (*Prunus persica* L. Batsch). Journal of Experimental Botany 53: 2431-2439.

González DH, Welchen E, Attallah CV, Comelli RN, Mufarrege EF. 2007. Transcriptional coordination of the biogenesis of the oxidative phosphorylation machinery in plants. The Plant Journal 51: 105-116.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acid Research 36: 3420-3435.

Goyer A, Haslekås C, Miginiac-Maslow M, Klein U, Le Marechal P, Jacquot JP, Decottignies P. 2002. Isolation and characterization of a thioredoxin-dependent peroxidase from *Chlamydomonas reinhardtii*. European Journal of Biochemistry 269: 272-282.

Griffith M, Antikainen M, Hon WC, Pihakaski-Maunsbach K, Yu XM, Chun JU, Yang DSC. 1997. Antifreeze proteins in winter rye. Physiologia Plantarum 100:327-332.

Griffith M, Yaish MWF. 2004. Antifreeze proteins in overwintering plants: a tale of two activities. Trends in Plant Science 9: 399-405.

Grimes HD, Transbarger TJ, Franceschi VR. 1993. Expression and accumulation patterns of nitrogen-responsive lipoxygenase in soybeans. Plant Physiology 103: 457-466.

Grossnickle SC. 2000. Ecophysiology of northern spruce species: the performance of planted seedlings, pp 1-407. Canada: NRC Research Press.

Gupta R. 2001. Prediction of glycosylation sites in proteomes: from post-translational modifications to protein function. Ph.D. thesis at the Center for Biological Sequence Analysis, Technical University of Denmark DTU.

Gyllenstrand N, Clapham D, Källman T, Lagercrantz U. 2007. A Norway spruce FLOWERING LOCUS T homolog is implicated in control of growth rhythm in conifers. Plant Physiology 144: 248-257.

Hamel F, Boivin R, Tremblay C, Bellemare G. 1997. Structural and evolutionary relationships among chitinases of flowering plants. Journal of Molecular Evolution 44: 614-624.

Hart PJ, Pfluger HD, Monzingo AF, Hollis T, Robertus JD. 1995. The refined crystal structure of an endochitinase from *Hordeum vulgare* L. seeds at 1.8 Å resolution. Journal of Molecular Biology 248: 402-413.

Heide OM, Prestrud AK. 2005. Low temperature, but not photoperiod, controls growth cessation and dormancy induction and release in apple and pear. Tree Physiology 25: 109-114.

Hennig M, Jansonius JN, van Scheltinga ACT, Dijkstra BW, Schlesier B. 1995. Crystal structure of concanavalin B at 1.65 Å resolution. An "inactivated" chitinase from seeds of *Canavalia ensiformis*. Journal of Molecular Biology 254: 237-246.

Henrissat B, Davies GJ. 2000. Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. Plant Physiology 124: 1515-1519.

Hietala AM, Kvaalen H, Schmidt a, Jøhnk N, Solheim H, Fossdal CG. 2004. Temporal and spatial profiles of chitinase expression by Norway spruce in response to bark colonization by *Heterobasidion annosum*. Applied and Environmental Microbiology 70: 3948-3953.

Hilhorst HWM, Koornneef M. 2007. Dormancy in plants. Encyclopedia of life sciences. DOI: 10.1002/9780470015902.a0002045. Retrieved July 8-2008.

Holm L, Sander C. 1994. Structural similarity of plant chitinase and lysozymes from animals and phage, an evolutionary connection. FEBS Letters 340: 129-132.

Hon WC, Griffith M, Mlynarz A, Kwok YC, Yang DSC. 1995. Antifreeze proteins in winter rye are similar to pathogenesis-related proteins. Plant Physiology 109: 879-889.

Hong JK, Jung HW, Kim YJ, Hwang BK. 2000. Pepper gene encoding a basic class II chitinase is inducible by pathogen and ethephon. Plant Science 159: 39-49.

Horton P, Park KJ, Obayashi T, Fujita N, Hatada H, Adams-Collier CJ, Nakai K. 2007. WoLF PSORT: protein localization prediction. Nucleic Acid Research 35(Web Server issue):W585-7. Epub 2007 May 21.

Hou HW, Zhou YT, Mwange KN, Li WF, He XQ, Cui KM. 2006. ABP1 expression regulated by IAA and ABA is associated with the cambium periodicity in *Eucommia ulmoides* Oliv. Journal of Experimental Biology 57: 3857-3867.

Howe GT, Gardner G, Hackett WP, Furnier GR. 1996. Phytochrome control of short-day-induced bud set in black cottonwood. Physiologia Plantarum 97: 95-103.

Howe GT, Saruul P, Davis J, Chen THH. 2000. Quantitative genetics of bud phenology, frost damage, and winter survival in an F_2 family of hybrid poplars. Theoretical and Applied Genetics 101: 632-642.

Howe GT, Aitken SN, Neale DB, Jermstad KD, Wheeler NC, Chen THH. 2003. From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. Canadian Journal of Botany 81: 1247-1266.

Huet J, Wyckmans J, Wintjens R, Boussard P, Raussens V, Vandenbussche G, Ruysschaert JM, Azarkan M, Looze Y. 2006. Structural characterization of two papaya chitinases, a family GH19 of glycosyl hydrolases. Cellular and Molecular Life Sciences 63: 3042-3054.

Huet J, Rucktooa P, Clantin B, Azarkan M, Looze Y, Villeret V, Wintjens R. 2008. X-ray structure of papaya chitinase reveals the substrate binding mode of glycosyl hydrolase family 19 chitinases. Biochemistry 47: 8283-8291.

Hüner NPA, Oquist G, Sarhan F. 1998. Energy balance and acclimation to light and cold. Trends in Plant Science 3: 224-230.

Iseli B, Boller T, Neuhaus JM. 1993. The N-terminal cysteine-rich domain of tobacco class I chitinase is essential for chitin binding but not for catalytic or antifungal activity. Plant Physiology 103: 221-226.

Image Master 2D Platinum Software, user manual for version 5.0. 2003. Swiss Institute of Bioinformatics (SIB).

Jeknic Z. and Chen THH. 1999. Changes in protein profiles of polar tissues during the induction of bud dormancy by short-day photoperiods. Plant Cell Physiology 40: 25-35.

Jøhnk N, Hietala AM, Fossdal CG, Collinge DB, Newman MA. 2005. Defense-related genes expressed in Norway spruce roots after infection with the root rot pathogen *Ceratobasidium bicorne* (anamorph: *Rhizoctonia* sp.). Tree Physiology 25: 1533-1543.

Kamauchi S, Wadahama H. Iwasaki K, Nakamoto K, Nishizawa K, Ishimoto M, Kawada T, Urade R. 2008. Molecular cloning and characterization of two soybean protein disulfide isomerases as molecular chaperones for seed storage proteins. FEBS Journal 275: 2644-2658.

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. 2008. KEGG for linking genomes to life and the environment. Nucleic Acid Research 36: 480-484.

Kasprzewska A. 2003. Plant chitinases – regulation and function. Cellular and Molecular Biology Letters 8: 809-824.

Keskitalo J, Bergquist G, Gardeström P, Jansson S. 2005. A cellular timetable of autumn senescence. Plant Physiology 139: 1635-1648.

Klein J, Saedler H, Huijser P. 1996. A new family of DNA binding proteins includes putative transcriptional regulators of the *Antirrhinum majus* floral meristem identity gene SQUAMOSA. Molecular and General Genetics 250: 7-16.

Koutaniemi S, Warinowski T, Kärkönen A, Alatalo E, Fossdal CG, Saranpää P, Laakso T, Fagerstedt KV, Simola LK, Paulin L, Rudd S, Teeri TH. 2007. Expression profiling of the lignin biosynthetic pathway in Norway spruce using EST sequencing and real-time RT-PCR. Plant Molecular Biology 65: 311-328.

Kropat J, Tottey S, Birkenbihl RP, Depège N, Huijser P, Merchant S. 2005. A regulator of nutritional copper signaling in *Chlamydomonas* is an SBP domain protein that recognizes the GTAC core of copper response element. Proceedings of the National Academy of Sciences USA 102: 18730-1835.

Kumar S. Tamura K, Nei M. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in Bioinformatics 5: 150-163.

Kumar S, Nei M, Dudley J, Tamura K. 2008. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Briefings in Bioinformatics 9: 299-306.

Lahm HW, Langen H. 2000. Mass spectrometry: a tool for the identification of proteins separated by gels. Electrophoresis. 21: 2105-2114.

Lang GA. 1987. Dormancy: a new universal terminology. HortScience 22: 817-820.

Langheinrich U, Tischner R. 1991. Vegetative Storage Proteins in Poplar: Induction and Characterization of a 32- and a 36-Kilodalton polypeptide. Plant Physiology 97: 1017-1025.

Launholt D, Merkle T, Houben A, Schulz A, Grasser KD. 2006. *Arabidopsis* chromatin-associated HMGA and HMGB use different nuclear targeting signals and display highly dynamic localization within the nucleus. Plant Cell 18: 2904-2918.

Lee SP, Chen THH. 1993. Molecular biology of plant cold hardiness development. In Li PH, Christersson L. (eds) Advances in plant cold hardiness, pp 1-30. USA: CRC press, Inc.

Li C, Junttila O, Heino P, Palva ET. 2003. Different responses of northern and southern ecotypes of *Betula pendula* to exogenous ABA application. Tree Physiology 23: 481-487.

Li C, Welling A, Puhakainen T, Vihera-Aarnio A, Ernstsen A, Junttila O, Heino P, Palva T. 2005. Differential responses of silver birch (*Betula pendula*) ecotypes to short-day photoperiod and low temperature. Tree Physiology 25: 1563-1569.

Li J, Wu XD, Hao ST, Wang XJ, Ling HQ. 2008. Proteomic response to iron deficiency in tomato root. Proteomics 8: 2299-2311.

Lippert D, Zhuang J, Ralph S, Ellis DE, Gilbert M, Olafson R, Ritland K, Ellis B, Douglas CJ, Bohlmann J. 2005. Proteome analysis of early somatic embryogenesis in *Picea glauca*. Proteomics 5: 461-473.

Lippert D, Chowrira S, Ralph SG, Zhuang J, Aeschliman D, Ritland C, Ritland K, Bolhmann J. 2007. Conifer defense against insects: Proteome analysis of Sitka spruce (*Picea sitchensis*) bark induced by mechanical wounding or feeding by white pine weevils. Proteomics 7: 248-270.

Marjamaa K, Lehtonen M, Lundell T, Toikka M, Saranpää P, Fagerstedt KV. 2003. Developmental lignification and seasonal variation in β-glucosidase and peroxidase activities in xylem of Scots pine, Norway spruce and silver birch. Tree Physiology 23: 977–986

Martinez –Zapater JM, Somerville CR. 1990. Effect of light quality and vernalization on late-flowering mutants of *Arabidopsis thaliana*. Plant Physiology 92: 770-776.

Mellerowicz EJ, Riding RT, Little CHA. 1989. Genomic variability in the vascular cambium of *Abies balsamea*. Canadian Journal of Botany 67: 990-996.

Meuriot F, Noquet C, Avice JC, Volenec JJ, Cunningham SM, Sors TG, Caillot S, Ourry A. 2004. Methyl jasmonate alters N partitioning, N reserves accumulation and induces gene expression of a 32-kDa vegetative storage protein that possesses chitinase activity in *Medicago sativa* taproots. Physiologia Plantarum 120: 113-123.

Möller I, Beatrix B, Kreibich G, Sakai H, Lauring B, Wiedmann M. 1998. Unregulated exposure of the ribosomal M-site caused by NAC depletion results in delivery of non-secretory polypeptides to the Sec61 complex. FEBS letters 441: 1-5.

Mølmann JA, Asante DKA, Jensen JB, Krane MN, Ernstsen A, Junttila O, Olsen JE. 2005. Low night temperature and inhibition of gibberellin biosynthesis override phytochrome action and induce bud set and cold acclimatation, but not dormancy in

PHYA overexpressors and wild-type of hybrid aspen. Plant, Cell and Environment 28: 1579-1588.

Morin X, Améglio T, Ahas R, Kurz-Besson C, Lanta V, Lebourgeois F, Miglietta F, Chuine I. 2007. Variation in cold hardiness and carbohydrate concentration from dormancy induction to bud burst among provenances of three European oak species. Tree Physiology 27: 817-825.

Mwange KN, Hou HW, Wang YQ, He XQ, Cui KM. 2005. Opposite patterns in the annual distribution and time-course of endogenous abscisic acid and indol 3-acetic acid in relation to the periodicity of cambial activity in *Eucommia ulmoides* Oliv. Journal of Experimental Botany 56: 1017-1028.

Nagy NE, Fossdal CG, Dalen LS, Lonneborg A, Heldal I, Johnsen O. 2004. Effects of *Rhizoctonia* infection and drought on peroxidase and chitinase activity in Norway spruce (*Picea abies*). Physiologia Plantarum 120: 465-473.

Neuhaus JM, Sticher L, Meins F, Boller T. 1991. A short C-terminal sequence is necessary and sufficient for the targeting of chitinases to the plant vacuole. Proceedings of the National Academy of Sciences USA 88: 10362-10366.

Nishitani H, Lygerou Z. 2002. Control of DNA replication licensing in a cell cycle. Genes to Cells 7: 523-534

Notredame C, Higgings DG, Heringa J. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. Journal of Molecular Biology 302: 205-217.

Odorizzi G, Babst M, Emr SD. 2000. Phosphoinositide signaling and the regulation of membrane trafficking in yeast. Trend in Biochemical Sciences 25: 229-235.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. 1999. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acid Research 27: 29-34.

Ogilvie RT. 1972. Speciation in the North American spruces and its relation to white spruce. In McMinn RG. (ed) White spruce: the ecology of a northern resource; (Proceedings of the symposium held June 21-1971; Canadian Botanical Association, Edmonton, Alberta). Information report NOR-X-40, pp. 1-44 Canada: Northern Forest Research Center, Canadian Forest Service, Department of Environment, Edmonton, Alberta, T6H 3S5.

Okubo H. 2000. Growth cycle and dormancy in plants. In Viemont J. (ed) Dormancy in plants, pp. 1-22. France: CABI publishing.

Olsen JE, Junttilla O, Nilsen J, Eriksson ME, Martinussen I, Olsson O, Sandberg G, Moritz T. 1997. Ectopic expression of the oat phytochrome A in hybrid aspen changes

critical daylength for growth and prevents cold acclimatization. Plant Journal 12: 1339-1350.

Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. 1997. CATH--a hierarchic classification of protein domain structures. Structure 5: 1093-1108.

Papin DJ, Hojrup P, Bleasby AJ. 1993. Rapid identification of proteins by peptide-mass fingerprinting. Current Biology 3: 327-332.

Park S, Keathley DE, Han KH. 2008. Transcriptional profiles of annual growth cycle in *Populus deltoides*. Tree Physiology 28: 321-329.

Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20: 3551-3567.

Peumans WJ, Proost P, Swennen RL, Van Damme EJM. 2002. The abundant class III chitinase homolog in young developing banana fruits behaves as a transient vegetative storage protein and most probably serves as an important supply of amino acids for the synthesis of ripening-associated proteins. Plant Physiology 130: 1063-1072.

Pihakaski-Maunsbach K, Moffatt B, Testillano P, Risueño M, Yeh S, Griffith M, Maunsbach AB. 2001. Genes enconding chitinase-antifreeze proteins are regulated by cold and expressed by all cell types in winter rye shoots. Physiologia Plantarum 112: 359-371.

Puig A, Gilbert HF. 1994. Protein disulfide isomerase exhibits chaperone and anti-chaperone activity in the oxidative refolding of lysozyme. Journal of Biological Chemistry 269: 7764-7771.

Raab T, López-Ráez JA, Klein D, Caballero JL, Moyano E, Schwab W, Muñoz-Blanco J. 2006. *FaQR*, required for the biosynthesis of the strawberry flavor compound 4-hydroxy-2,5-dimethyl-3(2H)-furanone, encodes an enone oxidoreductase. Plant Cell 18: 1023-1037.

Razem FA, El-Kereamy A, Abrams SR, Hill RD. 2006. The RNA-binding protein FCA is an abscisic acid receptor. Nature 439: 290-294.

Raynaud C, Perennes C, Reuzeau C, Catrice O, Brown S, Bergounioux C. 2005. Cell and plastid division are coordinated through the prereplication factor AtCDT1. Proceedings of the National Academy of Sciences USA 102: 8216-8221.

Richert S, Luche S, Chevallet M, Van Dorsselaer A, Leize-Wagner E, Rabilloud T. 2004. About the mechanism of interference of silver staining with peptide mass spectrometry. Proteomics 4: 909-916.

Ritchie G. 1959. Trees of knowledge: a handbook of maritime trees. 1996 582.1609715 C96-980159-9. Retrieved July 14-2008.

Roberts DR, Toivonen P, McInnis SM. 1991. Discrete proteins associated with overwintering of interior spruce and Douglas-fir seedlings. Canadian Journal of Botany 69: 437-441.

Rohde A, Howe GT, Olsen JE, Moritz T, Van Montagu M, Junttila O, Boerjan W. 2000. Molecular aspects of bud dormancy in trees. In Jain S, Minocha S. (ed) Molecular biology of woody plants, Vol 1, pp 89-134. Netherlands: Kluwer Academic Publishers.

Rohde A, Bhalerao RP. 2007. Plant dormancy in the perennial context. Trends in Plant Science 12: 217-223.

Romero-Puertas MC, Campostrini N, Mattè A, Righetti PG, Perazzolli M, Zolla L, Roepstorff P, Delledonne M. 2008. Proteomic analysis of S-nitrosylated proteins in *Arabidopsis thaliana* undegoing hypersensitive response. Proteomics 8: 1459-1469.

Ross ARS, Lee PJ, Smith DL, Langridge JI, Whetton AD, Gaskell SJ. 2002. Identification of proteins from two-dimensional polyacrylamide gels using acid-labile surfactant. Proteomics 2: 928-936.

Roux SJ. 1994. Signal transduction in phytochrome responses. In Kendrick RE, Kronenberg GHM. (eds) Photomorphogenesis in plants, pp 187-209. The Netherlands: Kluwer Academic Publishers.

Rüdiger W, Thümmler F. 1994. The phytochrome chromophore. In Kendrick RE, Kronenberg GHM. (eds) Photomorphogenesis in plants, pp 51-69. The Netherlands: Kluwer Academic Publishers.

Ruttink T, Arend M, Morreel K, Storme V, Rombauts S, Fromm J, Bhalerao RP, Boerjan W, Rohde A. 2007. A molecular timetable for apical bud formation and dormancy induction in poplar. Plant Cell 19: 2370-2390.

Salzer P, Hübner B, Sirrenberg A, Hager A. 1997. Differential effect of purified spruce chitinases and beta-1,3-glucanases on the activity of elicitors from ectomycorrhizal fungi. Plant Physiology 114: 957-968.

Sauter JJ, van Cleve B, Wellenkamp S. 1989. Ultrastructural and biochemical results on the localization and distribution of storage proteins in poplar tree and in twigs of other tree species. Holzforschung 43: 1-6.

Sauter JJ, van Cleve B. 1994. Storage, mobilization and interrelations of starch, sugars, protein and fat in the ray storage tissue of poplar trees. Trees Structure and Function 8: 297-304.

Sayle R, Milner-White EJ. 1995. RasMol: Biomolecular graphics for all. Trends in Biochemical Sciences 20: 374-376.

Schrader J, Baba K, May ST, Palme K, Bennett M, Bhalerao RP, Sandberg G. 2003. Polar auxin transport in the wood-forming tissues of hybrid aspen is under simultaneous control of developmental and environmental signals. Proceedings of the National Academy of Sciences USA 100: 10096-10101.

Schrader J, Moyle R, Bhalerao R, Hertzberg M, Lunderberg J, Nilsson P, Bhalerao RP. 2004. Cambial meristem dormancy in trees involves extensive remodeling of the transcriptome. Plant Journal 40: 173-187.

Sela-Buurlage MB, Ponstein AS, Bres-Vloemans SA, Melchers LS, van den Elzen PJM, Cornelissen BJC. 1993. Only specific tobacco (*Nicotiana tabacum*) chitinases and β-1,3-glucanases exhibit antifungal activity. Plant Physiology 101: 858-863.

Søgaard G, Johnsen Ø, Nilsen J, Junttila O. 2008. Climatic control of bud burst in young seedlings of nine provenances of Norway spruce. Tree Physiology 28: 311-320.

Sonhammer EL, Eddy SR, Durbin R. 1997. Pfam: a comprenhensive database of protein domain families based on seed alignments. Proteins 28: 405-420.

Staswick PE. 1994. Storage proteins of vegetative plant tissues. Annual Review of Plant Physiology and Plant Molecular Biology 45: 303-322.

Stepien V, Sauter JJ, Martin F. 1994. Vegetative storage proteins in woody plants. Plant Physiology and Biochemistry 32: 185-192.

Stitt M, Hurry V. 2002. A plant for all seasons, alterations in the photosynthetic carbon metabolism during cold acclimation in *Arabidopsis*. Current Opinion in Plant Biology 5: 199-206.

Svendsen E, Wilen R, Stevenson R, Liu R, Tanino KK. 2007. A molecular marker associated with low-temperature induction of dormancy in red osier dogwood (*Cornus sericea*). Tree Physiology 27: 385-397.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24: 1596-1599.

The Gene Ontology Consortium. 2000. Gene ontology: tool for the unification of biology. Nature genetics 25: 25-29.

Thompson JD, Higgins D, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acid Research 22: 4673-4680.

Tranberger TJ, Franceschi VR, Hildebrand DF, Grimes HD. 1991. The soybean 94-kilodalton vegetative storage protein is a lipoxygenase that is localized in paraveinal mesophyll cell vacuoles. Plant Cell 3: 973-987.

Tusher VG, Tibshirani R, Chu G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences USA 98: 5116-5121.

Ubhayasekera W, Tang CM, Ho SWT, Berglund G, Bergfors T, Chye ML, Mowbray SL. 2007. Crystal structures of a family 19 chitinase form *Brassica juncea* show flexibility of binding cleft loops. FEBS Journal 274: 3695-3703.

Van Damme EJM, Hause B, Hu J, Barre A, Rougé P, Proost P, Peumans WJ. 2002. Two distinct jacalin-related lectins with a different specificity and subcellular location are major vegetative storage proteins in the bark of the black mulberry tree. Plant Physiology 130: 757-769.

Vashney A, Brooks FP, Wright WV. 1994. Computing smooth molecular surfaces. IEEE Computer Graphics and Applications 14: 19-25.

Verburg JG, Rangwala SH, Samac DA, Luckow VA, Huynh QK. 1993. Examination of the Role of Tyrosine-174 in the Catalytic Mechanism of the *Arabidopsis thaliana*-Chitinase: Comparison of Variant Chitinases Generated by Site-Directed Mutagenesis and Expressed in Insect Cells Using Baculovirus Vectors. Archives of Biochemistry and Biophysics 300: 223-230

Wang MC, Peng ZY, Li Cl, Li F, Liu C, Xia GM. 2008. Proteomics analysis on a high salt tolerance introgression strain of *Triticum aestivum / Thinopyrom ponticum.* Proteomics 8: 1470-1489.

Wang S, Sakai H, Wiedmann M. 1995. Protective environment for regions of nascent chains just emerging from the peptidyl transferase center. The Journal of Cell Biology 130: 519-528.

Wareing PF. 1956. Photoperiodism in woody plants. Annual Review of Plant Physiology 7: 191-204.

Weber H, Bernhardt A, Dieterle M, Hano P, Mutlu A, Estelle M, Genschik P, Hallmann H. 2005. Arabidopsis AtCUL3a and AtCUL3b form complexes with members of the BTB/POZ-MATH protein family. Plant Physiology 137: 83-93.

Welling A, Moritz T, Palva ET, Junttila O. 2002. Independent activation of cold acclimation by low temperature and short photoperiod in hybrid aspen. Plant Physiology 129: 1633-1641.

Welling A, Rinne P, Viherä-Aarnio A, Kontunen-Soppela S, Heino P, Palva ET. 2004. Photoperiod and temperature differentially regulate the expression of two dehydrin genes during overwintering in birch (*Betula pubescens* Ehrh.). Journal of Experimental Biology 55: 507-516.

Welling A, Palva ET. 2006. Molecular control of cold acclimation in trees. Physiologia Plantarum 127: 167-181.

Wetzel S, Greenwood JS. 1989. Proteins as a potential nitrogen storage compound in bark and leaves of several softwoods. Trees 3: 149-153.

