Molecular analysis of lodgepole and jack pine seedlings response to inoculation by mountain pine beetle fungal associate *Grosmannia clavigera* under well watered and water deficit conditions

By

Elizabeth L. Mahon

A thesis submitted in partial fulfillment of the requirements for the degree of

Masters of Science

in

Plant Biology

Department of Biological Sciences
University of Alberta

Abstract

To date mountain pine beetle (MPB) has affected more than 19 million ha. of pine forests in Canada. The primary species affected by the current outbreak has been lodgepole pine (*Pinus contorta*), however as MPB range expands eastward beyond its historical habitat, the bark beetle has encountered a novel host: jack pine (*Pinus banksiana*). Ecological evidence has indicated that host trees originating from MPB's historic range have lower host quality compared to hosts from novel habitats, suggesting that co-evolved lodgepole pine may have acquired induced and constitutive defenses against MPB that are not present in jack pine. Ecological evidence has also suggested that trees subjected to abiotic stresses such as drought are more susceptible to MPB attack. MPB vectors a number of microbial symbionts, including *Grosmannia clavigera*, a fungal pathogen that contributes to tree mortality by growing into the host's xylem tissue and disrupting water transport.

Using data generated from a large scale microarray study, we examined the transcriptomic response of both lodgepole and jack pine seedlings inoculated with MPB fungal associate *G. clavigera.* In both species, activation of defense response pathways occurred through signaling action of jasmonic acid and ethylene. We identified qualitative differences between the secondary metabolite biosynthesis genes induced by lodgepole and jack pine in response to *G. clavigera,* with lodgepole pine seedlings exhibiting induction of more genes involved in flavonoid biosynthesis, and jack pine seedlings exhibiting induction of more genes involved in isoprenoid biosynthesis. In seedlings inoculated under water deficit conditions, we observed attenuation of inducible defense related genes. We also observed increased expression of some defense related genes in response to *G. clavigera* under water deficit relative to under well watered conditions. In

lodgepole pine these changes occurred to a greater degree than jack pine, suggesting water deficit impacts lodgepole pine defense response to a greater extent than jack pine.

Within the microarray data set, chitinase genes were amongst the most highly upregulated known defense associated genes. Chitinase genes are a family of well studied pathogenesis response proteins, some of which hydrolyze chitin, an important component of fungal cell walls. We conducted a phylogenetic analysis of the chitinase gene families of lodgepole and jack pine. We observed that expression patterns of the pine chitinases reflected phylogenetic relationships. We examined allelic variation of three putative orthologs pairs of chitinase genes from lodgepole and jack pine sampled from across Canada, and we identified several non synonymous substitutions. Some of these substitutions displayed spatially explicit patterns across the ranges of lodgepole and jack pine which may reflect adaptive variation.

Preface

Chapter two of this thesis decribes analysis of microarray data generated by Miranda Meents and Adriana Arango, using plant material from an experiment conducted by Adriana Arango, Miranda Meents, Charles Copeland and others from the Cooke Lab. RNA for the microarrays was extracted by Adriana Arango, Miranda Meents, Justin Khunkhun, William Peachman and Bulcha Dolal. Data extractions were carried out by Blaire Johnson and Dominik Royko. Statistical analyses of these data were carried out by Adriana Arango, with assistance from Walid El Kayal and Chelsea Ju. Annotation of the microarray with lodgepole and jack pine expressed gene data was carried out by Dominik Royko. Mapman annotations for the microarray annotations were provided by Jill Hamilton and Bjorn Usadel. The microarrays used for this experiment were provided by Dr. Jeff Dean and Walt Lorenz. I was responsible for the mining and analyses of this data which is described in chapter two of this thesis.

Chapter three of this thesis describes analysis of members of the chitinase gene family. Illumina transcriptome assemblies for lodgepole and jack pine were provided by Mack Yuen and Dr. Joerg Bohlmann.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Janice Cooke for providing me with the opportunity to carry out this research in her lab. I would like to thank her for her unwavering support, and for the considerable commitment of time and energy she has made to me through out my time as an undergraduate and graduate student in her lab. Her guidance and support was central to this research, and the experience I gained as a researcher under her intensive supervision has been invaluable.

I would like to thank the members of my advisory committee Dr. Jocelyn Hall and Dr. David Coltman, for their guidance and suggestions. I would like to thank Cheryl Nargang for her technical support with the large scale Sanger sequencing conducted as part of this research. Thank you to the past members of the Cooke lab who contributed to production of the microarray data analyzed in this research, and thank you to Dr. Catherine Cullingham and Dr. Chandra McAllister for their advice and mentorship throughout my degree. Finally, I would like to thank the University of Alberta and the Department of Biological Sciences for the financial support provided during my degree.

Table of Contents

LIST OF TABLES

List of Figures

List of Abreviations

ABF	abscisic acid response factor
ACC	aminocyclopropane carboxylic acid
AOS	allene oxidase synthase
bp	base pairs
CBD	chitin binding domaine
CHS	chalcone synthase
CTAB	hexadecyl-trimethyl ammonium bromide
DE	differentially expressed
DIR	dirigent like protiens
dpi	days post inoculation
EIL	ethylene insensitive-like
ERF	ethylene response factor
FC	fold change
GH	gylcosal hydrolase
HR	hypersensitive response
JA	jasmonic acid
JA-Ile	isoleucine
LOX	lipoxygenase
MAPK	mitogen activated protein kinase
MBS	MYB transcripton factor binding site
MeJA	methyl-jasmonate
MPB	mountain pine beetle

NB-LRR	nucleotide-binding/leucine-rich-repeat
NTSC	National Tree Seed Centre
ODPA	oxo-phytodeinoate
PAL	phenylalanine ammonia lyase
PAMP	pathogen associated molecular patterns
PCR	polymerase chain reaction
PP	polyphenolic parenchyma
PR	pathogenesis response
PRR	pattern recognition receptors
PTI	pattern triggered immunity
qRT-PCR	quantitative reverse transcription PCR
ROS	reactive oxygen species
SA	salicylic acid
SNV	single nucleotide variation
SP	signal peptide
SUE	SEUSS transcriptional co-regulator
TF	transcription factor
WD	water deficit
WW	well watered

1.0 Chapter 1: Introduction and Background

1.1 Pine forests in Canada

Canadian forests represent 10% of the world's overall forest cover, and these forests are composed primarily of conifer species. *Pinus* (Pinaceae, Coniferophyta) is one of the most widely distributed genera of trees found in North America (Critchfield and Little 1966). Pine species are characterized by long needle like leaves, typically found in bundles of two, three or five needles, but bundles can contain up to eight needles (Richardson 2000). Lodgepole pine, (*Pinus contorta* Dougl.), is a two needled pine found throughout British Columbia, extending north to the Yukon and Northwest territories, east across the Rocky Mountains into Alberta and south into the United States (Fig 1.1) (https://www.na.fs.fed.us /spfo/ pubs/silvics_manual/Volume_1/ accessed July 2, 2016). Lodgepole pine is considered mesophytic and is able to grow in a broad range of soil types, including bogs and clay soils (Carlson *et al.* 1999). Two varieties of lodgepole pine are found in Canada; *Pinus contorta* Dougl. ex. Loud. var. *latifolia,* which is commonly called lodgepole pine, and *Pinus contorta* Dougl. ex Loud. var. *contorta,* or shore pine which is only found is along the coast of British-Columbia, and is described as shorter and shrubbier than lodgepole pine (Richardson 2000). Here we examine one variety of lodgepole pine, *Pinus contorta* Dougl. ex. Loud. var. *latifolia.* Jack pine (*Pinus banksiana* Lamb.) is a two needled pine whose range extends east from the Northwest Territories and central Alberta to New Brunswick (Fig1.1) (https://www.na.fs.fed.us/ spfo/pubs /silvics_manual/Volume_1). Jack pine is often found growing in well drained, nutrient poor soils (Kenkel *et al.* 1997). The ranges of lodgepole and jack pine meet to form a hybrid zone in north central Alberta (Cullingham *et al.* 2012). Both species of pine are economically and

ecologically important to Canada. Pine species make up a large proportion of Canada's pulp and softwood lumber exports, and pine forests support wildlife, providing essential habitat for a diversity of native species (Natural Resources Canada, Canadian Forest Service 2015; Shore *et al.* 2006).

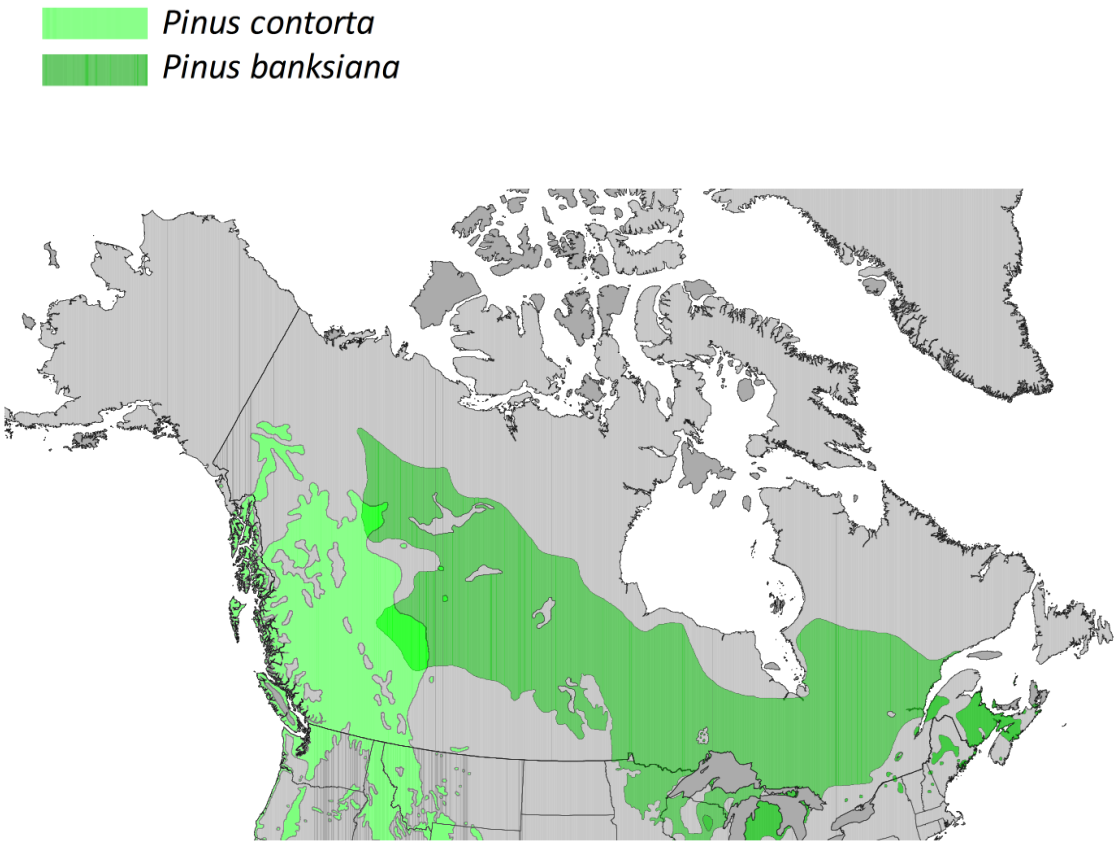


Figure 1.1: Approximate species ranges in Canada for lodgepole pine (*Pinus contorta)* shown in light green and jack pine (*Pinus banksiana*) shown in dark green. Range distributions were obtained from USGS (http://gec.cr.usgs.gov/data/little/ accessed July 29, 2016) and are based on Little (1971).

1.2 Mountain pine beetle outbreak

Mountain pine beetle (MPB; *Dendroctonus ponderosae* Hopkins) is an eruptive forest insect pest, meaning that populations alternating between endemic and epidemic phases.

Population densities during the endemic phase are typically very low, while population densities during epidemic (outbreak) phases can be very high (Raffa et al. 2008). The current MPB outbreak is estimated to have affected ca. 19 million ha of western Canadian forests, resulting in widespread tree mortality (Government of British Columbia 2012; Government of Alberta 2012). In Canada, MPB populations have been historically found in south-central British Columbia, primarily attacking lodgepole pine. The current MPB outbreak began in the late 1990s, leading to hyperepidemic levels and expansion of populations farther to the north in British-Columbia than previously reported (Safranyik & Carroll 2006). In 2006, MPB crossed the Rocky Mountains and was reported in the north central areas of Alberta (Safranyik & Carroll 2010). As a result of range expansion into northern Alberta, the beetle encountered the lodgepole x jack pine hybrid stands and pure jack pine stands of Alberta. In 2010, MPB was conclusively shown to have undergone a tree-host species shift from lodgepole pine into pure jack pine of Alberta (Cullingham *et al*. 2011).

Ecological studies have shown that host trees from MPB's historic range have lower host quality, resulting in lower beetle reproductive success, compared to hosts from novel habitats (Cudmore *et al.* 2010). Whitebark pine (*Pinus albicaulis* Engelm.) is found at higher elevations compared to lodgepole pine, and is considered a susceptible host. Studies show that naïve whitebark pine displayed a significantly lower investment in induced defenses than co-evolved lodgepole pine when attacked by MPB (Raffa *et al.* 2013). MPB has a higher rate of reproductive success in lodegpole pine forests found outside of MPB's historical range (Cudmore *et al.* 2010). Taken together, this evidence suggests that MPB is likely to encounter less well defended hosts as it expands into naïve tree host populations, and raises significant concerns about MPB's potential for expansion eastwards across the boreal forests of Canada. Eastward expansion of MPB through jack pine forests would likely result in timber loss, large scale socioeconomic

impacts on logging dependent communities and significant loss of wildlife habitat (Nealis & Peter 2008).

1.2.1 Life cycle of mountain pine beetle

MPB is a bark beetle native to the western forests of Canada. The beetle spends most of its life cycle (one to two years) under the bark of a host tree, emerging mid July to early September and dispersing to new tree hosts (Safranyik & Carroll 2006). MPB relies on a mass attack strategy to overcome a tree host's defenses and gain access to the nutrient-rich phloem tissue. MPB relies on chemical cues emitted from tree hosts, as well as the biosynthesis and perception of its own pheromones, in order to select suitable hosts and recruit other MPB to aid in mass attack (Hunum *et al.* 1980; Wood 1982). MPB can use defense compounds produced by the tree hosts to produce aggregation pheromones, such as in the case of α-pinene serving as a precursor for the aggregation pheromone trans-verbanol (Pierce 1987).

Once a tree host's defenses have been exhausted, MPB can then mine galleries beneath the host's bark and begin reproductive activity (Bentz *et al.* 2012; Bonnet *et al.* 2012). An important factor mediating MPB population growth is the rate of larval mortality which occurs during the winter months (Bentz *et al.* 1999). Once hatched, MPB larvae typically spend one year under the bark of a tree host, and produce cryoprotectants such as glycerol to survive winter temperatures (Bale *et al.* 2002). However, MPB is considered freeze susceptible, and extended periods of cold temperatures during the winter months can play a major role in reducing MPB populations which emerge in the following spring and summer (Bentz *et al.* 1999). MPB does not undergo diapause, an obligatory period of suspended development during winter months which allows insects to synchronize their life stages, but instead MPB is dependent on temperature cues to appropriately time continuation of development, as well as time the rate of

development (Logan & Bentz 1999; Bentz *et al.* 2001). This adaptation allows the beetle to synchronize its phenology, or life stages, over a broad range of climatic conditions, and is a factor that is considered to facilitate movement into novel habitat (Safranyik & Carrol 2007).

Host tree selection is an important factor affecting the beetle's reproductive success, which impacts their population growth, and the shift from endemic to epidemic status (Raffa et al. 2008). At lower densities, MPB will target weaker, physiologically stressed tree hosts, however at higher densities, MPB target will target healthier more vigorous trees, and gain access to phloem with greater nutritional quality (Boone *et al.* 2011). Populations are considered endemic when unable to reach the threshold density necessary to successfully colonize a large, vigorously growing tree (Safranyik & Carrol 2006). Epidemic populations have reached this threshold density and can successfully attack large, healthy trees (Safranyik & Carroll 2006). Rising winter temperatures have contributed to the frequency with which beetle populations have reached epidemic levels, by reducing larval overwinter mortality rates, and promoting completion of the beetle's life cycle in one year rather than in two (Safranyik & Carroll 2006). At epidemic populations, MPB is able to access healthier more nutritious trees resulting in higher reproductive success, and feeding forward into even greater population growth (Boone *et al.* 2011).

1.2.2 MPB fungal associate *Grosmannia clavigera*

MPB is associated with a number of different fungal species belonging to the order Ophiostomales within the phylum Ascomycota, including *Grosmannia clavigera* (Robinson-Jeffrey & R.W. Davidson) Zipfel. de Beer. & Wingf.*, Ophiostoma montium (*Rumbold) von Arx.*, Leptographium longiclavatum* (Lee, Kim and Breuil), and *Ceratocystiopsis* sp.1, and the basidiomycete *Entomocorticium* sp (Lee *et al.* 2006; Khadempour *et al.* 2012). *Grosmannia clavigera, L. longiclavatum* and *O. montium* are all considered fast growing, colonizing galleries

and quickly growing into phloem and xylem tissue causing blue grey discoloration of host tissue (Lee *et al.* 2006). *Ceratocystiopsis* sp.1 and the basidiomycete *Entomocorticium* sp. are slower growing and do not cause tissue discoloration, but may serve as an important nutritional source for the beetle (Khadempour *et al.* 2012). These fungi are thought to facilitate bark beetle attack by exhausting tree host defenses and providing nutrition for the beetle and emerging larvae (Safranyik & Carroll 2006; Six & Paine 1998). New MPB adults remain for a short period under the bark to feed on fungi which line the pupal chamber, collecting spores on their exoskeleton and mycangia before emerging and dispersing to a new tree host (Harrington 1993; Safranyik & Carroll 2006). MPB ophiostomatoid fungal associates produce sticky spores on long necked sexual and asexual fruiting bodies, making them well adapted to dispersal by beetles (Harrington 1993; Six & Klepzig 2004).

Grosmannia clavigera is considered to be the most pathogenic of MPB fungal associates, contributing to eventual tree host mortality by growing into host sapwood tissue and blocking transport of water and minerals through occlusion of ray parenchyma and tracheids (Solheim & Krokene 1998; Lee *et al.* 2006; Ballard *et al.* 1984). *Grosmannia clavigera is* also thought to aid in MPB attack by detoxifying terpenoid compounds produced by tree hosts (DiGuistini *et al.* 2011). *Grosmannia clavigera,* and its relative *Ceratocystis polonica* (Siemaszko) C. Moreau*,* associated with European spruce bark beetle, *Ips typographus* L.*,* are thought to be necrotrophic pathogens on the basis of tree host responses (Arango-Velez *et al.* 2016; Fossdal *et al.* 2012). When inoculated with *G. clavigera,* pine trees develop a darkened lesion which spreads vertically along the sapwood from the point of infection, and after time radially into the sapwood. This lesion is composed of cells which contain large quantities of defensive chemicals (Francheschi *et al.* 2005). It has been suggested that lesion length represents either the magnitude of defense response on the part of the tree, or the extent of the fungal invasion (Arango-Velez *et al.* 2016;

Lu *et al.* 2010). Water deficit condtions result in a reduction in lesion length compared with well-watered controls, which may suggest that lesion length is a measure of tree host defense response or that water limitation reduces fungal growth (Arango-Velez *et al.* 2016). PCR detection of *G. clavigera* identified along the length of the lesion as well which supports the latter explanation of lesion as a measure of fungal invasion, suggesting that the two explanations may not be mutually exclusive (C. McAllister & J.E.K. Cooke, in preparation). Regardless, differences in lesion length can be a useful indicator of differences in defense response between tree hosts species or between tree hosts under different abiotic stress. Understanding how naïve and co-evolved pine hosts interact with MPB fungal associates such as *G. clavigera,* is a critical aspect of understanding and predicting success of MPB attack in novel territories.

1.3 Components of conifer defense response

1.3.1 Constitutive defenses conifers

Conifers rely on an array of constitutive and induced defenses to defense against invading pests and pathogens. Constitutive defenses are pre-formed defenses which are present in a tree prior to challenge by pest or pathogen. MPB targets the nutrient-rich phloem and cambial layer found just beneath the bark of a tree stem (Franceschi *et al.* 2005). In order to protect the xylem and phloem tissue of the trunk, which form the major water and nutrient transport pathways respectively, conifer species allocate a considerable proportion of their resources to constitutive defenses surrounding these tissues (Franceschi *et al.* 2005). These generalized defenses include multiple layers of mechanical and chemical defenses which protect the tree from outside invaders. These protective layers begin with the periderm which forms a tough outer layer making it difficult for invaders to penetrate a tree host. This layer is followed, in younger trees,

by the cortex whose cells can contain large amounts of toxic phenolic compounds in their vacuoles (Franceschi *et al.* 2005). The cortex tissue in young pine can also contain axial resin ducts which transport toxic oleoresin to a site of attack. Beneath the layer of cortex tissue is secondary phloem tissue which grows outwards from the cambial meristem and is the site of multiple constitutive defenses (Franceschi *et al.* 2005). There are three principle constitutive defenses are found in the secondary phloem of most conifers; polyphenolic parenchyma cells, sclerenchyma cells and calcium oxalate crystals (Hudgins & Franceschi 2004). Polyphenolic parenchyma cells often occur tangential to one another, forming a ring within the secondary phloem tissue, and contain varying amounts of phenolic compounds which are thought to function as antifungal and anti-feedant agents (Krekling *et al.* 2000; Beckman 2000). In pine, sclerenchyma tissue often occurs as sclereids (stone cells) which are highly lignified and irregularly shaped cells that can from mechanical obstacles against bark boring insects (Hudgins *et al.* 2004). Calcium oxalate crystals occur as intracellular deposits in all pine species and considering their relative abundance in secondary phloem tissue and their physical toughness, they are thought to act as a deterrent against chewing animals and bark boring insects (Hudgins *et al.* 2003). All three of these layered constitutive defenses form a protective physical barrier surrounding the generative cambial tissue, which if reached by an invader can be easily damaged and result in the death of part or all of the tree (Franceschi *et al.* 2005).

Another critical constitutive defense in conifers is the constitutive presence of toxic oleoresin, composed of terpenoid compounds. Radial resin ducts arise from radial rays which extend outwards from the center of a tree. Axial resin ducts are similar in structure, but extend vertically along the stem. In pine species, resin producing structures are constitutively present in both xylem and phloem tissues (Hudgins *et al.* 2004). Resin ducts accumulate terpenoid resin under pressure as the epithelial cells which surround the duct synthesize and secrete terpenoid

resins into the duct (Franceschi *et al.* 2005). The pressurized resin is quickly released upon wounding or damage by an invading pest, and functions to repel invaders by either trapping the invading organism in sticky resin, killing the invader due to exposure to the toxic resin compounds or by flushing ("pitching") the invading organism out (Franceschi *et al.* 2005). Phenolic and terpenoid defense are discussed in more detail in the later sections describing inducible defenses.

1.3.2 Inducible defenses in conifers

Conifers induce an array of anatomical, chemical and molecular defenses in response to pathogen challenge. Induced anatomical defenses include the formation of wound periderm, a reinforced tissue composed of suberized phellem cells, phellogen (cork cambium) and parenchyma-like phelloderm, which form around invaded or wounded tissue in an effort to isolate the damage, and are often elicited in conifers in response to bark beetle or fungal infection. Another important induced anatomical defense is the formation of traumatic resin ducts, which are not constitutively present in tissues and which form after wounding above and below the site of injury in order to transport toxic terpenoid compounds to the point of invasion. In pine, axial traumatic resin ducts can form in the xylem tissue which can be interconnected with radial resin ducts found in the phloem (Nagy *et al.* 2004). The hypersensitive response (HR) is also considered to be an induced anatomical defense. In an HR, tissue at the point of infection undergoes rapid cell death through the production of reactive oxygen species, which serves to kill or contain an invading fungal organism (Franceschi *et al.* 2005). The lesion which develops in pine in response to inoculation by *G. clavigera* may in part indicate HR at the point of inoculation.

Chemical defenses in the form of phenolic and terpenoid compounds can be induced in response to attack through activation of biosynthetic pathways. Phenolics act as antifungal and anti-feedant agents, and are primarily produced within polyphenolic parenchyma (PP) cells, which expand and accumulate phenolic compounds in response to wounding or pathogen challenge (Franceschi *et al.* 2005). Polyphenolic parenchyma cells in lodgepole x jack pine hybrid seedlings were shown to swell in response to inoculation with *G. clavigera,* and bands of PP cells appeared in the xylem tissue of both jack and lodgepole pine seedlings, suggesting that accumulation of phenolic compounds plays a role in pine defense response (Arango-Velez *et al.* 2014; Arango-Velez *et al.* 2016). As previously discussed, oleoresin is constitutively present in pine, but can also be induced to accumulate in greater quantities in response to pathogen challenge (Keeling & Bohlmann 2006). Oleoresin production is increased through the formation of traumatic resin ducts, and activation of constitutive resin ducts (Eyles *et al.* 2010). In lodgepole x jack pine hybrids, expression of some terpene synthase genes was induced in response to inoculation with *G. clavigera* (Arango-Velez *et al.* 2014). Evidence suggests that the relative composition of terpenoid compounds produced by a tree may contribute to the perceived suitability of a host by MPB (Raffa *et al.* 2013). Profiling of defensive metabolites revealed differences in both qualitative and quantitative composition of terpenoid compounds produced in jack and lodgepole pine in response to *G. clavigera* (Arango-Velez *et al.* 2016). Jack pine was shown to contain higher amounts of – (α)-pinene, which is a precursor molecule to the female MPB aggregation pheromone (-) trans-verbenol (Arango-Velez *et al.* 2016). Lodgepole pine has been shown to release greater amounts of β-phellandrene compared to jack pine, which is an MPB attractant (Arango-Velez *et al.* 2016; Lusibrink *et al.* 2011). Jack pine was also reported to release greater levels of 3-carene, a montoperne known to play a role, along with other

monoterpenes in the formation of MPB aggregation pheromones (Lusibrink *et al.* 2011; Borden *et al.* 2008). Little is known concerning the regulatory pathways which activate the production of different terpenoid compounds between jack and lodgepole pine. From studies in pine and other conifers, we know upregulation of terpene synthase genes is positively correlated with terpenoid accumulation, and terpene synthase expression is induced by wounding, insect attack, pathogen challenge and treatment with the plant defense hormone methyl-jasmonic acid (MeJA) (Zulak *et al.* 2009; Keeling & Bohlmann 2006b; Arango-Velez *et al.* 2014).

A large number of antimicrobial proteins are induced by pine tree hosts in response to pathogen challenge. Many antimicrobial proteins are considered to be pathogenesis response (PR) proteins. PR proteins are typically undetectable in healthy tissues, but are rapidly induced both locally and systematically in response to pathogen challenge (Van Loon *et al.* 2006). PR proteins are grouped into 17 different families. Chitinases are members of PR families 3 and 8. Some members of the chitinase family break down chitin, a component in fungal cell walls (Neuhaus 1999). Peroxidases, members of PR family 9, have been shown to accumulate at high levels in pathogen infected Norway spruce, *Picea abies* Karst.*,* and are thought to aid in strengthening cell walls (Fossdal *et al.* 2001; Nagy *et al.* 2004). Osmotins, or thaumatin-like proteins, are important PR-5 proteins which inhibit hyphal growth by permeabilizing and degrading fungal cell walls of fungal invaders (Abad 1996, Osmond 2001). Expression of thaumatin-like genes is strongly induced in western white pine and Scots pine in response to fungal inoculation suggesting they play an important antimicrobial role in defense response (Piggott *et al.* 2004; Adomas *et al.* 2007). Although not considered PR proteins, dirigent-like proteins in conifers are highly up regulated in response to pathogen challenge and are thought to direct the stereospecific coupling of monolignols, which may act as precursors to phenolic defensive compounds (Ralph *et al.* 2006b; Kim *et al.* 2002). Expression of these well-known

defense associated proteins can be used as a marker of an active defense response in plants. Timing and magnitude of PR protein expression can serve as an indicator of the timing and magnitude of the larger defense response.

1.3.3 Pathogen recognition and activation of defense response

Plants and pathogens exist in a fluctuating co-evolutionary cycle of detection and evasion (Jones & Dangl 2006). Plants initially detect and respond to both necrotrophic and biotrophic pathogens by identifying specific molecular patterns characteristic of foreign invaders termed pathogen-associated molecular patterns (PAMPs) (Glazebrook 2005). Plants use surface pattern recognition receptors (PRRs) to recognize PAMP's, and once activated a PRR can elicit a generalized defenses response called pattern-triggered immunity (PTI) (Zipfel 2014). However, pathogens are able to adapt to particular host genotypes and release virulence factors, termed effectors, which can reduce basal defenses by interfering with PTI (Karasov *et al.* 2014; Raffaele *et al.* 2010; Dangl *et al.* 2013; Deslandes *et al.* 2012). Plants have, in turn, evolved to recognize pathogen released effector molecules through family of polymorphic intracellular nucleotide-binding/leucine-rich-repeat (NB-LRR) receptors which, once activated, can elicit effector-triggered immunity (ETI), a more robust and amplified defense response in comparison to PTI (Cui *et al.* 2015). Once activated, NB-LRR's invoke ETI by eliciting a wide array of defense responses including transcriptional reprogramming, production of reactive oxygen species (ROS), and mitogen activated protein kinase (MAPK) cascades, which ultimately transduce the signal to transcription factors, leading to altered transcriptional regulation (Cui *et al.* 2015). The defense responses induced during ETI are similar to those induced during PTI, however the responses are often amplified or of longer duration (Cui *et al.* 2015). A recently proposed model suggests that activated NB-LRRs function to enhance the defense response by reducing negative

constraints on PTI, allowing defense responses to be amplified (Cui *et al.* 2015). A large number of NB-LRR transcripts have been identified in western white pine, *Pinus monticola,* and variation in their sequences has been linked to resistance to white pine blister rust, *Cronartium ribicola* Fisch. (Liu & Ekramoddoullah 2004, 2007). It has been proposed that extensive gene expansion of the NB-LRR gene family in conifers and other long lived woody perennials such as poplar and grapevine is an adaptive mechanism which compensates for exposure to many different pathogens over extended generation times (Yang *et al.* 2008).

Expression analysis of NB-LRRs in Norway spruce has revealed that transcript abundance corresponding to these genes is maintained at low but relatively constant levels following pathogen challenge, suggesting NB-LRR genes play a role in early pathogen detection that is not related to changes in transcript abundance corresponding to these proteins (Fossdal *et al.* 2012). However, when comparing the defense response of co-evolved and naïve pine hosts to *G. clavigera,* it is important to note that an initial difference in detection and signaling via NB-LRR proteins could potentially contribute to the downstream differences observed between the responses of pine tree hosts who share a co-evolutionary relationship with MPB and *G. clavigera* and those who do not.

1.3.4 Defense hormones involved in eliciting plant defense response

Three plant hormones commonly associated with induction of defense associated gene expression are jasmonic acid (JA), ethylene and salicylic acid (SA) (Pieterse *et al.* 2009). In conifers, exogenous application of methyl-jasmontate (MeJA) induces defense responses similar to those induced by wounding and pathogen invasion (Miller *et al.* 2005; Zulak *et al.* 2009). Anatomical changes such as expansion of polyphenolic cells and formation of traumatic resin ducts are induced in conifers by exogenous MeJA treatment, along with accumulation of

terpenoid compounds and increase the expression of terpene synthases genes (Franceschi et al. 2002; Martin *et al.* 2002; Hudgins & Franceschi 2004). Microarray analysis of Sitka spruce (*Picea sitchensis* Bong.) responding to white pine weevil (*Pissodes strobi* Peck.) feeding demonstrated that genes putatively involved defense were expressed in patterns similar to those observed after MeJA treatment (Ralph *et al.* 2006a).

In slash pine, (*Pinus elliottii* Engelm.) and western white pine (*Pinus monticola* Dougl. Ex D. Don) xogenous application of MeJA led to increased expression of chitinases and PR-10 family gene members (Davis *et al.* 2002; Liu *et al.* 2003). Recently, increased levels of *in vivo* JA and JA-isoleucine (JA-Ile) – the active form of JA – were reported in response to inoculation with *G. clavigera (*Arango-Velez *et al.* 2016). JA signaling is often associated with defense response to necrotrophic pathogens in many plant species, and it has been proposed that *G. clavigera* is a necrotrophic pathogen along with other bark beetle-associated fungi (Arango-Velez *et al.* 2016; Fossdal *et al.* 2012). JA-Ile binds to its receptor, COI1, which constitutes part of the larger the SCF^{COI1} protein complex that targets members of the JAZ protein family for proteasomal degradation (Chini *et al.* 2007; Thines *et al.* 2007). Members of the JAZ protein family act as repressors of jasmonic acid-activated transcription factors such as MYC2 which function as major switches, inducing the expression of a diverse array of jasmonate-dependent responses (Dombrecht *et al.* 2007).

In conifers, production of ethylene has been correlated with biosynthesis of monoterpenes during fungal infections (Popp *et al.* 1995). Both wounding and exogenous application of MeJA increased expression of ethylene biosynthesis genes in Douglas-fir (*Psuedotsuga menzeisii* (Mirb.) Franco) and white spruce (*Picea glauca* (Moench) Voss.) through upregulation of ethylene biosynthesis enzymes, ACC oxidase and ACC synthases (Ralph *et al.* 2007; Hudgins *et al.* 2006). Exogenous application of ethylene induced defense responses similar to those induced

by MeJA treatment, promoting phenolic biosynthesis in polyphenolic parenchyma cells, lignification of sclerieds and formation of traumatic resin ducts (Hudgins & Franceschi 2004). This response suggests that the defenses induced by MeJA are mediated by ethylene, and that these two hormone signaling pathways act in concert to elicit a subset of plant defense responses. Ethylene acts to elicit defense-associated and developmental processes by binding ethylene receptors, a large family of transmembrane histidine kinase proteins found in the endoplasmic reticulum, which act as constitutive repressors of ethylene response pathways (Binder 2008). Once bound by ethylene, these repressors are deactivated resulting in a deactivation of corresponding repressors, CTR RAF-like kinases, which prevents the protein EIN2 from being targeted for proteasomal degradation (Qiao *et al.* 2009). Accumulation of EIN2 allows EIN3 transcription factors to accumulate and induce expression of ERF transcription factors which then induce expression of many different ethylene response genes (Solano *et al.* 1998).

Increased levels of JA are typically associated with response to necrotrophic pathogens, such as *G. clavigera,* which are pathogens that kill tissue and consume the remains (Glazebrook 2005; Tomma *et al.* 1998). Increased levels of SA are typically associated with response to biotrophic and hemibiotrophic pathogens, which are pathogens that consume living tissue (Glazebrook 2005; Tomma *et al.* 1998). The role of SA as an elicitor of conifer defense responses is not as well established. Salicylic acid has been shown to accumulate in Norway spruce seedlings responding to pathogen challenge by soil borne pathogen, *Pythium irregulae* Buisman., as well as when seedlings were treated with MeJA (Kozlowski & Metraux 1998; Kozlowski et al. 1999). SA did not elicit the same anatomical defenses in Douglas-fir as was seen with exogenous MeJA treatment (Hudgins & Franceschi 2004). In slash pine seedlings, application of SA induced expression of some chitinases, suggesting that SA does play a role in eliciting defense response pathways (Davis *et al.* 2002). In mature lodgepole and jack pine, *G. clavigera* inoculation also

impacted SA levels (Arango-Velez *et al.* 2016). In lodgepole pine seedlings inoculated with *G. clavigera*, there were no significant changes in SA acid levels, however in jack pine there was a significant increase in seedlings inoculated with *G. clavigera* under water deficit conditions and the authors suggests that SA does plays a greater a role in jack pine defense response to *G. clavigera* (Arango-Velez *et al.* 2016).

1.4 Effect of drought on pine defense response

Over the past two decades, Alberta has experienced extended periods of severe drought (Chhin *et al.* 2008). Current climate models project that increasing temperatures will be accompanied by reduced precipitation and the resulting drought stress experienced by Canada's forests (Seager 2007; IPPC 2013). This extended period of drought has had measurably negative impact on Alberta's forests, contributing to large scale aspen die back and is predicted to contribute to wide spread conifer mortality (Michaelian *et al.* 2011; Hogg & Michealian 2015; Adams *et al.* 2009). Ecological studies in piñon pine, *Pinus edulis* Engelm.*,* have demonstrated that water deficit leading to carbon starvation increases pine host susceptibility to other biotic stressors such as bark beetles (Breshears *et al.* 2008). The index of water deficit is an important climate variable used to model spread risk of MPB into new habitats, as it is correlated with increases in the rate of larval survival as well as rate of fungal colonization (Safranyik *et al.* 2010). Lodgepole x jack pine hybrid trees under well watered conditions and inoculated with *G. clavigera* developed significantly longer lesions at the point of inoculation than their water stressed counterparts, demonstrating that water deficit influences tree host defense responses, and suggesting that pine hosts under well watered conditions are able to mount a more vigorous defense response (Arango-Velez *et al.* 2014). Under drought stress, plants close stomata in order to prevent water loss via transpiration and become carbon limited due to a concomitant reduction

in photosynthesis. This is thought to cause a reduction in the allocation of carbon resources to defense, such as production of terpenoid and phenolic compounds, leading to increased susceptibility to attack by pests and pathogens (McDowell *et al.* 2008). Biochemical studies have demonstrated that jack and lodgepole pine seedlings inoculated with *G. clavigera* while under water deficit initially emitted more total monoterpenes than those inoculated under well watered conditions, however over time both species released significantly fewer total monoterpenes while under water deficit (Lusebrink *et al.* 2011). Molecular studies have shown that water deficit conditions altered the transcript abundance of several biotic stress associated genes in lodgepole x jack pine hybrids inoculated with *G. clavigera*. When inoculated under water deficit conditions, a subset of putative chitinases and terpene synthases genes decreased expression levels, while a separate subset of putative chitinase and terpene synthases displayed increased expression levels (Arango-Velez *et al.* 2014). The authors suggest that these patterns of co-expression between subsets of antimicrobial proteins and the enzymes involved in the biosynthesis of carbon rich chemical defenses, demonstrates that both carbon and nitrogen based defenses are impacted under water deficit (Arango-Velez *et al.* 2014). Lodgepole and jack pine exhibit species-specific differences in their defense responses to *G. clavigera* while under water deficit conditions. The increase in lodgepole pine total monoterpene levels resulting from G. clavigera inoculation was less under water deficit conditions than under well watered conditions, whereas in jack pine no significant differences in total monoterpene levels were observed between treatments (Arango-Velez *et al.* 2016). Water deficit was also shown to significantly alter the qualitative profile of monoterpenes produced by lodgepole pine, but had largely non-significant effects on monoterpene profiles in jack pine (Arango-Velez *et al.* 2016).

1.5 Chitinases

Chitinases are a well-studied family of PR proteins. The expression of many members of this gene family is highly induced in response to pathogen challenge. Some members of the chitinase family act as glycosyl hydrolases, catalyzing the hydrolytic cleavage of of *β-1-4* linked N-acetyl glucosamine units of chitin (Valuthakkal *et al.* 2012).

1.5.1 Biochemical classification of the chitinase gene family

Chitinases are traditionally grouped into seven classes based on the presence or absence of conserved amino acid motifs including a chitin binding domain, signal peptide domain, hinge domain and conserved loops found within the catalytic domain (Neuhaus 1999) (Figure 1.2). Class I, II, IV and VII chitinases are considered members of the glycosyl hydrolase family 19 (GH 19), a family of GHs found exclusively in plants, and are typically reported as being involved in defense response (Figure 1.2). Class III and class V chitinases are considered members of the GH 18 family, a broader family of GHs found in both bacteria and plants (Neuhaus 1999) (Figure 1.2).

Class I Chitinases

Class I chitinases are characterized by the presence of an N-terminal signal peptide sequence that targets them to the secretory pathway, followed by a proline-rich hinge domain, and a highly conserved chitin binding domain (CBD). The class I catalytic domain carries all four distinct loops. Loop 1 is found in the catalytic cleft, suggesting that it provides a sub-site for sugar binding. Loop 2 is held at the base by a sulfide bond and extends away from the side of the catalytic cleft. Loop 3 holds loop 4 in place on the outer surface of the enzyme away from the catalytic cleft. Loop 4 is also termed the C-terminal domain, and is responsible for targeting class I chitinases to the vacuole (Neuhaus 1999).

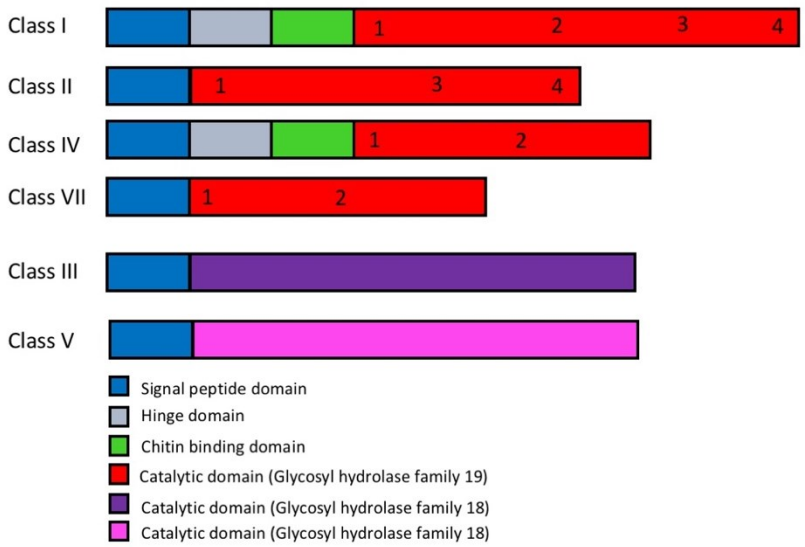


Figure 1.2: Schematic representation of the different classes of chitinases. Signal peptides are indicated in blue, hinge domain (grey), chitin binding domain (green), catalytic domain of glycosyl hydrolase family 18 (red), catalytic domain of glycosyl hydrolase family 18 class III chitinases (purple), and catalytic domain of glycosyl hydrolase family 18 class V chitinases (pink). Diagram adapted from Islam *et al.* 2011.

Class II chitinases

Class II chitinases are highly similar to class I chitinases; however, they lack a CBD and carry a deletion of loop 2 in the catalytic domain. Some class II chitinases are termed class IIa chitinases, or pathogenesis chitinases, and show a greater similarity to class I chitinases. Other class II chitinases are termed class IIb, or non pathogenesis chitinases and show less similarity to class I chitinases. It has been proposed that class IIa and class IIb were derived from class I chitinases following two independent deletion events, suggesting that the CBD domain may have been lost by class II chitinases on at least two separate occasions (Neuhaus 1999).

Class IV chitinases

Class IV chitinases are shorter than class I and II chitinases. They carry a signal peptide, hinge region, as well as a chitin binding domain and a catalytic domain both shortened by loop deletions. The loop deletion found in the CBD does not affect sugar binding properties of the enzyme, and may allow class IV chitinases to hydrolase chitin polymers closer to the surface of the invading pathogen. The catalytic domain carries a deletion of loops 3 and 4 (Neuhaus 1999). Due to the lack of a C-terminal vacuolar targeting signal, class IV chitinases may play an extracellular role in defense (Liu *et al.* 2005).

Class VII chitinases

Class VII chitinases carry no CBD, and a catalytic domain similar to the catalytic domain of class IV chitinases, with deletion of loops 3 and 4. Just as with class IV chitinases, class VII chitinases lack a C-terminal vacuolar targeting signal domain, suggesting they also play an extracellular role (Neuhaus 1999).

Class III and V

Class III and class V chitinases are both termed "bacterial" chitinases, and work exogenously on the non-reducing end of chitin to release chitbiose and chittriose, rather than working endogenously to release chitin polymers of varying length as with GH 19 chitinases (Neuhaus 1999). Class III and class V share one highly conserved DXDXE domain, but otherwise share very low sequence similarity. The catalytic domain of Class III chitinases has low sequence similarity but very high structural similarity to bacterial chitinases. Very few class V chitinases have been studied in gymnosperms, however a Class V chitinase, purified from *Cycas revoluta,* was reported as having transgylcosylation activity (Taira *et al.* 2009).

1.5.2 Phylogenetic relationship of chitinase gene family

A recent phylogenetic analysis of chitinases in spruce found that class I and most class II chitinases fall into one well-supported cluster while class IV and class VII group together in a separate cluster, suggesting that class I and II chitinases share a common ancestor, while class IV and class VII chitinases share a different common ancestor (Galindo-Gonzalez *et al.* 2015). It has been proposed that divergence into separate classes arose from multiple independent deletion events of the CBD during the evolutionary history of chitinases (Neuhaus 1999). Conifer class I and II chitinases group distinctly from angiosperm class I and II chitinases, and conifer class IV and VII chitinases group group distinctly from angiosperm class IV and VII chitinases. This indicates that the loss of the CBD in both class I and class IV chitinases to become class II and class VII respectively occurred following the split of angiosperms and gymnosperms (Galindo-Gonzalez *et al.* 2015). This pattern corresponds to previous reports of angiosperm and gymnosperm class IV chitinases falling into distinct clusters in phylogenetic analyses, and supports the concept that class IV chitinases likely diversified after the emergence of angiosperms and gymnosperms (Liu et al 2014).

1.5.3 Role of chitinase enzymes in conifer defense

Many conifer chitinases are highly induced in response to pathogen challenge. The spatial and temporal diversity of these induced expression patterns suggest that different chitinases play distinct and separate roles as part of a complex defensive response. Examination of a class I, class II and class IV chitinases in Norway spruce inoculated with *Heterobasidion annosum* (Fr.) Bref., a necrotrophic root rot, revealed upregulation of class II and class IV chitinases, but down regulation of a class I chitinase (Heitala *et al*. 2004). Class II and class IV chitinases were both strongly upregulated in roots and needles of Douglas-fir inoculated at the

roots with laminated root rot, *Phellinus sulphurascens* Pilát. However, the expression patterns appeared to be spatially distinct between classes, with class IV chitinases more highly upregulated in root tissue closer to the site of infection, and class II chitinases comparatively more highly upregulated in needle tissue (Islam *et al.* 2010). Transcript abundance corresponding to class I, VII and IV chitinases increased markedly in lodgepole pine infected with *G. clavigera,* while a class VII chitinase was not upregulated in response to *G. clavigera,* suggesting that some chitinases play a role in defense while others may play a more developmental role (Kolosova *et al*. 2014). Overall, the expression patterns of GH 19 chitinases in response to pathogen challenge suggest there is a degree of temporal and spatial specificity in their response to pathogen challenge. Enzymatic assays confirmed the chitinolytic activity of class I chitinases found in lodgepole pine and white spruce, yet the same class I chitinases failed to show antifungal activity against both blue stain beetle associated fungi, *Leptographium abietinum* (Peck) Wingf. and *G. clavigera (*Kolosova *et al.* 2014).