Wetzel S, Greenwood JS. 1991. The 32-kilodalton vegetative storage protein of *Salix microstachya* Turz : characterization and immunolocalization. Plant Physiology 97: 771-777.

Wisniewski ME, Bassett CL, Renaut J, Farell Jr. R, Tworkoski T, Artlip TS. 2006. Differential regulation of two dehydrin genes from peach (*Prunnus persica*) by photoperiod, low temperature and water deficit. Tree Physiology 26: 575-584.

Wiweger M, Farbos I, Ingouff M, Lagercrantz U, von Arnold S. 2003. Expression of *Chia4-Pa* chitinase genes during somatic and zygotic embryo development in Norway spruce (*Picea abies*): similarities and differences between gymnosperm and angiosperm class IV chitinases. Journal of Experimental Botany 54: 2691-2699.

Xu F, Fan C, He Y. 2007. Chitinases in *Oryza sativa* ssp. *Japonica* and *Arabidopsis thaliana*. Journal of Genetics and Genomics 34: 138-150.

Yadav SK, Singla-Pareek SL, Ray M, Reddy MK, Sopory SK. 2005. Methylglyoxal levels in plants under salinity stress are dependent on glyoxalase I and glutathione. Biochemical and Biophysical Research Communications 337: 61-67.

Yamamoto S, Minimikawa T. 1998. Isolation and characterization of cDNA encoding a high mobility group protein HMG-1 from *Canavalia gladiata* D.C. Biochimica et Biophysica Acta 1396: 47-50.

Yeh S, Moffatt BA, Griffith M, Xiong F, Yang DSC, Wiseman SB, Sarhan F, Danyluk J, Xue YQ, Hew CL, Doherty-Kirby A, Lajoie G. 2000. Chitinase genes responsive to cold encode antifreeze proteins in winter cereals. Plant Physiology 124: 1251-1263.

Young E, Hanover JW. 1977. Effects of quality, intensity and duration of light breaks during a long night on dormancy in blue spruce (*Picea pungens* Engelm.) seedlings. Plant Physiology 60: 271-273.

Yu XM, Griffith M. 1999. Antifreeze proteins in winter rye leaves form oligomeric complexes. Plant Physiology 119: 1361-1369.

Yuanqing J, Yang B, Harris N, Deyholos MK. 2007. Comparative proteomic analysis of NaCl stress-responsive proteins in *Arabidopsis* roots. Journal of Experimental Biology 58: 3591-3607.

Zar JH. 1999. Biostatistical Analysis, pp 1-663. USA: Prentice-Hall, Inc.

Zhu B, Coleman GD. 2001a. The poplar bark storage protein gene (*Bspa*) promoter is responsive to photoperiod and nitrogen in transgenic poplar and active floral tissues, immature seeds and germinating seeds of transgenic tobacco. Plant Molecular Biology 46: 383-394.

Zhu B, Coleman GD. 2001b. Phytochrome-mediated photoperiod perception, shoot growth, glutamine, calcium, and protein phosphorylation influence the activity of the poplar bark storage protein gene promoter (bspA). Plant Physiology 126: 342-351.

7.0 APPENDICES

Appendix 7.1

2D-SDS PAGE for sample under 0 SD (Figure 1) and 10 weeks of SD (Figure 2). Experiment S011, replicate 2. Annotated proteins that were upregulated upon each condition are outlined with their corresponding spot ID from Table 3.2. Although not all differentially expressed proteins could be annotated it is evident that upregulation of a larger number of proteins takes place upon 10 weeks of short days. Numbers on the right indicate MW.

166.0 kDa

66.2

45.0

35.0

25.0

18.4

14.4

pH 3-10

2

152
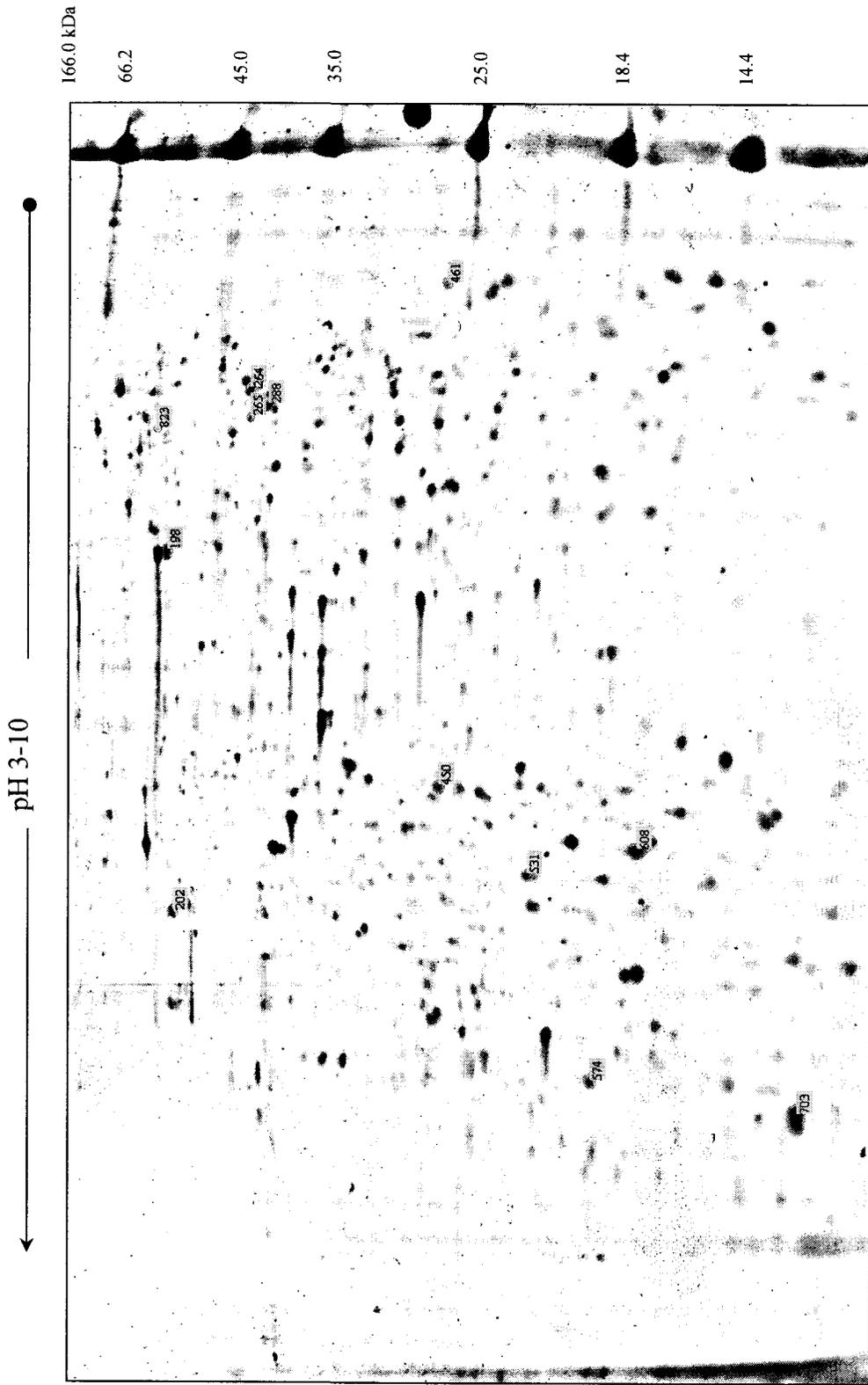
Appendix 7.2

Functional analysis table from Blast2GO. Top BLAST hit and BLAST annotation correspond to Table 3.2. The GO terms from third level correspond to the biological process bar graph of Figure 3.7. Lower GO terms correspond to biological process, molecular function and cellular compartment below third level. The Interpro annotation corresponds to all possible domains and motifs found using all the algorithms from Interpro.

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
143	ABA98741.2	retrotransposon protein, putative, Ty3-gypsy subclass [Oryza sativa (japonica cultivar-group)]	primary metabolic process; macromolecule metabolic process; cellular metabolic process	RNA-directed DNA polymerase activity;DNA binding;RNA-dependent DNA replication;RNA binding;aspartic-type endopeptidase activity;proteolysis;plastid;DNA recombination;mitochondrion	RNA-directed DNA polymerase (reverse transcriptase); Chromo domain; Integrase, catalytic core; Zinc finger, CCHC-type; Retrotransposon gag protein; Peptidase aspartic, catalytic; Polynucleotidyl transferase, Ribonuclease H fold; Retroviral aspartyl protease
486	ABG73467.1	6-phosphogluconolactonase [Oryza brachyantha]	primary metabolic process; macromolecule metabolic process; catabolic process; cellular metabolic process	pentose-phosphate shunt;plastid;6-phosphogluconolactonase activity	6-phosphogluconolactonase; Glucosamine/galactosamine-6-phosphate isomerase

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
105	ABH02600.1	ATP synthase beta subunit [Diarrhena obovata]	generation of precursor metabolites and energy; primary metabolic process; biosynthetic process; cellular metabolic process; transport	proton-transporting ATP synthase complex, catalytic core F(1);plasma membrane ATP synthesis coupled proton transport;hydrogen-exporting ATPase activity, phosphorylative mechanism;hydrogen ion transporting ATP synthase activity, rotational mechanism;integral to membrane;metal ion binding;thylakoid;ATP binding;hydrogen ion transporting ATPase activity, rotational mechanism;chloroplast	ATPase, F1/V1/A1 complex, alpha/beta subunit, nucleotide-binding; ATPase, F1/V1/A1 complex, alpha/beta subunit, C-terminal; AAA+ ATPase, core; ATPase, F1/V1/A1 complex, alpha/beta subunit, N-terminal; ATPase, F1 complex, beta subunit
90	Q43116.1	Protein disulfide-isomerase precursor (PDI)	cellular homeostasis; regulation of biological quality	endoplasmic reticulum lumen;cytoplasmic membrane-bounded vesicle;protein binding;cell redox homeostasis;protein disulfide isomerase activity	Disulphide isomerase; Protein disulphide isomerase; Thioredoxin-related; Thioredoxin fold; Thioredoxin-like fold; Thioredoxin domain
136	P17614.1	ATP synthase subunit beta, mitochondrial precursor	generation of precursor metabolites and energy; primary metabolic process; response to chemical stimulus; response to stress; biosynthetic process; cellular metabolic process; transport	response to oxidative stress;hydrogen-exporting ATPase activity, phosphorylative mechanism;hydrogen ion transporting ATP synthase activity, rotational mechanism;integral to membrane;ATP synthesis coupled proton transport;metal ion binding;ATP binding;mitochondrial proton-transporting ATP synthase, catalytic core;hydrogen ion transporting ATPase activity, rotational mechanism	ATPase, F1/V1/A1 complex, alpha/beta subunit, nucleotide-binding; ATPase, F1/V1/A1 complex, alpha/beta subunit, C-terminal; AAA+ ATPase, core; ATPase, F1/V1/A1 complex, alpha/beta subunit, N-terminal; ATPase, F1 complex, beta subunit

154

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
668	CAL50172.1	A Chain A, Crystal Structure Of Peroxisomal Acyl-Coa Oxidase-Ii (ISS) [Ostreococcus tauri]	primary metabolic process; sexual reproduction; cellular metabolic process	oxidoreductase activity, acting on the CH-CH group of donors;fatty acid metabolic process;mitochondrion;spermatogenesis;peroxisomal membrane	Acyl-CoA oxidase, C-terminal; Acyl-CoA dehydrogenase/oxidase, central region; Acyl-CoA dehydrogenase/oxidase C-terminal; Acyl-CoA dehydrogenase/oxidase, middle and N-terminal; Acyl-CoA oxidase; Acyl-CoA oxidase/dehydrogenase, type1/2, C-terminal; Acyl-CoA dehydrogenase/oxidase, N-terminal
314	CAL57543.1	Glucose-repressible alcohol dehydrogenase transcriptional effector CCR4 and related proteins (ISS) [Ostreococcus tauri]			Endonuclease/exonuclease/phosphatase
378	CAL56731.1	DNA repair protein-related (ISS) [Ostreococcus tauri]	primary metabolic process; macromolecule metabolic process; cellular metabolic process	DNA metabolic process;binding;hydrolase activity	Peptidase S16, Lon protease; AAA+ ATPase, core; DNA repair protein RadA
391	CAL54485.1	putative translational activator (ISS) [Ostreococcus tauri]			HEAT; Armadillo-like helical; Armadillo-type fold
597	CAL55258.1	putative chloroplast 1-hydroxy-2-methyl-2-(ISS) [Ostreococcus tauri]	primary metabolic process; response to other organism; response to biotic stimulus; biosynthetic process; cellular metabolic process	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase activity;response to bacterium;4 iron, 4 sulfur cluster binding;chloroplast;isopentenyl diphosphate biosynthetic process; mevalonate-independent pathway	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase, bacterial-type; Dihydropteroate synthase-like; 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase, atypical

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
156	CAL44986.1	putative Na/H antiporter [Cymodocea nodosa]	regulation of biological quality; transport	integral to membrane;regulation of pH;transport;solute:hydrogen antiporter activity	Cyclic nucleotide-binding; Sodium/hydrogen exchanger; RmlC-like jelly roll fold
723	ABJ55956.1	NBS-containing resistance-like protein [Prunus avium]			NB-ARC
263	ABK76304.1	chloroplast sedoheptulose-1,7-bisphosphatase [Morus alba var.multicaulis]	photosynthesis; primary metabolic process, macromolecular metabolic process; carbon utilization; biosynthetic process; cellular metabolic process	reductive pentose-phosphate cycle;sucrose biosynthetic process;sedoheptulose-bisphosphatase activity;protein binding;starch biosynthetic process;chloroplast	Inositol phosphatase/fructose-1,6-bisphosphatase
586	ABL97963.1	hydrogen-transporting ATP synthase [Brassica rapa]	generation of precursor metabolites and energy; primary metabolic process; biosynthetic process; cellular metabolic process; transport	hydrogen ion transporting ATP synthase activity, rotational mechanism;ATP synthesis coupled proton transport;metal ion binding;mitochondrial proton-transporting ATP synthase, catalytic core;hydrogen ion transporting ATPase activity, rotational mechanism	ATPase, F1 complex, delta/epsilon subunit

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
187	CAC19876.1	acetyl-CoA carboxylase [Brassica napus]	organization of an anatomical structure; response to chemical stimulus; anatomical structure development; reproductive developmental process; embryonic development; response to endogenous stimulus; anatomical structure morphogenesis; multicellular organismal development; primary metabolic process; biosynthetic process; cellular metabolic process	meristem organization;biotin carboxylase complex;embryonic development ending in seed dormancy;root development;ATP binding;shoot development;response to cytokinin stimulus;biotin carboxylase activity;metabolic process;biotin binding;acetyl-CoA carboxylase activity	Carboxyl transferase; Biotin/lipoyl attachment; Carbamoyl phosphate synthetase, large subunit, ATP-binding; Biotin carboxylase, C-terminal; Single hybrid motif; Rudiment single hybrid motif; ATP-grasp fold; Acetyl-coenzyme A carboxyltransferase, N-terminal; Biotin carboxylation region; Acetyl-CoA carboxylase, central region; ATP-grasp fold, subdomain 2; Pre-ATP-grasp fold; PreATP-grasp-like fold
800	AAC49545.1	alcohol dehydrogenase		cytoplasm;metabolic process;alcohol dehydrogenase activity;zinc ion binding	Alcohol dehydrogenase superfamily, zinc-containing; GroES-like; Alcohol dehydrogenase, zinc-binding; Alcohol dehydrogenase GroES-like; NAD(P)-binding
818	ABN09053.1	Cyclin-like F-box; Galactose oxidase, central [Medicago truncatula]		plastid	Cyclin-like F-box; Galactose oxidase/kelch, beta-propeller; Kelch-type beta propeller

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
223	CAA29003.1	P700 chlorophyll a-apoproteins 84 KD protein [Pisum sativum]	photosynthesis; primary metabolic process; macromolecule metabolic process; cellular metabolic process; transport	electron carrier activity;iron ion binding;electron transport;transport;magnesium ion binding;protein-chromophore linkage;chlorophyll binding;thylakoid membrane;4 iron, 4 sulfur cluster binding;integral to membrane;chloroplast;photosystem I;photosynthesis	Photosystem I psaA and psaB; Photosystem I psaA
71	EAZ11596.1	hypothetical protein OsJ_001421 [Oryza sativa (japonica cultivar-group)]		nucleic acid binding;mitochondrion	
693	EAZ32923.1	hypothetical protein OsJ_016406 [Oryza sativa (japonica cultivar-group)]			FAR1
561	ABO69625.1	synaptonemal complex protein ZYP1 [Brassica oleracea]	cell cycle process; primary metabolic process; macromolecule metabolic process; cellular component organization and biogenesis; cellular metabolic process; cell cycle	meiotic recombination;synapsis	
745	CAN71106.1	hypothetical protein [Vitis vinifera]			RNA-directed DNA polymerase (reverse transcriptase); Endonuclease/exonuclease/phosphatase; Sorting nexin, C-terminal; RNA-directed DNA polymerase (reverse transcriptase), related

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
600	CAN63377.1	hypothetical protein [Vitis vinifera]			Integrase, catalytic core; Transcription factor E2F/dimerisation partner (TDP); Winged helix repressor DNA-binding; Polynucleotidyl transferase, Ribonuclease H fold; Transcription factor DP, C-terminal; Transcription factor DP
350	NP_177466.1	nascent polypeptide-associated complex (NAC) domain-containing protein [Arabidopsis thaliana]	regulation of biological process; regulation of cellular process; regulation of metabolic process; cellular metabolic process	mitochondrion; transcription factor activity	Nascent polypeptide-associated complex NAC
798	NP_189551.1	glycine-rich protein [Arabidopsis thaliana]			
686	NP_180685.1	ATCDT1A/CDT1/CDT1A (ARABIDOPSIS HOMOLOG OF YEAST CDT1 A); cyclin-dependent protein kinase/ protein binding [Arabidopsis thaliana]			DNA replication factor CDT1-like
686	NP_190566.1	phototropic-responsive protein, putative [Arabidopsis thaliana]	response to abiotic stimulus	endomembrane system; response to light stimulus	NPH3; BTB/POZ fold

159

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
430	NP_191562.1	SPL12 (SQUAMOSA PROMOTER-BINDING PROTEIN-LIKE 12); transcription factor [Arabidopsis thaliana]	regulation of biological process; primary metabolic process; regulation of cellular process; regulation of metabolic process; cellular metabolic process	nucleus;regulation of transcription	Ankyrin; Transcription factor, SBP-box
961	NP_201224.1	unknown protein [Arabidopsis thaliana]		plastid	
166	BAF79940.1	receptor-like kinase [Marchantia polymorpha]	primary metabolic process; macromolecule metabolic process; cellular metabolic process; nitrogen compound metabolic process	protein amino acid phosphorylation;cytoplasmic membrane-bounded vesicle;ATP binding;protein serine/threonine kinase activity	Protein kinase, core; Tyrosine protein kinase; Protein kinase-like
347	NP_199328.1	RTL3 (RNASE THREE-LIKE PROTEIN 3); double-stranded RNA binding / ribonuclease III [Arabidopsis thaliana]	regulation of biological process; primary metabolic process; macromolecule metabolic process; response to chemical stimulus; regulation of cellular process; anatomical structure development; multicellular organismal development; cellular metabolic process; primary metabolic process	anatomical structure development;double-stranded RNA binding;multicellular organismal development;gene silencing by miRNA, production of miRNAs;nucleus	Ribonuclease III; Double-stranded RNA binding; Double-stranded RNA-binding-like

160

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
588	ABU75152.1	AtpB [Brucea javanica]			ATPase, F1/V1/A1 complex, alpha/beta subunit, nucleotide-binding; ATPase, F1/V1/A1 complex, alpha/beta subunit, C-terminal; AAA+ ATPase, core; ATPase, F1/V1/A1 complex, alpha/beta subunit, N-terminal; ATPase, F1 complex, beta subunit
701	CAO43835.1	unnamed protein product [Vitis vinifera]			
333	BAF81517.1	putative lactoylglutathione lyase [Brassica rapa]	primary metabolic process; response to abiotic stimulus; response to stress; cellular metabolic process	carbohydrate metabolic process; zinc ion binding; lactoylglutathione lyase activity; response to cold	Glyoxalase/bleomycin resistance protein/dioxygenase; Glyoxalase I; Glyoxalase/extradiol ring-cleavage dioxygenase
472	XP_001698933.1	hypothetical protein CHLREDRAFT_193403 [Chlamydomonas reinhardtii]			
836	EDQ48531.1	predicted protein [Physcomitrella patens subsp. patens]			
336	EDQ48928.1	predicted protein [Physcomitrella patens subsp. patens]			
853	EDQ64560.1	predicted protein [Physcomitrella patens subsp. patens]			Peptidase aspartic, catalytic

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
628	XP_001782328.1	predicted protein [Physcomitrella patens subsp. patens]		nucleic acid binding	Zinc finger, CCHC-type; Thaumatin, pathogenesis-related; Zinc finger, CCHC retroviral-type
709	AAL33589.1	methionine synthase [Zea mays]	primary metabolic process; nitrogen compound metabolic process; biosynthetic process; cellular metabolic process	cytoplasm;methionine synthase activity;methionine biosynthetic process;zinc ion binding;5-methyltetrahydropteroyltriglutama te-homocysteine S-methyltransferase activity	Methionine synthase, vitamin-B12 independent; 5-methyltetrahydropteroyltrig lutamate—homocysteine S-methyltransferase; Cobalamin (vitamin B12)-independent methionine synthase MetE, N-terminal
756	XP_001786655.1	predicted protein [Physcomitrella patens subsp. patens]			
734	AAL58654.1	ribulose-1,5-bisphosphate carboxylase/oxygenase [Hexalectris revoluta]	photosynthesis; primary metabolic process; carbon utilization, cellular metabolic process	reductive pentose-phosphate cycle;ribulose-bisphosphate carboxylase activity;monooxygenase activity;photorespiration;magnesi um ion binding;chloroplast ribulose bisphosphate carboxylase complex	Ribulose bisphosphate carboxylase, large subunit, C-terminal; Ribulose bisphosphate carboxylase, large subunit, ferrodoxin-like N-terminal; Ribulose bisphosphate carboxylase, large subunit, N-terminal
297	BAA19156.1	HMG-1 [Canavalia gladiata]	regulation of biological process; primary metabolic process; macromolecule metabolic process; regulation of cellular process; regulation of metabolic process; cellular component organization and biogenesis; cellular metabolic process	chromatin;DNA binding;chromatin assembly or disassembly;chromatin binding;structural constituent of chromatin;nucleus;regulation of transcription, DNA-dependent	High mobility group box, HMG1/HMG2; High mobility group box, HMG

162

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
838	NP_568662.1	O-acetyltransferase-related [Arabidopsis thaliana]	macromolecule metabolic process; biosynthetic process	endomembrane system;cytoplasmic membrane-bounded vesicle;plastid;O-acetyltransferase activity	Cas1p-like
796	ZP_02965050.1	NHL repeat containing protein [bacterium Ellin514]			NHL repeat; Six-bladed beta-propeller, TolB-like; NHL repeat, subgroup
451	AAL77589.1	chloroplast ribose-5-phosphate isomerase [Spinacia oleracea]	photosynthesis; primary metabolic process; macromolecule metabolic process; carbon utilization; catabolic process; biosynthetic process; cellular metabolic process	cytoplasm;pentose-phosphate shunt, non-oxidative branch;reductive pentose-phosphate cycle;uridine biosynthetic process;ribose-5-phosphate isomerase activity	Ribose 5-phosphate isomerase
959	AAL27805.1	ATPase beta subunit [Cycas revoluta]	generation of precursor metabolites and energy; primary metabolic process; biosynthetic process; cellular metabolic process; transport	proton-transporting ATP synthase complex, catalytic core F(1);hydrogen-exporting ATPase activity, phosphorylative mechanism;hydrogen ion transporting ATP synthase activity, rotational mechanism;integral to membrane;ATP synthesis coupled proton transport;metal ion binding;thylakoid;ATP binding;hydrogen ion transporting ATPase activity, rotational mechanism;chloroplast	ATPase, F1/V1/A1 complex, alpha/beta subunit, nucleotide-binding; ATPase, F1/V1/A1 complex, alpha/beta subunit, C-terminal; AAA+ ATPase, core; ATPase, F1/V1/A1 complex, alpha/beta subunit, N-terminal; ATPase, F1 complex, beta subunit
516	AAM77651.1	cpl0-like protein [Gossypium hirsutum]	primary metabolic process; macromolecule metabolic process; cellular metabolic process	protein folding;ATP binding;chloroplast	Chaperonin Cpn10; GroES-like; Chaperonin 21, chloroplast