It has been suggested that rather than directly inhibiting fungal growth, some chitinases work by releasing elicitors, in the form of chitin polymers, from the surface of fungal cells during the initial stages of infection, helping to activate early systemic defense response (Fossdal *et al.* 2007). In angiosperms, transgenically overexpressing chitinase genes led to enhanced resistance in angiosperm crop plants such as maize, rice, wheat, tomato, potato, grape, banana and others (Cletus *et al. 2013*). Silver birch, *Betula pendula* Roth.*,* transformed with a class IV chitinase originating from sugar beet also displayed enhanced resistance to fungal invaders (Pasonen *et al.* 2004). Little work has been done to examine the direct role pine chitinases play in pathogen containment, however evidence points to them having similar importance in conifer defense response.

Conifer chitinases may act as useful reporters of conifer defensive response. Early response to fungal invaders is critical in effective pathogen containment in any plant system. Chitinases are highly induced in response to pathogen challenge, and the timing and magnitude of chitinase expression has been shown to be important in the mounting a successful defensive response in conifers. In Norway spruce challenged with the root rot, *H. annosum,* class II and IV chitinases were upregulated earlier in the resistant seedling clone populations, and later but a much greater magnitude than in clone populations determined to be susceptible (Heitala *et al*. 2004). In slash pine, a class II chitinase, *PsChi4*, was expressed early at low levels in inoculation in slash pine seedlings resistant to *Fusarium subglutinans* (Wollenw. & Reinking) Nelson, Toussoun & Marasas., whereas in resistant seedlings *PsChi4* was expressed later after inoculation and at a much greater magnitude (Davis *et al.* 2004). Interestingly, the majority of chitinolytic activity was attributed to pine exudates not containing PsChi4, yet the authors propose that early, low level induction of the chitinase is an important indicator of rapid and more successful defensive response in seedlings. The authors speculate that class II chitinases may be acting as low level surveillance enzymes which release elicitors from invading pathogen cell walls helping to trigger a more specific and effective defense response from the host (Davis *et al.* 2004). Considering the size and scale of the current MPB outbreak, an effective measure of tree susceptibility is critical in helping to target management practices. The early expression of key chitinase enzymes could serve as a molecular marker of tree susceptibility to MPB.

1.5.4 High levels of evolutionary plasticity observed in chitinases

Chitinases display high levels of allelic variation across plant species, and this variation has been linked to resistant phenotypes in pine. Work done across multiple different species of *Arabis,* which are closely related to *Arabidopisis,* demonstrated that the rate of non-synonymous

substitutions in class I chitinases often exceeds the rate of non-synonymous substitutions of other gene families (Bishop *et al.* 2000). These non-synonymous mutations are often identified at the active site cleft, and the authors propose that this targeted form of mutation at the catalytic cleft is the product of strong selective pressure placed on chitinase enzymes in order to overcome the diverse forms of inhibitors produced by fungal pathogens (Bishop *et al.* 2000). Work done in western white pine identified high levels of induced expression of a class IV chitinase, *PmCh4*, in trees inoculated with white pine blister rust, *Cronartium ribicola,* Fisch. Two unique isozymes (27 kDa and 26 kDa) of PmCh4 were identified in resistant seedling tissue, while only one isozyme (27 kDa) was identified in susceptible seedlings (Liu et al 2005). This, along with the inducible expression patterns of *PmCh4*, suggests that chitinases play a defensive role in the quantitative resistance of western white pine against *C. ribicola.* Analysis of a class IV chitinase in western white pine demonstrated an association between allelic variants and quantitative levels of resistance against *C. ribicola* (Liu et al 2011). This evidence suggests that genetic variation in conifer chitinases contributes to resistant phenotypes. Chitinases may form part of a larger group of defense related genes which undergo highly specific selective pressure as part of the evolutionary "arms race" between pathogen and host, yet little work has been done to specifically examine genotypic variation in chitinases in lodgepole and jack pine across Canada.

1.6 Current Study

The objective of this study is to investigate the molecular mechanisms which contribute to lodgepole and jack pine defense responses to MPB fungal associate *G. clavigera,* and to investigate the influence of water deficit on these mechanisms. While differences in tree host quality and host defense response to MPB has been reported in evolutionarily co-evolved vs naïve pine hosts (Cudmore *et al.* 2010; Raffa *et al.* 2013), little is known about the underlying

regulatory genetic and molecular mechanisms responsible for these differences. The primary goal of this research was to identify some of these mechanisms by (1) comparing co-evolved lodgepole pine and naïve jack pine responses to inoculation with *G. clavigera* while under well watered and water deficit conditions, and (2) exploring expression profiles and allelic variation of chitinase genes, known defense response genes in lodgepole pine sampled from the MPB historic range as well as from lodegpole and jack pine sampled from outside this MPB historic range. The hypotheses tested in this research are: 1) that the evolutionarily co-evolved lodgepole pine has acquired induced and constitutive defenses to MPB that are not present in jack pine, 2) that defense response will be affected by water deficit conditions to a greater extent in lodgepole pine, which is a more drought responsive species than jack pine, and 3) that allelic variation putatively affecting protein function or transcriptional activation of known defense response genes, such as chitinases, will show spatially explicit patterns across the ranges of lodgepole and jack pine that might in turn reflect adaptive variation. The specific objectives of this study are: (1) to characterize the transcriptome-wide responses of lodgepole and jack pine seedlings inoculated with *G. clavigera* and grown under either well watered or water deficit conditions using microarray analysis, (2) to identify and characterize members of the lodgepole and jack pine chitinase gene families, and (3) to explore allelic variation of four putative orthologous pairs of chitinase genes in individuals from 11 provenances of jack and lodgepole pine sampled from across Canada. This thesis is organized into four chapters. Chapter 1 provides relevant back information on the MPB outbreak, the biology of MPB and its fungal associates. Chapter 2 describes analyses conducted on microarray data obtained from control or *G. clavigera*-inoculated lodgepole and jack pine seedlings grown under either well watered or water deficit conditions. Chapter 3 describes identification, phylogenetic analysis and *in silico* characterization of members of the lodgepole and jack pine chitinase gene families, expression profiling of a

subset of these genes in lodgepole and jack pine, as well as the allelic re-sequencing and variant analysis of four putative orthologous pairs of chitinase genes. The fourth chapter summarizes the findings of Chapters 2 and 3, outlines a conceptual model to explain differences in responses of lodgepole and jack pine to G. clavigera, and proposed future research directions.

Chapter 2: Transcriptome-wide analyses of lodgepole pine (*Pinus contorta*) and jack pine (*Pinus banksiana*) responses to *Grosmannia clavigera* under well watered versus water deficit conditions

2.1 Introduction

The current mountain beetle (MPB) outbreak is estimated to have affected approximately 19 million ha. of forests in the western Canadian provinces of British Columbia and Alberta, resulting in wide spread pine mortality (Arango-Velez *et al.* 2014). Over the course of the outbreak, MPB has undergone large scale range expansion moving from south central British Columbia, where it has historically attacked lodgepole pine, *Pinus contorta,* eastwards into the boreal forests of Alberta, where the beetle has encountered a novel host: jack pine, *Pinus banksiana* (Cullingham *et al.* 2011). MPB relies on a mass-attack strategy, aggregating in groups in order to concentrate an attack, and overcome the tree host's defenses. Ecological studies suggest that co-evolved lodgepole pine found in MPB's historic range have acquired induced and constitutive defenses against MPB that are not present in evolutionarily naïve jack pine hosts. Host trees from MPB's historic range have lower host quality compared to hosts from novel habitats, and MPB was reported to have a higher rate of reproductive success in lodgepole pine forests found outside of MPB's historical range (Cudmore *et al.* 2013; Cudmore *et al.* 2013 & 2010; Burke & Carroll 2016). If this is true, as MPB spreads eastwards into novel habitats, availability of naive tree hosts may augment reproductive success and growth of MPB populations, impacting the size and scale of the MPB outbreak. However, more work is necessary to test this hypothesis, and to identify molecular differences in defense response between naive and co-evolved pine hosts which underlie critical differences in tree host quality.

MPB is associated with a number of Ophiostomatoid fungal species, which assist in attack by weakening tree host defenses. *Grosmannia clavigera,* considered one the most pathogenic MPB fungal associates, contributes to eventual tree host mortality by growing into host sapwood tissue and blocking transport of water through occlusion of ray parenchyma and tracheids (Solheim & Krokene 1998; Lee *et al.* 2006; Ballard *et al.* 1984). Conifer trees rely on a number of different constitutive and induced defenses to defend against MPB and its fungal associate *G. clavigera*. Chemical defenses such as phenolic and terpenoid compounds help to form chemical and physical barriers against invaders, while expression of pathogenesis response (PR) proteins such as chitinases, osmotins and defensins all play important anti-microbial roles contributing to tree host defense response (Keeling & Bohlmann 2006; Neuhaus 1999). Early recognition and response to pathogen attack is critical in effective pathogen containment, and conifers rely on the defense related hormones salicylic acid (SA) and jasmonic acid (JA) to activate molecular and cellular defenses in response to pathogen challenge (Kolosova & Bohlmann 2012). SA is typically associated with challenge by biotrophic and hemi-biotrophic pathogens, and accumulates in conifers responding to pathogen challenge; however, the role that SA plays role as an elicitor of specific defense responses is less well established (Koslova & Bohlmann 2012). The ethylene-JA pathway is generally invoked in response to challenge by necrotrophic pathogens, and exogenous application of JA has been associated with the induction of anatomical and chemical defenses such as the formation of traumatic resin ducts and accumulation of terpenoid compounds in conifers (Franceschi et al. 2002; Martin et al. 2002; Hudgins & Franceschi 2004; Krokene et al. 2008; Gould et al. 2009). Earlier work has demonstrated that *in vivo* levels of JA-Ile increase in both lodgepole and jack pine seedlings in response to challenge by *G. clavigera,* and it has been suggested that *G. clavigera* along with other bark beetle

associated blue stain fungi, is a necrotrophic pathogen (Arango-Velez *et al.* 2016; Fossdal *et al.* 2012).

Higher densities of beetles are required to overcome the critical threshold of resistance in healthy and vigorously growing tree hosts, whereas lower densities are sufficient to overcome resistance of physiologically stressed trees (Berryman 1982; Kolb *et al.* 1998; Wallin & Raffa 2002; Raffa *et al.* 2005; Boone *et al.* 2011). Northern Alberta has experienced periods of drought over the past two decades, which have negatively impacted forests (Chhin *et al.* 2008, Michaelian *et al.* 2011, Hogg & Michaelian 2015). Ecological studies suggest that trees subjected to abiotic stresses such as drought are more susceptible to MPB attack, particularly at sub-epidemic populations (Breshears *et al.* 2009; McDowell *et al.* 2008; Safranyik *et al* 2010). Molecular studies have shown that water deficit conditions influence the defense response of lodgepole x jack pine hybrids to *G. clavigera*, altering the transcript abundance of biotic stress response genes and reducing the number traumatic resin ducts appearing in xylem tissue (Arango-Velez *et al.* 2014). Molecular studies have also shown that water deficit conditions attenuate some induced defense while transiently increasing some constitutive defenses differently between lodgepole and jack pine seedlings (Arango-Velez *et al.* 2016).

As MPB moves eastward across Alberta and into stands of naïve jack pine, factors such as drought, which weaken potential tree hosts and reduce the critical threshold for MPB attack, may influence the risk of continued MPB spread. Understanding how evolutionarily co-evolved lodgepole and naïve jack pine hosts interact with *G. clavigera* while under drought stress can aid in predicting how MPB will move through Alberta's pine forests. In this study, we aim to investigate the differences between the induced molecular defenses of evolutionarily co-evolved lodgepole and naïve jack pine seedlings in response to *G. clavigera,* and we aim to examine the effect of water deficit on these defenses. We hope to expand our understanding of the underlying

genetic and abiotic factors which influence tree host defense response and subsequent susceptibility to *G. clavigera*.

Using a two color cDNA microarray dataset (Arango-Velez *et al.*, in preparation), I have examined transcriptome-wide responses of lodgepole and jack seedlings either inoculated or not inoculated with *G. clavigera* under both well watered and water deficit conditions. We hypothesize that lodgepole and jack pine demonstrate differences in their responses to infection by *G. clavigera,* and that the co-evolved lodgepole pine seedlings will display qualitative differences in response to *G. clavigera* in comparison to naïve jack pine seedlings. We hypothesize that constitutive defenses will increase, and that induced defense responses will be attenuated under water deficit conditions. We predict that water deficit will exert more of an effect on lodegpole pine, since it is more sensitive, i.e. responds sooner, to water deficit than jack pine.

2.2 Materials and Methods

2.2.1 Plant materials and experimental design

One-year-old jack pine seedlings representing a provenance near Dryden Ontario, and one-year-old lodgepole pine seedlings representing a provenance near Hinton Alberta were shipped to Edmonton, and transplanted into 3.78L plastic pots filled with Sunshine Mix #4 (Sun Gro Horticulture Canada^{TM}). Seedlings were watered twice weekly, and fertilized once a week with 500 mg L^{-1} solution of 20-20-20 (N-P-K) fertilizer (Plant Products Company Ltd Brampton Ont.). Seedlings were grown in growth chambers under the following controlled conditions: 19°C, 20-25% relative humidity, and under incandescent lamps (200-250 μmol PAR, 16 h days / 8 h nights).

A full factorial experimental design was used to compare three treatment factors: species (lodgepole pine or jack pine), water availability (well-watered or water deficit) and inoculation (fungal inoculation, mechanical wound, control non-inoculated). Each individual seedling is defined as a unit of biological replication. Experimental treatments were applied based on a randomized complete block design. Seedlings were subjected to well watered conditions (soil water content, SWC >40%) or to water deficit conditions (SWC 20%) for two weeks prior to inoculation treatment, and continued throughout the duration of the experiment. Two weeks after the initiation of the water availability treatments, seedlings were subjected to (a) no further treatment (control), (b) mechanical wounding (mock), or (c) mechanical wounding plus inoculation with *G. clavigera*. Spore suspensions of the *G. clavigera* isolate M001-03-03-07-UC04DL09 *G. clavigera* isolate, a strain collected from Fox Creek Alberta (54°24'N, 116°48'W) (Roe *et al.* 2010, 2011) were used for inoculations. Inoculation was performed by injecting 5 μL of the spore suspension into the phloem tissue of seedlings using a 23G1 PrecisionGlide™ needle. Seedlings were inoculated at four equally distributed points around the stem. Bark was collected from 8 individual seedlings per treatment at 1 and 7 days post inoculation (dpi) by separating the bark away from the xylem at the cambial zone along the entire length of woody stem. Since most of the living tissue within the collected bark samples comprises the phloem, bark is hereafter referred to as phloem. Collected tissue was frozen immediately in liquid nitrogen and stored at -80°C.

2.2.2 RNA extractions, sample labeling and microarray hybridizations

Total RNA was extracted according to Pavy *et al.* (2008), quantified using a NanoQuant (Tecan Infinite, Morrisville NC, USA) and assessed using a 2100 Bioanalyzer (Agilent, Mississauga, ON, Canada). An aliquot of 2 μg of RNA from each of treatment combinations was

amplified using an amino allyl antisense RNA (aRNA) (Ambion® Amino Allyl Message AMP™ II aRNA Amplification Kit AM1753). Five micrograms of aRNA was labeled using Alexa Fluor® 555 or 647 dyes (Invitrogen, Carlsbad, CA, USA). Coupling efficiency was evaluated using the NanoQuant 2000.

Microarray experiments were conducted using PtGen2 loblolly arrays provided by Dr. Walter W. Lorenz and Dr. Jeffrey Dean (Lorenz *et al.* 2009), which contained 25,848 cDNAs amplified from loblolly pine (26,946 total spots minus buffer blanks and duplicate spots). We subsequently refer to each cDNA feature as a sequence, recognizing that due to a certain degree of redundancy represented on the array, not every sequence represents a unique gene. Microarrays on phloem tissue were carried out as described in El Kayal *et al.* (2011). For each species (lodgepole or jack pine), water availability (well watered or water limited) and day post-inoculation (1 dpi or 7 dpi) treatment combination, an inoculated sample was co-hybridized with a control uninoculated sample of the same treatment combination (Fig 2.1). Four biological replicates for each treatment combination were used, with two of the replicates representing dye swaps. Non-specific filtering was applied to reduce false discovery rate by removing invalid and low-intensity sequences (Fig 2.1). An empirical Bayes statistic was applied to obtain P-values, which were adjusted using the Benjamini-Hochberg procedure (Benjamini & Hochberg 1995). An adjusted P-value cut-off of 0.05 was used to obtain statistically significant differentially expressed (DE) sequences. This list was further filtered to include only sequences exhibiting fold changes greater than 1.5 or smaller than 0.67.

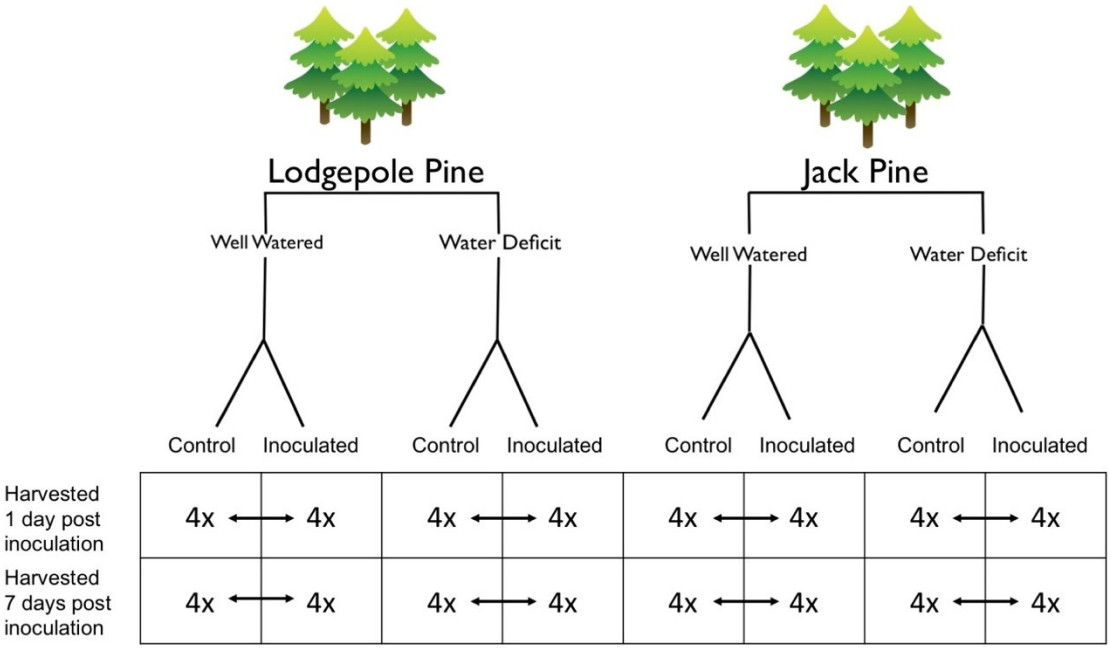


Figure 2.1: Experimental design depicting treatments applied to pine seedlings for use in microarray analysis. Lodgepole and jack pine seedlings were subjected to either well watered or water deficit conditions for two weeks; trees were then either not inoculated or inoculated with, with *G. clavigera* as described in Materials and Methods. Bark tissue was destructively sampled from eight seedlings at 1 dpi and from eight separate seedlings at 7 dpi. 4 biological replicates (4x), representing 4 individual seedlings subjected to each inoculated-water treatment combination, were co-hybridized on microarrays to one of the 4 biological replicates representing four individual seedlings subjected to the respective control-water treatment (co-hybridization indicated with black arrows).

2.2.3 Microarray annotation and data analysis

Lodgepole and jack pine sequences corresponding to the loblolly pine probe sequences represented on the microarray were identified from lodgepole and jack pine Illumina transcriptome assemblies (Hall *et al.* 2013) using BLASTx. Probe sequences with no apparent matches were excluded from downstream analyses. Sequences were annotated using BLASTx against NCBI and TAIR 7.0 (The Arabidopsis Information Resource) databases. The Mercator

annotation pipeline (Usadel *et al.* 2009) was used to assign MapMan BINs (functional categories) to each spot sequence.

The assigned MapMan BINs, annotation data associated with highly similar *Arabidopsis* gene sequences, and phylogenetic analyses were used to identify differentially expressed genes putatively involved in hormone biosynthesis and regulation, and with biotic and abiotic defense responses. Further manual sequence characterization was used to provide additional evidence for annotations. Potential transcription factors were identified from *Arabidopsis* gene sequences using the AGRIS (Arabidopsis Gene Regulatory Information Server) (Palaniswamy *et al.* 2006) transcription factor database, and from assigned MapMan annotation. Heatmaps of expression data for genes putatively involved in hormone biosynthesis and signaling were generated using MeV Multi-Experiment Viewer (Saeed *et al.* 2003).

Gene enrichment analyses were performed according to Galindo-Gonzalez et al. (2015) by comparing functional categories assigned by MapMan corresponding to a subset of genes to the functional categories corresponding to all sequences on the complete PtGen array. A hypergeometric distribution statistic was used to detect significant differences between frequency of functional categories, and a Bonferroni correction was applied to obtain adjusted P-values.

2.3 Results

2.3.1 Broad transcriptome-scale responses of lodgepole and jack pine seedlings to G. clavigera

Inoculation with *G. clavigera* induced large scale changes in the transcriptomes of lodgepole and jack pine at 1 and 7 dpi relative to control (untreated) plants at the same time points, under both well watered and water deficit conditions. Under well watered conditions,

inoculated jack pine seedlings exhibited 2991 DE sequences in comparison to control seedlings under well watered conditions across both 1 and 7 dpi (Fig 2.2). Of these, 1137 sequences were DE in inoculated vs. control seedlings of both lodgepole and jack pine. Under water deficit conditions, inoculated jack pine seedlings exhibited 1678 DE sequences in comparison to control seedlings. Under well watered conditions, inoculated lodgepole pine seedlings exhibited 1368 DE sequences relative to controls, and under water deficit conditions inoculated lodgepole pine seedlings exhibited 3009 DE sequences compared to controls (Fig 2.2). Under well deficit conditions, inoculated lodgepole and jack pine shared 1192 sequences that were DE relative to controls.

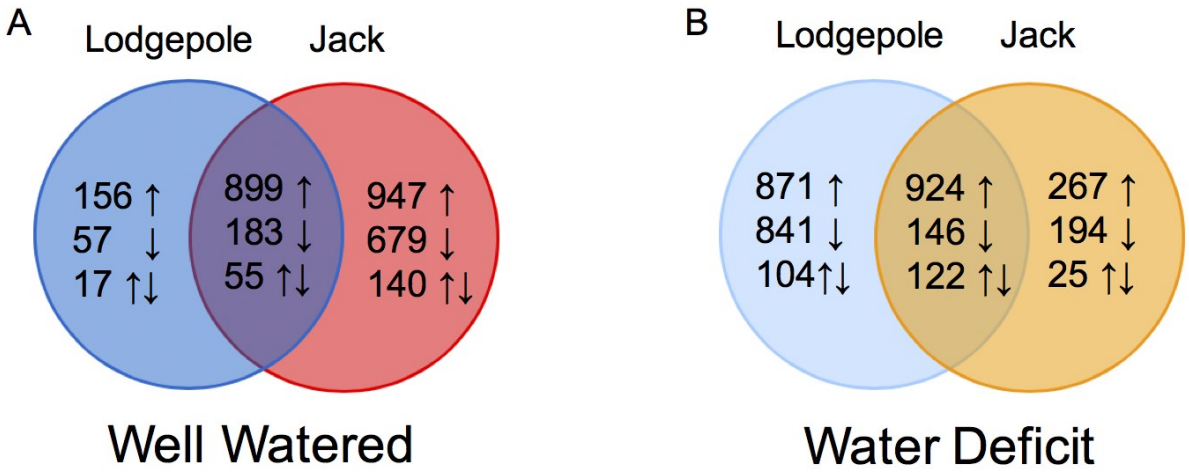


Figure 2.2: Venn diagrams comparing sequences DE in lodgepole pine and jack pine inoculated vs. control treatments at both 1 and 7 dpi. A. Well watered conditions. B. Water deficit conditions. Within each cell of the Venn diagram, DE sequences are further divided into those that were upregulated at 1 and/or 7 dpi (↑), downregulated at 1 and/or 7 dpi (↓), or upregulated at 1 or 7 dpi, but downregulated on the other day (↑↓).

2.3.2 Genes DE in both lodgepole and jack pine in response to G. clavigera

Four way Venn diagrams were used to examine DE sequences in lodgepole pine versus jack pine by day to reveal (a) DE sequences that constitute a core response common to both species, (b) sequences DE uniquely in one species versus the other, and (c) sequences DE earlier (1 dpi) versus later (7dpi) in the response to *G. clavigera* inoculation (Fig 2.3).

There were 229 sequences that were DE only at 1 dpi in both lodgepole and jack pine under well watered conditions, while just four sequences were DE only at 1 dpi in both species under water deficit conditions (Fig 2.3). Among the upregulated sequences under well watered conditions were three putative α-pinene synthases (spotID:65.9.8, spotID:32.9.2, spotID:28.20.11) (Table 2.1). Six putatively defense associated transcription factors (TF) were upregulated under well watered conditions in inoculated versus control samples in both lodgepole and jack pine at 1 dpi, including sequences with similarity to two putative ethylene response factors, ERFs, (spotID:2.13.13, spotID:43.3.3), two ethylene insensitive-like (EIL) factor (spotID:53.21.10, spotID:65.22.9), two JAZ-like TF's (spotID:7.14.2, spotID:63.5.4), one NAC domain containing sequence (spotID:53.21.10), and one MYC2-like sequence (spotID:38.13.8) (Table 2.1). Among the 289 sequences upregulated under well watered conditions in inoculated vs. control conditions for both lodgepole and jack pine only at 7 dpi were sequences encoding for 21 putative osmotins (spotID:31.19.9, spotID:32.4.6, spotID:41.22.12, spotID:45.3.1, spotID:5.1.12, spotID:61.10.8, spotID:61.14.12, spotID: 61.19.1, spotID:61.20.4, spotID:61.6.1, spotID:62.21.8, spotID:63.14.6, spotID:64.14.3, spotID:64.15.7, spotID:65.8.13, spotID:66.10.3, spotID:66.24.6, spotID: 67.1.6, spotID: 67.16.4, spotID: 67.9.10, spotID:9.1.12), three putative basic secretory proteins (spotID:45.15.15, spotID:61.9.1, spotID:7.6.7) and three putative chitinases (spotID:6.19.1, spotID:2.22.8, spotID:9.6.8) (Table 2.1).

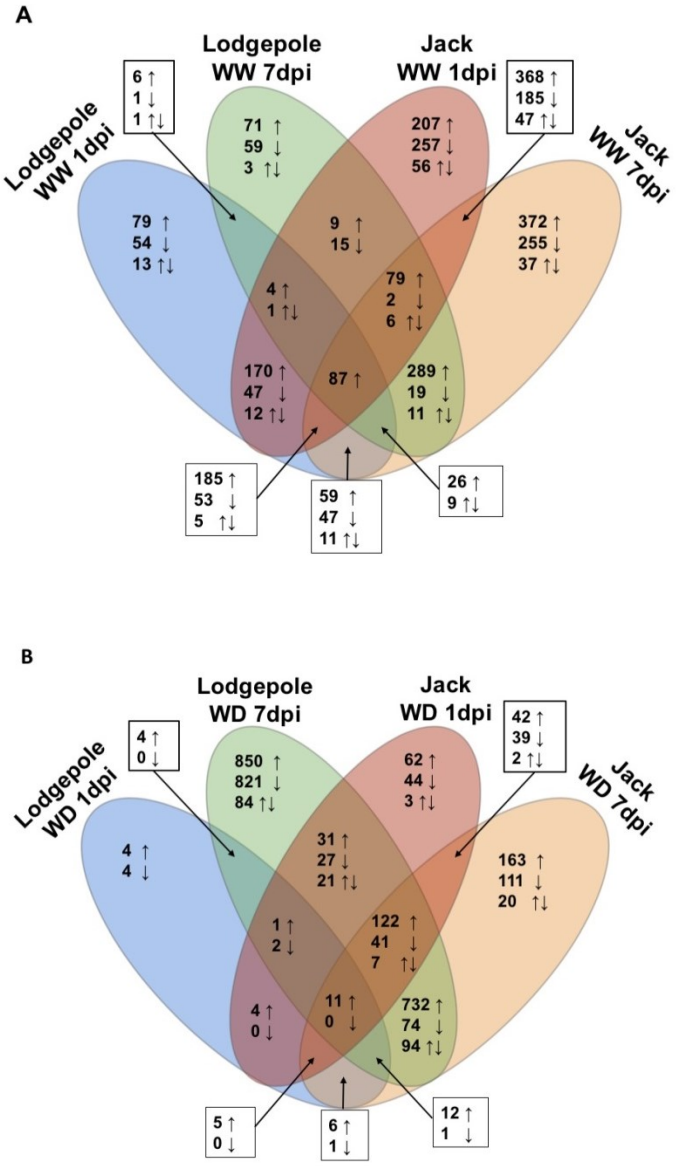


Figure 2.3: Four-way Venn diagrams comparing DE sequences between lodgepole pine and jack pine inoculated vs. control treatments and further comparing between sequences DE at 1 dpi and 7 dpi. A. Well watered conditions (WW). B. Water deficit conditions (WD). Within each cell of the Venn diagram, DE sequences are further divided into those that were upregulated at 1 and/or 7 dpi (↑), downregulated at 1 and/or 7 dpi (↓), or upregulated at 1 or 7 dpi, but downregulated on the other day (↑↓).

Three defense associated transcriptional regulators - two defense associated putative, WRKY TFs (spotID:67.8.5, spotID:67.21.8), and a putative wounding-responsive MYB TF (spotID:66.17.4) - were also upregulated under well watered conditions in both species at only 7 dpi, along with a putative aminocyclopropane carboxylic acid (ACC) synthase involved in biosynthesis of ethylene (spotID:68.8.5) (Table 2.1).

Eighty-seven sequences were DE under well watered conditions in both species at both 1 and 7 dpi, including seven putative chitinases (spotID:38.14.8, spotID:4.9.8, spot:40.21.10, spotID:53.7.5, spotID:6.13.15, spotID:7.15.16, spotID:7.24.4), and two putative dirigent-like protein (DIR) (spotID: 66.11.5, spotID:27.20.7) (Fig 2.3; Table 2.1).

Enrichment analysis of sequences DE in both well watereded lodgepole pine and well watered jack pine revealed significant overrepresentation of genes involved in secondary metabolism, stress response, protein processing and photosynthesis (Table 2.2). Further examination of genes involved in secondary metabolism revealed 19 sequences involved in biosynthesis of isoprenoids, including terpenes, six sequences involved in the biosynthesis of phenylpropanoids, and four sequences involved in the biosynthesis of flavonoids (Table 2.2 & Table 2.3).

In contrast to well watered conditions only four sequences were upregulated at 1 dpi in both lodgepole and jack pine under water deficit, none of which play roles in defense response or transcriptional regulation (figure 2.3). Among the 732 sequences upregulated under water deficit conditions at only 7 dpi in both lodgepole and jack pine were sequences encoding three putative BSP-like sequences (spotID:4.15.15, spotID:67.23.8, spotID:19.22.11), 14 putative chitinase genes (spotID:63.22.12, spotID:66.24.14, spotID:9.6.8, spotID:30.2.12, spotID:22.23.9, spotID:6.5.12, spotID:17.12.14, spotID:66.23.7, spotID:2.22.8, spotID:61.14.6, spotID:52.6.6, spotID:64.6.10, spotID:67.21.1, spotID:63.24.13, spotID:25.22.16, spotID:3.17.16), 16 putative

Table 2.1: Sequences on the PtGen2 microarray differentially expressed in both lodgepole and jack pine phloem tissue between inoculated vs. control treatments under well watered conditions at 1 and/ or 7 dpi with *G. clavigera*. Sequence putative identity was predicted on the basis of sequence similarity to sequences within the TAIR database on NCBI nr database (not shown). FC: fold change. P: adjusted p-value.

PtGen2 spot ID	**TAIR ID**	**BLAST X E-value**	**Gene Symbols**	**Putative identity**	Lodgepole 1 dpi WW		Lodgepole 7 dpi WW		Jack 1 dpi WW		Jack 1 dpi WW	
					FC	P	FC	P	FC	P	FC	P
DE at 1 dpi in both lodgepole and jack pine under well watered conditions												
65.9.8	AT2G41710.1	2.E-66	Integrase-type DNA-binding superfamily protein	pinene synthase-like	**2.09**	0.04	0.99	0.98	**2.31**	0.01	0.94	0.53
32.9.2	AT2G41710.1	2E-66	Integrase-type DNA-binding superfamily protein	pinene synthase-like	**1.96**	0.05	0.82	0.82	**2.40**	0.03	0.73	0.06
65.9.8	AT2G41710.1	2E-66	Integrase-type DNA-binding superfamily protein	pinene synthase-like	**2.09**	0.04	0.99	0.98	**2.31**	0.01	0.94	0.53
2.13.13	AT5G47220.1	4E-12	ATERF2| ethylene responsive element binding factor 2	ERF-like TF	**2.59**	0.04	0.82	0.50	**2.76**	0.01	1.38	0.07
44.3.3	AT5G47220.1	4E-12	ATERF2| ethylene responsive element binding factor 2	ERF-like TF	**0.47**	0.03	0.77	0.28	**0.49**	0.00	0.71	0.03
53.21.10	AT2G27050.1	2.E-141	AtEIL1 | ETHYLENE-INSENSITIVE3-like 1	EIL-like TF	**1.87**	0.04	1.24	0.38	**1.78**	0.02	1.32	0.05
65.22.9	AT2G27050.1	2E-141	AtEIL1 | ETHYLENE-INSENSITIVE3-like 1	EIL-like TF	**1.66**	0.04	1.13	0.52	**1.58**	0.03	1.10	0.49

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative identity	Lodgepole 1 dpi WW		Lodgepole 7 dpi WW		Jack 1 dpi WW		Jack 1 dpi WW	
					FC	P	FC	P	FC	P	FC	P
7.14.2	AT1G74950.1	6E-6	JAZ2 | TIFY domain/Divergent CCT motif family protein	JAZ-like TF	**3.05**	0.05	0.85	0.78	**8.21**	0.00	1.10	0.80
63.5.4	AT3G17860.1	2E-21	JAZ3 | jasmonate-zim-domain protein 3	JAZ-like TF	**2.06**	0.03	1.24	0.18	**1.86**	0.02	1.23	0.09
53.21.9	AT3G10500.1	1E-75	NAC053 | NAC domain containing protein 53	NAC-like TF	**1.67**	0.04	1.00	1.00	**1.97**	0.01	1.23	0.22
38.13.1	AT1G32640.1	6E-64	AtMYC2 | Basic helix-loop-helix (bHLH) DNA-binding family protein	MYC2-like	**4.43**	0.01	1.13	0.71	**2.77**	0.00	1.07	0.73
DE at 7 dpi in both lodgepole and jack pine under well watered conditions												
31.19.9	AT4G11650.1	2E-64	OSM34 | osmotin 34	osmotin-like	1.10	0.88	**14.40**	0.02	1.15	0.60	**5.22**	0.00
32.4.6	AT4G11650.1	1E-75	OSM34 | osmotin 34	osmotin-like	0.96	0.81	**11.80**	0.01	1.03	0.87	**9.14**	0.00
41.22.12	AT4G11650.1	1E-75	OSM34 | osmotin 34	osmotin-like	2.49	0.13	**23.90**	0.01	2.44	0.12	**40.98**	0.00
45.3.1	AT4G11650.1	4E-55	OSM34 | osmotin 34	osmotin-like	0.88	0.43	**4.61**	0.00	1.23	0.29	**7.39**	0.00
5.1.12	AT4G11650.1	3E-49	OSM34 | osmotin 34	osmotin-like	1.27	0.58	**2.85**	0.05	1.76	0.18	**10.32**	0.00
61.10.8	AT4G11650.1	1E-75	OSM34 | osmotin 34	osmotin-like	1.24	0.45	**8.52**	0.01	1.18	0.22	**7.91**	0.00
61.14.12	AT4G11650.1	1E-75	OSM34 | osmotin 34	osmotin-like	1.42	0.38	**10.60**	0.02	1.21	0.45	**16.66**	0.01
61.19.1	AT4G11650.1	4E-55	OSM34 | osmotin 34	osmotin-like	0.92	0.73	**2.62**	0.06	0.93	0.79	**4.01**	0.00
61.20.4	AT4G11650.1	1E-75	OSM34 | osmotin 34	osmotin-like	1.26	0.20	**4.70**	0.02	1.21	0.10	**5.55**	0.00

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative identity	Lodgepole 1 dpi WW		Lodgepole 7 dpi WW		Jack 1 dpi WW		Jack 1 dpi WW	
					FC	P	FC	P	FC	P	FC	P
61.6.1	AT4G11650.1	1E-75	OSM34 | osmotin 34	osmotin-like	1.60	0.28	**7.18**	0.01	1.23	0.18	**13.84**	0.00
62.21.8	AT4G11650.1	1E-75	OSM34 | osmotin 34	osmotin-like	1.38	0.40	**16.20**	0.01	1.50	0.13	**23.88**	0.00
63.14.6	AT4G11650.1	1E-75	OSM34 | osmotin 34	osmotin-like	1.80	0.16	**13.10**	0.01	1.51	0.26	**20.96**	0.00
64.14.3	AT4G11650.1	1E-75	OSM34 | osmotin 34	osmotin-like	3.26	0.27	**15.60**	0.01	1.57	0.24	**22.50**	0.00
64.15.7	AT4G11650.1	2E-16	OSM34 | osmotin 34	osmotin-like	0.98	0.94	**2.45**	0.01	1.34	0.12	**3.93**	0.00
65.8.13	AT4G11650.1	1E-75	OSM34 | osmotin 34	osmotin-like	1.50	0.24	**10.20**	0.01	1.37	0.05	**12.84**	0.00
66.10.3	AT4G11650.1	1E-26	OSM34 | osmotin 34	osmotin-like	1.18	0.46	**6.96**	0.01	1.39	0.04	**7.53**	0.00
66.24.6	AT4G11650.1	1E-75	OSM34 | osmotin 34	osmotin-like	1.46	0.24	**7.41**	0.03	1.97	0.09	**14.22**	0.01
67.1.6	AT4G11650.1	4E-55	OSM34 | osmotin 34	osmotin-like	0.85	0.31	**3.45**	0.03	0.76	0.14	**4.73**	0.01
67.16.4	AT4G11650.1	1E-75	OSM34 | osmotin 34	osmotin-like	2.73	0.29	**14.80**	0.01	2.87	0.14	**36.06**	0.00
67.9.10	AT4G11650.1	1E-75	OSM34 | osmotin 34	osmotin-like	1.12	0.47	**4.50**	0.03	0.93	0.64	**6.56**	0.00
9.1.12	AT4G11650.1	2E-16	OSM34 | osmotin 34	osmotin-like	0.58	0.18	**3.80**	0.01	1.08	0.68	**7.71**	0.00
45.15.15	AT2G15220.1	5E-48	Plant basic secretory protein (BSP) family protein	BSP-like	1.15	0.45	**7.59**	0.04	1.48	0.02	**7.91**	0.01
61.9.1	AT2G15220.1	5E-48	Plant basic secretory protein (BSP) family protein	BSP-like	1.00	1.00	**2.47**	0.02	1.52	0.15	**1.74**	0.03
7.6.7	AT2G15220.1	6E-59	Plant basic secretory protein (BSP) family	BSP-like	1.75	0.22	**11.10**	0.00	0.97	0.83	2.02	0.03

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative identity	Lodgepole 1 dpi WW		Lodgepole 7 dpi WW		Jack 1 dpi WW		Jack 1 dpi WW	
					FC	P	FC	P	FC	P	FC	P
			protein									
6.19.1	AT2G43590.1	8E-18	Chitinase family protein	chitinase-like	1.29	0.38	**2.01**	0.04	1.09	0.48	**1.87**	0.02
2.22.8	AT2G43590.1	2E-40	Chitinase family protein	chitinase-like	1.55	0.08	**2.89**	0.02	1.25	0.12	**4.01**	0.00
9.6.8	AT2G43590.1	3E-15	Chitinase family protein	chitinase-like	2.47	0.08	**5.73**	0.03	1.86	0.05	**10.33**	0.00
67.21.8	AT5G64810.1	3E-17	AtWRKY51 | WRKY DNA-binding protein 51	WRKY-like TF	0.91	0.67	**3.89**	0.01	1.43	0.19	**6.25**	0.00
67.8.5	AT5G64810.1	3E-17	AtWRKY51 | WRKY DNA-binding protein 51	WRKY-like TF	1.36	0.15	**1.98**	0.02	1.30	0.20	**3.08**	0.00
66.17.4	AT1G22640.1	1E-26	AtMYB3 | myb domain protein 3	MYB-like TF	0.68	0.05	**1.72**	0.11	1.13	0.62	**2.65**	0.00
68.8.5	AT4G26200.1	2E-26	AtACS7 | ACC synthase	ACC synthase-like	0.86	0.39	**1.82**	0.03	0.79	0.06	**1.84**	0.01
64.10.1	AT3G12500.1	2E-60	ATHCHIB| basic chitinase	chitinase-like	**6.21**	0.02	**12.10**	0.01	**9.21**	0.00	**6.49**	0.00
40.21.10	AT3G12500.1	8E-89	ATHCHIB| basic chitinase	chitinase-like	**5.17**	0.03	**7.06**	0.01	**6.33**	0.01	**7.93**	0.00
53.7.5	AT3G12500.1	6E-104	ATHCHIB| basic chitinase	chitinase-like	**3.76**	0.03	**4.87**	0.04	**4.42**	0.00	**4.34**	0.00
4.9.8	AT3G12500.1	8E-89	ATHCHIB| basic chitinase	chitinase-like	**10.20**	0.02	**8.63**	0.02	**7.66**	0.01	**11.17**	0.00
7.24.4	AT3G12500.1	6E-104	ATHCHIB| basic chitinase	chitinase-like	**2.21**	0.04	**3.85**	0.02	**2.95**	0.01	**4.18**	0.00
6.13.15	AT3G12500.1	6E-104	ATHCHIB| basic chitinase	chitinase-like	**2.93**	0.03	**4.63**	0.01	**3.06**	0.02	**4.08**	0.00
38.14.8	AT3G12500.1	2E-60	ATHCHIB| basic chitinase	chitinase-like	**3.26**	0.01	**3.45**	0.03	**3.40**	0.01	**4.30**	0.00
27.20.7	AT1G64160.1	9E-47	Disease resistance-responsive (dirigent-like	DIR-like	**15.39**	0.01	**4.84**	0.01	**18.34**	0.00	**13.71**	0.00

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative identity	Lodgepole 1 dpi WW		Lodgepole 7 dpi WW		Jack 1 dpi WW		Jack 1 dpi WW	
					FC	P	FC	P	FC	P	FC	P
			protein) family protein									
66.11.5	AT1G64160.1	9E-47	Disease resistance-responsive (dirigent-like protein) family protein	DIR-like	**7.89**	0.03	**3.91**	0.02	**11.30**	0.00	**7.09**	0.00

osmotins (spotID:63.17.3, spotID:61.19.1, spotID:61.10.8, spotID:65.8.13, spotID:67.16.12, spotID:4.8.10, spotID:32.4.6, spotID:33.10.1, spotID:63.16.13, spotID:45.3.1, spotID:31.8.10, spotID:64.15.7, spotID:25.22.16, spotID:55.22.13, spotID:9.1.12, spotID:41.22.12), two putative members of pathogenesis response family 4 (PR-4) (spotID:7.6.3, spotID:3.24.8), and eight sequences encoding 4 different putative terpene synthases, including two β-farnesene synthases (spotID:50.4.11, spotID:67.3.1), two levoprimaradiene synthases (spotID:36.22.5, spotID:61.19.10), a Δ-selinene synthase (spotID:64.4.2), and three sesquiterpene synthase (spotID:21.2.12, spotID:40.5.7, spotID:63.20.12) (Table 2.4).

Table 2.2: Enrichment analysis of DE genes in both lodgepole and jack pine inoculated vs. control seedlings under well watered conditions. Table displays functional categories significantly over represented in sequences DE in both lodgepole and jack pine inoculated vs. control seedlings under well watered conditions using a hypergeometric distribution statistic.

Functional Category	Total PtGen elements in category	Elements DE in lodgepole and jack	Probability Density Function (Hypergeometric distribution)	Adjust P value (Bonferroni)
Secondary metabolism	3831	38	5.30E-34	1.54E-32
Miscellaneous cellular processes	1544	143	3.00E-23	8.69E-22
Stress response	1167	84	2.25E-08	6.53E-07
Protein processing	3463	96	1.02E-05	2.96E-04
Photosynthesis	526	6	7.82E-05	2.27E-03

Table 2.3: Count of sequences putatively involved in secondary metabolism and determined by microarray analysis to be differentially expressed in inoculated vs control tissues under well watered or water deficit conditions in both lodgepole and jack pine, exclusively in lodgepole pine or exclusively in jack pine. Sequences are further classified into those involved in flavonoid biosynthesis, isoprenoid biosynthesis and phenylpropanoid biosynthesis on the basis of sequence identity using MapMan functional categorization.

Functional category	**DE in both lodgepole and jack pine**	**DE exclusively in lodgepole pine**	**DE exclusively in jack pine**
DE under well watered conditions			
flavonoid biosynthesis	4	5	24
isoprenoids biosynthesis	19	0	38
phenylpropanoids biosynthesis	6	2	11
DE under water deficit conditions			
flavonoid biosynthesis	39	24	1
isoprenoids biosynthesis	22	11	14
phenylpropanoids biosynthesis	13	6	4

Interestingly, 21 putative chalcone synthase sequences (CHS), involved in phenolic biosynthesis were DE in lodgepole and jack pine at only 7 dpi exclusively under water deficit conditions. Seventeen of these genes of these were upregulated in both species under water deficit conditions (spotID:11.18.6, spotID:11.8.9, spotID:17.2.11, spotID:23.16.14, spotID:24.17.5, spotID:27.17.3, spotID:32.16.12, spotID:33.10.7, spotID:40.15.16, spotID:59.23.9, spotID:63.19.6, spotID:63.8.11, spotID:7.18.5, spotID:19.21.5, spotID:25.14.7, spotID:36.1.7, spotID:6.4.5), while four were upregulated in lodgepole but down regulated in jack pine under water deficit conditions (spotID:31.20.16, sotID:34.9.15, spotID:40.12.10, spotID56.8.8) (Table 2.4). Among the defense associated transcriptional regulators DE in both species only at 7 dpi under water deficit conditions were three putative JAZ-like TFs (spotID:16.8.10, spotID:62.10.9), two putative WRKY TFs (spotID:54.14.15, spotID:67.8.5), three putative ERFs (spotID:15.24.6, spotID:25.5.14, spotID:7.13.5), and one putative MYB TF

(spotID:66.17.4) (Table 2.4). Five putative ACC oxidase sequences involved in ethylene biosynthesis (spotID:46.23.12, spotID:4.4.16, spotID:28.2.5, spotID:40.14.9, spotID:56.18.8). One putative jasmonate methyl-transferase sequence was upregulated in both lodgepole and jack pine at only 7 dpi under water deficit conditions along with three putative phenylalanine ammonia lyase (PAL) sequences central to biosynthesis of phenolic compounds, including salicylic acid (spotID:3.21.4, spotID:63.19.7, spotID:35.4.12) (Table 2.4). There were nine sequences DE under both well watered and water deficit conditions in lodgepole and jack at both 1 and 7 dpi. One of these sequences encode a putative chitinase (spotID:64.10.1), and two encode putative DIRs (spotID:16.18.12, spotID:27.20.7) (Table 2.1 &4).