163

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
928	NP_195035.2	APM1 (Aminopeptidase M1) [Arabidopsis thaliana]	transport	auxin polar transport;tyrosine aminopeptidase activity	Peptidase M1, membrane alanine aminopeptidase; Peptidase M1, membrane alanine aminopeptidase, N-terminal; Peptidase M1, puromycin-sensitive aminopeptidase
605	NP_680770.1	FAT domain-containing protein / phosphatidylinositol 3- and 4-kinase family protein [Arabidopsis thaliana]			Phosphatidylinositol 3- and 4-kinase, catalytic; PIK-related kinase, FATC; Protein kinase-like; PIK-related kinase; Armadillo-type fold
614	CAA10904.1	asparaginyl-tRNA synthetase [Arabidopsis thaliana]	primary metabolic process; macromolecule metabolic process; nitrogen compound metabolic process; biosynthetic process; cellular metabolic process	asparagine-tRNA ligase activity;asparaginyl-tRNA aminoacylation;mitochondrion;nucleotide binding	Aspartyl-tRNA synthetase, class IIb; Aminoacyl-tRNA synthetase, class II (D, K and N); Nucleic acid binding, OB-fold, tRNA/helicase-type; Asparaginyl-tRNA synthetase, class IIb; Aminoacyl-tRNA synthetase, class II; Nucleic acid-binding, OB-fold-like
688	BAC16059.1	putative cis-zeatin O-glucosyltransferase [Oryza sativa Japonica group]		transferase activity, transferring hexosyl groups;mitochondrion;plastid	UDP-glucuronosyl/UDP-glucosyltransferase
574	BAA22288.1	polyprotein [Oryza australiensis]	primary metabolic process; macromolecule metabolic process; cellular metabolic process	mitochondrion;DNA binding;DNA recombination	Integrase, catalytic core; Zinc finger, CCHC-type; Polynucleotidyl transferase, Ribonuclease H fold; Zinc finger, CCHC retroviral-type; Reverse transcriptase, RNA-dependent DNA polymerase

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
609	BAC53939.1	MCM protein-like protein [Nicotiana tabacum]	regulation of biological process; primary metabolic process; macromolecule metabolic process; regulation of cellular process; regulation of metabolic process; cellular metabolic process	regulation of DNA replication initiation;DNA unwinding during replication;MCM complex;DNA binding;ATP binding;chromatin;DNA-dependent ATPase activity;chloroplast;chromatin binding	MCM; MCM protein 2; Nucleic acid-binding, OB-fold; Nucleic acid-binding, OB-fold-like
531	Q9GI85.1	Maturase K (Intron maturase)	primary metabolic process; macromolecule metabolic process; cellular metabolic process	mRNA processing;RNA splicing;chloroplast	Intron maturase, type II; Maturase-related, N-terminal
285	AAO22131.1	quinone oxidoreductase [Fragaria x ananassa]	response to abiotic stimulus; response to stress	zinc ion binding;metabolic process;oxidoreductase activity;response to cold	Alcohol dehydrogenase superfamily, zinc-containing; GroES-like; Alcohol dehydrogenase, zinc-binding; Alcohol dehydrogenase GroES-like; NAD(P)-binding
833	CAA34161.1	ribulose-1,5-carboxylase/oxygenase [Larix laricina]	photosynthesis; primary metabolic process; carbon utilization; response to abiotic stimulus; response to stress; cellular metabolic process	reductive pentose-phosphate cycle;ribulose-bisphosphate carboxylase activity;response to cold;monooxygenase activity;photorespiration;chloroplast ribulose bisphosphate carboxylase complex	Ribulose bisphosphate carboxylase, small chain

165

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
559	AAP73860.1	putative DNA gyrase subunit [Oryza sativa (japonica cultivar-group)]	primary metabolic process; macromolecule metabolic process; cellular metabolic process	DNA topoisomerase (ATP-hydrolyzing) activity;DNA unwinding during replication;ATP binding;chromosome;DNA topological change;chloroplast;mitochondrion	DNA topoisomerase, type IIA, subunit A or C-terminal; DNA gyrase, subunit A; DNA gyrase/topoisomerase IV, subunit A, C-terminal beta-pinwheel; DNA topoisomerase, type IIA, subunit A, alpha-helical; DNA topoisomerase, type IIA, subunit A or C-terminal, alpha-beta; DNA topoisomerase, type IIA, central
169	AAP84375.1	mutant FCA-D1 [Triticum aestivum]		protein binding;nucleotide binding;RNA binding	RNA recognition motif, RNP-1; Nucleotide-binding, alpha-beta plait; Ribonucleoprotein, BRUNO-like
778	BAA31157.1	leghemoglobin [Pisum sativum]	nitrogen compound metabolic process; transport	heme binding;oxygen transporter activity;oxygen transport;iron ion binding;nitrogen fixation;oxygen binding	Globin, subset; Leghaemoglobin; Globin-like; Globin

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
459	AAP03088.1	class Ia chitinase [Galega orientalis]	immune response; death; cellular developmental process; primary metabolic process; macromolecule metabolic process; catabolic process; cellular component organization and biogenesis; defense response; cell development; response to biotic stimulus; cellular metaboli process	chitin catabolic process; plant-type hypersensitive response; chitin binding; cell wall catabolic process; chitinase activity; vacuole; response to biotic stimulus	Glycoside hydrolase, family 19, catalytic; Chitin-binding, type 1; Glycoside hydrolase, family 19
838	AAR20302.1	maturase [Euonymus fortunei]	primary metabolic process; macromolecule metabolic process; cellular metabolic process	RNA splicing; RNA-directed DNA polymerase activity; RNA-dependent DNA replication; RNA binding; mitochondrion	Intron maturase, type II
467	BAD02824.1	putative class I chitinase [Taxodium distichum]	primary metabolic process; macromolecule metabolic process; catabolic process; cellular component organization and biogenesis; cellular metabolic process	chitinase activity; chitin catabolic process; cell wall catabolic process; chitin binding	Glycoside hydrolase, family 19, catalytic; Chitin-binding, type 1; Glycoside hydrolase, family 19
935	AAR24912.1	fructokinase 3 [Solanum lycopersicum]	primary metabolic process; cellular metabolic process; macromolecule metabolic process; catabolic process	fructokinase activity; D-ribose metabolic process; chloroplast; ribokinase activity	Ribokinase; Carbohydrate/purine kinase

167

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
455	BAD07864.1	putative secretory carrier membrane protein [Oryza sativa Japonica group]	cellular component organization and biogenesis; secretion; establishment of cellular localization; cellular localization; transport	early endosome;transmembrane transporter activity;plasma membrane;secretory pathway;integral to membrane;receptor-mediated endocytosis	SCAMP
658	NP_179846.2	kinesin motor protein-related [Arabidopsis thaliana]			Kinesin, motor region
590	AAD25591.1	Mutator-like transposase [Arabidopsis thaliana]			Transposase, MuDR, plant; Zinc finger, PMZ-type
97	BAD16023.1	hypothetical protein [Oryza sativa Japonica Group]			
473	AAT09427.1	class II chitinase [Picea abies]	primary metabolic process; macromolecule metabolic process; response to other organism; catabolic process; cellular component organization and biogenesis;response to biotic stimulus; cellular metabolic process	endochitinase activity;chitin catabolic process;beta-N-acetylglucosaminidase activity;cytoplasmic membrane-bounded vesicle;cell wall catabolic process;response to other organism;chitin binding	Glycoside hydrolase, family 19, catalytic; Glycoside hydrolase, family 19

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
930	CAA83548.1	PsHSC71.0 [Pisum sativum]	primary metabolic process; macromolecule metabolic process;response to other organism; response to biotic stimulus; response to stress; cellular metabolic process	protein folding;cytosol;ATP binding;response to stress;response to virus;protein binding	Heat shock protein Hsp70; Heat shock protein 70
106	AAT38562.1	chloroplast serine acetyltransferase [Thlaspi goesingense]	primary metabolic process; nitrogen compound metabolic process; biosynthetic process; cellular metabolic process; macromolecule metabolic process	serine O-acetyltransferase activity;cysteine biosynthetic process from serine;chloroplast;cytosol	Bacterial transferase hexapeptide repeat; Serine O-acetyltransferase; Serine acetyltransferase, N-terminal; Trimeric LpxA-like
457	BAD29556.1	hydroxyproline-rich glycoprotein family protein-like [Oryza sativa Japonica group]			
463	BAD44910.1	zinc knuckle domain containing protein-like [Oryza sativa Japonica group]		nucleic acid binding	Zinc finger, CCHC-type
696	BAD45437.1	speckle-type POZ protein-like [Oryza sativa (japonica cultivar-group)]		protein binding	BTB/POZ-like; BTB/POZ fold; BTB/POZ; Kelch related
675	CAH58634.1	thioredoxin-dependent peroxidase [Plantago major]	response to chemical stimulus; response to stress	peroxidase activity	Thioredoxin fold; Thioredoxin-like fold; Redoxin

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
400	AAA21277.1	2-phospho-D-glycerate hydrolase	primary metabolic process; macromolecule metabolic process; catabolic process;response to abiotic stimulus; cellular metabolic process; nitrogen compound metabolic process; biosynthetic process	acyltransferase activity;phosphopyruvate hydratase complex;phosphopyruvate hydratase activity;magnesium ion binding;glycolysis;nucleus;response to light stimulus	Enolase
417	BAD69015.1	putative auxin-independent growth promoter [Oryza sativa (japonica cultivar-group)]		mitochondrion	Protein of unknown function DUF246, plant
371	AAD52863.1	maturase-like protein [Adesmia volckmannii]	primary metabolic process; macromolecule metabolic process; cellular metabolic process	mRNA processing;RNA splicing;chloroplast	Intron maturase, type II; Maturase-related, N-terminal
491	BAD90814.1	thaumatin-like protein [Cryptomeria japonica]	response to other organism; response to biotic stimulus	response to other organism	Thaumatin, pathogenesis-related
673	AAF08180.1	NADH dehydrogenase subunit F [Cercidiphyllum japonicum]	generation of precursor metabolites and energy; cellular metabolic process; biosynthetic process; transport	quinone binding;NADH dehydrogenase (ubiquinone) activity;photosystem II;electron transport;integral to membrane;ATP synthesis coupled electron transport;chloroplast	NADH-Ubiquinone oxidoreductase (complex I), chain 5/L, N-terminal; NADH/Ubiquinone/plastoquinone (complex I); NADH dehydrogenase (ubiquinone), chloroplast chain 5, C-terminal; NADH-ubiquinone oxidoreductase, chain 5; NADH-plastoquinone oxidoreductase, chain 5

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
108	BAA92724.1	putative chaperonin 60 beta precursor [Oryza sativa Japonica group]	immune response; death; cellular developmental process; primary metabolic process; macromolecule metabolic process; response to other organism; defense response; cell development; response to biotic stimulus; cellular metabolic process	systemic acquired resistance;chaperone cofactor-dependent protein folding;ATP binding;cell death;unfolded protein binding;chloroplast	Chaperonin Cpn60; Chaperonin Cpn60/TCP-1; GroEL-like chaperone, ATPase; Chaperonin GroEL
425	BAA94510.1	protein kinase 2 [Populus nigra]	primary metabolic process; macromolecule metabolic process; cellular metabolic process; nitrogen compound metabolic process	protein amino acid phosphorylation;ATP binding;protein serine/threonine kinase activity;protein-tyrosine kinase activity	Protein kinase, core; Tyrosine protein kinase; Protein kinase-like
298	ABA46972.1	cysteine protease Mir1 [Zea diploperennis]	primary metabolic process; macromolecule metabolic process; cellular metabolic process	proteolysis;cysteine-type endopeptidase activity	Peptidase C1A, papain C-terminal; Peptidase C1A, papain; Proteinase inhibitor I29, cathepsin propeptide
562	ABA97297.1	hypothetical protein LOC_Os12g15610 [Oryza sativa (japonica cultivar-group)]			
98	NP_001031936.1	hydrolase, hydrolyzing O-glycosyl compounds [Arabidopsis thaliana]	primary metabolic process	hydrolase activity, hydrolyzing O-glycosyl compounds;carbohydrate metabolic process;cytoplasmic membrane-bounded vesicle	Glycoside hydrolase, family 17; X8; Glycoside hydrolase, subgroup, catalytic core; Glycoside hydrolase, catalytic core

171

Spot ID	Top BLAST hit	BLASTp annotation	GO Terms (third level)	GO Terms (lower terms)	Interpro annotation
604	AAF76340.1	NADH dehydrogenase subunit F [Penaea mucronata]	generation of precursor metabolites and energy; cellular metabolic process; biosynthetic process; transport	quinone binding;NADH dehydrogenase (ubiquinone) activity;photosystem II;electron transport;ATP synthesis coupled electron transport;chloroplast	NADH/Ubiquinone/plastoquinone (complex I); NADH dehydrogenase (ubiquinone), chloroplast chain 5, C-terminal; NADH-plastoquinone oxidoreductase, chain 5; Calycin-like
328	ABD32884.1	SH2 motif; Alcohol dehydrogenase, zinc-containing; Resolvase, Rnase H-like fold; Nucleic acid binding, OB-fold, subgroup [Medicago truncatula]	cellular component organization and biogenesis	chromatin assembly or disassembly	SH2 motif; S1, RNA binding; Resolvase, RNase H-like fold; Nucleic acid-binding, OB-fold; Nucleic acid-binding, OB-fold-like; Transcription elongation factor Spt6
376	CAB96173.1	enolase [Spinacia oleracea]	primary metabolic process; macromolecule metabolic process; catabolic process; cellular metabolic process; nitrogen compound metabolic process; biosynthetic process	acyltransferase activity;phosphopyruvate hydratase complex;phosphopyruvate hydratase activity;magnesium ion binding;glycolysis	Enolase

172

Appendix 7.3

Chitinase nucleotide sequences representing the three non-redundant hits from nine spots found upregulated after 10 weeks of short days. The sequences correspond to the contig ID numbers found in the spruce gene catalog (GCAT, https://genome.ulaval.ca/arborea/gcat/login).

```
>GQ0134.B7.1_C02:53:Contig1:1-1185:+
taagatggaaactagtgtccgacaaatgaaatcgatgaggttctccctgatggtgatggccttggttacaatggcgacgatgaatttct
attttgtgtctgcagagcaatgcggacggcaagcaagcggagtgctctgtaccggcggcgtacagggagtctcatctgttatcactgaa
agcattttcaatgatttattgaagcacagaaacgacgccgtttgcaaggccaaaggattctactcctacgccgccttcattgcggctgc
cgatgctttccccggtttcggcaccaccggcgatctcactactcagaaaagagaactcgctgctttctttggccaaacctcccacgaaa
ccacaggcggatggccgacggccccagacggcccgtacgcgtggggttattgtttcaaagaggagcaaggcaatcctcccggcgagtac
tgccaggcgtcctctgaatggccctgtgcatctgggaagagatactacggacgagggcccgttcaattatcatggaactacaactatgg
accggccgggaaggcagttggattcgacggcataaacaacccggacattgttgctaacgatgcgacggtctcttttaagacggcggtgt
ggttctggatgaccgcacagtctccaaagccttcctgccacaacgtgatggccgggggatggggtccgtccagcagcgacaccgccgcc
gggagagcggcgggatatggagtggtaaccaacatcatcaacggcgggttggagtgcgggaaaggcagcgactcgaggcaggaggaccg
catcgggttctacaaaagatactgcgacattctcggcgtgagctacggatcaaacctggactgcaacagccagaagcctttcggctttg
cggctcaatctcaacctcgccttatcaaaaccgtggtttaaatactgttcataatctgtgtttttgtggtggcctttgcgttgttagtct
tgcagtgcgatctccatctgtttgccgcataataaaatagaataaaatccatataaattaactctagcattcgacgagcgtggtgtggt
tttttaactgttgagaatcatgtgggtggatttggtattaaccatgattaggatccatatatccaatctccattgaggtttcgagcatg
tcaatatcttcctgaatttaaaatcaaa
```

```
>GQ03602.B7_D17:5-:Contig3:1-897:+
aacagaatttgaattttatattcgaagtctaatatggcgtccgtgacaatgaagtggatgagggtgctgttggcactcacggtagtggc
gatgatgaatactttctgttatgtttctgcacaacagggagtcgcttttatcataactgaagatattttttaatgaatttctgaagcaca
gaacgacggcgcatgcgcggccaaaggattctacacctacaactccttcattctggctgccaatcgcttccccagcttcggtaatgtc
gggaacctcgaaacacgcaaaagagagctcgctgctttctttgcacagacctcgcacgaaaccaccggcgggtgggagacggccccaga
cggcccatacgcgtggggttactgcttcaaagaggagcaagatgccgatcctgaagaaatgtactacggacgaggacccattcagctaa
ccggaaagaacaactacgaagcggccgggaaagctttaggatacgatctgataaacaacccgaacattgtggcgtctaatcccaccatt
tcgttcaagacggcggtgtggttctggatgaacccgcagcctccgaagccttcctcccacatcgtgatgattggaaaatggagtccgtc
gggcagcgatacggccgcgggcagagtgccgggatatggagtggtgaccaatataatcaacggcggggtggagtgcgggagaggcagcg
actccaatcaggaggaccgcatcggtttctacaagaaatactgtgacattctcggcgcgagctatggatccaacatcgactgcaacagc
cagaggtctttcggaggttaatctgaacttcaactcaactcgtcacatcataacgactgatgttcatgttatcagcacattcatagtct
acttata
```

```
>GQ03602.B7_D17:53:Contig1:1-1101:+
ggtaaggccattacggccaggggagcagaatttgaatttcatattcgaagtctaatatggcgtccgtgacaatgaagtggatgagggtg
ctgttggcactcacggtagtggcgatgatgaatactttctgttatgtttctgcacaacagggagtcgctttcatcataactgaagatgt
tttcaatcgatttctgaagcacagaaacgacggtgcatgcgcggccaaaggattctacacctacaactccttcattctggctgccaatc
gtttccccagcttcggcaatgtcgggaacctcgagacacgcaaaagagagctcgctgctttctttgcacagacctcgcacgaaaccacc
ggcgggtgggagaccgccaccagacggccccatacgcgtggggttattgcttcaaagaggagcaagatgccgatcccgaagaaatgtacta
cggacgaggacccattcaactaaccggaaagagcaactacgaagctgccgggaaagcgttagggtacgatctgttaaacaatccgggca
ttgtggcgtccaatcccaccatttcgtttaagacggcggtctggttctggatgaccacgcagtctccgaagccttctcccacactgtg
atgattggaaaatggagtccgtcgggcagtgatacggccgcgggcagagtgccgggatatggagtggtgaccaacataatcaacggcgg
ggtggagtgcgggagaggcagcgactcaaatcaagaggaccgcatcggtttctacaaggcatactgtgacattctcggcgcgagctacg
gatcgaacgtcgactgcaatactcagaggtctttcggaggttaatctgaacctcaactcaactcgtcacatcaacactcactcagtcgc
tctctctcgcataacaactactgttcactgttatcagcacattcatagtctacttataggacagatgtgctcactgctatgcatggatg
tatagtacgtctaataaataatccgagatttgtggtggcgactgtgtagccagggtgcttcgagttgctgagaagcatattataaatat
gaatataataaaatggaaattaattacaaaaaa
```

Appendix 7.4

Chitinase phylogenetic trees constructed using two different alignment programs (CLUSTAL and MUSCLE) and 4 different phylogenetic algorithms neighbor joining (NJ), minimum evolution (ME), unweighted pair group method with arithmetic mean (UPGMA) and maximum parsimony (MP). Sequences in the trees include the putative translation products of chitinase-like sequences from the spruce gene catalog (GCAT), chitinase protein sequences previously reported by Xu *et al.* (2007) and additional previously characterized chitinases from other plant species. These sequences are reported at the end of this appendix. Descriptions of each tree are given with each tree. The names of each sequence correspond to the genus and species first letters followed by the genbank identifier or the GCAT contig identifier. Pg (*Picea glauca*), Cj (*Cryptomeria japonica*), Pa (*Picea abies*), At (*Arabidopsis thaliana*), Bn (*Brassica napus*), Os (*Oryza sativa*), Dj (*Dioscorea japonica*), Dc (*Daucus carota*), Vv (*Vitis vinifera*), Cd (*Cynodon dactylon*), Hv (*Hordeum vulgare*), Sc (*Secale cereale*), St (*Solanum tuberosum*), Ca (*Capsicum annum*), Nt (*Nicotiana tabacum*), Ta (*Triticum aestivum*), Zm (*Zea mays*), Gh (*Gossypium hirsutum*), Mc (*Momordica charantia*), Rg (*Rehmmania glutinosa*), La (*Lupinus albus*), Sr (*sesbania rostrata*), Ms (*Medicago sativa*), Ma (*Musa acuminata*).

Phylogenetic trees using alignments from clustal

NJ
1000 bootstrap

The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 581 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.

175

ME
1000 bootstrap

Cluster IV

Cluster I

Cluster VI

Cluster V

Cluster II

Cluster III

The evolutionary history was inferred using the Minimum Evolution method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm at a search level of 3. The Neighbor-joining algorithm was used to generate the initial tree. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 581 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.

176

UPGMA
1000 bootstrap

The evolutionary history was inferred using the UPGMA method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 581 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.