Enrichment analysis of sequences DE in both lodgepole and jack pine under water deficit conditions revealed over representation of genes involved in miscellaneous cellular processes including peroxidase activity, secondary metabolism, stress response, protein processing, biodegradation of xenobiotics and the oxidative pentose phosphate pathway (Table 2.5). Examination of the genes involved in secondary metabolism revealed 39 sequences involved in the metabolism of flavonoids, 22 sequences involved in the biosynthesis of isoprenoids, and 13 sequences involved in the biosynthesis of phenylpropanoids (Table 2.3).

Table 2.4: Sequences differentially expressed on the PtGen2 microarray in both lodgepole and jack pine phloem tissue between inoculated vs. control treatments under water deficit at 1 and/ or 7 dpi with *G. clavigera*. Sequence putative identity was predicted on the basis of sequence similarity to sequences within the TAIR database on NCBI nr database (not shown). FC: fold change. P: adjusted p-value.

PtGen2 spot ID	**TAIR ID**	**BLAST X E-value**	**Gene Symbols**	**Putative Identity**	Lodgepole 1 dpi WD		Lodgepole 7 dpi WD		Jack 1 dpi WD		Jack 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
DE in lodgepole and jack pine at 7 dpi under water deficit conditions												
45.15.15	AT2G15220.1	5E-48	Plant basic secretory protein (BSP) family protein	BSP-like	0.83	0.46	**3.65**	0.00	1.32	0.18	**8.82**	0.01
67.23.8	AT2G15220.1	5E-48	Plant basic secretory protein (BSP) family protein	BSP-like	1.07	0.94	**8.64**	0.01	1.73	0.28	**13.81**	0.02
19.22.11	AT2G15220.1	6E-59	Plant basic secretory protein (BSP) family protein	BSP-like	NA	NA	**2.28**	0.01	NA	NA	**2.70**	0.05
63.22.12	AT3G54420.1	7E-10	ATCHITIV | homolog of carrot EP3-3 chitinase	Chitinase-like	0.90	0.61	**2.67**	0.03	1.73	0.12	**2.02**	0.04
66.24.14	AT2G43590.1	3E-67	Chitinase family protein	Chitinase-like	1.12	0.78	**3.54**	0.00	1.19	0.31	**3.14**	0.04
9.6.8	AT2G43590.1	3E-15	Chitinase family protein	Chitinase-like	1.11	0.90	**6.09**	0.01	0.79	0.75	**3.68**	0.02
30.2.12	AT3G54420.1	1E-51	ATCHITIV | homolog of carrot EP3-3 chitinase	Chitinase-like	1.76	0.23	**3.39**	0.02	0.68	0.19	**4.35**	0.01
22.23.9	AT2G43590.1	2E-40	Chitinase family protein	Chitinase-like	1.66	0.72	**10.90**	0.00	1.63	0.41	**16.22**	0.03
6.5.12	AT2G43590.1	6E-16	Chitinase family protein	Chitinase-like	1.08	0.74	**2.53**	0.01	1.11	0.67	**3.29**	0.01
17.12.14	AT2G43590.1	2E-40	Chitinase family	Chitinase-	1.56	0.71	**7.23**	0.00	1.43	0.20	**9.40**	0.02

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative Identity	Lodgepole 1 dpi WD		Lodgepole 7 dpi WD		Jack 1 dpi WD		Jack 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
			protein	like								
66.23.7	AT3G12500.1	2E-60	ATHCHIB | basic chitinase	Chitinase-like	1.74	0.10	**3.71**	0.01	4.72	0.05	**6.04**	0.01
2.22.8	AT2G43590.1	2E-40	Chitinase family protein	Chitinase-like	1.14	0.64	**4.65**	0.01	1.24	0.33	**3.83**	0.03
61.14.7	AT3G54420.1	9E-52	ATCHITIV | homolog of carrot EP3-3 chitinase	Chitinase-like	1.54	0.39	**9.03**	0.00	1.69	0.23	**6.54**	0.00
52.6.6	AT3G12500.1	6E-104	ATHCHIB | basic chitinase	Chitinase-like	2.86	0.17	**2.57**	0.01	2.49	0.12	**2.55**	0.02
64.6.10	AT3G12500.1	8E-89	ATHCHIB | basic chitinase	Chitinase-like	1.67	0.11	**2.69**	0.01	4.59	0.07	**5.62**	0.00
67.21.1	AT3G12500.1	2E-60	ATHCHIB | basic chitinase	Chitinase-like	2.07	0.09	**3.83**	0.01	4.67	0.05	**5.72**	0.00
63.24.13	AT3G12500.1	2E-60	ATHCHIB | basic chitinase	Chitinase-like	1.31	0.44	**2.13**	0.01	1.51	0.16	**1.83**	0.01
25.22.16	AT3G54420.1	4E-46	ATCHITIV | homolog of carrot EP3-3 chitinase	Chitinase-like	1.10	0.61	**1.93**	0.01	1.26	0.23	**3.08**	0.01
3.17.16	AT2G43590.1	2E-40	Chitinase family protein	Chitinase-like	1.00	1.00	**6.94**	0.00	1.32	0.34	**5.38**	0.03
63.17.3	AT4G11650.1	4E-55	OSM34 | osmotin 34	Osmotin-like	1.07	0.88	**4.21**	0.02	1.32	0.45	**11.68**	0.01
61.19.1	AT4G11650.1	4E-55	OSM34 | osmotin 34	Osmotin-like	0.66	0.34	**3.75**	0.02	0.68	0.42	**4.81**	0.01
61.10.8	AT4G11650.1	1E-75	OSM34 | osmotin 34	Osmotin-like	1.10	0.57	**6.73**	0.01	1.19	0.46	**4.05**	0.03
65.8.13	AT4G11650.1	1E-75	OSM34 | osmotin 34	Osmotin-like	1.36	0.29	**10.00**	0.02	1.23	0.55	**5.39**	0.04
67.16.12	AT4G11650.1	2E-64	OSM34 | osmotin 34	Osmotin-like	1.05	0.89	**1.82**	0.04	1.06	0.90	**2.40**	0.02
4.8.10	AT4G11650.1	1E-75	OSM34 | osmotin 34	Osmotin-like	1.00	0.99	**1.92**	0.04	1.60	0.08	**2.11**	0.05

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative Identity	Lodgepole 1 dpi WD		Lodgepole 7 dpi WD		Jack 1 dpi WD		Jack 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
32.4.6	AT4G11650.1	1E-75	OSM34 | osmotin 34	Osmotin-like	1.06	0.85	**9.25**	0.01	1.07	0.77	**8.19**	0.02
33.10.1	AT4G11650.1	1E-75	OSM34 | osmotin 34	Osmotin-like	1.48	0.54	**3.23**	0.00	0.84	0.46	**2.25**	0.01
63.16.13	AT4G11650.1	2E-16	OSM34 | osmotin 34	Osmotin-like	0.82	0.64	**2.95**	0.04	1.26	0.77	**16.83**	0.00
45.3.1	AT4G11650.1	4E-55	OSM34 | osmotin 34	Osmotin-like	0.93	0.65	**3.34**	0.01	1.03	0.94	**9.52**	0.01
31.8.10	AT4G11650.1	2E-16	OSM34 | osmotin 34	Osmotin-like	0.79	0.34	**2.90**	0.04	1.10	0.81	**13.44**	0.00
64.15.7	AT4G11650.1	2E-16	OSM34 | osmotin 34	Osmotin-like	0.91	0.68	**1.98**	0.02	1.83	0.06	**6.63**	0.00
25.22.16	AT3G54420.1	4E-46	ATCHITIV | homolog of carrot EP3-3 chitinase	Chitinase-like	1.10	0.61	**1.93**	0.01	1.26	0.23	**3.08**	0.01
55.22.13	AT4G11650.1	2E-16	OSM34 | osmotin 34	Osmotin-like	0.78	0.42	**4.21**	0.02	1.30	0.68	**18.21**	0.00
9.1.12	AT4G11650.1	2E-16	OSM34 | osmotin 34	Osmotin-like	0.93	0.92	**2.45**	0.01	1.34	0.46	**8.05**	0.01
41.22.12	AT4G11650.1	1E-75	OSM34 | osmotin 34	Osmotin-like	2.04	0.33	**19.80**	0.02	NA	NA	**12.82**	0.05
7.6.3	AT3G04720.1	4E-31	PR-4 | pathogenesis-related 4	PR-4-like	0.90	0.64	**3.15**	0.01	1.27	0.41	**3.57**	0.00
3.24.8	AT3G04720.1	2E-49	PR-4 | pathogenesis-related 4	PR-4-like	NA	NA	**12.40**	0.00	8.70	0.07	**32.15**	0.02
50.4.11	AT1G70080.1	4.5	Terpenoid cyclases/Protein prenyltransferases superfamily protein |	Farnesene Synthase-like	1.36	0.39	**4.20**	0.03	1.29	0.32	**6.25**	0.01
67.3.1	AT1G70080.1	4.5	Terpenoid cyclases/Protein prenyltransferases superfamily protein	Farnesene Synthase-like	1.42	0.23	**2.67**	0.02	1.19	0.54	**4.60**	0.03
36.22.5	AT1G61120.1	4E-19	TPS4 | terpene synthase 04	Levoprima radiene	0.87	0.80	**3.55**	0.00	1.55	0.29	**11.88**	0.01

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative Identity	Lodgepole 1 dpi WD		Lodgepole 7 dpi WD		Jack 1 dpi WD		Jack 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
				synthase-like								
61.19.10	AT4G16730.1	6E-50	TPS02 | terpene synthase 02	Levoprimaradiene synthase-like	0.93	0.77	**2.89**	0.00	2.12	0.13	**15.27**	0.01
64.4.2	AT3G25810.1	4E-63	Terpenoid cyclases/Protein prenyltransferases superfamily protein	Selinene Synthase-like	1.59	0.44	**5.14**	0.05	4.39	0.08	**15.24**	0.01
21.2.12	AT4G02780.1	1E-106	CPS1 | Terpenoid cyclases/Protein prenyltransferases superfamily protein	Sesquiterpene Synthase-like	1.00	0.99	**4.77**	0.01	1.30	0.23	**8.36**	0.02
40.5.7	AT4G02780.1	1E-106	CPS1 | Terpenoid cyclases/Protein prenyltransferases superfamily protein	Sesquiterpene Synthase-like	0.97	0.89	**3.54**	0.00	0.96	0.89	**6.93**	0.01
63.20.12	AT4G02780.1	1E-106	CPS1 | Terpenoid cyclases/Protein prenyltransferases superfamily protein	Sesquiterpene Synthase-like	NA	NA	**4.37**	0.01	NA	NA	**6.49**	0.03
11.18.6	AT5G13930.1	2E-84	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.83	0.43	**2.14**	0.01	0.72	0.42	**1.61**	0.02
11.8.9	AT5G13930.1	3E-69	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.45	0.58	**2.30**	0.01	0.75	0.41	**1.63**	0.03
17.2.11	AT5G13930.1	2E-69	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.42	0.54	**2.24**	0.01	0.77	0.37	**1.67**	0.03
19.21.5	AT5G13930.1	0.00009	CHS | Chalcone and stilbene synthase	Chalcone synthase-	1.87	0.46	**2.05**	0.01	0.66	0.33	**1.61**	0.03

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative Identity	Lodgepole 1 dpi WD		Lodgepole 7 dpi WD		Jack 1 dpi WD		Jack 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
			family protein	like								
23.16.14	AT5G13930.1	2E-69	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.59	0.38	**2.48**	0.01	1.06	0.82	**1.80**	0.05
24.17.5	AT5G13930.1	2E-84	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.83	0.47	**2.15**	0.01	0.67	0.42	**1.71**	0.02
25.14.7	AT5G13930.1	4E-71	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.67	0.39	**2.72**	0.01	0.95	0.88	**1.80**	0.04
27.17.3	AT5G13930.1	2E-69	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.53	0.44	**3.00**	0.01	1.13	0.44	**1.99**	0.02
32.16.12	AT5G13930.1	2E-69	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.60	0.47	**2.17**	0.01	0.72	0.39	**1.58**	0.04
33.10.7	AT5G13930.1	2E-69	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.32	0.54	**2.29**	0.02	0.87	0.42	**1.66**	0.02
36.1.7	AT5G13930.1	0.00006	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.61	0.56	**2.20**	0.01	0.79	0.38	**1.66**	0.04
40.15.16	AT5G13930.1	2E-84	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.70	0.51	**2.22**	0.01	0.75	0.33	**1.70**	0.02
59.23.9	AT5G13930.1	2E-84	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.45	0.57	**1.84**	0.02	0.87	0.67	**1.55**	0.02
6.4.5	AT5G13930.1	2E-36	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.57	0.11	**1.67**	0.02	1.78	0.12	**1.89**	0.02
63.19.6	AT5G13930.1	2E-69	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.34	0.55	**2.51**	0.01	1.00	1.00	**1.80**	0.02
63.8.11	AT5G13930.1	2E-69	CHS | Chalcone and	Chalcone	1.34	0.54	**2.15**	0.01	0.85	0.50	**1.56**	0.03

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative Identity	Lodgepole 1 dpi WD		Lodgepole 7 dpi WD		Jack 1 dpi WD		Jack 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
			stilbene synthase family protein	synthase-like								
7.18.5	AT5G13930.1	2E-69	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.25	0.61	**2.07**	0.01	0.89	0.65	**1.71**	0.02
31.20.16	AT5G13930.1	2E-69	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.44	0.54	**2.40**	0.01	0.76	0.36	**1.74**	0.01
34.9.15	AT5G13930.1	2E-69	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.46	0.50	**2.53**	0.01	0.84	0.33	**2.02**	0.02
40.12.10	AT5G13930.1	2E-69	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.36	0.53	**2.24**	0.01	0.80	0.34	**1.56**	0.03
56.8.8	AT5G13930.1	4E-71	CHS | Chalcone and stilbene synthase family protein	Chalcone synthase-like	1.51	0.52	**2.63**	0.00	0.72	0.39	**1.85**	0.03
16.8.10	AT1G74950.1	5E-16	JAZ2 | TIFY domain/Divergent CCT motif family protein	JAZ-like TF	2.01	0.11	**3.94**	0.03	2.15	0.10	**3.58**	0.01
62.10.9	AT1G74950.1	5E-16	JAZ2 | TIFY domain/Divergent CCT motif family protein	JAZ-like TF	2.42	0.15	**3.92**	0.04	2.13	0.18	**3.47**	0.01
65.20.3	AT5G13220.1	0.0000001	JAZ10 | jasmonate-zim-domain protein 10	JAZ-like TF	1.17	0.41	**3.72**	0.00	1.29	0.21	**2.86**	0.02
56.14.15	AT5G28650.1	1E-49	WRKY74 | WRKY DNA-binding protein 74	WRKY-like TF	0.96	0.93	**1.71**	0.02	1.61	0.05	**1.67**	0.02
67.8.5	AT5G64810.1	3E-17	WRKY51 | WRKY DNA-binding protein 51	WRKY-like TF	0.91	0.65	**2.66**	0.02	1.34	0.19	**2.93**	0.01

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative Identity	Lodgepole 1 dpi WD		Lodgepole 7 dpi WD		Jack 1 dpi WD		Jack 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
15.24.6	AT1G28370.1	1E-11	ERF11 | ERF domain protein 11	ERF-like	0.81	0.54	**1.58**	0.01	0.92	0.83	**2.65**	0.02
25.5.14	AT1G28370.1	1E-11	ERF11 | ERF domain protein 11	ERF-like	1.29	0.24	**1.57**	0.03	1.22	0.23	**1.60**	0.03
7.13.5	AT3G20310.1	1E-20	ERF7 | ethylene response factor 7	ERF-like	1.38	0.30	**1.68**	0.01	1.58	0.07	**2.02**	0.01
66.17.4	AT1G22640.1	1E-26	MYB3 | myb domain protein 3	MYB-like TF	1.09	0.85	**3.08**	0.03	0.77	0.39	**3.29**	0.02
46.23.12	AT1G77330.1	3E-88	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	ACC oxidase-like	1.11	0.63	**1.87**	0.02	1.34	0.20	**2.73**	0.01
4.4.16	AT3G11180.1	2E-53	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	ACC oxidase-like	1.32	0.51	**2.23**	0.02	1.54	0.09	**2.41**	0.01
28.2.5	AT1G77330.1	3E-88	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	ACC oxidase-like	1.73	0.08	**4.90**	0.01	1.74	0.24	**4.80**	0.01
40.14.9	AT3G11180.1	2E-53	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	ACC oxidase-like	1.25	0.32	**2.85**	0.01	1.23	0.20	**2.31**	0.04
56.18.8	AT3G11180.2	7E-40	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	ACC oxidase-like	1.03	0.90	**2.00**	0.03	1.11	0.64	**1.69**	0.03
3.21.4	AT3G53260.1	0	PAL2| phenylalanine ammonia-lyase 2	PAL-like	1.30	0.34	**3.19**	0.00	1.42	0.09	**2.54**	0.01

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative Identity	Lodgepole 1 dpi WD		Lodgepole 7 dpi WD		Jack 1 dpi WD		Jack 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
63.19.7	AT3G53260.1	0	PAL2| phenylalanine ammonia-lyase 2	PAL-like	1.24	0.46	**2.46**	0.00	1.45	0.09	**2.39**	0.01
35.4.12	AT3G10340.1	0	PAL2| phenylalanine ammonia-lyase 2	PAL-like	0.96	0.89	**2.95**	0.01	0.76	0.16	**1.71**	0.04

Table 2.5: Enrichment analysis of sequences differentially expressed in both lodgepole and jack pine inoculated vs. control seedlings under water deficit conditions. Table displays functional categories significantly over represented in sequences DE in both lodgepole and jack pine inoculated vs. control seedlings under well watered conditions using a hypergeometric distribution statistic.

Functional Category	Total PtGen elements in category	Elements DE in lodgepole and jack	Probability Density Function (Hypergeometric distribution)	Adjust P value (Bonferroni)
Miscellaneous cellular processes	1544	154	4.66E-27	1.26E-25
Secondary metabolism	3831	86	3.42E-12	9.24E-11
Stress response	1167	85	4.11E-08	1.11E-06
Protein processing	3463	97	3.64E-06	9.84E-05
Biodegradation of xenobiotics	34	8	4.35E-05	1.17E-03
Oxidative pentose phosphate path	51	9	1.51E-04	4.10E-04

2.3.3 Gene DE exclusively in lodgepole pine in response to *G. clavigera*

There were 146 and 133 sequences DE exclusively in lodgepole pine under well watered conditions, at 1 and 7 dpi respectively (Fig 2.3). Unlike sequences exclusively DE in 1 dpi jack pine, there were no well known defense associated genes DE exclusively in 1 dpi lodgepole pine for either well watered or water deficit treatments. A putative PAL sequence (spotID:62.16.17) was differentially upregulated at 1 dpi in well watered conditions (Table 2.6). Among the lodgepole genes uniquely upregulated in inoculated samples only at 7 dpi under well watered conditions were a basic secretory-like protein (spotID:4.22.10), a PR-like protein (spotID:29.23.12), and a putative NB-LRR (spotID:33.6.10) (Table 2.6). There was no known

defense associated TF differentially expressed exclusively in lodegpole inoculated vs control tissue under well watered conditions. Only one putative stress responsive MYB TF (spotID:35.18.6) was upregulated only in 7 dpi lodgepole pine under well watered conditions. Only 8 sequences are uniquely DE in lodgepole pine at both 7 and 1 dpi under well watered conditions (Fig 2.3). None of these sequences are typical defense associated genes.

Enrichment analysis of sequences DE only in lodgepole pine under well watered conditions revealed over representation of genes involved in secondary metabolism, protein processing, hormone metabolism, miscellaneous cellular processes, stress response, signalling, RNA regulation and amino acid metabolism (Table 2.7). Further examination of sequences involved in secondary metabolism revealed five sequences involved in flavonoid biosynthesis, and two sequences involved in phenylpropanoid biosynthesis (Table 2.3).

There were eight and 1755 sequences DE exclusively in inoculated vs. control lodgepole pine phloem tissue under water deficit conditions, at 1 and 7 dpi respectively (Fig 2.3). There was no typical defense associated sequences DE exclusively in lodgepole pine under water deficit conditions.

Table 2.6: Sequences differentially expressed on the PtGen2 microarray exclusively in lodgepole tissue between inoculated vs. control treatments under well watered conditions at 1 and/ or 7 dpi with *G. clavigera*. Sequence putative identity was predicted on the basis of sequence similarity to sequences within the TAIR database and within the NCBI nr database (not shown). FC: fold change. P: adjusted p-value.

PtGen2 spot ID	**TAIR ID**	**BLAST X E-value**	**Gene Symbols**	**Putative Identity**	Lodgepole 1 dpi WW		Lodgepole 7 dpi WW		Jack 1 dpi WW		Jack 7 dpi WW	
					FC	P	FC	P	FC	P	FC	P
DE at 1 dpi in lodgepole under well watered conditions												
63.17.16	AT3G53260.1	0	PAL2 | phenylalanine ammonia-lyase 2	PAL-like	**1.63**	0.03	0.55	0.18	1.36	0.25	1.06	0.81
DE at 7 dpi in lodgepole under well watered conditions												
33.6.10	AT1G27170.2	0.63	transmembrane receptors ATP binding	NB-LRR-like	1.20	0.30	**0.59**	0.04	1.09	0.75	0.87	0.52
29.23.12	AT1G78780.2	1E-75	pathogenesis-related family protein	PR-like	1.33	0.23	**2.96**	0.04	1.02	0.84	0.98	0.94
4.22.10	AT2G15220.1	5E-48	Plant basic secretory protein (BSP) family protein	BSP-like	1.11	0.56	**4.69**	0.04	1.01	0.94	3.22	0.06
35.18.6	AT3G13540.1	4E-54	MYB5 | myb domain protein	MYB-like TF	0.75	0.17	**0.49**	0.04	1.08	0.66	0.64	0.03

Table 2.7: Enrichment analysis of sequences differentially expressed only in lodgepole pine inoculated vs. control seedlings under well watered conditions. Table displays functional categories significantly over represented in sequences DE only in lodgepole pine inoculated vs. control seedlings under well watered conditions using a hypergeometric distribution statistic.