Cluster IV
Cluster I
Cluster VI
Cluster II
Cluster V
Cluster III

MP
1000 bootstrap

Cluster III

Cluster II

Cluster V

Cluster VI

Cluster IV

Cluster I

The evolutionary history was inferred using the Maximum Parsimony method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The MP tree was obtained using the Close-Neighbor-Interchange algorithm with search level 3 in which the initial trees were obtained with the random addition of sequences (10 replicates). All alignment gaps were treated as missing data. There were a total of 581 positions in the final dataset, out of which 437 were parsimony informative. Phylogenetic analyses were conducted in MEGA4.
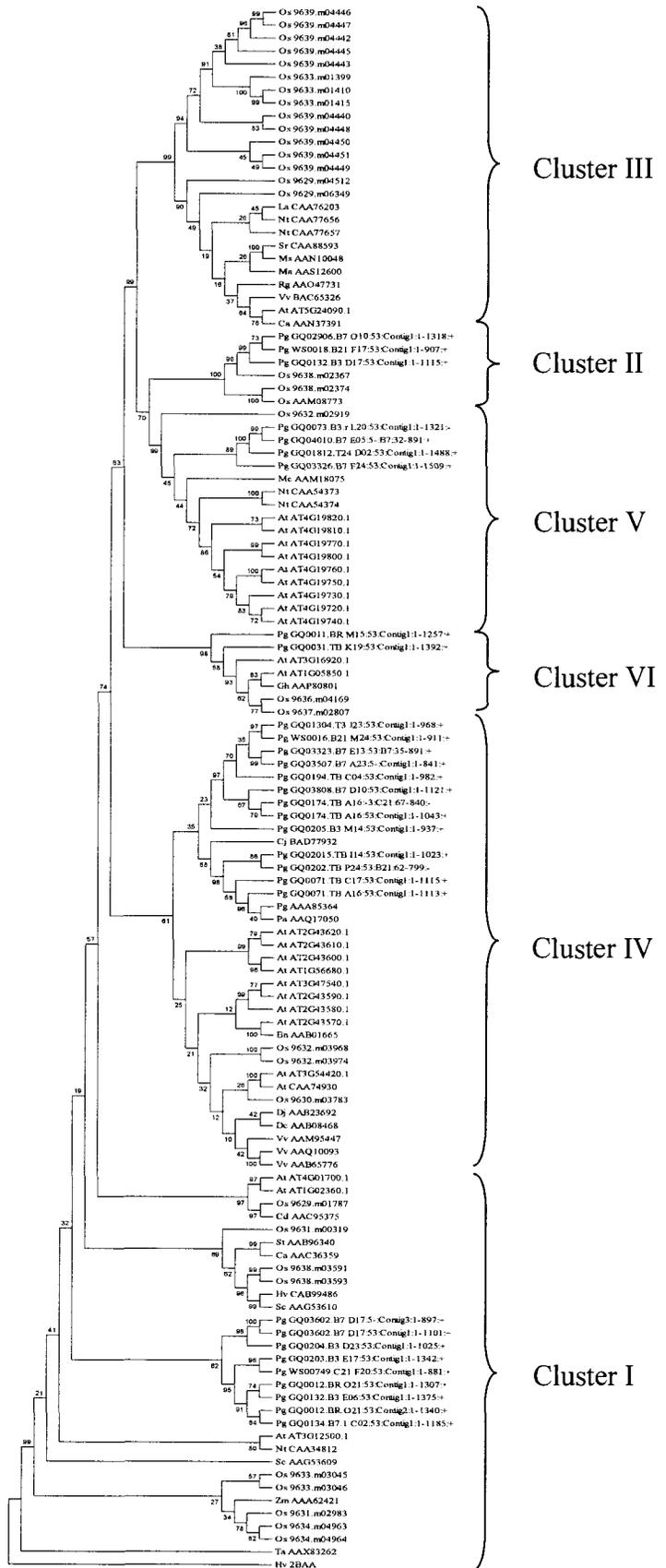
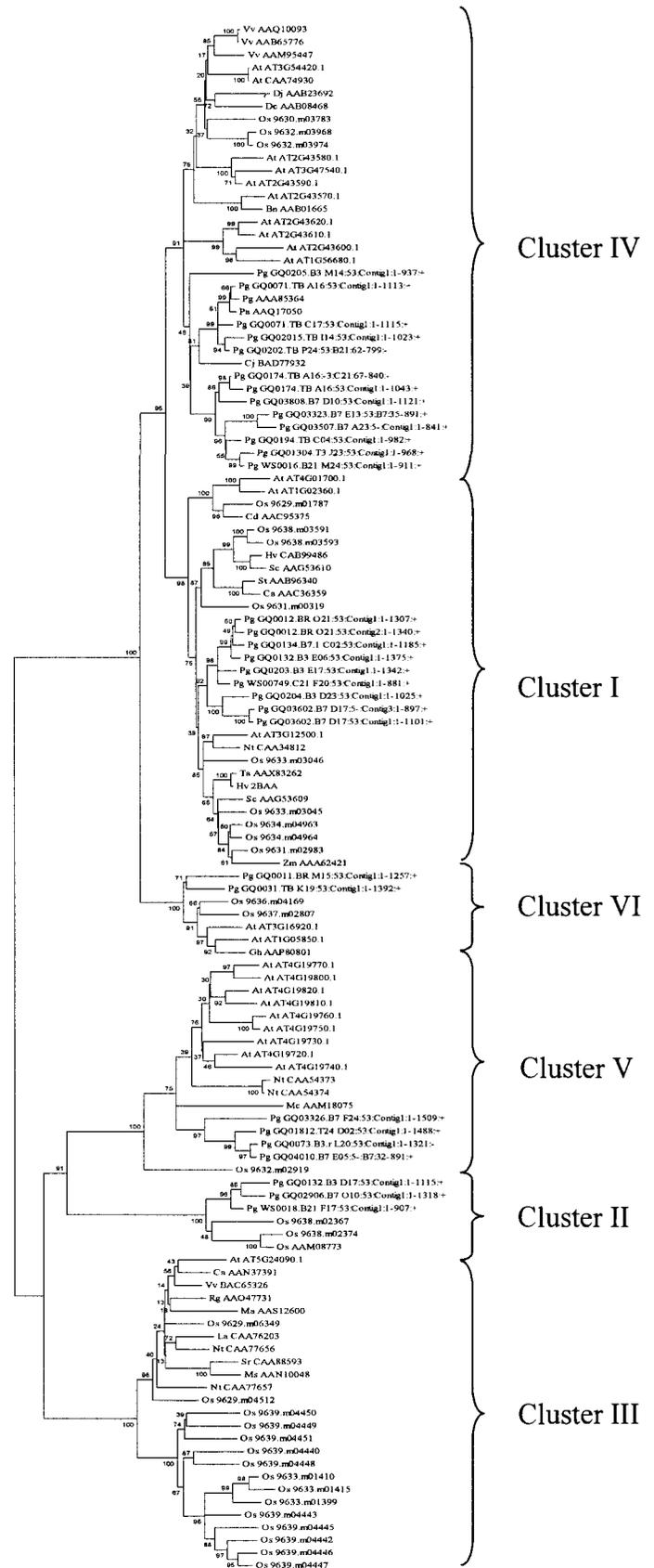# Phylogenetic trees using alignments from MUSCLE

NJ
1000 bootstrap

The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 593 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.

Cluster IV

Cluster I

Cluster VI

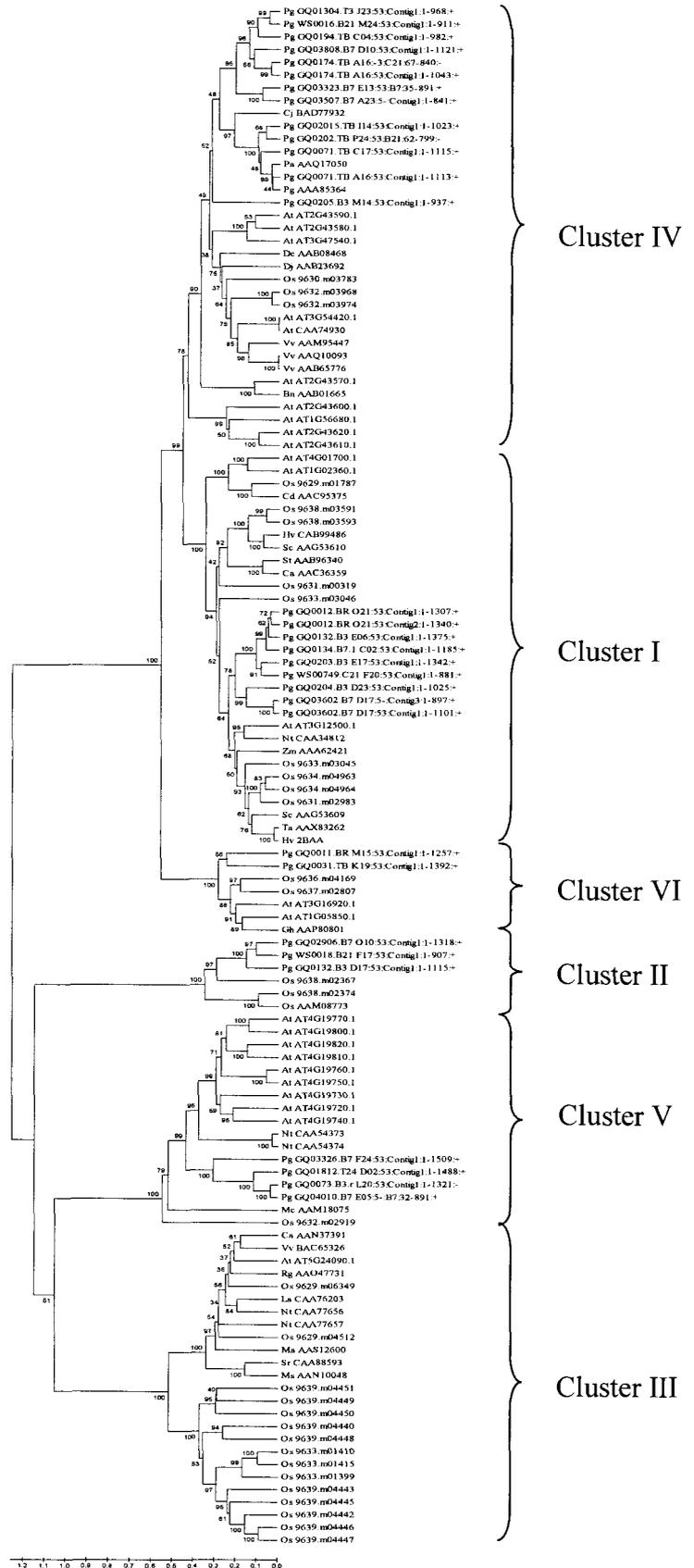Cluster V

Cluster II

Cluster III

179

ME
1000 bootstrap

The evolutionary history was inferred using the Minimum Evolution method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm at a search level of 1. The Neighbor-joining algorithm was used to generate the initial tree. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 593 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.

Cluster IV

Cluster I

Cluster VI

Cluster V

Cluster II

Cluster III

0.2

180

UPGMA
1000 bootstrap

Cluster IV

Cluster I

Cluster VI

Cluster II

Cluster V

Cluster III

The evolutionary history was inferred using the UPGMA method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 593 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.
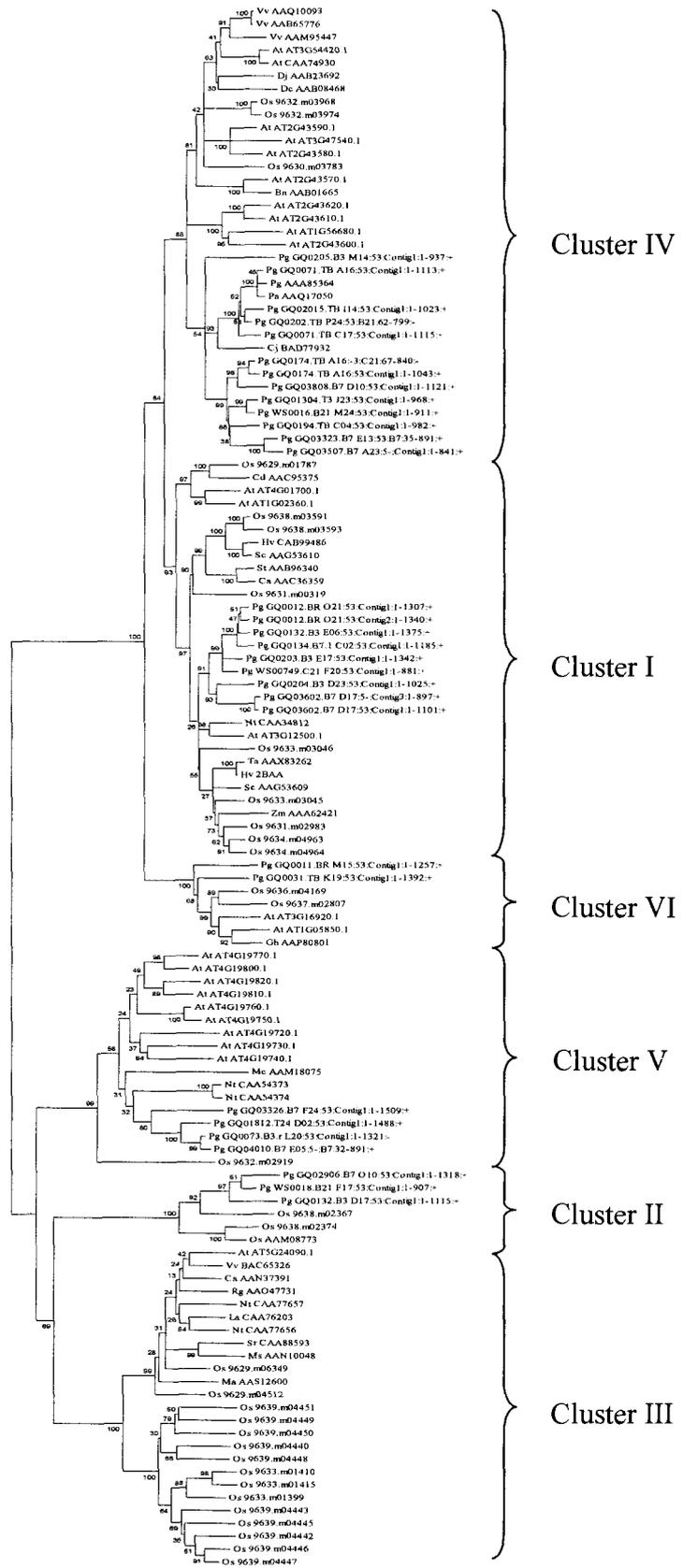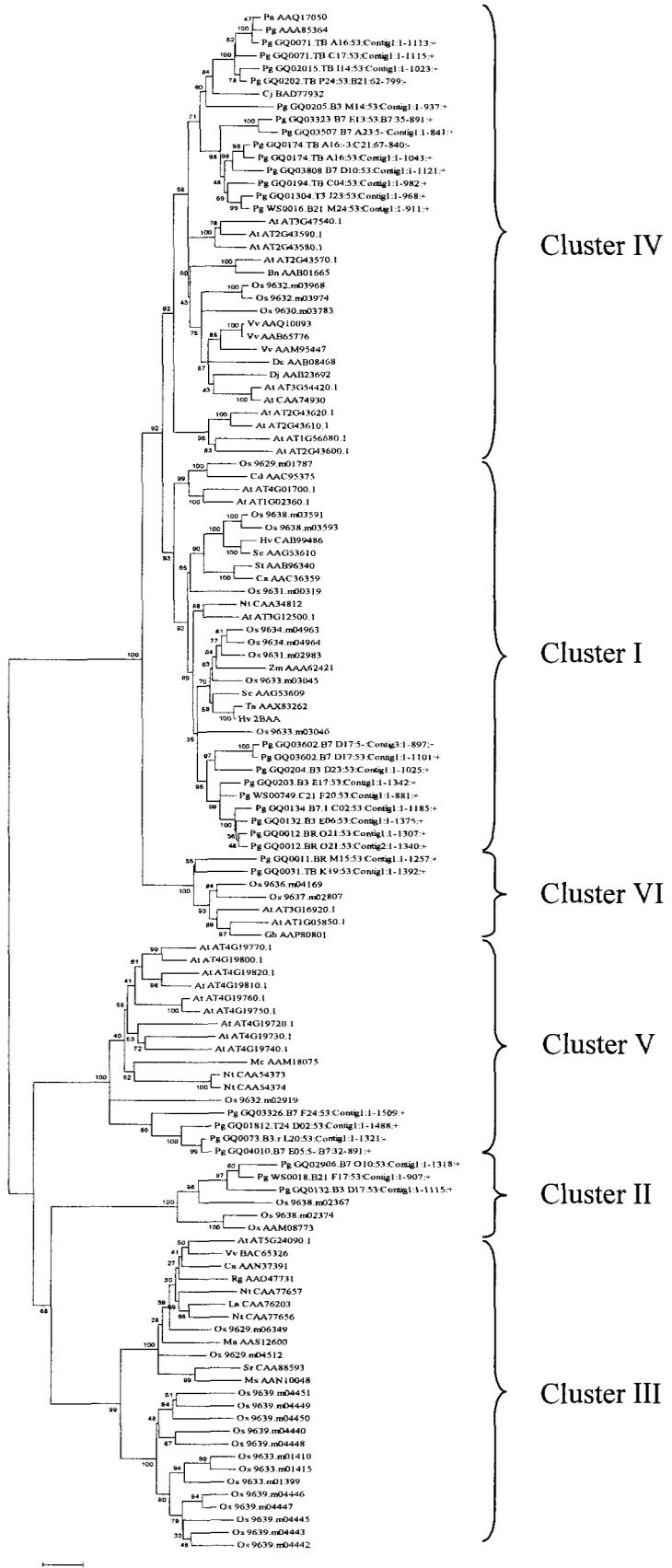
MP
1000 bootstrap

The evolutionary history was inferred using the Maximum Parsimony method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The MP tree was obtained using the Close-Neighbor-Interchange algorithm with search level 3 in which the initial trees were obtained with the random addition of sequences (10 replicates). All alignment gaps were treated as missing data. There were a total of 593 positions in the final dataset, out of which 455 were parsimony informative. Phylogenetic analyses were conducted in MEGA4.

Cluster IV

Cluster I

Cluster VI

Cluster V

Cluster II

Cluster III

Amino acid sequences used for phylogenetic trees.

```
>Pg_GQ0011.BR_M15:53:Contig1:1-1257:+
METKQIIGSVNRMFVVVGICLMISSWLCCIEPAAAADEDQTTKSKKISCIKGAECKNKTISELFTVDQFESLFSHRNAPMAHAQGFWD
YHSFITAAAHFEPKGFGATGGDLVQKKELAAFFAHVATETSCESLMAQSSTATTDSPTKWGLCYKEELSPDSTYCESSLVYPCAPGVS
YHGRGALPVYWNYNYGQLGQALKVDLLHHAEDLSQNATLAFQAAMWRWMNPIKVKQPSAHQVMVGKWVPTKNDTNSLRHPGFGMTINI
LKGEAECGAGSDDKQMNKRIAHYLYFLDQLDVGRDNAGDNLDCSDQKVLNPSSAST

>Pg_GQ0012.BR_O21:53:Contig1:1-1307:+
MKSMRFSVIAIALVTMATMNLYFVSAEQCGRQASGALCPGGLCCSKWGWCGNTEAHCGQDCQSQCGGSPPSGQGVASVITENIFNELL
NHRNDAGCKAKGFYSYAAFIAAANAFPGFGTTGDLATRKRELAAFFGQTSHETTGGWPTAPDGPYAWGYCFKEEQGNPPAEYCQASSQ
WPCAPGKRYYGRGPVQISWNYNYGPAGRAIGFDGINNPDIVANDATVSFKTAVWFWMTSQSPKPSCHDVMTGRWSPSGSDSAAGRAAG
YGVVTNIINGGLECGRGSDSRQQDRIGFYKRYCDILGVSYGSNLDCNNQRPFGFAAQSQPLLIKTVV

>Pg_GQ0012.BR_O21:53:Contig2:1-1340:+
MKSIRFSAMAIALVTMGTMNLYFASAEQCGRQASGALCPGGLCCSKWGWCGNTEAHCGQDCQSQCGGSTPTPPSPTPGGQGVSSVITE
SIFNDLLKHRNDAACKAKGFYSYAAFIAAANAFPAFGTTGDLSTRKRELAAFFGQTSHETTGGWPAAPDGPYAWGYCFKEEQGNPPGE
YCQASSQWPCASGKRYYGRGPVQISWNYNYGPAGRAIGFDGINNPDIVANDATVSFKTAMWFWMTAQSPKPSCHDVMTGRWSPSGSDS
AAGRAAGYGVVTNIINGGLECGKGSDSRQVDRIGFYKRYCDILGVSYGSNLDCNSQKPFGFAAQSQPHLIKTVV

>Pg_GQ0031.TB_K19:53:Contig1:1-1392:+
MATGRVFLVVLLALAVSFNVSLAKTKICDKGWECKGVYCCNQTISQIFTVDNFEELFSKRNSPVAHAVGFWDYYSFINAAAQFEGIGF
GTTGGQLMQQKELAAFFGHVAAETSCGYSVAVGGPYAWGLCYKEEMSPDQLYCDQNYLFPCSPGASYHGRGALPLYWNYDYGQIGEAL
KLDLLNYPDLLSNNATIGFQTAIWRWMNPIKPKQPSAHDVLVGNWKPTKNDTASFRFPGFGMTINVLNGGLECGQGDIEAMSNRVSHY
LHFLDLMGVGRELAGDHLDCGEQVALNPVSNSATSR

>Pg_GQ0071.TB_A16:53:Contig1:1-1113:+
MGSSSSDKSVMALVLVLLLVGVSVNAQNCGCASGVCCSQYGYCGTTSAYCGKGCKSGPCYSSGGVSPSAGGGSVGGIISQSFFNGLAG
GAGSSCEGKGFYTYNAFIAAANAYSGFGTTGSNDVKKRELAAFFANVMHETGGLCSINEINPKIIYCQSSSTWPCTSGKSYHGRGPLQ
LSWNYNYGAAGKSIGFDGLNNPEKVGQDSTISFKTAVWFWMKNSNCHSAITSGQFGGTIKAINNMECGGNSGQVKSRVNYYKNICSQ
LGVDPGANVSC

>Pg_GQ0071.TB_C17:53:Contig1:1-1115:+
MAGSSGKFASPRGRVVVRMSLVLLLVVGVSVNVVNAQNCGCASGLCCSKWGYCGTTSAYCGNGCQSGPCSGGGSPSGGGGNVGTIISQ
NFFNSLASGAGSSCKGKGFYTYNAFIAAANAYSGFGTTGSNEVQKRELAAFFANVMHETGGLCYINEISPSSNYCQSSSTWPCTSGKS
YHGRGPLQISWNYNYGAAGQSIGFDGLNNPEKVGQDATISFKTAVWFWMKNSNCHSAITSGQGFGGTIKAINSGECNGGNSGQVNSRV
TYYKKFCSQLGVDAGTNVSC

>Pg_GQ0073.B3.r_L20:53:Contig1:1-1321:-
MAQPQKYLLIILLSMSTSLVYSSSSQPASHRGHYSNKNNGGGVKAAYWPSYTVSYSPPSSIDAALFTHIFYAFADLNDQTFQVELPAT
DNPAEFTPTLLKKNPSLKTLLSIGGGGSNSTAFAIMASNASHRKVFIKSAIAVARKYGFHGLDLDWEFPTDAAEMKNLGRLFAEWRHA
IDLEAKASGRPRLLLTAAVYFAQFFFLSGEKRAYPATSIARNLDWVNAMCFDYHGSWDTSATGSHAALYDSSSNISTSFGIGSWLHSG
IPPNKVVMGMPLYGRSWILKSLNETGIGAPAVAAGPKQNLSSETGVMFFSEIREFIRQKKATEVFDKKSGSAYSYSSDLMWVGYDNKQ
SVAAKVSFAKKRHLLGYFFWAISQDSNWMLSTQASESWNRAR

>Pg_GQ01304.T3_J23:53:Contig1:1-968:+
MATHFRVNVIFLWLAFALSALSICRGAVSDIATQDFFNGILSAATNGCAGKTFYTYTDFINAANSFSSFGTTGTSDDNKREIAAFFAN
VAHETTNLCYVEEIAKSDYCSSNTQYPCASGQQYYGRGPLQLTGNGNYGAAGDYLGVDLLNNPGLVAQDDLTSWKTALWFWNVNSNCH
TAITSGQGFGATIQAINGAIECNGGNTDQVNDRISRYTNYCSQFGVDPGSNLSC

>Pg_GQ0132.B3_D17:53:Contig1:1-1115:+
GANGHLFREYIGAMSNGVKFSDVPINPAVEVHFILSFAIDYTGSNSPSPTNGKFNIFWDTGSLTPKAVQAIKAQHKNVKVALSLGGDS
VGNGHPVQFKPLSVSSWVNNAVSSLSKIIKEYHLDGIDIDYEHFQADDNMFADCIGQLITKLRQSKAISFASIAPFDDGGAVQSHYRA
LWSKYGNVIDYVNFQFYAYDASTTVSQFIQHYNTQASLYKGGRVLVSFNTESGAGGLLPTHGFFQACKMLKQKGELQGIFVWSADSSK
SNGFKYEKQAEALLA

>Pg_GQ0132.B3_E06:53:Contig1:1-1375:+
MKSVSLSAMAIALVTMAMMNLYIVSAQQCGRQAGGALCSGGLCCSKWGFCGNTQSYCGQDCQSQCGGSSPTPASPTSSGQGVASVITE
NIFNELLKHRNDAGCKAKGFYSYNAFIAAANAFPGFGTTGDLATRKRELAAFFGQTSHETTGGWPTAPDGPYAWGYCFKEEQGNPPAE
YCQASSQWPCVSGKRYYGRGPVQISWNYNYGPAGKAIGFDGINNPDIVANDATVSFKTAVWFWMTAQSPKPSCHDVMTGRWNPSGSDS
AAGRAAGYGVVTNIVNGGLECGKGSDSRQQDRIGFYKRYCDILGVSYGSNLDCNNQKPFGSTA

>Pg_GQ0134.B7.1_C02:53:Contig1:1-1185:+
MKSMRFSLMVMALVTMATMNFYFVSAEQCGRQASGVLCTGGVQGVSSVITESIFNDLLKHRNDAVCKAKGFYSYAAFIAAADAFPGFG
TTGDLTTQKRELAAFFGQTSHETTGGWPTAPDGPYAWGYCFKEEQGNPPGEYCQASSEWPCASGKRYYGRGPVQLSWNYNYGPAGKAV
GFDGINNPDIVANDATVSFKTAVWFWMTAQSPKPSCHNVMAGGWGPSSSDTAAGRAAGYGVVTNIINGGLECGKGSDSRQEDRIGFYK
RYCDILGVSYGSNLDCNSQKPFGFAAQSQPRLIKTVV
```

>Pg_GQ0174.TB_A16:-3:C21:67-840:-
NGILSGASDTCTGKNFYKYSDFINAANAFSGFGTTGTSDDNKRELAAFFANVAHETGSLCYVEEINKSDYCDSTNTQYPCVAGKQYYG
RGPLQLTWNYNYGAAGDYLGFDGLNNPEIVAEDDSISWKTAVWFWMLDSNCHSAITSGQGFGATIQAINGGECNGGNPDAVTDRVNYY
NSYCSQFGVDPGSNVSC

>Pg_GQ0174.TB_A16:53:Contig1:1-1043:+
MATHLQSVLSLWLALALYVISICRGDVGDIATQDFFNGLISGAADSCTGKNFYTYSDFINAASAFSGFGTTGTSDDNKRELAAFFANV
AHETGSLCYVEEINKSDYCDSTNTQYPCVAGKQYYGRGPLQLTWNYNYGAAGNYLAFDGLNNPEIVAQDDSISWKTAVWFWMLNSNCH
SAITSGQGFGATIQAINGGECNGGDSDSVTDRVNYYNNYCSQFGVDPGSNVRC

>Pg_GQ01812.T24_D02:53:Contig1:1-1488:+
MAQPPRNPLIILLVVTSSLIHSSSSSLNNSINNGGVKAAYWPSHSILPPSSINTDLFTHTFYAFAHLNNQTFEVEVPSTDDPEEFTST
LLMKKPSLKTLLSIGGGGSNTTAFALMASNASHRKVFINSAIALARKYGFHGLDLDWEFPKDAVEMENLGTLFAEWRRAIDLEATASG
RPRILLTAAVYFAQYFFLRGEKRAYPATSIARNLDWVNVMCYDYHGSWNISATGAHAALYDPSSNISTSFGIGSWLHSGVPPDNIVMG
MPLYGRSWILKSLEETGIGAPAVAAGPKQTLSKEKGVMFFSEIRDLINQKNATEVFDEETVSAYSYSSDMIWVGYDNQDSVATKVSFA
KERRLLGYFFWAISQDSNWMLSTQASESWEQVQ

>Pg_GQ0194.TB_C04:53:Contig1:1-982:+
MAIHLRVMTASVFVLWLALALSICRGDVGDIATEDFFNGILSGAADTCAGKKFYTYSDFITAANGFSGFGTTGTSDDEKREIAAFFAN
VAHETIGFCYIEEIQKDDYCDSNNTQYPCAAGKQYYGRGPIQLSWNFNYGPAGSYLGVDLLNNPETVAQDALIAWKTALWYWNVNSDC
HSAITSGQGFGATIQAINGALECNGGNTDEVNDRVSYYTKYCSQFGVDPGSNLSC

>Pg_GQ02015.TB_I14:53:Contig1:1-1023:+
MGTSNGRSVMALVLVLVLLLVGVTVNDAECGCASGLCCSQYGYCGTTSDFCGDGCKSGPCSSSGGGSPTGGGGNVGTIISQSFFDGLA
SGAASSCEGKGFYTHDAFIEAASTYSGFGTTGADDVQKRELAAFFANVMHETGGLCSINEINPPQIYCDQTNTAWPCTSGKSYHGRGP
LQLSWNYNYGAAGQSIGFDGLNNPEKVGQDPTIAFKTAVLFWMQNSECHSAITSGQGFGATIKAINSGECSGGNSDEVSGRVNYYNTI
CSQLGVDPGANVSC

>Pg_GQ0202.TB_P24:53:B21:62-799:-
ISQNFFNSLASGAGSSCKGKGFYTYNAFIAAANAYSGFGTTGADDVRKRELAAFFANVMHETGGLCSINEIDPKQIYCDPSYTAWPCT
SGKSYHGRGPLQLTWNYNYGAAGQNIGFDGLNNPEKVGQDATIAFKTAVWFWMQNSNCHSAITSGQGFGATIKAINSGECNGGNSGEV
SSRVNYYNTICSQLGVDPGANVSC

>Pg_GQ0203.B3_E17:53:Contig1:1-1342:+
MDSTHTPPSLLLPMAFSGFGTAGDLTTNKRELAAFLAQTSHETTGGWQSAPDGPYAWGYCFKEENDPPSDYCQASSQWACASGKRYYG
RGPLQISWNYNYGPAGSALGFDGINNPDIVASDATVSFKTAIWFWMTAQSPKPSCHDVITGTWSPSGSDSAAGRAAGYGMVTNIINGG
VECGKGSDSRQEDRIGFYKRYCDILGVSYGSNLDCNNQTPFPS

>Pg_GQ0204.B3_D23:53:Contig1:1-1025:+
MASATIGRMKSMRVLSALTALAMMGTLCCQVSAQQGVASIISEDVFNQFLKHRNDDACPAKGFYTYSAFIAAANSFPEFGNNGDLETS
KRELAAFFGQTSQETTGGWATAPDGPYAWGYCFKEENSADRYHGRGPIQLTGDYNYKAAGDALGYDLINNPELVVTDATVSFKTAVWF
WMTPQSPKPSCHDVILGRWSPSDTDTAAGRVPGYGMVTNIINGGVECGQGTSSATQQGRIGFYQTFCNKLGVDSGSNLDCNNQKHFGN

>Pg_GQ0205.B3_M14:53:Contig1:1-937:+
GQCCSKWGYCGTTKAYCGEGCQSGACYNKPPVSPAPTGTGVASIISDTFFNRIKAEISSACKGKAFYTYQGFITAANAFSNFGRTGTL
EVRKRELAAFFANVARETGSLCYVEEISKPRYCTPSAVWPCKSGESYYGRGPLQLTGNYNYGEAGKYLGIDLLSNPDMVAQSPAIAFK
TSIWYWSVYSNCHKAMTSTAGKGFAGTIRSINPGECGNPAYRSAVQSRVNYYEKFCSWLGITTGPNLYC

>Pg_GQ02906.B7_O10:53:Contig1:1-1318:+
MGVKGNLFADYIGALFKGVKFSDVPINPAVEFHFILAFAIDYTTESGSPVPTDGNFNIFWDTGNLSPEAVQAIKAQHTNVKVALSLGG
DSVDSQPAEFKPSSVSSWVDNAVSSLSKIIEEYHLDGIDIDYEHFQANNDTFAECIGQLITRLKQSNVISFASIAPYDDQEVQSHYLA
LWNKYEHVIDYVNFQFYAYDSSTTISQFIGYYNTQASQYSGGKVLVSFSSATGAGGLLPENGFFEACRMLKQNGELKGIFVWCADWSK
ANGFKYEIESQDLLAS

>Pg_GQ03323.B7_E13:53:B7:35-891:+
MSDMATHLCVMTRSVFILWLSLALSVVSIGRGDVRGFVTEDFFNKIVSGDVLLSGGVGDIATEYLFKGILSSASRGCAGKSLYTYSNF
ITATNSFSGFASNRTSSNNKREIAAFFAHVAHETVSLCYVEEVDKSIDYCDSSSTQYPCAARKQYYGRGPLQLSWNYNYGAAGDFLGV
DLLNNPGLVARDDLIAWKTALWFWNVNSKCHSAMASGQGFMGTLQAINVGECSGGNIAQINDRINNYLAYCSRLGVSQGYNVVC

>Pg_GQ03326.B7_F24:53:Contig1:1-1509:+
MAALHRLLAFTFLCIIAPYTASAVKGGYWPAYSYSYFPPSQINTSLYTHLYYAFVDLDNQTFEVGVSLENQQSIQEFTATLLTNNASV
KTLVSIGGGNANISTFAEMAADASLRQTFIDSSIALARNYSFHGLDLDWEYPQTVVEMNNLGQLFSEWSSAIQAEAQDTGNDPLLLTA
AVYFSHEFFAWGDVREYPTAVDNLDWINIMTYDFHGPWQPDQTGAPAALYDPESNFSTSFGVQSWLDAGFSSSKVVMGMPLYGHSWKL
QSSEEVGIGAPATGVGIDDGSITYSDIRDFIEQNTATEVFDTTTVSAYCYAGLDWIGYDNDQSVAGKVVFAKQEGLLGYFFWNVAQDT
ADWALSGTASDTWDNGTDDAGIPILRLSHLYQPEYQIS

>Pg_GQ03507.B7_A23:5-:Contig1:1-841:+
MSDMATHLCVMTRSVFILWLSLALSVVSIGRGDVRGFVTEDFFNKIVSGDVLLSGGVGDIATQDLFKGILSSASWGCAGKSFYTYSNF
ITATKLFSGFASNSNSSNNKREIAAFFAHVAHDTSSLCYVEEIDKSNDYCDSSSTQYPCVARKKYYGRGPLQLRWNNNYGGAGEFLGV
DLLNNPGLVAQDELLAWKTTLWFWNVNSNCHVAMASRQGFKGTIQAINGAYECNGGNTDEVTDRINNYLAYCSRLRVSPGNNVAC

>Pg_GQ03602.B7_D17:5-:Contig3:1-897:+
MASVTMKWMRVLLALTVVAMMNTFCYVSAQQGVAFIITEDIFNEFLKHRNDGACAAKGFYTYNSFILAANRFPSFGNVGNLETRKREL
AAFFAQTSHETTGGWETAPDGPYAWGYCFKEEQDADPEEMYYGRGPIQLTGKNNYEAAGKALGYDLINNPNIVASNPTISFKTAVWFW
MNPQPPKPSSHIVMIGKWSPSGSDTAAGRVPGYGVVTNIINGGVECGRGSDSNQEDRIGFYKKYCDILGASYGSNIDCNSQRSFGG

>Pg_GQ03602.B7_D17:53:Contig1:1-1101:+
MASVTMKWMRVLLALTVVAMMNTFCYVSAQQGVAFIITEDVFNRFLKHRNDGACAAKGFYTYNSFILAANRFPSFGNVGNLETRKREL
AAFFAQTSHETTGGWETAPDGPYAWGYCFKEEQDADPEEMYYGRGPIQLTGKSNYEAAGKALGYDLLNNPGIVASNPTISFKTAVWFW
MTTQSPKPSPHTVMIGKWSPSGSDTAAGRVPGYGVVTNIINGGVECGRGSDSNQEDRIGFYKAYCDILGASYGSNVDCNTQRSFGG

>Pg_GQ03808.B7_D10:53:Contig1:1-1121:+
MTPRVIILWLVLAFDIVSICQADVGDFATEDFFTGISSGASDTCPGKNLYRYNDFINAATAFSGFGTTGTDVDHKRELAAFFANVAHE
TDRLCYVEEIEKSDYCDFSNTQYPCTAGKQYYGRGPLQLRWNYNYGAAGDDLGFDGLNHPEIVARDGSISWKTAVWFWMKDSKCHSAI
TSGQGFRATIKAIDGEECNGGNSNAVDERVNYYTNYCSEFGVDPGSNLSC

>Pg_GQ04010.B7_E05:5-:B7:32-891:+
MAQPQKYLLIILLSMSTSLVYSSSSQPASHRGHYSNKNNGGGVKAAYWPSYTVSYSPPSSIDAALFTHIFYAFADLNDQTFQVELPAT
DNPAEFTPTLLKKNPSLMTLLSIGGGGSNSTAFALMASNAFRRKVFINSAIAVARKYGFHGLDLDWEFPKNVVEMQNLGSLFAEWRHA
IDLEATASGRPRLLLTAAVYFAQYFFVWGEKRAYPATSIAKNLDWVNAMCFDYHGSWDTSATGAHAALYDPSSNISTSFGIGSWLHS

>Pg_WS0016.B21_M24:53:Contig1:1-911:+
MATHLRVSVFILWLALALSALSICRGAVGDIATQDFFNGILSAASDGCAGKTFYTYTDFINAANSFSSFGTTGTSDDNKREIAAFFAN
VAHETGSLCYVEEINKSDYCDSTNTKYPCAAGKQYYGRGPLQLSWNYNYGAAGDYLGVDLLNNPETVAQDDITSWKTALWFWNVNSNC
HTAITSGQGFGATIQAINGAIECNGGNTDQVNDRISRYTNYCSQFGVDPGSNLSC

>Pg_WS0018.B21_F17:53:Contig1:1-907:+
SVQAIKAQHKNVKVALSVGGDSIDGQPVEFKPSSVSSWVDNAVSSLSKIIRDYHLDGIDIDYEHFQADNTTFAECIGQLITRLKKSKA
ISFASIAPFDNQEVQSHYLALWKKYGHAIDYVNFQFYAYSANTTISQFIKYYNTQASLYSGGKVLISFSSASSGGGLSPKNSFFEAAR
MLKQSGKLEGIFLWCADWSKGNGFKYDTKSQTLLAS

>Pg_WS00749.C21_F20:53:Contig1:1-881:+
GVASIMTEDLFNQLLKYKDDSRCKANGFYSYAAFIAAASAFSGFGTAGDLTTNKRELAAFLAQTSHETTGGWQSAPDGPYAWGYCFKE
EQDPVSDFCQASSQWPCASGKRYYGRGPIQISWNYNYGQAGSALQFDGINNPDIVASDPTVSFKTAVWFWMTAQSPKPSCHDVMTGTW
SPSGSDSAAGRAAGYGLVTNIINGGLECGKGSNVKQDDRIGFYKRYCDILGVSYGSNIDCNSQTPYGG

>At_AT4G19770.1
MASSSYGRKSFILSTISIARSYGFDGLDLDWEYPRNAAEMSDFAELLKEWRYAVQGEAYSSELPVLILTATVYYSSNYNGVVYPVKFI
SELLDWVNIKAYDFYGPGCTEVTGPPAALYLQSDGPSGDSGVKDWIDAGLPAEKAVLGFPYYGWAWTLADPKNHGYYVDTTGPAISDD
GEISYSQLKTWIVDNKATTVHDNIVIGDYCYAGTTWIGYDSEESIVTKVIYAKQKGLLGYFSWQVGGDDKSELSSAGSSPYHILI

>At_AT4G19800.1
MSESDVVKASYWFPATDFPATDIDSSLFTHLFCTFADLEAESYEITIATWNQAPFHAFTETVQQRNPHVKTLLSIGGGNADKDAFASM
ASNPDSRASFIQSTITVARSYGFHGLDLDWEYPRNEEEMYDFGKLLEEWRSAVEAESNSSGTTALILTAAVYYSSNYQGVPYPVLAIS
NSLDWINLMAYDFYGPGWSTVTGPPASLYLPTDGRSGDSGVRDWTEAGLPAKKAVLGFPYYGWAWTLADPDVNGYDANTTGPAISDDG
EISYRQLQTWIVDNGATKVHDDMMVGDYCYAGTTWIGYDSEKSIVTKVIYAKQKGLLGYFSWHVGGDDNSELSSAGSTPYHILILCMS
WKY
SKIYHPIHVINSTNWTIIYNSRKFDILNFVFIASQAWDITEME

>At_AT2G43620.1
MATLRAMLKNAFILFLFTLTIMAKTVFSQQCGTTGCAANLCCSRYGYCGTTDAYCGTGCRSGPCSSSTTPIPPTPSGGAGGLNADPRD
TIENVVTPAFFDGIMSKVGNGCPAKGFYTRQAFIAAAQSFDAYKGTVAKREIAAMLAQFSHESGSFCYKEEIARGKYCSPSTAYPCTP
GKDYYGRGPIQITWNYNYGAAGKFLGLPLLTDPDMVARSPQVAFQCAMWFWNLNVRPVLDQGFGATTRKINGGECNGRRPAAVQSRVN
YYLEFCRTLGITPGANLSC

>At_AT4G19820.1
MSSTKLISITFFLSLLLRFSSAQTVVKATYWFAESESPLAQIDSSLFTHLFCAFADINTLTYQVIVSSRNKPKFSTFTQTVRRRNPTV
KTLLSIGGDFTYNFAFASMASNPTSRKLFISSSIKLARSCGFHGLDLNWKYPSITTEMDNFGKLLREWRLAVEAEARSSGKPRLLLTA
AVFYSYSYYSVLHPVNAVADSLDWVNLVAYDFYESGSSRVTCSPAPLYDPITTGPSGDAGVRAWTQAGLPAKKAVLGFPLYGYAWCLT
DAKNHNYYANSSGPAISPDGSIGYDQIRRFIVDNKATMVYNSNLVQNYCYAKKTWIGYDDNQSIVMKVKYAKQRGLLGYFSWHIGADD
NSR
LSRAGSVLFQF

>At_AT4G19730.1
MTSTVILGSPSAEVKASYWFPDGETQDPITSAETIPSALFTHLFCAFADLDANSHKVFVSQAHEFIFSTFTETVKIRNPQVKTLLSIG
GKNANNSAFASMASNHQSRKTFIDSWIFIARSNGFHGLDLAWEYPYSDHEMTDFGNLVGELRAAVEAESRRSSKPTLLLTAAVYYSSV
YKTFTYPVQVMRESLDWVNIIAYDFYGPVSSSKFTVPTAGLHVSSNNEGPSGDSGLKQWIKDGLPEKKAVLGFSYVGWAWTLQNDKDT
GYNAAAAGVAKSEDDVSEDGSINYAQINKFIRDEEAAKVYDPKVVGHYCFAKKIWIGYEDTQSVEAKV

>At_AT2G43570.1
MAKPTSRNDRFALFFITLIFLILTVSKPVASQNCGCASDFCCSKYGYCGTTDEFCGEGCQAGPCRSSGGGGDPAVSLEGTVTPDFFNS
ILNQRGDCPGKGFYTHDTFMAAANSYPSFGASISKREIAAFFAHVAQETGFMCYIEEIDGPAKAASGEYCDTEKPEFPCAQGKGYYGR

GAIQLSWNYNYGLCGKALDENLLASPEKVAQDQVLAFKTAFWFWTTNVRTSFKSGFGATIRAVNSRECSGGDSTAKAANRIKYFQDYC
GKLGVAPGDNLTC

>At_AT2G43610.1
MATQNAILKKALIIFLFTLTIMTGTAFSQNCGTNGCKGNMCCSRWGYCGTTKAYCGTGCQSGPCNSKPKPTPTPSGSGGLNAGPRGTI
ASVITPAFFNSIMSKVGSGCPAKGFYTRQAFIAAAESFAAYKGTVAKREIAAMLAQFSHESGSFCYKEEIARGRYCSPSTTYPCQPGK
NYYGRGPIQITWNYNYGAAGKFLGLPLLKDPDMVARSPTVAFQCAMWFWNKNVRPVLSQGFGATTRRINGGECNGGRPAAVQSRVNHY
LDFCKKLGVTPGTNLSC

>At_AT2G43600.1
MTIKNVIFSLFILAILAETVFSQNCMDTSCPGLKECCSRWGFCGTKDEYCGFFCFSGPCNIKGKSYGYDYNVDAGPRGKIETVITSAL
FDSIMSKVESNCSAKGFYTYEAFITAFKSFGAYKGKVAKREIAAILAHFSYGSKSFCYKEEISNERYCSKSKKYPCEPGKNYYGRGLL
QSITWNEYYGAGKHLGLPLLKDPDLVSRSPEVAFKFAMWFWNRNVRPALYLGFGEITKRVDGRECGNWRRDDTKNKVKQYIEFCEMLG
VTPDQGLDC

>At_AT1G56680.1
MATQNKIQKNSLIIFLFTLVVIAQTATSQNCLTIGCPGINECCSHTGYCGTNVEHCGFWCLSGPCQLSKSSSSYRLNDGPRGKIESIV
TPALFHRLMRKVGSNCTGKGFYTREAFITAVKSFEGYKGTVAKREIAAILAQFSYESGNFCYKEEVTSETYCSSSKTYPCQSGKKYYG
RGLLQSIKWNEFYGEAGKYLGLPLLKDPDMVARSPEVAFKFAMWFWKTEVGPSLRLGFGATTMRINGIECGGMSWNAEAMQNRINQYL
EICKWFGVNPGKDLYC

>At_AT4G01700.1
MEKQISLLLCLLLFIFSISSSLHETEARKHNKYKPAPIMSLVPRTLYDQIFIHKDNNACPAKGFYPYEAFVEATRSFPKFGSVGNFWT
RRREVAAFLAQISHETTGGWATAPDGPYAWGLCFKEEVSPQSNYCDASNKDWPCVSGKSYKGRGPIQLSWNYNYGQAGRALGFDGLQN
PELVANNSVLAFKTALWFWMTEQTPKPSCHNVMVNRYRPTKADRAANRTVGYGLVTNIINGGLECGIPGDGRVTDRVGYFQRYAQLFK
VTTGPNLDCENQRPFS

>At_AT4G19760.1
MADQTQFLRTNSIVKASYWFPDGKSQSPECLSQGTPSSFIDSTLFTHLFCAFADVDSSTHEVTISAANSYQFSSFTETVKEKNTDVQT
LLSIGGKDADKAVLASMASNSKNRKAFIDSSIDIARKKDFYGLDLAWEYPSNDVEMTNFGKLLEEWRAAVVEESDKTNQLPLLLTAAV
YYSPQYDGVEYPVKAIADNLDFVNIMAYDFYGPGWSPVTGPPAALFHDPSNPAGRSGNSGLRKWLDEAKLPPKKAVLGFPYCGWAWTL
EDAENNGYDAATDGAAISPDGSITYAKIRNYIVDNGAATFHDPAVIGFYCYVGNTWIGYDDNQSIVYKVKYAKFTGLLGYFSWHVGAD
YNCGLSRAGSFSF

>At_AT4G19810.1
MSSTKLISLIVSITFFLTLQCSMAQTVVKASYWFPASEFPVTDIDSSLFTHLFCAFADLNSQTNQVTVSSANQPKFSTFTQTVQRRNP
SVKTLLSIGGGIADKTAYASMASNPTSRKSFIDSSIRVARSYGFHGLDLDWEYPSSATEMTNFGTLLREWRSAVVAEASSSGKPRLLL
AAAVFYSNNYYSVLYPVSAVASSLDWVNLMAYDFYGPGWSRVTGPPAALFDPSNAGPSGDAGTRSWIQAGLPAKKAVLGFPYYGYAWR
LTNANSHSYYAPTTGAAISPDGSIGYGQIRKFIVDNGATTVYNSTVVGDYCYAGTNWIGYDDNQSIVTKVRYAKQRGLLGYFSWHVGA
DDNSGLSRAASQAWDATTATTRTIQKV

>At_AT3G47540.1
MASTKISLVFFLCLVGPCIGAGTTSAAESVSSIVTQGFFNNIINQAGNGCTRKRFNTRDSFIDAANNFPNFASSVTRREIASMFAHVT
HETGHFCYIEEIKKRSRGRCDENVEQRPCPSQSKGHSGRGHQSLGLDLLRQPELVGSNPTVAFRKGLSFWINSVRPVLNQGFGATIRA
INGMECNGGNSGAVKARIGYYRDYCRQLGVDPGHNLSC

>At_AT2G43590.1
MAFTKISLVLLLCLLGFFSETVKSQNCGCAPNLCCSQFGYCGTDDAYCGVGCRSGPCRGSGTPTGGSVGSIVTQGFFNNIINQAGNGC
AGKRFYTRDSFVNAANTFPNFANSVTRREIATMFAHFTHETGHFCYIEEINGATRNYCQSSNTQYPCAPGKGYFGRGPIQLSWNYNYG
ACGQSLGLDLLRQPELVGSNPTVAFRTGLWFWMNSVRPVLNQGFGATIRAINGMECNGGNSGAVNARIGYYRDYCGQLGVDPGPNLSC

>At_AT1G02360.1
MAQQHSFLLLCFFLSISYLLSSAQTEATSIERLVPRDLYNKIFIHKDNTACPANGFYTYESFVQATRRFPRFGSVGSPVTQRLEVAAF
LAQISHETTGGWATAPDGPYAWGLCFKEEVSPQSTYCDSSDTQWPCFPNKTYQGRGPIQLSWNYNYGPAGRALGFDGLRNPETVSNNS
VIAFQTALWFWMTPQSPKPSCHDVMIGKYRPTAADLAANRTGGFGLTTNIINGGLECGIPGDGRVNDRIGFFQRYTGLFKVATGPNLD
CENQRPYA

>At_AT1G05850.1
MVTIRSGSIVILVLLAVSFLALVANGEDKTIKVKKVRGNKVCTQGWECSWWSKYCCNQTISDYFQVYQFEQLFSKRNTPIAHAVGFWD
YQSFITAAALFEPLGFGTTGGKLMGQKEMAAFLGHVASKTSCGYGVATGGPLAWGLCYNREMSPMQSYCDESWKFKYPCSPGAEYYGR
GALPIYWNFNYGAAGEALKADLLNHPEYIEQNATLAFQAAIWRWMTPIKRAQPSAHDIFVGNWKPTKNDTLSKRGPTFGSTMNVLYGE
YTCGQGSIDPMNNIISHYLYFLDLMGIGREDAGPNDELSCAEQKPFNPSTVPSSSSS