Functional Category	Total PtGen elements in category	Elements DE only in lodgepole	Probability Density Function (Hypergeometric distribution)	Adjust P value (Bonferroni)
Secondary metabolism	3831	8	2.40E-44	4.74E-43
Protein processing	3463	34	1.08E-18	2.17E-17
Hormone metabolism	767	1	4.82E-10	9.65E-09
Miscellaneous cellular processes	1544	15	5.68E-09	1.18E-07
Stress response	1167	9	2.14E-08	4.28E-07
Signaling	746	3	9.36E-08	1.87E-06
RNA regulation	1242	14	1.54E-06	3.8E-05
Amino acid metabolism	797	7	1.21E-05	2.42E-04
Transport	586	6	4.65E-04	9.30E-03

Among the known defense sequences DE uniquely in 7 dpi lodgepole pine under water deficit conditions were sequences encoding two DIR-like proteins (spotID:20.5.11, spotID:6.16.8), one putative basic secretory protein (spotID:51.6.15), 14 putative osmotins (spotID:33.14.1, spotID:62.4.7, spotID:67.16.4, spotID:66.24.6, spotID:64.14.3, spotID: 63.14.6, spotID:62.21.8, spotID:61.6.1, spotID:7.22.11, spotID:61.20.4, spotID:61.14.12, spotID:27.23.10, spotID:42.10.2, spotID:66.10.3) and six putative terpene synthases (spotID:61.23.12, spotID:63.12.7, spotID:14.10.13, spotID:63.24.6, spotID: 25.1.15, spotID:67.10.13) (Table 2.8). Transcriptional regulators differentially upregulated in 7 dpi lodgepole pine under water deficit conditions included a putative JA responsive JAZ-like TF

(spot ID:62.15.10, spotID:67.16.14, spotID:5.5.4), two putative defense associated ERF (spotID:62.6.15, spotID:66.5.11), as well as two putative defense associated b-ZIP TFs (spotID:56.2.6, spotID:29.16.10) (Table 2.8). Transcriptional regulators differentially down regulated in lodgepole at 7 dpi under water deficit conditions included two putative defense associated ERFs (spotID:49.18.7, 20.11.11), a putative bZIP TF involved in defense response to bacteria (spotID:56.2.6, spotID:29.16.10), along with three TFs putatively involved in drought stress response: an abscisic acid response factor ABF (spotID:52.22.3) and a putative SEUSS transcriptional co-regulator (SEU) (spot ID:47.16.12) (Table 2.8). A large number of sequences involved in the biosynthesis of JA are DE in 7 dpi lodgepole under water deficit conditions, including an upregulated putative lipoxygenase (LOX) (spotID:7.4.9), two downregulated putative allene oxidase synthase (AOS)(spotID:12.11.4, spotID:12.12.14), and a downregulated putative oxo-phytodeinoate (ODPA) reductase (spotID:50.4.9) (Table 2.8). A putative ACC oxidase (spotID:68.8.5) involved in the biosynthesis of ethylene, and a putative PAL (spotID:38.8.15) involved in the biosynthesis of SA are both differentially upregulated at 7 dpi under water deficit conditions exclusively in lodegpole pine seedlings. (Table 2.8).

Only four sequences were uniquely DE in lodgepole pine at both 7 and 1 dpi under water deficit conditions, none of which were typical defense associated genes. Unlike jack pine, very few sequences in lodgepole pine were DE across both 1 and 7 dpi for either well watered or water deficit conditions.

Enrichment analysis of sequences DE only in lodgepole pine under water deficit conditions revealed overrepresentation of genes involved in secondary metabolism, photosynthesis, miscellaneous cellular processes, metal handling, transport and the oxidative pentose phosphate pathway (Table 2.9).

Table 2.8: Sequences differentially expressed on the PtGen2 microarray exclusively in lodgepole phloem tissue between inoculated vs. control treatments under water deficit conditions at 1 and/ or 7 dpi with *G. clavigera.* Sequence putative identity was predicted on the basis of sequence similarity to sequences within the TAIR database and within the NCBI nr database (not shown). FC: fold change. P: adjusted p-value.

PtGen2 spot ID	**TAIR ID**	**BLAST X E-value**	**Gene Symbols**	**Putative Identity**	Lodgepole 1 dpi WD		Lodgepole 7 dpi WD		Jack 1 dpi WD		Jack 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
DE at 7 dpi in lodgepole under water deficit conditions												
6.16.8	AT1G58170.1	6E-24	Disease resistance-responsive (dirigent-like protein) family protein	DIR-like	1.11	0.64	**1.50**	0.04	1.32	0.33	1.07	0.76
20.5.11	AT2G21100.1	2E-46	Disease resistance-responsive (dirigent-like protein) family protein	DIR-like	0.71	0.51	**2.12**	0.01	0.68	0.34	0.70	0.15
51.6.15	AT2G15220.1	6E-59	Plant basic secretory protein (BSP) family protein	BSP-like	0.87	0.49	**1.61**	0.04	0.66	0.21	2.55	0.06
67.16.4	AT4G11650.1	1E-75	OSM34 | osmotin	Osmotin-like	1.67	0.47	**12.00**	0.02	6.67	0.17	11.91	0.06
66.24.6	AT4G11650.1	1E-75	OSM34 | osmotin	Osmotin-like	1.12	0.84	**3.27**	0.01	NA	NA	5.19	0.05
64.14.3	AT4G11650.1	1E-75	OSM34 | osmotin	Osmotin-like	1.55	0.54	**11.70**	0.02	3.34	0.26	8.12	0.06
63.14.6	AT4G11650.1	1E-75	OSM34 | osmotin	Osmotin-like	1.09	0.89	**10.70**	0.02	NA	NA	6.78	0.05
62.21.8	AT4G11650.1	1E-75	OSM34 | osmotin	Osmotin-like	1.14	0.67	**12.70**	0.02	1.27	0.39	9.58	0.06
61.6.1	AT4G11650.1	1E-75	OSM34 | osmotin	Osmotin-like	1.34	0.58	**8.69**	0.01	1.06	0.77	6.31	0.05
7.22.11	AT4G11650.1	4E-67	OSM34 | osmotin	Osmotin-like	1.64	0.35	**1.90**	0.02	1.14	0.78	3.03	0.08

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative Identity	Lodgepole 1 dpi WD		Lodgepole 7 dpi WD		Jack 1 dpi WD		Jack 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
61.20.4	AT4G11650.1	1E-75	OSM34 | osmotin	Osmotin-like	0.96	0.83	**3.69**	0.02	1.20	0.27	2.53	0.06
61.14.12	AT4G11650.1	1E-75	OSM34 | osmotin	Osmotin-like	1.00	1.00	**5.28**	0.03	1.24	0.63	4.74	0.11
27.23.10	AT4G11650.1	1E-75	OSM34 | osmotin	Osmotin-like	2.78	0.36	**14.90**	0.02	NA	NA	8.77	0.06
42.10.2	AT4G11650.1	1E-75	OSM34 | osmotin	Osmotin-like	NA	NA	**8.55**	0.03	NA	NA	5.90	0.09
66.10.3	AT4G11650.1	1E-26	OSM34 | osmotin	Osmotin-like	0.97	0.94	**5.67**	0.01	1.04	0.92	3.23	0.11
33.14.1	AT4G11650.1	4E-67	OSM34 | osmotin	Osmotin-like	2.18	0.20	**3.08**	0.02	1.41	0.75	7.21	0.07
62.4.7	AT4G11650.1	1E-75	OSM34 | osmotin	Osmotin-like	0.98	0.93	**2.20**	0.03	0.56	0.08	1.18	0.74
61.23.12	AT1G61680.2	4E-20	TPS14 | terpene synthase 14	Terpene Synthase-like	1.06	0.83	**1.74**	0.01	1.08	0.76	1.03	0.81
62.24.6	AT1G70080.1	4.5	Terpenoid cyclases/Protein prenyltransferases superfamily protein	Terpene Synthase-like	1.67	0.18	**2.49**	0.00	NA	NA	2.64	0.08
25.1.15	AT4G02780.1	1E-106	CPS1 | Terpenoid cyclases/Protein prenyltransferases superfamily protein	Terpene Synthase-like	1.05	0.90	**2.67**	0.01	0.83	0.35	6.00	0.06
67.10.13	AT4G02780.1	1E-106	CPS1 | Terpenoid cyclases/Protein prenyltransferases superfamily protein	Terpene Synthase-like	0.91	0.82	**0.42**	0.00	0.65	0.40	0.75	0.41
63.12.7	AT2G41710.1	2E-66	Integrase-type DNA-binding superfamily protein	Terpene Synthase-like	1.69	0.22	**3.74**	0.00	2.32	0.12	1.01	0.98
14.10.13	AT1G61680.2	4E-20	TPS14 | terpene synthase 14	Terpene Synthase-like	1.57	0.20	**2.76**	0.03	3.20	0.13	2.53	0.06

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative Identity	Lodgepole 1 dpi WD		Lodgepole 7 dpi WD		Jack 1 dpi WD		Jack 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
62.15.10	AT5G13220.1	0.0000001	JAZ10 | jasmonate-zim-domain protein 10	JAZ-like TF	1.24	0.30	**3.36**	0.01	1.39	0.15	1.43	0.09
67.16.14	AT1G74950.1	0.005	JAZ2 | TIFY domain/Divergent CCT motif family protein	JAZ-like TF	1.58	0.49	**1.62**	0.02	5.95	0.10	2.69	0.06
5.5.4	AT5G13220.1	0.0000001	JAZ10 | jasmonate-zim-domain protein 10	JAZ-like TF	1.05	0.83	**3.18**	0.02	1.34	0.14	1.46	0.43
62.6.15	AT4G17500.1	0.00003	ERF-1 | ethylene responsive element binding factor 1	ERF-like	0.90	0.83	**1.82**	0.03	1.34	0.38	1.29	0.19
56.2.6	AT5G06960.2	4E-110	TGA5 | OCS-element binding factor 5	bZIP-like TF	0.88	0.58	**2.05**	0.01	1.07	0.77	1.47	0.05
29.16.10	AT5G65210.5	9E-50	TGA1 | bZIP transcription factor family protein	bZIP-like TF	1.10	0.83	**1.87**	0.04	0.89	0.48	1.41	0.05
66.5.11	AT4G17500.1	5E-30	ERF-1 | ethylene responsive element binding factor 1	ERF-like	0.96	0.88	**2.02**	0.01	1.41	0.23	1.42	0.08
49.18.7	AT3G23240.1	1E-13	ERF1| ethylene response factor 1	ERF-like	0.58	0.24	**0.53**	0.01	0.54	0.28	1.22	0.49
20.11.11	AT1G28360.1	3E-20	ERF12 | ERF domain protein 12	ERF-like	1.13	0.51	**0.56**	0.04	0.83	0.27	0.71	0.10
52.22.3	AT1G45249.1	2E-21	ABF2 | abscisic acid responsive elements-binding factor 2	ABF-like TF	0.79	0.41	**0.55**	0.02	0.78	0.25	0.88	0.54
47.16.12	AT1G43850.2	4E-73	SEU | SEUSS transcriptional co-regulator	SUESS-like TF	0.98	0.95	**0.53**	0.03	0.80	0.40	0.85	0.41
7.4.9	AT1G55020.1	0	LOX1 | lipoxygenase 1	LOX-like	1.11	0.62	**1.25**	0.18	1.19	0.49	1.63	0.02
12.12.14	AT5G42650.1	5E-128	AOS | allene oxide synthase	AOS-like	0.94	0.86	**0.39**	0.01	0.73	0.35	0.63	0.20
12.11.4	AT5G42650.1	5E-128	AOS| allene oxide synthase	AOS-like	1.05	0.91	**0.38**	0.01	0.48	0.06	0.75	0.26
50.4.9	AT3G05390.1	0.64	molecular_function	ODPA-like	1.05	0.87	**0.57**	0.03	1.18	0.30	1.37	0.45

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative Identity	Lodgepole 1 dpi WD		Lodgepole 7 dpi WD		Jack 1 dpi WD		Jack 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
			unknown									
68.8.5	AT4G26200.1	2E-26	ACS7 | 1-amino-cyclopropane-1-carboxylate synthase 7	ACC-oxidase like	0.94	0.87	**1.66**	0.02	0.98	0.92	1.39	0.41
38.8.15	AT3G30775.1	2E-101	ERD5 | Methylenetetrahydrofolate reductase family protein	PAL-like	0.98	0.96	**0.46**	0.04	0.66	0.22	0.71	0.36

Examination of the sequences involved in secondary metabolism reveals 24 sequences involved in flavonoid biosynthesis, 11 sequences involved in isoprenoids biosynthesis and six sequences involved in biosynthesis of phenylpropanoids (Table 2.3).

Table 2.9: Enrichment analysis of sequences differentially expressed only lodgepole pine inoculated vs. control seedlings under water deficit conditions. Table displays functional categories significantly over represented in sequences DE only in lodgepole pine inoculated vs. control seedlings under water deficit conditions using a hypergeometric distribution statistic.

Functional Category	**Total PtGen elements in category**	**Elements DE in lodgepole and jack**	**Probability Density Function (Hypergeometric distribution)**	**Adjust P value (Bonferroni)**
Secondary metabolism	3831	46	6.81E-50	1.98E-48
Photosynthesis	526	70	8.57E-13	2.49E-11
Miscellaneous cellular processes	1544	125	5.36E-07	1.56E-05
Metal handling	49	13	9.92E-07	2.88E-05
Transport	586	52	1.50E-04	3.04E-03
Oxidative pentose phosphate pathway	51	9	1.05E-03	03.06E-02

2.3.4 Genes DE exclusively in jack pine in response to *G. clavigera*

Venn diagrams comparing genes DE in lodegpole versus jack pine under either well watered or watered deficit conditions revealed sequences DE uniquely in one species or the other. In well watered conditions, 520 sequences were DE exclusively in 1 dpi jack pine and 664 sequences were DE exclusively in 7 dpi jack pine (Fig 2.3). Under well watered conditions, DE sequences with strong similarity to known defense associated genes included sequences encoding five putative α-pinene synthases (spotID:30.12.2, spotID:66.9.12, spotID:15.14.3, spotID:68.5.1,

spotID:63.12.7), three putative sesquiterpene synthases (spotID:67.10.13, spotID:47.2.7, spotID:67.20.8) and a putative defensin (spotID:3.23.11) (Table 2.10). Analysis of defense related transcription factors in this set of DE sequences revealed an upregulated putative ethylene activated WRKY (spotID:38.15.12). Two putative LOX sequences (spotID:69.6.13, and spotID:48.17.7) involved in the biosynthesis of JA were upregulated exclusively in 1 dpi well watered jack pine as well (Table 2.10). Known defense genes upregulated under well watered conditions at 7 dpi included two putative terpene synthases (spotID:25.2.15, spotID:62.24.6), two putative chitinases (spotID: 66.24.4, spotID:64.7.12), four osmotins (spotID:7.22.11, spotID:67.16.12, spotID:33.10.1, spotID:67.1.6), four basic secretory proteins (spotID:1.23.8, spotID:2.7.12, spotID:1.23.8, spotID:19.22.11), and seven chalcone/stilbene synthases (spotID:3.24.15, spotID:32.14.15, spotID:35.13.12, spotID:69.2.2, spotID:64.2.4, spotID:63.8.3, spotID:49.2.4, spotID:69.2.2) which encode enzymes of the phenolic biosynthesis pathway (Table 2.10). Potential regulators upregulated only in inoculated vs. control 7 dpi jack pine under well watered conditions included a putative defense associated sirtuin TF (spotID:63.3.4) and putative wounding responsive MYB TF (spot ID:22.2.14), as well as three ACC oxidases involved in the biosynthesis of ethylene (spotID:46.23.12, spotID:4.4.16, spotID:40.14.9, spotID:56.18.8) (Table 2.10). There were 600 sequences DE at both 1 and 7 dpi in only jack pine under well watered and water deficit conditions, respectively (Fig 2.3). Among the defense associated sequences upregulated under well watered conditions were four sequences encoding three different putative terpene synthases - two putative α-pinene synthases (spotID: 62.16.9, spotID:69.22.15), a putative farnesene synthase (spotID:14.10.13), and a putative selinene synthase (spotID:64.4.2) - three putative osmotins (spotID: 63.17.3, spotID:62.4.7, spotID:4.8.10), and one putative chitinase (spotID: 63.24.13) (Table 2.10). Defense associated

transcriptional regulators upregulated only in under well watered conditions uniquely in jack pine at both 1 and 7 dpi included three putative JAZ like TF (spot ID:16.8.10,62.10.9, spotID:62.23.13), and one putative MYB TF (spotID:3.22.16). (Table 2.10). Two PAL sequences, (spotID:3.21.4, spotID:63.19.7) were upregulated under well watered conditions only in jack pine at both 1 and 7dpi, along with one ODPA reductase gene (spotID: 50.4.9) involve in the biosynthesis of JA (Table 2.10).

Enrichment analysis of sequences DE only in jack pine under well watered conditions revealed overrepresentation of genes involved in secondary metabolism, miscellaneous cellular processes including peroxidase activity, the oxidative pentose phosphate pathway, transport, photosynthesis and cell wall biosynthesis (Table 2.11). Examination of sequences involved in secondary metabolism reveals 38 sequences involved in isoprenoid biosynthesis, 24 sequences involved in flavonoid biosynthesis, and 11 sequences involved in phenylpropanoid biosynthesis (Table 2.3).

Under water deficit conditions, 109 sequences were DE exclusively in 1 dpi jack pine and 294 genes were DE exclusively in 7 dpi jack pine (Fig 2.3). Among the 62 genes upregulated only in 1 dpi jack pine under water deficit is a wounding responsive NAC containing TF (spotID:53.21.9) (Table 2.12). Under water deficit conditions, very few defense associated genes are DE only in inoculated vs. control 7 dpi jack pine; these include one putative basic secretory protein (spotID:61.9.1), two PR-5 family proteins osmotin-like proteins (spotID:63.11.13, spotID:14.19.10), and no defense associated transcriptional regulators. There were 80 sequences DE at both 1 and 7 dpi only in jack pine under well watered and water deficit conditions, respectively (Fig 2.3). Among the known defense response genes upregulated in only jack pine at both 1 and 7dpi under water deficit conditions was a NB-LRR PR protein (spotID:54.5.11) (Table 2.12).

Table 2.10: Sequences differentially expressed on the PtGen2 microarray exclusively in jack pine phloem tissue between inoculated vs. control treatments under well watered conditions at 1 dpi and/ or 7 dpi with *G. clavigera.* Sequence putative identity was predicted on the basis of sequence similarity to sequences within the TAIR database and within the NCBI nr database (not shown). FC: fold change. P: adjusted p-value.

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative Identity	Lodgepole 1 dpi WW		Lodgepole 7 dpi WW		Jack 1 dpi WW		Jack 7 dpi WW	
					FC	P	FC	P	FC	P	FC	P
DE at 1 dpi in jack under well watered conditions												
15.14.3	AT1G70080.1	8E-7	Terpenoid cyclases/Protein prenyltransferases superfamily protein	Pinene Synthase-like	1.23	0.62	0.55	0.35	**2.12**	0.03	1.10	0.86
30.12.2	AT1G70080.1	8E-7	Terpenoid cyclases/Protein prenyltransferases superfamily protein	Pinene Synthase-like	1.62	0.07	0.81	0.55	**2.10**	0.01	1.16	0.64
66.9.12	AT2G24210.1	1E-64	TPS10 | terpene synthase 10	Pinene Synthase-like	1.69	0.09	0.87	0.42	**1.53**	0.05	0.80	0.21
68.5.1	AT4G16730.1	5E-59	TPS02 | terpene synthase	Pinene Synthase-like	1.43	0.28	0.66	0.55	**2.62**	0.00	1.21	0.69
63.12.7	AT2G41710.1	2E-66	Integrase-type DNA-binding superfamily protein	Pinene Synthase-like	1.54	0.10	1.97	0.49	**2.26**	0.02	0.77	0.24
67.10.13	AT4G02780.1	1E-106	CPS1 | Terpenoid cyclases/Protein prenyltransferases superfamily protein	Sesquiterpene Synthase-like	1.24	0.26	0.69	0.15	**0.58**	0.03	1.29	0.23
47.2.7	AT4G02780.1	4E-138	CPS1 | Terpenoid cyclases/Protein prenyltransferases superfamily protein	Sesquiterpene Synthase-like	1.58	0.25	0.79	0.35	**2.38**	0.01	0.61	0.05
67.20.8	AT4G02780.1	4E-138	CPS1 | Terpenoid	Sesquiterpen	1.86	0.10	0.85	0.53	**2.73**	0.01	0.67	0.12

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative Identity	Lodgepole 1 dpi WW		Lodgepole 7 dpi WW		Jack 1 dpi WW		Jack 7 dpi WW	
					FC	P	FC	P	FC	P	FC	P
			cyclases/Protein prenyltransferases superfamily protein	e Synthase-like								
3.23.11	AT2G02120.1	2E-10	PDF2.1 | Scorpion toxin-like knottin superfamily protein	Defensin-like	1.18	0.32	0.63	0.10	**1.78**	0.01	0.95	0.91
38.15.12	AT1G62300.1	1E-36	WRKY6 | WRKY family transcription factor	WRKY-like TF	2.16	0.06	0.86	0.53	**2.68**	0.01	0.66	0.04
48.17.7	AT1G72520.1	0	LOX4 | PLAT/LH2 domain-containing lipoxygenase family protein	LOX-like	1.47	0.12	1.31	0.26	**1.69**	0.01	1.32	0.08
69.6.13	AT3G22400.1	0	LOX5 | PLAT/LH2 domain-containing lipoxygenase family protein	LOX-Like	1.47	0.11	1.17	0.65	**1.83**	0.03	1.04	0.80
DE at 7 dpi in jack under well watered conditions												
62.24.6	AT1G70080.1	4.5	Terpenoid cyclases/Protein prenyltransferases superfamily protein	Terpene Synthase-like	2.43	0.18	3.24	0.10	NA	NA	**7.55**	0.04
25.1.15	AT4G02780.1	1E-106	CPS1 | Terpenoid cyclases/Protein prenyltransferases superfamily protein	Terpene Synthase-like	0.77	0.10	2.88	0.09	0.96	0.83	**4.50**	0.02
64.7.12	AT2G43590.1	2E-40	Chitinase family protein	Chitinase-like	1.13	0.55	1.56	0.07	1.10	0.37	**1.84**	0.02
66.24.4	AT3G12500.1	6E-104	HCHIB | basic chitinase	Chitinase-like	2.51	0.08	3.45	0.05	2.65	0.05	**3.25**	0.00
7.22.11	AT4G11650.1	4E-67	OSM34 | osmotin 34	Osmotin-like	1.54	0.10	2.53	0.07	1.59	0.31	**4.74**	0.00
67.16.12	AT4G11650.1	2E-64	OSM34 | osmotin 34	Osmotin-like	1.06	0.82	2.37	0.02	1.26	0.25	**1.71**	0.02
33.10.1	AT4G11650.1	1E-75	OSM34 | osmotin 34	Osmotin-like	0.96	0.84	1.31	0.27	1.31	0.50	**2.11**	0.01

PtGen2 spot ID	TAIR ID	BLAS TX E-value	Gene Symbols	Putative Identity	Lodgepole 1 dpi WW		Lodgepole 7 dpi WW		Jack 1 dpi WW		Jack 7 dpi WW	
					FC	P	FC	P	FC	P	FC	P
67.1.6	AT4G11650.1	4E-55	OSM34 | osmotin 34	Osmotin-like	0.85	0.31	3.45	0.03	0.76	0.14	**4.73**	0.01
2.7.12	AT2G15220.1	6E-59	Plant basic secretory protein (BSP) family protein	BSP-like	1.37	0.23	2.68	0.08	1.28	0.19	**2.73**	0.01
1.23.8	AT2G15220.1	5E-48	Plant basic secretory protein (BSP) family protein	BSP-like	1.18	0.53	2.09	0.14	1.28	0.28	**2.82**	0.01
19.22.11	AT2G15220.1	6E-59	Plant basic secretory protein (BSP) family protein	BSP-like	NA	NA	3.71	0.12	NA	NA	**4.93**	0.03
69.2.2	AT5G05270.2	6E-58	Chalcone-flavanone isomerase family protein	Chalcone Synthase-like	1.02	0.91	1.19	0.28	1.22	0.34	**1.68**	0.02
64.2.4	AT5G13930.1	2E-32	Chalcone-flavanone isomerase family protein	Chalcone Synthase-like	0.84	0.43	2.51	0.06	1.23	0.55	**4.00**	0.01
63.8.3	AT5G13930.1	2E-36	Chalcone-flavanone isomerase family protein	Chalcone Synthase-like	0.74	0.08	2.32	0.10	1.13	0.70	**4.13**	0.00
49.2.4	AT5G13930.1	2E-36	Chalcone-flavanone isomerase family protein	Chalcone Synthase-like	0.79	0.14	1.67	0.20	1.02	0.92	**2.57**	0.01
32.14.15	AT5G13930.1	3E-15	Chalcone-flavanone isomerase family protein	Chalcone Synthase-like	0.95	0.79	1.43	0.37	1.05	0.75	**2.90**	0.01
3.24.15	AT5G13930.1	2E-36	Chalcone-flavanone isomerase family protein	Chalcone Synthase-like	0.71	0.10	1.96	0.12	1.10	0.75	**3.39**	0.01
35.13.12	AT5G05270.2	6E-58	Chalcone-flavanone isomerase family protein	Chalcone Synthase-like	0.95	0.79	1.33	0.19	1.47	0.15	**1.77**	0.01
63.3.4	AT5G09230.7	6E-118	SRT2 | sirtuin 2	Situin-like TF	1.12	0.52	1.25	0.28	1.07	0.62	**1.56**	0.01
22.2.14	AT5G35550.1	6E-43	MYB123 | Duplicated homeodomain-like	MYB-like TF	1.02	0.93	1.18	0.60	1.28	0.05	**1.68**	0.02

PtGen2 spot ID	TAIR ID	BLAS TX E-value	Gene Symbols	Putative Identity	Lodgepole 1 dpi WW		Lodgepole 7 dpi WW		Jack 1 dpi WW		Jack 7 dpi WW	
					FC	P	FC	P	FC	P	FC	P
			superfamily protein									
46.23.12	AT1G77330.1	3E-88	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	ACC oxidase-like	1.22	0.24	1.64	0.15	1.21	0.20	**2.01**	0.03
4.4.16	AT3G11180.1	2E-53	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	ACC oxidase-like	2.03	0.05	1.68	0.11	1.36	0.10	**2.05**	0.00
40.14.9	AT3G11180.1	2E-53	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	ACC oxidase-like	1.07	0.73	1.53	0.31	1.32	0.08	**1.80**	0.04
56.18.8	AT3G11180.2	7E-40	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	ACC oxidase-like	1.29	0.34	1.76	0.14	1.11	0.35	**1.70**	0.02
DE at 1 and 7 dpi in jack under well watered conditions												
62.16.9	AT2G41710.1	2E-66	Integrase-type DNA-binding superfamily protein	Pinene Synthase-like	1.23	0.27	0.97	0.93	**2.06**	0.01	**1.59**	0.01
69.22.15	AT2G41710.1	2E-66	Integrase-type DNA-binding superfamily protein	Pinene Synthase-like	1.31	0.37	0.75	0.50	**1.97**	0.04	**0.55**	0.02
14.10.13	AT1G61680.2	4E-20	TPS14 | terpene synthase 14	Farnesene Synthase-like	2.44	0.09	1.89	0.30	**2.88**	0.00	**2.59**	0.01
64.4.2	AT3G25810.1	4E-63	Terpenoid cyclases/Protein prenyltransferases superfamily protein	Selinene Synthase like	2.16	0.39	6.52	0.09	**2.39**	0.01	**11.50**	0.00
62.4.7	AT4G11650.1	1E-75	OSM34 | osmotin 34	Osmotin-like	0.65	0.11	3.18	0.08	**0.57**	0.01	**2.21**	0.02
63.17.3	AT4G11650.1	4E-55	OSM34 | osmotin 34	Osmotin-like	1.34	0.34	3.85	0.02	**1.90**	0.04	**8.29**	0.00
4.8.10	AT4G11650.1	1E-75	OSM34 | osmotin 34	Osmotin-like	1.30	0.26	3.14	0.04	**1.51**	0.03	**2.95**	0.00

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative Identity	Lodgepole 1 dpi WW		Lodgepole 7 dpi WW		Jack 1 dpi WW		Jack 7 dpi WW	
					FC	P	FC	P	FC	P	FC	P
63.24.13	AT3G12500.1	2E-60	HCHIB | basic chitinase	Chitinase-like	0.95	0.79	1.40	0.16	**1.70**	0.01	**1.58**	0.04
62.10.9	AT1G74950.1	5E-16	JAZ2 | TIFY domain/Divergent CCT motif family protein	JAZ-like TF	2.03	0.07	1.67	0.16	**2.87**	0.01	**3.32**	0.00
16.8.10	AT1G74950.1	5E-16	JAZ2 | TIFY domain/Divergent CCT motif family protein	JAZ-like TF	1.94	0.07	1.72	0.14	**2.76**	0.01	**3.05**	0.00
62.23.13	AT1G74950.1	6E-10	JAZ2 | TIFY domain/Divergent CCT motif family protein	JAZ-like TF	NA	NA	1.65	0.25	**5.53**	0.00	**3.59**	0.00
3.22.16	AT5G67300.1	6E-49	MYB44 | myb domain protein r1	MYB-like TF	1.14	0.55	0.99	0.95	**2.04**	0.01	**1.59**	0.05
63.19.7	AT3G53260.1	0	PAL2 | phenylalanine ammonia-lyase 2	PAL-like	1.38	0.10	1.41	0.23	**1.87**	0.01	**1.58**	0.02
3.21.4	AT3G53260.1	0	PAL2 | phenylalanine ammonia-lyase 2	PAL-like	1.45	0.06	1.67	0.17	**1.75**	0.02	**1.60**	0.03
50.4.9	AT3G05390.1	0.64	molecularfunction unknown	ODPA-like	0.97	0.88	0.80	0.61	**1.59**	0.02	**1.93**	0.00

Table 2.11: Enrichment analysis of sequences differentially expressed only jack pine inoculated vs. control seedlings under well watered conditions. Table displays functional categories significantly over represented in sequences DE only in jack pine inoculated vs. control seedlings under well watered conditions using a hypergeometric distribution statistic.

Functional Category	Total PtGen elements in category	Elements DE only in jack	Probability Density Function (Hypergeometric distribution)	Adjust P value (Bonferroni)
Secondary metabolism	3831	86	3.78E-26	1.06E-24
Miscellaneous	1544	157	4.00E-16	1.12E-14
Oxidative pentose phosphate pathway	51	14	2.40E-07	6.73E-06
Transport	586	52	1.01E-04	2.84E-03
Photosynthesis	526	44	9.03E-04	2.53E-02
Cell wall	503	42	1.16E-03	3.26E-02

Enrichment analysis of sequences DE only in jack pine under water deficit conditions revealed overrepresentation of genes involved in secondary metabolism, protein processing, transport, miscellaneous cellular processes including peroxidase activity, and cell wall biosynthesis (Table 2.13). Examination of sequences involved in secondary metabolism reveals 14 sequences involved in isoprenoid biosynthesis, four sequences involved in phenlypropanoid biosynthesis, and one sequence involved in flavonoid biosynthesis (Table 2.3).

Table 2.12: Sequences differentially expressed on the PtGen2 microarray exclusively in jack pine phloem tissue between inoculated vs. control treatments under water deficit conditions at 1 and/ or 7 dpi with *G. clavigera.* Sequence putative identity was predicted on the basis of sequence similarity to sequences within the TAIR database and within the NCBI nr database (not shown). FC: fold change. P: adjusted p-value.

PtGen2 spot ID	TAIR ID	BLASTX E-value	Gene Symbols	Putative Identity	Lodgepole 1 dpi WD		Lodgepole 7 dpi WD		Jack 1 dpi WD		Jack 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
DE at 1 dpi in jack under water deficit conditions												
53.21.9	AT3G10500.1	1E-75	NAC053 | NAC domain containing protein 53	NAC-like TF	1.81	0.09	1.05	0.79	**2.20**	0.03	1.19	0.16
DE at 7 dpi in jack under water deficit conditions												
61.9.1	AT2G15220.1	5E-48	Plant basic secretory protein (BSP) family protein	BSP-like	1.13	0.58	1.29	0.33	1.29	0.41	**1.70**	0.04
63.11.13	AT1G75040.1	4E-76	PR5| pathogenesis-related gene 5	PR5-like	1.09	0.73	1.90	0.06	1.56	0.14	**3.74**	0.01
14.19.10	AT1G75040.1	4E-76	PR5| pathogenesis-related gene 5	PR5-like	1.13	0.59	1.39	0.07	1.56	0.21	**2.24**	0.03
54.5.11	AT2G14080.1	3E-6	Disease resistance protein (TIR-NBS-LRR class) family	NB-LRR-like	0.88	0.64	1.28	0.29	**1.78**	0.03	**1.74**	0.04

Table 2.13: Enrichment analysis of sequences differentially expressed only in jack pine inoculated vs. control seedlings under water deficit conditions. Table displays functional categories significantly over represented in sequences DE only jack pine inoculated vs. control seedlings under water deficit conditions using a hypergeometric distribution statistic.

Functional Category	Total PtGen elements in category	Elements DE in lodgepole and jack	Probability Density Function (Hypergeometric distribution)	Adjust P value (Bonferroni)
Secondary metabolism	3831	23	6.60E-09	1.58E-07
Protein processing	3463	24	9.95E-07	2.39E-05
Transport	586	20	0.000348	0.00834
Miscellaneous	1544	39	0.000419	0.0100
Cell wall	503	17	0.000990	0.0238

2.3.5 Temporal and qualitative differences between lodgepole and jack pine response to *G. clavigera*

Four-way Venn diagrams revealed sequences whose patterns of DE were temporally distinct between species, i.e. DE earlier or later in one species compared to the other. There were 117 sequences DE in lodgepole pine at 1 dpi and later in jack pine at 7 dpi under well watered conditions (Fig 2.3). Among defense associated sequences DE early under well watered conditions in lodgepole and later in jack pine were five putative osmotin-like sequences (spotID:63.22.5, spotID:7.7.6, spotID: 17.19.1, spotID:63.16.13, spotID:55.22.13, spotID:31.8.10) and a putative DIR-like protein. Among defense related transcriptional regulators DE at under well watered conditions in the same category was a putative ERF (spotID:7.13.5) and a putative JAZ-like TF (spotID:67.23.2) (Table 2.14). One upregulated ACC oxidase

sequence (spotID:28.2.5), involved in the biosynthesis of ethylene was also DE in lodgepole pine at 1 dpi and in jack pine at 7 dpi under well watered conditions (Table 2.14). There were 24 sequences DE in jack pine at 1 dpi but later in lodgepole at 7 dpi under well watered conditions (Fig 2.3). No well known defense associated genes, including genes involved in defense hormone biosynthesis or transcriptional regulation were found in this category under well watered conditions.

There were 7 sequences DE in lodgepole at 1 dpi and later in jack pine at 7 dpi under water deficit conditions (Fig 2.3). No defense associated transcriptional regulators, nor hormone biosynthesis and signaling genes were DE under water deficit conditions early at 1 dpi in lodgepole pine and later at 7 dpi in jack pine. There were 79 sequences DE in jack pine at 1 dpi but later in lodgepole at 7 dpi under water deficit conditions (Fig 2.3). Among the sequences upregulated early in jack and later in lodgepole under water deficit conditions were two putative DIR-like PR proteins (spotID:39.23.11, spotID:44.9.1), one putative chitinase (spotID:4.20.9), and two α-pinene synthases (spotID:68.2.15, spotID:65.9.8) (Table 2.15). No defense associated transcriptional regulators, nor hormone biosynthesis and signaling genes were DE under water deficit conditions early at 1 dpi in jack pine and later at 7 dpi in lodgepole pine.

Table 2.14: Sequences differentially expressed on the PtGen2 microarray in lodgepole and jack pine phloem tissue between inoculated vs. control treatments under well watered conditions at contrasting time points between species, either 1 or 7 dpi with *G. clavigera*. Sequence putative identity was predicted on the basis of sequence similarity to sequences within the TAIR database and within the NCBI nr database (not shown). FC: fold change. P: adjusted p-value.

PtGen2 spot ID	**TAIR ID**	**BLAST X E-value**	**Gene Symbols**	**Putative Identity**	Lodgepole 1 dpi WW		Lodgepole 7 dpi WW		Jack 1 dpi WW		Jack 7 dpi WW	
					FC	P	FC	P	FC	P	FC	P
DE at 1 dpi in lodgepole and at 7 dpi in jack pine												
63.22.5	AT1G75040.1	4E-76	PR5 | pathogeneis-related gene 5	osmotin- like	**1.02**	0.89	2.96	0.09	1.33	0.04	**3.81**	0.00
7.7.6	AT4G11650.1	0.00	OSM34 | osmotin 34	osmotin-like	**0.41**	0.01	3.25	0.07	1.13	0.50	**6.23**	0.01
17.19.1	AT4G11650.1	0.00	OSM34 | osmotin 34	osmotin-like	**0.41**	0.02	2.47	0.06	1.14	0.53	**4.63**	0.01
63.16.13	AT4G11650.1	0.00	OSM34 | osmotin 34	osmotin-like	**0.40**	0.02	3.82	0.03	1.31	0.35	**8.26**	0.00
55.22.13	AT4G11650.1	0.00	OSM34 | osmotin 34	osmotin-like	**0.56**	0.02	6.43	0.01	1.26	0.36	**12.61**	0.00
31.8.10	AT4G11650.1	0.00	OSM34 | osmotin 34	osmotin-like	**0.51**	0.02	3.68	0.03	0.94	0.73	**6.29**	0.00
3.17.8	AT5G42510.1	0.00	Disease resistance-responsive	DIR-like	**1.57**	0.04	4.05	0.01	1.45	0.09	**4.16**	0.00
67.23.2	AT1G74950.1	9E-14	JAZ2 | TIFY domain/Divergent CCT motif family protein	JAZ-like TF	**1.94**	0.03	1.71	0.11	1.38	0.07	**1.82**	0.03
7.13.5	AT3G20310.1	1E-20	ERF7 | ethylene response factor 7	ERF-like	**1.53**	0.05	1.40	0.09	1.45	0.02	**2.03**	0.01
28.2.5	AT1G77330.1	3E-88	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	ACC oxidase	**1.69**	0.03	2.57	0.15	1.35	0.15	**2.81**	0.03

Table 2.15: Sequences differentially expressed on the PtGen2 microarray in lodgepole and jack pine phloem tissue between inoculated vs. control treatments under water deficit conditions at contrasting time points between species, either at 1 or 7 dpi with *G. clavigera*. Sequence putative identity was predicted on the basis of sequence similarity to sequences within the TAIR database and within the NCBI nr database (not shown). FC: fold change. P: adjusted p-value.

PtGen2 spot ID	**TAIR ID**	**BLAST X E-value**	**Gene Symbols**	**Putative Identity**	Lodgepole 1 dpi WD		Lodgepole 7 dpi WD		Jack 1 dpi WD		Jack 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
DE at 1 dpi in jack and at 7 dpi in lodgepole pine under water deficit conditions												
44.9.1	AT1G64160.1	9E-47	Disease resistance-responsive (dirigent-like protein) family protein	DIR-like	1.62	0.17	**1.98**	0.01	**2.30**	0.03	2.03	0.05
39.23.11	AT1G64160.1	1E-11	Disease resistance-responsive (dirigent-like protein) family protein	DIR-like	2.13	0.09	**2.25**	0.01	**2.76**	0.03	2.25	0.17
4.20.9	AT3G09890.1	7E-51	Ankyrin repeat family protein	Chitinase-like	1.52	0.26	**2.08**	0.01	**2.54**	0.03	4.62	0.08
68.2.15	AT2G41710.1	2E-66	Integrase-type DNA-binding superfamily protein	Pinene Synthase	2.25	0.14	**1.55**	0.03	**3.72**	0.02	0.90	0.65
65.9.8	AT2G41710.1	2E-66	Integrase-type DNA-binding superfamily protein	Pinene Synthase	1.57	0.10	**1.61**	0.04	**3.09**	0.05	1.01	0.97

2.3.6 Influence of water deficit on gene expression patterns

Transcript abundance profiles were markedly different in inoculated vs. control samples under well watered or water deficit conditions (Fig. 2.4). Enrichment analysis of lodgepole sequences DE in inoculated vs. control seedlings under both well watered and water deficit conditions revealed significant overrepresentation of genes involved in secondary metabolism, miscellaneous cellular processes including peroxidase activity, stress response, protein processing, and the oxidative pentose phosphate pathway (Table 2.16). Enrichment analysis of sequences in lodgepole DE only in inoculated vs. control seedlings under well watered conditions revealed significant over representation of sequences putatively involved in miscellaneous cellular processes including peroxidase activity, and sequences putatively involved in secondary metabolism (Table 2.16). In contrast, enrichment analysis of sequences in lodgepole DE only in inoculated vs. control seedlings under water deficit conditions revealed significant over representation of genes involved in secondary metabolism, photosynthesis, miscellaneous cellular processes including peroxidase activity, metal handling, transport and the oxidative pentose phosphate pathway (Table 2.16).

While the categories of significantly overrepresented DE sequences in lodgepole pine inoculated vs. control seedlings were similar for these three comparisons, there were important qualitative differences in the sequences represented within these categories. For example, secondary metabolism was significantly overrepresented in lodgepole pine sequences DE under well watered and water deficit conditions, only DE under well watered conditions, and only DE under water deficit conditions, but there were different proportions of DE genes involved in distinct secondary metabolic pathways. Secondary metabolism sequences DE under both well watered and water deficit conditions were predominantly those involved in isoprenoid

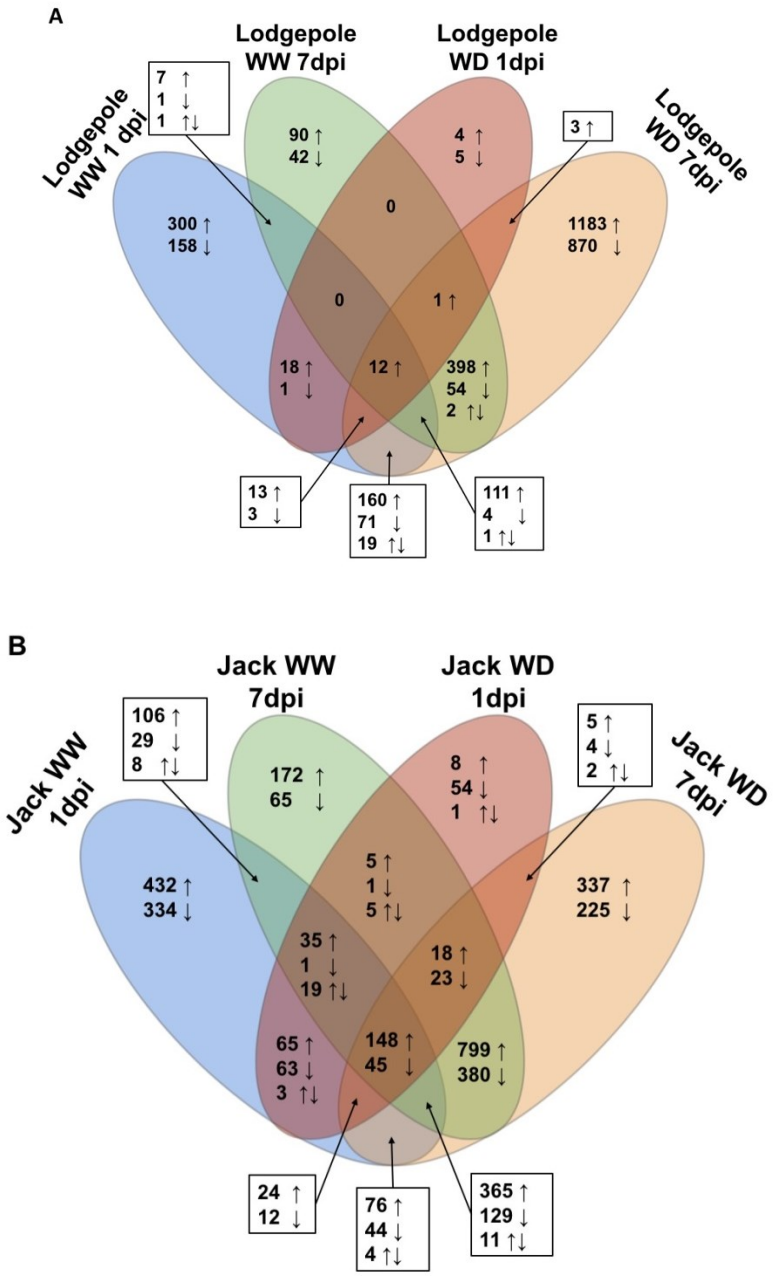


Figure 2.4: Four way Venn diagrams comparing DE sequences between well watered (WW) and water deficit (WD) inoculated vs. control treatments and further comparing between sequences DE at 1 dpi and 7 dpi A. Lodgepole pine sequences. B. Jack pine sequences. Within each cell of the Venn diagram, DE sequences are further divided into those that were upregulated at 1 and/or 7 dpi (↑), downregulated at 1 and/or 7 dpi (↓), or upregulated at 1 or 7 dpi, but downregulated on the other day (↑↓).

Table 2.16: Enrichment analysis of sequences differentially expressed in lodgepole pine inoculated vs. control seedlings under well watered and water deficit conditions, only under well watered conditions or only under water deficit conditions. Table displays functional categories significantly over represented in sequences DE in lodgepole pine inoculated vs. control seedlings under well watered and water deficit conditions, only under well watered conditions or only under water deficit conditions, using a hypergeometric distribution statistic.

Functional Category	Total PtGen elements in category	Elements DE in lodgepole	Probability Density Function (Hypergeometric distribution)	Adjust P value (Bonferroni)
DE in lodgepole pine under well watered and water deficit conditions				
Secondary metabolism	3831	34	3.36E-21	9.40E-20
Miscellaneous cellular processes	1544	113	4.27E-21	1.20E-19
Stress response	1167	81	7.14E-14	2.00E-12
Protein processing	3463	74	0.000484	0.0135
Oxidative pentose phosphate pathway	51	7	0.000520	0.0146
DE in lodgepole pine only under well watered conditions				
Miscellaneous cellular processes	1544	38	2.20E-09	4.61E-08
Secondary metabolism	3831	11	4.54E-07	9.53E-06
DE in lodgepole pine only under water deficit conditions				
Secondary metabolism	3831	100	1.33E-32	3.58E-31
Photosynthesis	526	72	3.71E-10	1.00E-08
Miscellaneous cellular processes	1544	157	7.76E-10	2.10E-08
Metal handling	49	12	3.57E-05	0.000964
Transport	586	61	4.95E-05	0.00134
Oxidative pentose phosphate pathway	51	12	5.39E-05	0.00146

biosynthesis (15 sequences), flavonoid biosynthesis (6 sequences), and phenylpropanoid biosynthesis (5 sequences) (Table 2.3). In comparison, secondary metabolism sequences DE in

lodgepole pine only under only well watered conditions were involved in phenylpropanoid biosynthesis (6 sequences) and isoprenoid biosynthesis (3 sequences), but not flavonoid biosynthesis. Finally, secondary metabolism sequences DE in lodgepole pine under only water deficit conditons were involved in flavonoid biosynthesis (57 sequences), isoprenoid biosynthesis (18), and phenylpropanoid biosynthesis (15 sequences) (Table 2.3).

Enrichment analysis of sequences DE in jack pine inoculated vs. control seedlings under both well watered and water deficit conditions revealed significant overrepresentation of genes involved in miscellaneous cellular processes including peroxidase activity, stress response, signalling, cell wall biosynthesis, biodegradation of xenobiotics, transport, the oxidative pentose phosphate pathway, hormone metabolism, metal handling, C1-metabolism, RNA regulation, and amino acid metabolism (Table 2.17). Enrichment analysis of sequences DE in jack pine inoculated vs. control seedlings exclusively under water deficit conditions revealed significant overrepresentation of sequences involved in secondary metabolism, miscellaneous cellular processes including peroxidase activity, and photosynthesis and the oxidative pentose phosphate pathway (Table 2.17). By comparison, enrichment of sequences DE in jack pine inoculated vs. control seedlings exclusively under water deficit conditions revealed significantly over representation of sequences involved in secondary metabolism, transport, miscellaneous cellular processes including peroxidase activity, and cell wall biosynthesis (Table 2.17).