>At_AT5G24090.1
MTNMTLRKHVIYFLFFISCSLSKPSDASRGGIAIYWGQNGNEGNLSATCATGRYAYVNVAFLVKFGNGQTPELNLAGHCNPAANTCTH
FGSQVKDCQSRGIKVMLSLGGGIGNYSIGSREDAKVIADYLWNNFLGGKSSSRPLGDAVLDGIDFNIELGSPQHWDDLARTLSKFSHR
GRKIYLTGAPQCPFPDRLMGSALNTKRFDYVWIQFYNNPPCSYSSGNTQNLFDSWNKWTTSIAAQKFFLGLPAAPEAAGSGYIPPDVL
TSQILPTLKKSRKYGGVMLWSKFWDDKNGYSSSILASV

```
>At_AT4G19720.1
MYTEVVKASYWFPDGASSPTTGSVVPQSSAVLIDSTLFTHLFCAFADLDPQTNSVVVSGAHEQEFSNFTKIVKKKNPHVQTLLSIGGR
NADKSAFASMASNPTSRKSFIWSAISSARYYRFDGLDLVWKYPKDDVEMRNFGQLLEQWREAIEDDAERTERMPLLLTAAVYYSPVYD
SVSYPIREIKKKLDWVNLIAYDFYSSSTTIGPPAALFDPSNPKGPCGDYGLKEWIKAGLPAKKAVLGFPYVGWTWSLGSGNDAATSRV
ATSAEGSINYDQIKRLIVDHKARPVFDSTVVGDYCFAGTSLIGYDDHQSVVAKVKYAKQKGLLGYFSWHVGADDNFGLSRAGSSFCTY
ESLFFSLFTFS

>At_AT2G43580.1
MALTKIFLILLLSLLGLYSETVKSQNCDCAPNLCCSQFGYCGTTADYCGSTCQSGPCRVGGPPTGAGLVGNIVTQIFFNNIINQAGNG
CAGKSFYTRDSFINATNTFPSFANTVTRREIATMFAHFTYETGHFCYIEEINGASRVMCDQNNRQYPCAPAKSYHGRGPLLLSWNFNY
GACGQSLGLDLLRQPELVSSNPVVAFRTALWFWMKSVRPVLNQGFGATIRAISGFDCDGRNLGGVNARIGYYRDYCGQLGLDPGANIT
C

>At_AT3G54420.1
MLTPTISKSISLVTILLVLQAFSNTTKAQNCGCSSELCCSQFGFCGNTSDYCGVGCQQGPCFAPPPANGVSVAEIVTQEFFNGIISQA
ASSCAGNRFYSRGAFLEALDSYSRFGRVGSTDDSRREIAAFFAHVTHETGHFCYIEEIDGASKDYCDENATQYPCNPNKGYYGRGPIQ
LSWNFNYGPAGTAIGFDGLNAPETVATDPVISFKTALWYWTNRVQPVISQGFGATIRAINGALECDGANTATVQARVRYYTDYCRQLG
VDPGNNLTC

>At_AT3G12500.1
MKTNLFLFLIFSLLLSLSSAEQCGRQAGGALCPNGLCCSEFGWCGNTEPYCKQPGCQSQCTPGGTPPGPTGDLSGIISSSQFDDMLKH
RNDAACPARGFYTYNAFITAAKSFPGFGTTGDTATRKKEVAAFFGQTSHETTGGWATAPDGPYSWGYCFKQEQNPASDYCEPSATWPC
ASGKRYYGRGPMQLSWNYNYGLCGRAIGVDLLNNPDLVANDAVIAFKAAIWFWMTAQPPKPSCHAVIAGQWQPSDADRAAGRLPGYGV
ITNIINGGLECGRGQDGRVADRIGFYQRYCNIFGVNPGGNLDCYNQRSFVNGLLEAAI

>At_AT4G19740.1
MASNRTSRESFISSSISIARSLGFYGLDLAWEYPNNDVEMNNFGKLLQEWRSAVEVESQRTGIRPLLLTAAVYYTSDYNSVSYPVQAI
NRSLDWVNLIAYEFYGLTTEIGPPAGLYDPSIKGPCGDTGLKHWLKAGLPEKKAVFGFPYVGWSWTLDDDKDHGDDVAVTHRVAVTAN
GSINYDQIVKFITEYKARTVYDSEVVGFYCIAGTT

>At_AT4G19750.1
MADQTQFLRTNSIVKASYWVVKPENDFPAGNIDSTRFTHLFCAFADVDSSTHEVTISAANSCQVSSFTHTVKDKNTDVQTLLSIGGKD
ADKAVLASMASNSKNRKAFIDSSIDIARKKDFYGLDLAWEYPSNDVEMANFGKLVKEWRAAVVEESDRTNQLPLLLTAAVYYSPDYYG
EEYPVQAIADNLDFVNIMAYDFYGPGWSPVTGPPAALFDPSNPAGRSGDSGLSKWLEAKLPAKKAVLGFSYCGWAWTLEDAENNGYDA
ATDGAAISSDGSITYAKIRNYIIDNGAATFHDPAVIGFYCYVGTTWIGYDDNQSIVSKVRYAKLKGLLGYFSWHVGADYNCGLSRAGS
FSLTFFLLYP

>At_AT3G16920.1
MVTLIDEEKDKKKSTMVSKPLFSLLLLTVALVVFQTGTLVNAEDSEPSSSTRKPLVKIVKGKKLCDKGWECKGWSEYCCNHTISDFFE
TYQFENLFSKRNSPVAHAVGFWDYRSFITAAAEYQPLGFGTAGEKLQGMKEVAAFLGHVGSKTSCGYGVATGGPLAWGLCYNKEMSPD
QLYCDDYYKLTYPCTPGVSYHGRGALPVYWNYNYGQTGEALKVDLLSHPEYLENNATLAFQAAIWRWMTPPKKHLPSAHDVFVGKWKP
TKNDTAAKRTPGFGATINVLYGDQICNSGFDNDEMNNIVSHYLYYLDLIGVGREEAGPHEKLSCADQEPFSSSSSAPPSSGSSS

>Os_9634.m04963
MRALALAVVAMAVVAVRGEQCGSQAGGALCPNCLCCSQYGWCGSTSDYCGAGCQSQCSGGCGGGPTPPSSGGGSGVASIISPSLFDQM
LLHRNDQACAAKGFYTYDAFVAAANAYPDFATTGDADTCKREVAAFLAQTSHETTGGWPTAPDGPYSWGYCFKEENNGNAPTYCEPKP
EWPCAAGKKYYGRGPIQITYNYNYGPAGQAIGSDLLNNPDLVASDATVSFKTAFWFWMTPQSPKPSCHAVITGQWTPSADDQAAGRVP
GYGEITNIINGGVECGHGADDKVADRIGFYKRYCDMLGVSYGDNLDCYNQRPYPPS

>Os_9634.m04964
MRALAVVVVATAFAVVAVRGEQCGSQAGGALCPNCLCCSQYGWCGSTSAYCGSGCQSQCSGSCGGGGPTPPSGGGGSGVASIVSRSLF
DQMLLHRNDAACPAKNFYTYDAFVAAANAFPSFATTGDAATRKREVAAFLAQTSHETTGGWATAPDGPYSWGYCFKEENNGNVGSDYC
VQSSQWPCAAGKKYYGRGPIQISYNYNYGPAGQAIGSDLLNNPDLVASDATVSFKTAFWFWMTPQSPKPSCHAVMTGQWTPNGNDQAA
GRVPGYGVVTNIINGGVECGHGADSRVADRIGFYKRYCDMLGVSYGANLDCYNQRPFNS

>Os_9631.m02983
MRALAVVAMVATAFLAAAVHAEQCGSQAGGAVCPNCLCCSQFGWCGSTSDYCGAGCQSQCSAAGCGGGGPTPPSGSGGGSGVASIVSRS
LFDQMLLHRNDAACPASNFYTYDAFVAAASAFPGFAAAGGDADTNKREVAAFLAQTSHETTGGWATAPDGPYAWGYCFKEENGGAAGP
DYCQQSAQWPCAAGKKYYGRGPIQLSYNFNYGPAGQAIGADLLGDPDLVASDATVSFDTAFWFWMTPQSPKPSCHAVATGQWTPSADD
QAAGRVPGYGVITNIINGGLECGHGEDDRVADRIGFYKRYCDILGVSYDANLDCYSQRPFGS

>Os_9633.m03045
MAAQVVAEQCGSQAGGALCPNCLCCSSYGWCGSTSDYCGDGCQSQCDGCGGGGGGGGGGGGGGGGGAVEAVVSKELFEQLLLHRND
AACPARGFYTYDALVTAAAAFPDFAATGDDEARKREVAAFLGQTSHETTGGWATAPDGPYSWGYCFKEEIGATASYCVPSADNYNYGP
AGEAIGEDLLNNPELVASDPVVSFKTALWFWMTPQSPKPSCHDVITGQWTPSSGDIAAGRVPGYGVITNIINGGLECGFGPDDRVANR
IGFYQRYCDVLGIGYGSNLDCYDQRPFNSGLTAAQ
```

187

>Os_9633.m03046
MIAARAANLQVAMKALALALAVLALAYAAATARAEQCGRQAGGARCPNRLCCSRWGWCGLTDDYCKGGCQSQCRVSRDGGDDDVAAVLLT
APGGGRAGVASVVTSDQFERMLPHRDDAACPARGFYAYRAFVAAAGAFPAFAATGDADTRKREVAAFLAQTSHATSGGPYSWGYCYKE
VKGATSDFCVPNARWPCAPGKAYHARGPMQIAYNYNYGAAGEAIGADLLGNPELVATDPTVAFKTALWLWMTARSPSQPSPHAVVTGQ
WTPTPADSAAGRAPGYGLTTNILTGGLQCAGGNGGADRVAFYKRYCDVLGVGYGPNLDCFGQAPFDGDIMSASAAK

>Os_9631.m00319
MRRLLPLAGATLLIAAAGGASGQQAGVGSIITRAMFESMLSHRGDQGCQGAFYTYDAFIKAAGDFPRFGTTGNDETRRRELAAFFGQT
SHETTGGWATAPDGPFAWGYCRVNEITPSDPPYYGRGPIQLTHKYNYQLAGDALGLDLVNNPDLVSSDPVVAFRTAIWFWMTAQSPKP
SCHDVITNQWTPSGDDRSSGRLPGYGMATNIINGGEECGKGYSTDNAKDRVGYYKRYCDMFRVGYGDNIACRDQKPYGGG

>Os_9638.m03591
MTTTTTRFVQLAACAAASLLAVAASGAAAQGVGSVITQAVFNSMLPNRDNSQCPARGFYTYDAFIAAANSFPAFGTSGGSAELIRREL
AAFFGQTSHETTGGTRGSSDQFQWGYCFKEEINKATSPPYYGRGPIQLTGQSNYQAAGNALGLDLVGNPDLVSTDAVVSFKTAIWFWM
TAQGNKPSCHDVILGRWTPSAADTAAGRVPGYGVITNIINGGIECGVGQNDANVDRIGYYKRYCDMLGAGYGSNLDCYNQRNFAS

>Os_9638.m03593
MTTTTTTTRRFVQLAACAVAVLLAVAASGAAAQGVGSVITEAVFNSMLPNRDNSLCPARGFYTYDAFIAAANSFPAFGTSGGSAELIR
RELAAFFGQTSHETTDGTRGSSDQFQWGYCFKEEINKATSPPYYGRGPIQLTGFFIINTYMFLNFATSLAVYYPYSINYESHAIELPS
SQERAGAGPGRRPDLVSTNAVVSFKTAIWFWMTGQYNKPSSHDVILGRWTPSAADTAAGRVPGYGVITNIINGRFECDVGQNDANVDR
IGYYKRYCDMLGADPGSNLDCYNQRDFDSQPS

>Os_9629.m01787
MAKPTPAPRATPFLLAAVLSIVVVAASGAEARWYGGGGGGGYSPSPSPVSSIVSEQLYASLFLHKDDAACPARGFYTYASFVRAATRF
PRFAATGCADARKREVAAFLAQISHETTGGWATAPDGPYAWGLCYKEEINPQSSYCDATDKQWPCYPGKSYHGRGPIQISWNFNYGPA
GQALGFDGLRNPEIVANCSDIAFQTALWFWMTPRDTKPSCHQVMVGEYRPGPADVAANRTAGFGLVTNIVNGGLECNRAGDARVNNRI
GFYRRYCQVLGVDVGPNLDCEHQQPF

>Os_9632.m03968
MANSPTLTMLVFLAIGLSLVLSAAGVSVESVVTEAFFNGIKNQAPNGCAGKSFYTRQSFLNAARSYSGFANDRTNDDSKREIAAFFAH
VTHETGHMCYINEINGANMDYCDKSNKQWPCQPGKKYYGRGPLQISWNFNYGPAGKNIGFDGLRDPDKVAQDPTISFKTALWFWMNNV
HQVMSQGFGATIRAINGALECNGKNPGAVNARVNYYKDYCRQFGVSPGGNLYC

>Os_9632.m03974
MANSPTPTMLAFLALGLALLLSATGQASAQNCGCQSNMCCSKWGYCGTGKDYCGDGCRSGPCYGGGGGGGGGGGGGGGGGGGGSGVSVE
SVVTEAFFNGIKNQAPNGCAGKNFYTRQSFLNAAHSYSGFARDRTNDDSKREIAAFFAHVTHETGHMCYINEINGASMDYCDKNNKQW
PCQPGKKYYGRGPLQISWNYNYGPAGQNIGFDGLRDPDRVAQDPTISFKTALWFWMNNVHQVMLQGFGATIRAINGALECNGKNPGAV
NARVNYYKDYCRQFGVDPGGNLYC

>Os_9630.m03783
MARRLSLLAVVLAMVAAVSASTAAAQSCGCASDQCCSKWGFCGTGSDYCGTGCQAGPCDVPATNDVSVASIVTPEFFAALVAQADDGC
AAKGFYTRDAFLTAAGGYPSFGRTGSVDDSKREIAAFFAHANHETIKFCYIEEIDGPSKNYCDETSTQWPCMAGKGYYGRGPLQISWN
FNYGPAGQSIGFDGLGDPDAVARSPVLAFQTALWYWTNNVHDAFVSGQGFGATIRAINGALECDGKNPTAVSNRVAYYQQFCQQFGVD
PGSNLTC

>Os_9636.m04169
MRTSRAAAAAASLPLLLLVALLVAAEGRRHKDGSGDEEAKACDKGWECSGSRFCCNDTITDYFKAYQFEELFAHRNDRSLAHAAGFWD
YHAFITAAALFEPRGFGTTGGKEVGMKEVAAFLGHVGAKTSCGYSVATGGPLAWGLCYNHELSPSQSYCDNSNELYPCVEGVEYYGRG
ALPVYWNYNYGIIGQGIKQDLLNHPELLEQNATLAFEAAIWRWMTPMKRKQPSAHDVFVGNWKPTKNDTLSKRYPGFGATMNILYGDL
ICGQGSIDKMNVIVSHYQHYLDLMGVGSDKAGDNLDCADQVAFNPSSKNLDS

>Os_9637.m02807
MKRKTRNKIILTTLLVSAAAILIGGTVALILTAGTWKVKMKESREKICDKGWECSGSKYCCNDTITDFFKVYQFENLFSKRNSPVAHA
VGFWDYQSFITAAALFEPLGFCTTGGKQMQMMELCAFLGHVGSKTSCGFGVATGGPTAWGLCYNHEMSPKEDYCDKTNLQYPCVEGAE
YYGRGAIPVFWNYNYGAAGDGIHEDLLHHPEYLEQNATMAFMAAMWRWMTPMKKKQPSAHDVFVGNWKPTKNDTLAKRLPGFGATMNV
LYGDQICGKGYIDDMNVIISHYQYYLDLMGVGREHSGDNRDCAEQAAFNPSYKKPDDQQQQS

>Os_9632.m02919
MADKNGLLLLSTIAAVTLSSLSLAAVASPATRRSQESREAVRAGYYLAADAHLRPLAALDASMYTHLYYYAVAVHPARRTLLLPPDPA
AASLLGDFSRAVKAKNAAVKTVLSIGRGGGAGGAAAVAGSGSDPAFAAMAADPASRAAFIGAAVKVARENGFDGLDVAWRFPASAVEM
AEFGFLVAEWRAAVPRGFLLTATVYFSNHVFDAPFAGVDYPSETVARSLDWVNVMAFGLRPPGAANANATAFDAPLYDRASHYSASYG
VVSWLDAGVPASKVVMGIPLYGRSWFLRNKANSGVGAPVVAAGPKQRGSNATGAMSYAEVQWLAATATRGSRAVITAYDNASVASYVS
VGDVWVAFDGVAVVAEKLAFAARCGLLGYFLWPVNYDDANLTVSRRASQVWTQTKISPEFKNVTGGARQTQAPVQRPPALQSPAPTTA
PMSSSSTTSSFSRLSWRMLDVRLHLGALLLLLLLVCYQI

>Os_9638.m02367
MGSAKLIAVVLLPALLAFQAPMATAVNSNLFRDYIGAIFNGVKFTDVPINPKVRFDFILAFIIDYTTETNPPTPTNGKFNIFWQNTVL
TPSAVASIKQSNPNVRVAVSMGGATVNDRPVFFNITSVDSWVNNAVESLTGIIQDNNLDGIDIDYEQFQVDPDTFTECVGRLITVLKA
KGVIKFASIAPFGNAEVQRHYMALWAKYGAVIDYINFQFYAYGASTTEAQYVDFFNQQIVNYPGGNILASFTTAATTTSVPVETALSA
CRTLQKEGKLYGIFIWAADHSRSQGFKYETESQALL
ANATISY

188

>Os_9638.m02374
MTNGYLFREYIGAQFTGVRFSDVPINPNLSFNFILSFAIDYTSPAGGATPAPTNGVFSPYWDTANLSPADVAAVKAAHPNVSVMVGLG
GDSVQDTAKVFFSPTSVDSWVANAVASVSGIIDAYGLDGVDVDYEHFNDDGGAGVDTFVECIGRLLTELKARHPNITTSIAPFEDAVV
QRYYQPLWRRYAGVIDLVNFQFYGYGDNTDVPTYVMFYDEQAANYPGGKVLASFKTGDVAGLLWPEQGIAGAKELQRQGKLPGLFIWS
ADSSKVSSYGFEYEIKAQEIIANH

>Os_9629.m06349
MQMLIMVVVALAGLAAGARAGDIAIYWGQNGNEGTLAQTCATGNYRFVIVAFLPVFGKGQTPVLNLAGHCDPASNGCTGVGADIKSCQ
SLGIKVMFSIGGGVGNYGLSSRDDAKQVAAYLWNNYLGGTSPSRPLGDAVMDGIDFDIESGGGMYWDDLARYLKAYSRQGSSKKPVYL
TAAPQCPFPDASLGVALSTGLFDYVWVQFYNNPPCQYSSSNGVGNLASAWKQWTSIPAGRVFLGLPAAAEAAGSGFVETSDLVSKVLP
VVKKSPKYGGIMLWSRYYDGLTGYSDKVKSSV

>Os_9629.m04512
MAANKLKFSPLLALFLLAGIAVTSRAGDIAVYWGQNGDEGSLADACNSGLYAYVMVAFLSTFGNGQTPVLNLAGHCEPSSGGCTGQSS
DIQTCQSLGVKVILSIGGGAGSYGLSSTQDAQDVADYLWNNFLGGSSGSRPLGDAVLDGVDFDIETGNPAHYDELATFLSRYSAQGGG
KKVILTAAPQCPYPDASLGPALQTGLFDSVWVQFYNNPPCQYANGDASNLVSAWNTWTGGVSAGSFYVGVPAAEAAAGSGYVAPGDLT
SAVLPAVQGNAKYGGIMVWNRFYDVQNNFSNQVKSSV

>Os_9639.m04451
MSKLQLRPPLLATLHCSLLVLLIINGAAAAGKTGELTVIWGRNKDEGSLRSTCDTGLYTTVVISFLTVFGHGRYRTDLAGHPLAGVGA
DVKHCQKAKNVTVLLSIGGAGDQYSLPTAKSAQDVAEHLWHAYLGGGRRGVSRPFGDAVLDGVDVYVDRGRWGHYDELARRLRSFGRE
KPAVRLTASPACSLALFDDEVETMKTLSLFERLHVRFYNESSCDYNYFETRPFWGAWRTWTSRFPAARVQVGWPAMEEMSGFVDPQTL
RESVLSSVQDDANYGGVMLWDRYYDKITGFGRAIKDIV

>Os_9639.m04450
MARHQLCLLTTTMLVAVVVFLPCLATATGKTGQIAVFWGRNKTEGSLKEACDTGLYTTVIISFFSVFGHGRYWTDLSGHDVSRVGADV
KHCQSKNIPVLLSVGGDGYQYSLPTANSAKDVADHLWHAYLGGGRRGVFRPFGDAVLDGVDLYIDHGGPANYDVLVRRLAGYRGKPVL
LTATPRCVYPDANAAAALGTGLVRRIHPRFYGDAACTNKTDGEGRRSLFDWEDWDAWTSRFPASQVYVGLPAEETAADWINPESLYYA
VMQRAQTASNYGGAMLWDRGADKAYDNYYGRALKDFV

>Os_9639.m04449
MASSSQRRRALPLSFVVIVLLILAGPGPVAGDKTGELTVFWGRNKDEGSLREACDTGIYNTVIISFLTVFGHGRYWADLSGHPVAGVG
ADIKHCQHAKNVTVLLSIGGDGDQYSLPTPRSAKDVADHLWHAYLGGGRHGVFRPFGDAVVDGIDLYIDHGGSANYDELAKRLGEHGG
VLLTATVRCMDGQETSGEAAVATGLIGRIHVRFYDDRRCSYDSSERRPFYGAWLGWTARYANASVHVGLPAAWDAASDGWINPAALVF
DALPLVRGTPNYGGVVLWNRHFDRRSRYGQTIKGML

>Os_9639.m04440
MASQRRRSSATAVLLSLLLLLQLAAAYPQGKRQNVAVFWGRNKAEGSLRQTCDTGDYNIVIISFLSVFGHGKYWLDLSGHDLRDVGAD
IRHCQSKGVYMLLSIGGDGYQYSLPSSKSAADVAENLYYSVLGGDRPGAFHPFGDDTIVNGVDFFIDNGPADHYDDLANRINDYNQNI
RDPIGIMLTATVRCSYPDPRMKAALDTKLFRRIHVRFYDDPSCSYNHAGLAGVMAQWNKWSATYPDGQIFLGVVAANLTGKNDMVAVG
ELRDKLLPAVQNTDTYGGVMLWNSYYDSLTHYGRYVKDLA

>Os_9639.m04448
MVALGRRSWLVPLAMVLAVSSCLAGPAMAAGKTGQMTVFWGRNKNEGTLKETCDTGLYTTVVISFYSVFGHGRYWGDLSGHDLRVIGA
DIKHCQSKNIFVFLSIGGAGKDYSLPTSKSAADVADNIWNAHMDGRRPGVFRPFGDAAVDGIDFFIDQGAPDHYDDLARNLYAYNKMY
RARTPVRLTATVRCAFPDPRMKKALDTKLFERIHVRFYDDATCSYNHAGLAGVMAQWNKWTARYPGSHVYLGLAAANVPGKNDNVFIK
QLYYDLLPNVQKAKNYGGIMLWDRFYDKQTGYGKTVKYWA

>Os_9633.m01410
MALRRLAALLSLAVLLSAGLAAVSATSQNTGDTVIIWGRNKDEGSLREACDAGRYTTVIISFLSAFGYIPGTYKLDISGHQVSAVGPD
IKYCQSKGKLILLAIGGQGGEYSLPSSQAAVDLHDHLWYSYLGGRRNGVYRPFGDANVNGIDFFIDQGAREHYNELAKMLYDHNKDYR
ATVGVMVTATTRCGYPDHRLDEALATGLFHRIHVKMFSDGRCPAWSRRQSFEKWAKTYPQSRVLIGVVASPDVDKDAYMPPEALNNLL
QFINKQPNFGGVMVWDRFYDKKTGFTAHL

>Os_9633.m01415
MALRRHAALLSLAVVLLFAGLAAASPTSQNTGDTVIFWGRNKDEGSLREACDTGLYTTVIISFLSAFGYKPGYYKVDISGHPVSAVGP
DIKYCQSKGILILLAIGGQGGEYSLPTPQAAVELNDHLWYSYLGGHRNGVYRPFGDAIVNGIDFFIDQGGRENYNKLAKLLYAHNKDY
RGTVGVMLTATTRCEYPDHRLDEALATGLFHRIHVKKFSDGRCPASSWIQSFQKWAKMYPQSRVLVGVVASREVDREAYISPEDLKKL
MQYVFSKLPNFGGVMVWDRFYDEKTGFTGRLRA

>Os_9633.m01399
MASRRLAPLLVLLLSSSLAAGQSTGDTVVFWGRNTDQLEGSLREACDTGLYTTVIISFLSAFGYIPGTYKLDISGHTVSAVGPDIKHC
QSRGILVLLAIGGQGGEYSLPTSQAAVDLEDYLWNAFLGGGRPGVARPFGDAVVDGIDFFIDQGATEHYDELARRLYAHNKDYKGRLG
VMLTATARCVFPDQRLQAALSTGLFSRIHVKVFGDGRCASRREELEKWMAAYPQSRVLVGVVASPEADRDGYVSHKDLYYDVLQFINK
LPNYGGIMVWNRYWDKKTGWTAGNEPGLLDQVL

>Os_9639.m04443
MAFGRRSLFLPVVGVAAILLLAAGHATAVNTGETVVFWGRNKDEGSLREACDTGLYTSVIISFLAVFGHGRYSLDLSGHDVSAVGADI
KHCQSKYIPVLLSIGGQGGAYSLPTNASAADVADHLWDSFLGGGRAGVPRPFGDAVVDGVDLFIDQGGAEHYDELARRLFSHYKFEML
LTATTRCSYPDHRLDMALATGLFTHIHVRVFGGGGDAGCTTRHRASWERWAAAYPGSLVYLGVVASPEQDANAYLPRKVLFSDVLSHI
VEKPNYGGLMIWDRYYDKKTGYSAGKVF

>Os_9639.m04445
MAFRRRSCIPAALAVFFLLLAGQSTAGEDQTAVIWGRHGDEGTLREACDTGHYNTVIISFLSVFGHGRYSLDLSGHDLRRVGNDIKHC
QRKGIVVLLSIGGQGGDYSLPSSRSAADVADNLWNAFLAGRRKGVLRPFGNAAVDGIDFFIDRGSGDHYDELARKLYSYRNNKGKGVM
LTATPRCRFPDRRLEKALATGVFARIHVRMFGDDVNCTAAPRESWEKWAAAYPASQVYLGLVASSEQDPGYLSPKPLYYTLVMYIRDR
LNYGGKMIWDRYYDKKTDYSIGKLI

>Os_9639.m04442
MGLVHALLPFAAAAALLLLAAPPPATADDPGLAVYWGRHKEEGSLREACDTGRYTTVIITFYNAFGHGRYSLDISGHPLAAVGADIKH
CQSRGITVLLSIGGQGGAYSLPTNASAADVADNLWNAYLGGHRAGVARPFGDDAAVDGIDFFIDQGGADHYDDLARRLDGYNKYYRGR
VGVLLTATTRCSYPDHRLEKALATGVFARIHVRMFGDEQCTMSPRYSWEKWAAAFPGSKVYIGLVASPEQDSAWMFQKDLYYEMLQFV
RSLPNYGGLAIYDRYFDKKANYTGEG

>Os_9639.m04446
MKMKALLPVAAMLLLVSGQLAAPVTADGYVGQLAVFWGRHKEEGSLREACDTGRYNIVVITFYNVFGYQRGRYGLDFSGHPVAAVGAD
IKHCQSKGVQVLLSIGGQGGEYSLPSSQSASDVADNLWNAYLGGRRAGVPRPFGDAVVDGIDFFIDQGGADHYEQLARQLHGRGVLLT
ATVRCAYPDSRMEAALATGVFARIHVRIFGDDQCTMFPKDAWEKWAAAYPRCTVFLTVVASPEQDEGYMFQKDLYYGVQQFIDKEPNY
GGIAIWDRYYDKKANYSGEG

>Os_9639.m04447
MMGLLSLLLVVVSCLAAPATADWYGPLAVYWGRHKDYEGSLREACDTGRYNTVIITFYSVFGYVKGRYGLDISGHPVAAVGADIKHCQ
SKGVQVLLSIGGQGGGYSLPSSQSAADVADNLWNAYLGGRRAGVPRPFGDAAVDGIDFFIDQGGADHYDELARQLHGRGVALTATVRC
SYPDSRLQKALATGLLGRIHVRIFGDNQCTMLPLDAWEKWAAAYPRSKVWLALVASWEQDEVGYMFQKDLYYGVLQFILNKPNYGGIA
IWDRYYDKKANYSGEG

>Zm_AAA62421
MMRALAWWPCWPAFFAVPARAEQCGSQAGGALCPNCLCCSQFGWCGSSDYCGSGCQSQCSAACSTPNPPSSGGVASIIPESLFNQMLL
HRNDAACPANGFYTYAGFIAAANAFPGLAPRVRPDVQKRELAAFLAQTSHETTGGWATARRPYAWGYCFKEEQGGASGPDYCEPSAQW
PCAAGKKYYGPRAIQISININIGPAGRPSAPGILANPDLVATRPHRVVRRPPSGSGMTPQSPKPSCTDVMTGQWTPSAADTAAGRLPG
YGVVTNISNGGLECGHGADSRVADRIGFYKRYCDLLGVSYGDNLDCANQTPFNG

>Bn_AAB01665
LTTLFFLILTVSKPVASQNCGCASGLCCSSAGYCGTTDAYCGEGCKEGPCKNSGPGDPTVSLEETVTPEFFNSILSQATESDCKGRGF
YTHETFMAAANAYPSFGATISKREIAAFFAHVAQETGFMCHIEEVDGPAKAARGDYCDTTKTEFPCVPGKGYYGRGAIQLSWNYNYGP
CGRDLNEGDLLATPEKVAQDQVLAFKASFWYWTTNVRSSFKSGFGATIRAVNSMECTGGGVPSETAANRIRYFQDYCTKLGVQPGENL
SC

>Dj_AAB23692
ENCQCDTTIYCCSQHGYCGNSYDYCGPGCQAGPCWDPCEGDGTLTVSDIVTQEFWDGIASQAAANCPGKSFYTRSNFLEAVSAYPGFG
TKCTDEDRKREIAAYFAHVTHETGHLCYIEERDGHANNYCQESQQYPCNPNKEYFGRGPMQLSWNYNYIDAGKELHFDGLNDPDIVGR
DPIISFKTSLWFWIRKGVQYVILDPNQGFGATIRIINGGQECDGHNTAQMMARVGYYQEYCAQLGVSPGNNLPC

>St_AAB96340
FLTETLAQNAGSIVTRELFEQMLSFRNNDVCPGKGFYTYDAFIAAANSFPAFGTTGDDTARKKEMAAFFGQTSHETNGGSAGTFTGGY
CFVRQIDQSERYYGRGPIQLTHQSNYERAGQGIGVGQDLVNNPDLVATDAIISFKTAIWFWMTAQDNKPSCHNVIIGQWTPSPADTAA
NRVPGYGVITNIINGGLECGMGQNTAVESRIGFYRRYCGMLNVPTGENLDCNNQRNFAQG

>Ca_AAC36359
MEFSVSPVALSCLFFLFLITGTLAQNIGSIVTRDLFERMLSFRNNAACPGKGFYTYEAFITAANSFPAFGTTGDNTARKKEVAAFFGQ
TSHETNGGRAGTFNGGYCFVRQIDQSDRYYGRGPIQLTHRSNYERAGRGIGVGQDLVNNPDKVATNPVISFKTAIWFWMTAQDNKPSC
HDVIIGRWKPSPADRSANRVPGYGVITNIINGGIECGKGRNGAVESRIGFYKRYCGMLNVPTGNNLDCYNQKNFAQG

>Os_AAM08773
MVNGYLFREYIGAQFTGVRFSDVPVNPGLSFHFILAFAIDYFMATQSSKPAPANGVFAPYWDTANLSPAAVAAAKAAHPNLSVILALG
GDTVQNTGVNATFAPTSSVDAWVRNAADSVSGLIDAYGLDGVDVDYEHFAAGVDTFVECIGRLLTELKARHPNIATSIAPFEHPVVQR
YYQPLWRRYAGVIDYVNFQFYGYGANTDVATYVMFYDEQAANYPGSKLLASFKTGNVTGLLSPEQGIAGAKELQRQGKLPGLFIWSAD
SSMVSSYKFEYETKAQEIVANH

>Mc_AAM18075
MSHPTIVNGGYWLSSTAGKLPASNIDATLYTHLFCGYGEIDATSHEVTIPSTEVSHFSEKVKVSNPSVKTLLSIRGDSKVFSDIAKVP
TARDTFVNSSIKAAVDCGFDGLDLQWLYPSSPTDMQNFKTLIVKWRQAAVDDARNNNRPERILVASVPNLPYIEKNVEYPGEAIKQNL
DWVNLVCYDFYTPAEGSSPDYTGPSSALFNRRSNSLSAHFGIESWTKYVPAKKIVFGIPFHGWAWTLANAAEHDMFSKATGSCVWPMD
YSEVKSHRRARQCEYNIPNDPRFGIAYAYVETTWIGYESESTIPVKIKFAKQNQLLGYFAWNISGDDDVNTLANAAQSNWQ

>Vv_AAM95447
MESKALALILLAGILAGALPGSTVAQDCGCSSDQCCSQWGYCGTGDDYCGTGCQSGPCTSTSNNVSVSDVVTQEFFNGIIDQADASCA
GKNFYTRAAFLDALNSYPGFGTEGPTDDSLREIAAFFGHVTHETGHFCYIEEIDGASNDYCDEDNTQYPCVSGKDYYGRGPIQLSWNY
NYGAAGDSIGFDGLNNPETVATDVDISFKTALWYWMTNVHSVIDQGFGATIRAINSIECDGGNTAAVNARVEYYTEYCNELGVSTGDN
LTC

>Ca_AAN37391
MTINLLLPSILFLALIQTSIARSGIAIYWGQNGNEATLNDTCASGNYAYVNLSFLNKFGNGQTPEINLAGHCNPAVNGCTILGPQIKF
CQKLGVKVMLSMGGGVGNYSLASKKDAKDVARYLYNNFLGGRSSFRPLGNARLDGIDFDIELGSSLYYEDLAQYLKRYSKLGRKMYLT
AAPQCPFPDRLLGTALNTGLFDNVWIQFYNNPSCQYTTNNVDDLKNSWTRWTTSVNARRIFLGLPAAPQAAGSGFIPAEVLTGGILPV
IKKSRKYGGVMLWSKFWDEQTGYSASIVKSV

>Rg_AAO47731
MAAVRAALLFILPVFSALSSRNASGKISIYWGQNGNEGTLAETCATGNYHYVNIAFLPTFGNGQTPMINLAGHCDPTITNGCTHLSSQ
IKSCQAKGIKVMLSIGGGAGSYYLSSSQDAKQVATYLFNNFLSGKSSPRPLGDAILDGIDLDIEGGTDLYWDDLARYLSNYGKRGRKV
YLTAAPQCPFPDYYIGNALQTGLFDYVWVQFYNNPPCQYSSGMDSFEKAWKDWNSIPAGEIFLGLPASAQAAGTGFVPAGDLTSQVLP
AIKGSAKYGGVMLWDKYHDTGYSSSIKKDV

>Gh_AAP80801
MADFKKATALILTVALLVNLALMVDADGDDDKKIRVRKHKGEKQCIQGWECSYWSKYCCNKTVSDVFQVYQFEDLFAKRNSPVAHAVG
FWDYHSFILAASIYEPLGFGTTGGKRMQMKEVAAFLAHVGAKTSCGDGVIDGGPLAWGLCFKREMSPSQDYCDDYYKYMYPCAPGAQY
YGRGALPIYWNYNYGAAGDGIKVDLLHHPEYLEQNATIAFQAAIWRWMTPIKKNQPSAHDIFVGNWKPTKNDTEEKRGPTFGSTMNVL
YGDYTCGQGDIDPMNIIISHYLHYLDLLGVGREEAGPHEELSCAEQKAFNPTPAPPAASAS

>Vv_AAQ10093
MAAKLLTVLLVGALFGAAVAQNCGCASGLCCSKYGYCGTGSDYCGDGCQSGPCDSSSGSGSSVSDIVTQSFFDGIINQAASSCAGKNF
YTRAAFLSALNSYSGFGNDGSTDANKREIAAFFAHVTHETGHFCYIEEINGASHNYCDSSNTQYPCVSGQNYYGRGPLQLTWNYNYGA
AGNSIGFNGLSNPGIVATDVVTSFKTALWFWMNNVHSVIGQGFGATIRAINGAVECNGGNTAAVNARVQYYKDYCSQLGVSPGDNLTC

>Ta_AAX83262
MRSLAVVAVVVATVAMAIGAARGSVSSVVSRAQFDRMLLHRNDGACQAKGFYTYDAFVAAAAAFPGFGTTGSADAQKREVAAFLAQTS
HETTGGWATAPDGAFAWGYCFKQERGAASDYCTPSAQWPCAPGKRYYGRGPIQLSHNYNYGPAGRAIGVDLLSNPDLVATDPTVSFKT
AMWFWMTAQAPKPSSHAVITGQWSPSGADRAAGRVPGFGVITNIINGGIECGHGQDSRVADRIGFYKRYCDILGVGYGDNLDCYNQRP
FA

>Vv_BAC65326
MARTPQSTPLLISLSVLALLQTSYAGGIAIYWGQNGNEGTLTQTCNTGKYSYVNIAFLNKFGNGQTPEINLAGHCNPASNGCTSVSTG
IRNCQNRGIKVMLSIGGGVGSYSLSSSNDAQNVANYLWNNFLGGQSSSRPLGDAVLDGIDFDIELGSTLHWDDLARALSGFSKRGRKV
YLTAAPQCPFPDKFLGTALNTGLFDYVWVQFYNNPQCQYSSGNTNNLLNSWNRWTSSINSQIFMGLPASSAAAGSGFIPANVLTSQIL
PVIKRSPKYGGVMLWSKYYDDQSGYSSSIKSSV

>Cj_BAD77932
MQIMATQNSKSNIFWSSSASVVLVLLLLVDVGVCQNCGCNGLCCSQYGYCGSGEAYCGAGCKEGPCSSSSPPSTGTGVGSIVSSDVFN
SIVGGAASGCAGNGFYTYDSFISAANAFNGFGTSGSSDVNKREIAAFFANAAHETGGFCYIEEQNPTSIYCDASNTQYPCASGKTYHG
RGPLQLSWNYNYGAAGSYIQFDGLNNPEIVGTDSTISFKTAVWFWMVNSNCHTAITSGQGFGATIRAINSMECDGGNAATVASRVNYY
QKFCQQLNVDTGSNLQC

>Nt_CAA54373
IIFSCFLLRQLVCTNSQNVIKGGYWFKNSGLALNNIDSTLFTHLFCAFADLNPQSNQLIISPENQDSFSQFTSTVQRKNPSVKTFLSI
AGGRADTTAYGIMARQPNSRKSFIDSSIRLARQFGFHGLDLDWEYPLSATDMTNLGILLNEWRTAINMEARNSGRAALLLTAAVSYSP
RVNGLNYPVESVARNLNWINLMAYDFYGPNWSPSQTNSHAQLFDPVNHISGSDGINAWIQAGVPTKKLVLGIPFYGYAWRLVNPNIHD
LRAPAAGKSNVGAVDDGSMTYNRIRDYIVQSRATTVYNATIVGDYCYSGSNWISYDDTQSVRNKVNYVKGRGLLGYFAWHVAGDQNWG
LSRTASQTWGVSSQEMK

>At_CAA74930
MLTPTISKSISLVTILLVLQAFSNTTKAQNCGCSSELCCSQFGFCGNTSDYCGVGCQQGPCFAPPPANGVSVAEIVTQEFFNGIISQA
ASSCAGNRFYSRGAFLEALDSYSRFGRVGSTDDSRREIAAFFAHVTHETGRNFCYIEEIDGASKDYCDENATQYPCNPNKGYYGRGPI
QLSWNFNYGPAGTAIGFDGLNAPETVATDPVISFKTALWYWTNRVQPVISQGFGATIRAINGALECDGANTATVQARVRYYTDYCRQL
GVDPGNNLTC

>La_CAA76203
MASLKQVSLILFPLLLLISSSFKLSNAAGIVIYWGQNGNEGSLADACNTNNYQYVNIAFLSTFGNGQTPELNLAGHSRDGLNADIKGC
QGKGIKVLLSLGGGAGSYSLNSADDATNLANYLWNNFLGGTSDSRPFGDAVLDGIDFDIEAGGAQHYDELARALNGFSSQKKVYLGAA
PQCPIPDAHLDAAINTGLFDYVWVQFYNNPQCQYACGNTNNLINSWNQWTSSQAKQVFLGLPASEAAAPSGGFIPTDVLISQVLPTIK
TSPKYGGVMLWNGFNDIQTGYSDAIKASV

>Nt_CAA77656
MIKYSFLLTALVLFLRALKLEAGDIVIYWGQNGNEGSLADTCATNNYAIVNIAFLVVFGNGQNPVLNLAGHCDPNAGACTGLSNDIRA
CQNQGIKVMLSLGGGAGSYFLSSADDARNVANYLWNNYLGGQSNTRPLGDAVLDGIDFDIEGGTTQHWDELAKTLSQFSQQRKVYLTA

```
APQCPFPDTWLNGALSTGLFDYVWVQFYNNPPCQYSGGSADNLKNYWNQWNAIQAGKIFLGLPAAQGAAGSGFIPSDVLVSQVLPLIN
GSPKYGGVMLWSKFYDNGYSSAIKANV


>Nt_CAA77657
MNIKVSLLFILPIFLLLLTSKVKAGDIVVYWGQDVGEGKLIDTCNSGLYNIVNIAFLSSFGNFQTPKLNLAGHCEPSSGGCQQLTKSI
RHCQSIGIKIMLSIGGGTPTYTLSSVDDARQVADYLWNNFLGGQSSFRPLGDAVLDGIDFDIELGQPHYIALARRLSEHGQQGKKLYL
TAAPQCPFPDKLLNGALQTGLFDYVWVQFYNNPECEFMSNSENFKRRWNQWTSIPAKKLYIGLPAAKTAAGNGYIPKQVLMSQVLPFL
KGSSKYGGVMLWNRKFDVQCGYSSAIRGAV


>Hv_CAB99486
MLRSPRSCSPWRWAAPRRRAWAPSSRNRCYASMLPNRDNSQCPGQGFYTYVRFIAAANTFPGFGTTGSADDVKRELAAFFGQTSHETT
GGTRGAADQFQWGYCFKEEINKATSPPYYGRGPIQLTGQSNYDLAGRAIGKDLVSNPDLVSTDAVVSFRTAIWFWMTAQGNKPSSHDV
ALGRWTPTAADTAAGRVPGYGVITNIINGGLECGMGQNDANVDRIGYYTRYCGMLGTATGGNLDCYTQRNFAS


>Pg_AAA85364
MGSSSSDKSVMALVLVLLLVGVSVNAQNCGCASGVCCSQYGYCGTTSAYCGKGCKSGPCYSSGGGSPSAGGGSVGGIISQSFFNGLAG
GAGSSCEGKGFYTYNAFIAAANAFSGFGTTGSNDVKKRELAAFFANVMHETGGLCYINEKNPPMKYCQSSSTWPCTSGKSYHGRGPLQ
LSWNYNYGAVGKSIGFDGLNNPEKVGKDPTTSSKTAVWFWMKNRNCHSAITSGKGLGGTIKAINSMECNGGNSGEVNSRVNYYKKICS
QLGVDPGANVSC


>Dc_AAB08468
MKTSFATKFIFLIFIAAPLVSAQNCNCAAGLCCSKYGYCGTTSDYCGEGCQAGPCTNSAPSGGGNAVSVADIVTDDFFNGIISQATGD
CDGKNFYTRSAFLEALQSYSSFGTSGSADDSKREIAAFFAHATHETGYFCHKEETNGRDHNYCQSTAEYPCNPNVNYFGRGPLQLTWN
YNYIDAGKSNQFDGLNNPDIVASDAVVSFKTALWYWKVKVQSVTSQGFGATIRAINSIECNSGSPDAVNSRVSLYNSYCSKFGVAPGD
NQGC


>Vv_AAB65776
MAAKLLTVLLVGALFGAAVAQNCGCASGLCCSKYGYCGTGSDYCGDGCQSGPCDSSGSSVSDIVTQSFFDGIISQAASSCAGKNFYTR
AAFLSALNSYSGFGNDGSTDANKREIAAFFAHVTHETGHFCYIEEINGASHNYCDSSNTQYPCVSGQNYYGRGPLQLTWNYNYGAAGN
SIGFNGLSNPGIVATDVVTSFKTALWFWMNNVHSVIGQGFGATIRAINGAVECNGGNTAAVNARVQYYKDYCSQLGASPGDNLTC


>Cd_AAC95375
MAYSDALLFAVTAVASLVTSGGFFAEARWYGPGGKCSSVEALAARAFPKFAGTGDLATRKRELAAFFAQISHETTGGWATAPDGPYSW
GLCYKEEISPASNYCDATDKQWPCYPGKSYHGRGPIQLSWNFNYGPAGQALGFDGLRNPEIVANCSDTAFRTALWFWMTPRRPKPSCH
EVMVGEYRPTATDVAGNRMPGFGLVTNIVNGGLECNRTDDARVNNRIGFYRRYCQIFNVDTGPNLDCAHQQPY


>Sc_AAG53609
MRGVVVVAMLAAAFAVSAHAEQCGSQAGGATCPNCLCCSKFGFCGSTSEYCGDGCQSQCNRCGGTPVPVPTPTGGGVSSIISQSLFDQ
MLLHRNDAACLAKGFYNYGAFIAAANSFSAFATTGGTDVRKREVAAFLAQTSHETTGGWPTAPDGPYSWGYCFNQERGAPSDYCSPSS
QWPCAPGKKYFGRGPIQISYNYNYGPAGRAIGTDLLNNPDLVATDATVSFKTALWFWMTPQSPKPSSHDVITGRWSPSGADQAAGRVP
GYGVITNIINGGLECGRGQDARVADRIGFYKRYCDLLGVSYGHNLDCYNQRPFA


>Sc_AAG53610
MARFAALAALLLAVAVGGAAAQSVGSVITQSMYASMLPNRDNSLCPARGFYTYDAFIAAANTFPGFGTTGSTDDVKREVAAFFGQTSH
ETTGGTRGAADQFQWGYCFKEEINKATSPPYYGRGPIQLTGRSNYDLAGRAIGKDLVSNPDLVSTDAVVSFRTAMWFWMTAQGNKPSS
HDVALRRWTPTAADNAAGRVPGYGVITNIINGGLECGMGRNDANVDRIGYYTRYCGMLGTATGGNLDCYTQRNFAS


>Ma_AAS12600
MAIRSPASLLLFAFLMLALTGRLQARRSSCIGVYWGQNTDEGSLADACATGNYEYVNIATLFKFGMGQTPEINLAGHCDPRNNGCARL
SSEIQSCQERGVKVMLSIGGGGSYGLSSTEDAKDVASYLWHSFLGGSAARYSRPLGDAVLDGIDFNIAGGSTEHYDELAAFLKAYNEQ
EAGTKKVHLSARPQCPFPDYWLGNALRTDLFDFVWVQFFNNPSCHFSQNAINLANAFNNWVMSIPAQKLFLGLPAAPEAAPTGGYIPP
HDLISKVLPILKDSDKYAGIMLWTRYHDRNSGYSSQVKSHVCPARRFSNILSMPVKSSK


>Nt_CAA54374
MANSVTLFAIIFSCFLLQQLVCTNSQNVKGGYWFKDSGLALNNIDSTLFTHLFCAFADLNPQLNQLIISPENQDSFRQFTSTVQRKNP
SVKTFLSIAGGRANSTAYGIMARQPNSRKSFIDSSIRLARQLGFHGLDLDWEYPLSAADMTNLGTLLNEWRTAINTEARNSGRAALLL
TAAVSNSPRVNGLNYPVESLARNLDWINLMAYDFYGPNWSPSQTNSHAQLFDPVNHVSGSDGINAWIQAGVPTKKLVLGIPFYGYAWR
LVNANIHGLRAPAAGKSNVGAVDDGSMTYNRIRDYIVESRATTVYNATIVGDYCYSGSNWISYDDTQTVRNKVNYVKGRGLLGYFAWH
VAGDQNWGLSRTASQTWGVSFQEMK


>Sr_CAA88593
MAPKRQALILIVLSLLTINTSEAATGGIAIYWGQNNGDGTLTSTCDTGNYEIVVLSFLTTFGCSRTPQWNFAGHCGDWSPCTKLQPEI
QHCQQKGVKVFLSLGGASGSYSLCSPQDAKEVADYLFSNFLTGRYGPLGSVTLDGIDFDIEGGSNLYWDDLAKELDALRQTNNYFYLS
AAPQCPIPDYYLDKAIKTGLFDYVFVQFYNNPPCQYSNGNTGPLLGSWDAWTSLVLPNNTVFMGLPASREAAPSGGYIPPNVLISEVL
PYIKQASNYGGIMLWSRFQDVTNHYSDQIKYYVTKYMLRFVKAVSNAISDCVSAALHRFLPKPY


>Ms_AAN10048
TRLLLLLVLTIFPFTIKASSSGGIAIYWGQNLGDGTLTSTCDTGNYEIVLLAFLNVFGGGRVPNWNFAGHCGDWSPCTKLEPEIKHCQ
QKGVKVLLSIGGAVGSYSLSSPEDAKNVADYLHTNFLSGQFGPLGSVTLDGVDFDIEGGTNLYWDDLARDLDNLRQQNRYFYLSAAPQ
CFMPDYYLDKAIKTGLFDYVLVQFYNNPPCQYDIKNSDPKLLLQSWNAWTSLVLPNNTVFMGLPAAPNAAPSGGYIPPDDLISKVPPS
IKPTSNYGGIMLWDRFHDVTGNNDYIYSDQIKEHVKRSVLRFKTQVSEAISRCISAALNPMLPN
```

>Pa_AAQ17050
MGIIIIDKSVMARVLVLLLVGFIVNAQNCGCATGLCCSQYGYCGTTSAYCGKGCKTGPCYSSGGGSPSAGGGSVGGIISQSFFNGLAG
GAGSSCEGKGFYTYNAFIAAANAYSGFGTTGSNDVKKRELAAFFANVMHETGGLCYINEKNPPINYCQSSSTWPCTSGKSYHGRGPLQ
LSWNYNYGAAGKSIGFDGLNNPEKVGQDSTISFKTAVWFWMKNSNCHSAITSGQGFGGTIKAINSMECNGGNSGEVSSRVNYYKKICS
QLGVDPGANVSC

>Nt_CAA34812
MRLCKFTALSSLLFSLLLLSASAEQCGSQAGGARCPSGLCCSKFGWCGNTNDYCGPGNCQSQCPGGPTPTPPTPPGGGDLGSIISSSM
FDQMLKHRNDNACQGKGFYSYNAFINAARSFPGFGTSGDTTARKREIAAFFAQTSHETTGGWATAPDGPYAWGYCWLREQGSPGDYCT
PSGQWPCAPGRKYFGRGPIQISHNYNYGPCGRAIGVDLLNNPDLVATDPVISFKSALWFWMTPQSPKPSCHDVIIGRWQPSAGDRAAN
RLPGFGVITNIINGGLECGRGTDSRVQDRIGFYRRYCSILGVSPGDNLDCGNQRSFGNGLLVDTM

>Hv_2BAA
SVSSIVSRAQFDRMLLHRNDGACQAKGFYTYDAFVAAAAAFPGFGTTGSADAQKREVAAFLAQTSHETTGGWATAPDGAFAWGYCFKQ
ERGASSDYCTPSAQWPCAPGKRYYGRGPIQLSHNYNYGPAGRAIGVDLLANPDLVATDATVGFKTAIWFWMTAQPPKPSSHAVIAGQW
SPSGADRAAGRVPGFGVITNIINGGIECGHGQDSRVADRIGFYKRYCDILGVGYGNNLDCYSQRPFA

Appendix 7.5

SignalP prediction of signal peptides in the three non-redundant chitinase-like amino acid sequences from spruce. The software predicts signal peptides and its cleavage sites in eukaryotes using neural networks and hidden Markov models. A high *S-score* corresponds to amino acids that are part of a signal peptide. The *C-score* corresponds to the cleavage site. For each position in the submitted sequence, a C-score is reported, which should only be significantly high at the cleavage site. *Y-max* is a derivative of the C-score combined with the S-score resulting in a better cleavage site prediction than the raw C-score alone. The *S-mean* is the average of the S-score. The *D-score* is average of the S-mean and Y-max score. This score gives a good assessment for discrimination of secretory and non-secretory proteins.

```
>Pg_GQ0134.B7.1_C02_5  length = 70
# Measure   Position  Value  Cutoff  signal peptide?
   max. C    27        0.792  0.32    YES
   max. Y    27        0.765  0.33    YES
   max. S    14        0.996  0.87    YES
   mean S    1-26      0.872  0.48    YES
        D    1-26      0.818  0.43    YES
# Most likely cleavage site between pos. 26 and 27: VSA-EQ
```


SignalP-NN prediction (euk networks): Pg GQ0134.B7.1 C02 53 Contig1 1-1185
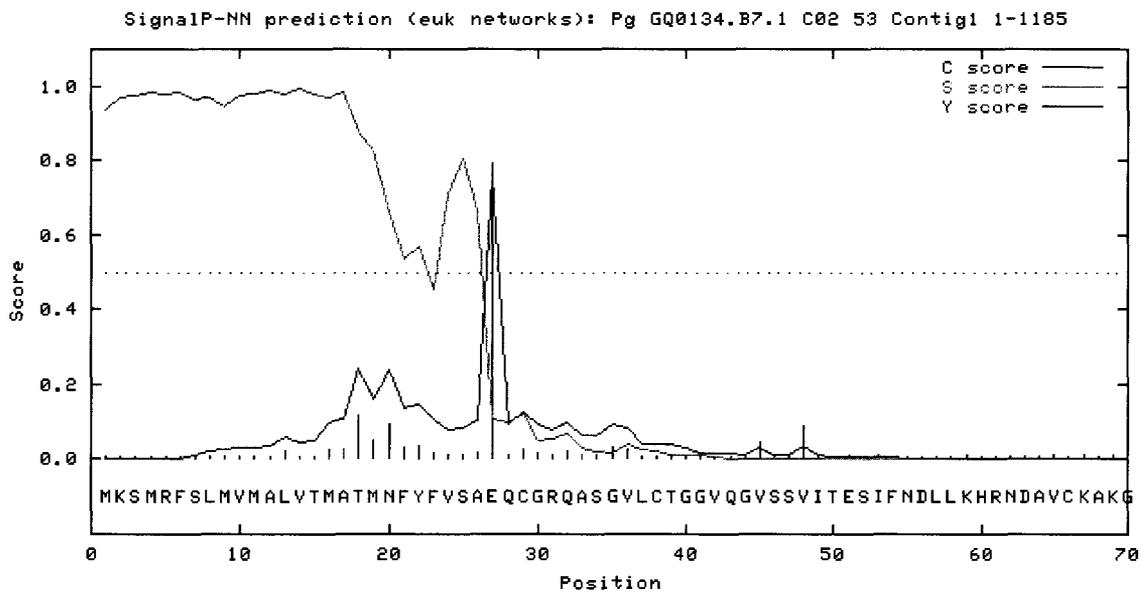
>Pg_GQ03602.B7_D17_5- length = 70

```
# Measure   Position   Value   Cutoff   signal peptide?
  max. C     30         0.420   0.32     YES
  max. Y     30         0.560   0.33     YES
  max. S     13         0.997   0.87     YES
  mean S     1-29       0.892   0.48     YES
       D     1-29       0.726   0.43     YES
# Most likely cleavage site between pos. 29 and 30: VSA-QQ
```


SignalP-NN prediction (euk networks): Pg GQ03602.B7 D17 5- Contig3 1-897
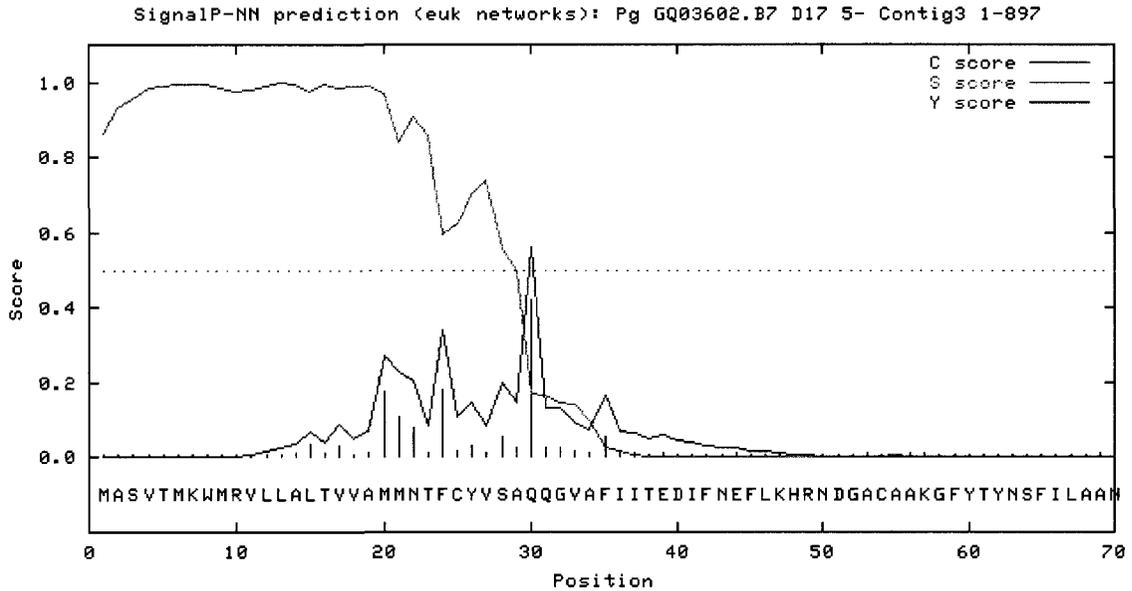
>Pg_GQ03602.B7_D17_53 length = 70

```
# Measure   Position   Value   Cutoff   signal peptide?
  max. C     30         0.420   0.32     YES
  max. Y     30         0.560   0.33     YES
  max. S     13         0.997   0.87     YES
  mean S     1-29       0.894   0.48     YES
       D     1-29       0.727   0.43     YES
# Most likely cleavage site between pos. 29 and 30: VSA-QQ
```


SignalP-NN prediction (euk networks): Pg GQ03602.B7 D17 53 Contig1 1-1101

195
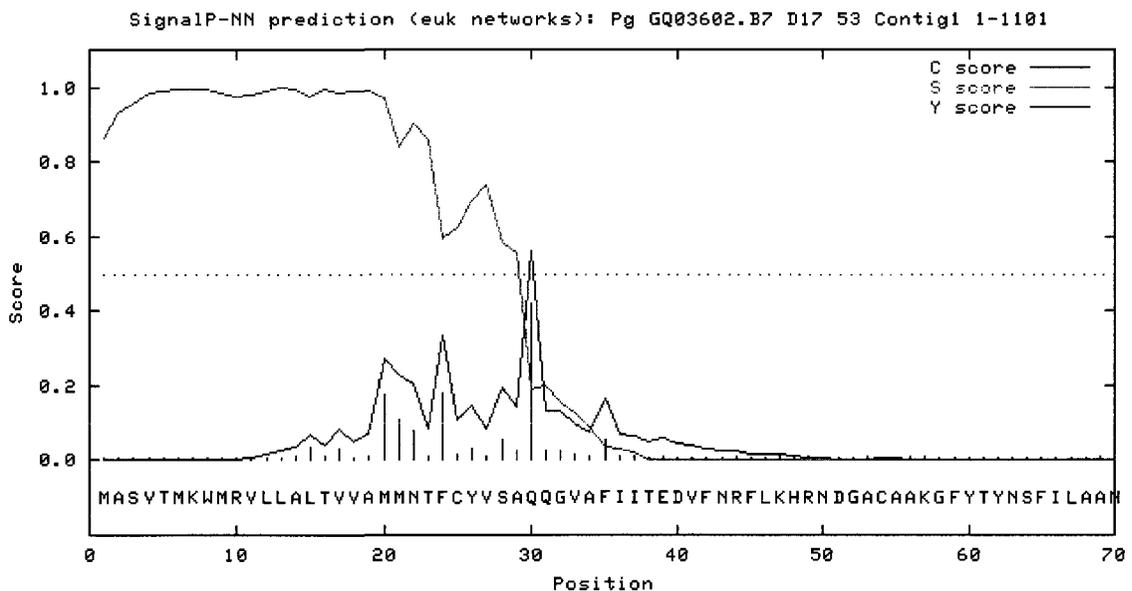
Appendix 7.6

InterProScan search of protein domains, families and motifs in the three non-redundant chitinase-like amino acid sequences from spruce. InterProScan, allows to search different features in primary sequence data by searching across more than 15 databases using different search algorithms.

PgGQ0134.B7.1_C02:53

PRODOM	PD354900	Q9FS45_VITVI_Q9FS45;	7e-106 [52-286]T
PFAM	PF00182	Glyco_hydro_19	2.3e-166 [48-280]T
PROSITE	PS00773	CHITINASE_19_1	NA [66-88]T
PROSITE	PS00774	CHITINASE_19_2	NA [193-203]T
PROSITE	PS00773	CHITINASE_19_1	8e-5 [66-88]T
PROSITE	PS00774	CHITINASE_19_2	8e-5 [193-203]T

PgGQ03602.B7_D17:5

PRODOM	PD354900	Q6E6M9_PICAB_Q6E6M9;	5e-94 [42-261]T
PFAM	PF00182	Glyco_hydro_19	1.8e-142 [36-254]T
PROSITE	PS00773	CHITINASE_19_1	NA [54-76]T
PROSITE	PS00774	CHITINASE_19_2	NA [167-177]T
PROSITE	PS00773	CHITINASE_19_1	8e-5 [54-76]T
PROSITE	PS00774	CHITINASE_19_2	8e-5 [167-177]T

PgGQ0134.B7.1_C02:53

PRODOM	PD354900	Glyco_hydro_19	0.0 [42-261]T
PANTHER	PTHR22595	Glyco_hydro_19_cat	6.1000265614980995E-90 [112-261]T
PFAM	PF00182	Glyco_hydro_19	0.0 [36-254]T

Appendix 7.7

WoLF PSORT results in the three non-redundant chitinase-like amino acid sequences from spruce. WoLF PSORT predicts protein sorting signals by comparing sequences to the closest neighbors bearing localization signals (sites).

PgGQ0134.B7.1_C02:53

id	site	distance	identity	Comments
FBH1_CUPWR	chlo	281.5	16%	[Uniprot] SWISS-PROT45:Chloroplast.
At4g35350.1	vacu	282.1	14%	[Arath]
GST1_LYCES	extr	291.5	8%	[Uniprot] SWISS-PROT45:Secreted.
GAS4_ARATH	extr	295.6	8%	[Uniprot] SWISS-PROT45:Secreted.
At2g44920.1	chlo	296.6	14%	[Arath] Subclass:thylakoid
RBS5_ACECL	chlo	297.5	14%	[Uniprot] SWISS-PROT45:Chloroplast.
RBS2_LEMGI	chlo	300.8	11%	[Uniprot] SWISS-PROT45:Chloroplast.
CHI4_BRANA	extr	304.8	23%	[Uniprot] SWISS-PROT45:Extracellular.
AMP_IMPBA	extr	306.2	13%	[Uniprot] SWISS-PROT45:Secreted.
At2g30290.1	golg	311.3	14%	[Arath]
FBH2_CUPWR	chlo	312	14%	[Uniprot] SWISS-PROT45:Chloroplast.
At3g50820.1	chlo	315.9	13%	[Arath] Subclass:thylakoid
RBSA_SOLTU	chlo	319.6	14%	[Uniprot] SWISS-PROT45:Chloroplast.
RBS1_LYCES	chlo	323.9	13%	[Uniprot] SWISS-PROT45:Chloroplast.

PgGQ03602.B7_D17:5

id	site	distance	identity	comments
MPH1_HOLLA	extr	265.7	11%	[Uniprot] SWISS-PROT45:Secreted.
NTPA_PEA	nucl	273.3	11%	[Uniprot] SWISS-PROT45:Nuclear.
At2g30290.1	golg	296.9	12%	[Arath]
At4g35350.1	vacu	309.9	15%	[Arath]
ASPR_HORVU	vacu	311.4	14%	[Uniprot] SWISS-PROT45:Vacuolar.
AMP1_MACIN	extr	319.2	10%	[Uniprot] SWISS-PROT45:Secreted.
DEF_NICAL	vacu	324.6	9%	[Uniprot] SWISS-PROT45:Vacuolar.

FBH1_CUPWR	chlo	330.4	15%	[Uniprot] SWISS-PROT45:Chloroplast.
At4g38240.1	golg	330.7	11%	[Arath]
FER_CAPAN	chlo	335.2	14%	[Uniprot] SWISS-PROT45:Chloroplast. GO:0009507; C:chloroplast; Evidence:ISS
CHI4_BRANA	extr	339.3	29%	[Uniprot] SWISS-PROT45:Extracellular.
ASPR_CUCPE	vacu	348	12%	[Uniprot] SWISS-PROT45:Vacuolar.
AFP4_RAPSA	extr	350.4	11%	[Uniprot] SWISS-PROT45:Secreted.
AKH2_MAIZE	chlo	352.7	8%	[Uniprot] SWISS-PROT45:Chloroplast.

PgGQ0134.B7.1_C02:53

id	site	distance	identity	comments
MPH1_HOLLA	extr	223.5	11%	[Uniprot] SWISS-PROT45:Secreted.
At4g35350.1	vacu	288.4	15%	[Arath]
NTPA_PEA	nucl	297.4	11%	[Uniprot] SWISS-PROT45:Nuclear.
FER_CAPAN	chlo	311.7	14%	[Uniprot] SWISS-PROT45:Chloroplast. GO:0009507; C:chloroplast; Evidence:ISS
AMP_IMPBA	extr	316.5	12%	[Uniprot] SWISS-PROT45:Secreted.
GST1_LYCES	extr	320	11%	[Uniprot] SWISS-PROT45:Secreted.
CHI4_BRANA	extr	322.1	29%	[Uniprot] SWISS-PROT45:Extracellular.
At2g30290.1	golg	322.3	12%	[Arath]
AMP1_MACIN	extr	324.5	11%	[Uniprot] SWISS-PROT45:Secreted.
RBS2_SPIOL	chlo	331.4	17%	[Uniprot] SWISS-PROT45:Chloroplast.
ASPR_HORVU	vacu	336	12%	[Uniprot] SWISS-PROT45:Vacuolar.
AFP3_BRANA	extr	340.7	7%	[Uniprot] SWISS-PROT45:Secreted.
RBS2_LEMGI	chlo	340.8	15%	[Uniprot] SWISS-PROT45:Chloroplast.
FBH1_CUPWR	chlo	345.3	14%	[Uniprot] SWISS-PROT45:Chloroplast.

198

Appendix 7.8

YingOYang results in the three non-redundant chitinase-like amino acid sequences from spruce. The software uses neural networks to predict sites of N-acetylglucosamine glycosylation. When the predicted values are above the two thresholds the site has a higher probability of being glycosylated.

PgGQ0134.B7.1_C02:53

SeqName	Residue		O-GlcNAc result	Potential	Thresh. (1)	Thresh. (2)
Sequence	15	T	+	0.5812	0.5267	0.6604
Sequence	46	S	+	0.5098	0.4686	0.5820
Sequence	145	S	++	0.4697	0.3835	0.4674
Sequence	151	S	+	0.4070	0.3856	0.4701
Sequence	207	S	++	0.4640	0.3719	0.4517
Sequence	211	S	+	0.3897	0.3849	0.4692
Sequence	223	S	+	0.4324	0.3851	0.4695
Sequence	224	S	+++	0.6022	0.3783	0.4603
Sequence	292	S	+	0.4283	0.3983	0.4872
Sequence	299	T	+	0.4972	0.4061	0.4978

PgGQ03602.B7_D17:5

SeqName	Residue		O-GlcNAc result	Potential	Thresh. (1)	Thresh. (2)
Sequence	185	S	++	0.5672	0.3937	0.4811
Sequence	186	S	+	0.5042	0.4265	0.5252

PgGQ0134.B7.1_C02:53

SeqName	Residue		O-GlcNAc result	Potential	Thresh. (1)	Thresh. (2)
Sequence	179	T	++	0.5431	0.4200	0.5165
Sequence	181	S	++	0.4353	0.3546	0.4284
Sequence	185	S	+	0.4374	0.3618	0.4381
Sequence	195	S	+	0.3920	0.3891	0.4749
Sequence	259	S	+	0.3971	0.3344	0.4011

Appendix 7.9

Alignment of the proteins with ordered surface carbons (OSCs). The three non-redundant chitinase-like amino acid sequences from spruce and a rye chitinase were aligned using AlignX from VectorNTI. The sequence upstream of the catalytic domain was trimmed in all sequences. Aminoacid positions correspond to the positions found for the OSCs in Figure 3.10.

```
                                        1                                      40
Pg_GQ0134.B7.1_C02:53_(trimmed)    (1)  IT    LL  R   VCKA GFY YA FIAAADAFPGF
  Pg_GQ03602.B7_D17:5_(trimmed)    (1)  IT D   L  R   C A GFY Y  FI AA  FP  F
 Pg_GQ03602.B7_D17:53_(trimmed)    (1)  IT D  R L  R   C A GFY Y  FI AA  FP  F
            ScAAG53610(trimmed)    (1)  ITQ  ASMLPNRDNSLCPA GFY YD FIAAA TFPGF
                      Consensus    (1)  IITESIFNDFLKHRNDGACAAKGFYTYNAFILAANRFPSF
                                        41                                     80
Pg_GQ0134.B7.1_C02:53_(trimmed)   (41)  G TGD T QKRE AAFF QTSHETTGG P A D   WGY
  Pg_GQ03602.B7_D17:5_(trimmed)   (41)  GN G    KRE AAFF QTSHETTGG   A D   WGY
 Pg_GQ03602.B7_D17:53_(trimmed)   (41)  GN G    KRE AAFF QTSHETTGG   A D   WGY
            ScAAG53610(trimmed)   (41)  G TGST DVKRE AAFF QTSHETTGGTRGAAD-Q QWGY
                      Consensus   (41)  GTVGNLETRKRELAAFFAQTSHETTGGWETAPDGPYAWGY
                                        81                                     120
Pg_GQ0134.B7.1_C02:53_(trimmed)   (81)  CFKEE GNP GEYCQASSEWPCASGKRYYGRGP QL WNY
  Pg_GQ03602.B7_D17:5_(trimmed)   (81)  CFKEE  --------------  YYGRGP QL   N
 Pg_GQ03602.B7_D17:53_(trimmed)   (81)  CFKEE  --------------  YYGRGP QL
            ScAAG53610(trimmed)   (80)  CFKEEINKAT-------------SPPYYGRGP QL
                      Consensus   (81)  CFKEEQDADP                EEMYYGRGPIQLTGKS
                                        121                                    160
Pg_GQ0134.B7.1_C02:53_(trimmed)  (121)  NYGPAG A G DG NP V NDA SF TA WFWM  Q
  Pg_GQ03602.B7_D17:5_(trimmed)  (107)  NY AG A G D  NPN V    SF TA WFWMNPQP
 Pg_GQ03602.B7_D17:53_(trimmed)  (107)  NY AG A G D  NPG V    SF TA WFWM TQ
            ScAAG53610(trimmed)  (106)  NY LAG A GKD SNP V DAV SF TA WFWM  QG
                      Consensus  (121)  NYEAAGKALGYDLINNPDIVASNPTISFKTAVWFWMTAQS
                                        161                                    200
Pg_GQ0134.B7.1_C02:53_(trimmed)  (161)  KPSCHNV A GWGP S D AAGRAAGYGV TNIINGG E
  Pg_GQ03602.B7_D17:5_(trimmed)  (147)  KPS HIV    W P  D AAGR GYGV TNIINGG E
 Pg_GQ03602.B7_D17:53_(trimmed)  (147)  KPSPHTV   W P  D AAGR GYGV TNIINGG E
            ScAAG53610(trimmed)  (146)  NKPS HDVA R W P  DNAAGR GYGV TNIINGG E
                      Consensus  (161)  PKPSSH VMIGKWSPSGSDTAAGRVPGYGVVTNIINGGLE
                                        201                                    240
Pg_GQ0134.B7.1_C02:53_(trimmed)  (201)  CG G  R DRIG Y YC LGV G N DC Q PF
  Pg_GQ03602.B7_D17:5_(trimmed)  (187)  CG G    DRIG Y YC LG  G N DC Q F
 Pg_GQ03602.B7_D17:53_(trimmed)  (187)  CG G    DRIG Y AYC LG  G N DC Q F
            ScAAG53610(trimmed)  (186)  CGMGRNDA VDRIG YT YCG LGT TGGN DCY Q NF
                      Consensus  (201)  CGRGSDSNQEDRIGFYKRYCDILGASYGSNLDCNSQRSFG
                                        241           254
Pg_GQ0134.B7.1_C02:53_(trimmed)  (241)  FAAQSQPRLIKTVV
  Pg_GQ03602.B7_D17:5_(trimmed)  (227)  ------------
 Pg_GQ03602.B7_D17:53_(trimmed)  (227)  ------------
            ScAAG53610(trimmed)  (226)  S------------
                      Consensus  (241)  G
```