Similar to lodgepole pine, secondary metabolism was significantly overrepresented in jack pine inoculated vs. control seedlings for all three comparisons, but there were qualitative differences in the DE sequences within the secondary metabolism category for each.

Table 2.17: Enrichment analysis of sequences differentially expressed in jack pine inoculated vs. control seedlings under well watered and water deficit conditions, only under well watered conditions or only under water deficit conditions. Table displays functional categories significantly over represented in sequences DE in jack pine inoculated vs. control seedlings under well watered and water deficit conditions, only under well watered conditions or only under water deficit conditions, using a hypergeometric distribution statistic.

Functional Category	Total PtGen elements in category	Elements DE in jack pine	Probability Density Function (Hypergeometric distribution)	Adjust P value (Bonferroni)
DE in jack pine under well watered and water deficit conditions				
Miscellaneous cellular processes	1544	273	5.44E-94	1.52E-92
Stress response	1167	121	4.07E-19	1.14E-17
Signalling	746	81	2.91E-14	8.15E-13
Cell wall	503	58	1.19E-11	3.33E-10
Biodegradation of xenobiotics	34	13	7.10E-10	1.99E-08
Transport	586	55	5.39E-08	1.51E-06
Oxidative pentose phosphate pathway	51	13	1.72E-07	4.82E-06
Hormone metabolism	767	64	3.16E-07	8.85E-06
Secondary metabolism	3831	117	2.01E-06	5.62E-05
Metal handling	49	11	5.62E-06	0.000157
C1-metabolism	19	6	0.000103	0.00289
RNA	1242	79	0.000191	0.00535
Amino acid metabolism	797	53	0.000649	0.0182
DE in jack pine under well watered conditions				
Secondary metabolism	3831	56	2.24E-38	6.04E-37
Miscellaneous cellular processes	1544	131	6.36E-10	1.72E-08
Photosynthesis	526	58	9.97E-09	2.69E-07
Oxidative pentose phosphate pathway	51	15	1.58E-08	4.22E-07
DE in jack pine under water deficit conditions				
Secondary metabolism	3831	16	5.38E-17	1.29E-15

Functional Category	Total PtGen elements in category	Elements DE in jack pine	Probability Density Function (Hypergeometric distribution)	Adjust P value (Bonferroni)
Transport	586	25	3.75E-05	0.000900
Miscellaneous cellular processes	1544	44	0.000631	0.0151
Cell wall	503	19	0.00112	0.0269

Secondary metabolism sequences DE in jack pine under both well watered and water deficit conditions were predominantly involved in flavonoid biosynthesis (43 sequences), isoprenoid biosynthesis (38 sequences), and phenylpropanoid biosynthesis (19 sequences) (Table 2.3). Secondary metabolism sequences DE in jack pine under only well watered conditions were predominanty involved isoprenoid biosynthesis (30 sequences), flavonoid biosynthesis (18 sequences), and phenylpropanoid biosynthesis (3 sequences) (Table 2.3). Finally, among the secondary metabolism sequences DE in jack pine under only water deficit conditions the two most predominant categories were those involved in isoprenoid biosynthesis (14 sequences) and phenylpropanoid biosynthesis (4 sequences), followed by those involved in flavonoid biosynthesis (1 sequence) (Table 2.3).

For both lodgepole pine and jack pine, there were instances in which *G. clavigera* inoculation under water deficit conditions altered the expression patterns by delaying significant changes in transcript abundance from 1 dpi to 7 dpi. In lodgepole pine, 884 sequences were DE at 1 dpi under well watered conditions, and 137 of these sequences were also DE at 7dpi (Fig 2.4). Inoculation under water deficit conditions delayed DE for 366 of the sequences, which were expressed early at 1 dpi under well watered conditions, but only later at 7 dpi under water deficit conditions (Fig 2.4). These sequences included four defense associated putative chitinases (spotID:38.14.8, spotID:53.7.5, spotID:7.24.4, spotID:6.13.15, spotId:4.9.8, spotID:61.14.7), five

putative DIRs (spotID:3.17.8, spotID:66.11.5, spotID:8.10.8, spotID:39.23.11, spotID:44.9.1), four putative osmotins (spotID:33.14.1, spotID:7.15.16, spotID:25.17.10, spotID:44.13.3), two PR-10 likes (spotID:4.7.1, spotID:48.10.10) and one α-pinene synthase (spotID:65.9.8) (Table 2.18). There were four defense associated transcriptional regulators in lodgepole pine that showed temporally delayed changes in transcript abundance under water deficit conditions, including two putative JAZ TF (spotID:63.5.4, spotID:6.22.1), and one ERF-like TF (spotID:7.13.5) (Table 2.18).

In jack pine, 1868 genes were DE at 1 dpi under well watered conditions, with 869 of these genes also DE at 7 dpi (Fig 2.4). Under water deficit conditions, 629 of these genes were DE in inoculated vs. control samples later 7 dpi, including 10 putative chitinase genes (spotD:67.21.1, spotID:7.15.6, spotID:52.6.6, spotID:66.23.7, spotID:3.17.16, spotID:64.6.10, spotID:61.14.7, spotID:17.12.14, spotID:22.23.9, spotID:63.24.13), a DIR-like protein (spotID:8.10.8) four putative terpene synthases encoding three different enzymes - a Δ-selinene synthase (spotID:64.24.13), an α-farnesene synthase (spotID:61.14.5) and a putative limonene synthase (spotID:69.22.15) - five putative osmotins (spotID:63.17.3, spotID:4.8.10, spotID:48.24.5, spotID:66.6.5, spotID:14.19.10), and two NBS like proteins (spotID:41.2.10, spotID:22.22.9) (Table 2.19). There were three putative JAZ TFs (spotID:16.8.10, spotID:62.10.9, spotID:7.14.2) whose expression was delayed under water deficit conditions (Table 2.19). The expression of some genes involved in the biosynthesis of JA was delayed under water deficit conditions, including two putative LOX (spotID:58.4.1), a putative ODPA reductase (spotID:3.9.13) and a JA methyltransferase (spotID:3.4.13) (Table 2.19).

Table 2.18: Sequences differentially expressed on the PtGen2 microarray in lodgepole phloem tissue between inoculated vs. control treatments early under well watered conditions at 1 dpi and later under water deficit conditions at 7 dpi. Sequence putative identity was predicted on the basis of sequence similarity to sequences within the TAIR database and within the NCBI nr database (not shown). FC: fold change. P: adjusted p-value.

PtGen2 spot ID	**TAIR ID**	**BLASTX E-value**	**Gene Symbols**	**Putative Identity**	Lodgepole 1 dpi WW		Lodgepole 1 dpi WD		Lodgepole 7 dpi WW		Lodgepole 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
DE in lodgepole 1 dpi under well watered conditions and 7 dpi under water deficit conditions												
38.14.8	AT3G12500.1	2E-60	HCHIB | basic chitinase	Chitinase-like	**3.26**	0.01	1.70	0.10	**3.45**	0.03	**2.22**	0.02
53.7.5	AT3G12500.1	6E-104	HCHIB | basic chitinase	Chitinase-like	**3.76**	0.03	2.54	0.07	**4.87**	0.04	**2.51**	0.03
7.24.4	AT3G12500.1	6E-104	HCHIB | basic chitinase	Chitinase-like	**2.21**	0.04	2.01	0.10	**3.85**	0.02	**2.27**	0.03
6.13.15	AT3G12500.1	6E-104	HCHIB | basic chitinase	Chitinase-like	**2.93**	0.03	1.86	0.14	**4.63**	0.01	**2.75**	0.01
4.9.8	AT3G12500.1	8E-89	HCHIB | basic chitinase	Chitinase-like	**10.20**	0.02	5.14	0.12	**8.63**	0.02	**4.43**	0.00
61.14.7	AT3G54420.1	9E-52	CHIV | basic chitinase	Chitinase-like	**3.40**	0.04	1.54	0.39	**9.45**	0.02	**9.03**	0.00
3.17.8	AT5G42510.1	2E-20	Disease resistance-responsive (dirigent-like protein) family protein	DIR-like	**1.57**	0.04	1.41	0.14	**4.05**	0.01	**4.12**	0.01
66.11.5	AT1G64160.1	9E-47	Disease resistance-responsive (dirigent-like protein) family protein	DIR-like	**7.89**	0.03	5.31	0.06	**3.91**	0.02	**5.71**	0.01
8.10.8	AT5G42510.1	2E-20	Disease resistance-responsive (dirigent-like protein) family protein	DIR-like	**5.74**	0.01	3.79	0.08	**23.00**	0.02	**27.70**	0.01

PtGen2 spot ID	TAIR ID	BLASTX E-value	Gene Symbols	Putative Identity	Lodgepole 1 dpi WW		Lodgepole 1 dpi WD		Lodgepole 7 dpi WW		Lodgepole 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
39.23.11	AT1G64160.1	1E-11	Disease resistance-responsive (dirigent-like protein) family protein	DIR-like	**3.79**	0.03	2.13	0.09	**1.59**	0.07	**2.25**	0.01
44.9.1	AT1G64160.1	9E-47	Disease resistance-responsive (dirigent-like protein) family protein	DIR-like	**2.41**	0.01	1.62	0.17	**1.52**	0.12	**1.98**	0.01
33.14.1	AT4G11650.1	4E-67	OSM34 | osmotin 34	Osmotin-like	**3.62**	0.04	2.18	0.20	**6.95**	0.03	**3.08**	0.02
7.15.16	AT4G11650.1	7E-41	OSM34 | osmotin 34	Osmotin-like	**1.63**	0.04	1.01	0.98	**9.91**	0.01	**5.14**	0.01
25.17.10	AT4G11650.1	4E-55	OSM34 | osmotin 34	Osmotin-like	**0.59**	0.04	0.71	0.36	1.71	0.19	**1.55**	0.02
44.13.3	AT1G20030.1	6E-84	Pathogenesis-related thaumatin superfamily protein LENGTH=1382	Osmotin-like	**1.93**	0.02	2.47	0.06	1.70	0.07	**2.20**	0.04
4.7.1	AT1G24020.2	1.E-07	MLP-like protein 423	Osmotin-like	**4.87**	0.03	1.12	0.93	3.55	0.13	**2.67**	0.03
48.10.10	AT1G24020.2	1.E-07	MLP423 | MLP-like protein 423	Osmotin-like	**2.88**	0.04	1.36	0.57	2.14	0.21	**2.02**	0.02
65.9.8	AT2G41710.1	2E-66	Integrase-type DNA-binding superfamily protein	Pinene Synthase-like	**2.09**	0.04	1.57	0.10	0.99	0.98	**1.61**	0.04
63.5.4	AT3G17860.1	2E-21	JAZ3 | jasmonate-zim-domain protein 3	JAZ-like TF	**2.06**	0.03	1.48	0.22	1.24	0.18	**1.56**	0.01
6.22.1	AT1G74950.1	9E-14	JAZ2 | TIFY domain/Divergent CCT motif family protein	JAZ-like TF	**6.36**	0.01	2.72	0.15	5.31	0.05	**16.50**	0.02
7.13.5	AT3G20310.1	1E-20	ERF7 | ethylene response factor 7	JAZ-like TF	**1.53**	0.05	1.38	0.30	1.40	0.09	**1.68**	0.01

In order to further investigate defense response genes attenuated under water deficit conditions in lodgepole and jack pine, we compared sequences upregulated in inoculated vs. control seedlings under well watered conditions to sequences down regulated or not DE in inoculated vs. control seedlings under water deficit conditions in lodgepole and jack pine (Fig 2.5).

Enrichment analysis of genes significantly upregulated in lodgepole inoculated vs. control seedlings under well watered conditions, and also downregulated or not DE in lodgepole inoculated vs. control seedlings under water deficit conditions revealed significant overrepresentation of genes involved in secondary metabolism, miscellaneous cellular processes including peroxidase activity, photosynthesis, and protein processing (Table 2.20). Enrichment analysis of genes upregulated in jack pine inoculated vs. control seedlings under well watered conditions, and also downregulated or not DE in lodgepole inoculated vs. control seedlings under water deficit conditions revealed significant over representation of genes involved in secondary metabolism, miscellaneous cellular processes including peroxidase activity, protein processing, photosynthesis, TCA cycling, C1-metabolism, the oxidative pentose phosphate pathway, and cell wall biosynthesis (Table 2.20).

Given the established importance of terpenoid biosynthesis to conifer defense against pests and pathogens, and given the prominence of putative terpene synthases in our above analyses, we used heat maps to compare expression profiles of all DE putative terpene synthases. While expression profiles of the putative di- and sesquiterpene synthases were similar between species,jack pine seedlings displayed differential expression of more putative α-pinene synthases than lodgepole pine, and upregulation was most apparent at 1 dpi (Fig 2.6).

Table 2.19: Sequences differentially expressed on the PtGen2 microarray in jack pine phloem tissue between inoculated vs. control treatments early under well watered conditions 1 dpi and later under water deficit conditions 7 dpi. Sequence putative identity was predicted on the basis of sequence similarity to sequences within the TAIR database and within the NCBI nr database (not shown). FC: fold change. P: adjusted p-value.

PtGen2 spot ID	**TAIR ID**	**BLAST X E-value**	**Gene Symbols**	**Putative Identity**	Jack 1 dpi WW		Jack 1 dpi WD		Jack 7 dpi WW		Jack 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
DE in jack pine 1 dpi under well watered conditions and 7 dpi under water deficit conditions												
67.21.1	AT3G12500.1	2E-60	HCHIB | basic chitinase	Chitinase-like	**4.34**	0.00	4.67	0.05	**2.93**	0.01	**5.72**	0.00
52.6.6	AT3G12500.1	6E-104	HCHIB | basic chitinase	Chitinase-like	**3.72**	0.01	2.49	0.12	**3.47**	0.00	**2.55**	0.02
66.23.7	AT3G12500.1	2E-60	HCHIB | basic chitinase	Chitinase-like	**4.26**	0.00	4.72	0.05	**3.48**	0.00	**6.04**	0.01
3.17.16	AT2G43590.1	2E-40	Chitinase family protein	Chitinase-like	**1.86**	0.03	1.32	0.34	**7.20**	0.00	**5.38**	0.03
64.6.10	AT3G12500.1	8E-89	HCHIB | basic chitinase	Chitinase-like	**4.51**	0.00	4.59	0.07	**3.46**	0.01	**5.62**	0.00
61.14.7	AT3G54420.1	9E-52	CHIV | homolog of carrot EP3-3 chitinase	Chitinase-like	**3.39**	0.00	1.69	0.23	**18.51**	0.00	**6.54**	0.00
17.12.14	AT2G43590.1	2E-40	Chitinase family protein	Chitinase-like	**2.53**	0.02	1.43	0.20	**16.94**	0.00	**9.40**	0.02
22.23.9	AT2G43590.1	2E-40	Chitinase family protein	Chitinase-like	**4.04**	0.00	1.63	0.41	**28.16**	0.00	**16.22**	0.03
63.24.13	AT3G12500.1	2E-60	HCHIB | basic chitinase	Chitinase-like	**1.70**	0.01	1.51	0.16	**1.58**	0.04	**1.83**	0.01
8.10.8	AT5G42510.1	2E-20	Disease resistance-responsive (dirigent-like protein) family protein	DIR-like	**6.83**	0.00	4.92	0.06	**30.33**	0.00	**47.56**	0.01
64.23.13	AT5G48540.1	6E-28	receptor-like protein kinase-related family protein	Selinene Synthase-like		0.08	0.57	0.37	1.32	0.32	**0.62**	0.35

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative Identity	Jack 1 dpi WW		Jack 1 dpi WD		Jack 7 dpi WW		Jack 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
61.14.5	AT4G02780.1	2E-20	CPS1 | Terpenoid cyclases/Protein prenyltransferases superfamily protein	Farnesene Synthase-like	**1.52**	0.01	1.61	0.08	**2.06**	0.04	**1.96**	0.02
69.22.15	AT2G41710.1	2E-66	Integrase-type DNA-binding superfamily protein	Limonene Synthase	**1.97**	0.04	2.23	0.14	**0.55**	0.02	**0.42**	0.03
63.17.3	AT4G11650.1	4E-55	OSM34 | osmotin 34	Osmotin-like	**1.90**	0.04	1.32	0.45	**8.29**	0.00	**11.68**	0.01
4.8.10	AT4G11650.1	1E-75	OSM34 | osmotin 34	Osmotin-like	**1.51**	0.03	1.60	0.08	**2.95**	0.00	**2.11**	0.05
48.24.5	AT1G75040.1	4E-76	PR-5 | pathogenesis-related gene 5	Osmotin-like	**2.29**	0.02	2.80	0.14	**9.85**	0.00	**8.57**	0.00
66.6.5	AT1G75040.1	4E-76	PR-5 | pathogenesis-related gene 5	Osmotin-like	**1.78**	0.02	2.08	0.23	**1.90**	0.02	**2.12**	0.01
14.19.10	AT1G75040.1	4E-76	PR-5 | pathogenesis-related gene 5	Osmotin-like	**1.58**	0.03	1.56	0.21	**2.13**	0.02	**2.24**	0.03
41.2.10	AT1G72910.1	5E-10	Toll-Interleukin-Resistance (TIR) domain-containing protein	NB-LRR like	**0.40**	0.01	0.38	0.09	0.69	0.07	**0.41**	0.01
22.22.9	AT1G69550.1	6E-13	Toll-Interleukin-Resistance (TIR) domain-containing protein	NB-LRR like	**1.61**	0.02	1.44	0.11	**1.57**	0.01	**1.96**	0.02
16.8.10	AT1G74950.1	5E-16	JAZ2| TIFY domain/Divergent CCT motif family protein	JAZ-like	**2.76**	0.01	2.15	0.10	**3.05**	0.00	**3.58**	0.01
62.10.9	AT1G74950.1	5E-16	JAZ2| TIFY domain/Divergent CCT motif family protein	JAZ-like	**2.87**	0.01	2.13	0.18	**3.32**	0.00	**3.47**	0.01
7.14.2	AT1G74950.1	6.E-07	JAZ2| TIFY domain/Divergent CCT motif family	JAZ-like	**8.21**	0.00	2.54	0.26	1.10	0.80	**2.40**	0.04

PtGen2 spot ID	TAIR ID	BLAST X E-value	Gene Symbols	Putative Identity	Jack 1 dpi WW		Jack 1 dpi WD		Jack 7 dpi WW		Jack 7 dpi WD	
					FC	P	FC	P	FC	P	FC	P
			protein									
58.4.1	AT1G72520.1	0	LOX4 | PLAT/LH2 domain-containing lipoxygenase family protein	LOX-like	**4.67**	0.00	2.45	0.06	**1.63**	0.01	**2.16**	0.02
3.9.13	AT1G76690.1	2E-154	OPR2 | 12-oxophytodienoate reductase 2	ODPA like	**2.22**	0.01	2.10	0.12	**1.81**	0.02	**2.26**	0.04
3.4.13	AT1G48850.1	2E-168	EMB1144 | chorismate synthase, putative / 5-enolpyruvylshikimate-3-phosphate phospholyase, putative	JA Metransferase-like	**1.53**	0.01	1.40	0.13	1.07	0.65	**1.65**	0.02

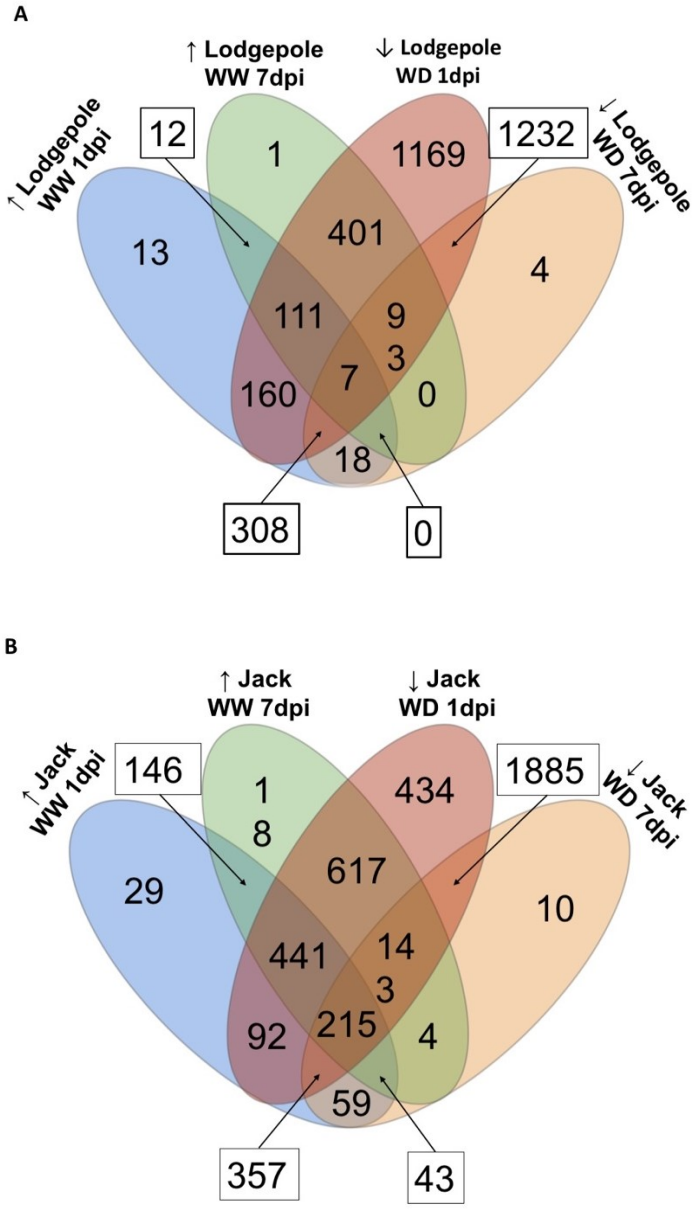


Figure 2.5. Four way Venn diagrams comparing sequences upregulated (up facing arrows) in inoculated vs. control treatments under well watered (WW) and sequences down regulated or non DE (down facing arrows) in inoculated vs. control treatments under water deficit (WD) and further comparing between sequences DE at 1 and 7 dpi A. Lodgepole pine sequences. B. Jack pine sequences.

Table 2.20: Enrichment analysis of sequences which are differentially expressed and upregulated in inoculated vs. control lodgepole and jack pine seedlings under well watered conditions, but down regulated or not DE in inoculated vs. control in lodgepole and jack pine seedlings under water deficit conditions. Table displays functional categories significantly over represented in these sequences, using a hypergeometric distribution statistic.

Functional Category	Total PtGen elements in category	Elements DE in lodgepole and jack	Probability Density Function (Hypergeometric distribution)	Adjust P value (Bonferroni)
Upregulated in lodgepole pine under well watered conditions and downregulated or not DE under water deficit conditions				
Secondary metabolism	3831	26	7.74E-15	2.17E-13
Misc	1544	78	8.46E-14	2.37E-12
PS	526	1	1.48E-04	4.13E-03
Protein	3463	50	5.19E-04	1.45E-02
Upregulated in jack pine under well watered conditions and downregulated or not DE under water deficit conditions				
Secondary metabolism	3831	78	2.61E-24	7.06E-23
Misc	1544	133	8.21E-17	2.22E-15
protein	3463	90	1.02E-12	2.75E-11
PS	526	2	7.78E-09	2.10E-07
TCA / org transformation	222	22	1.84E-04	4.97E-03
C1-metabolism	19	5	1.00E-03	2.71E-02
OPP	51	8	1.24E-03	3.36E-02
cell wall	503	10	1.33E-03	3.60E-02

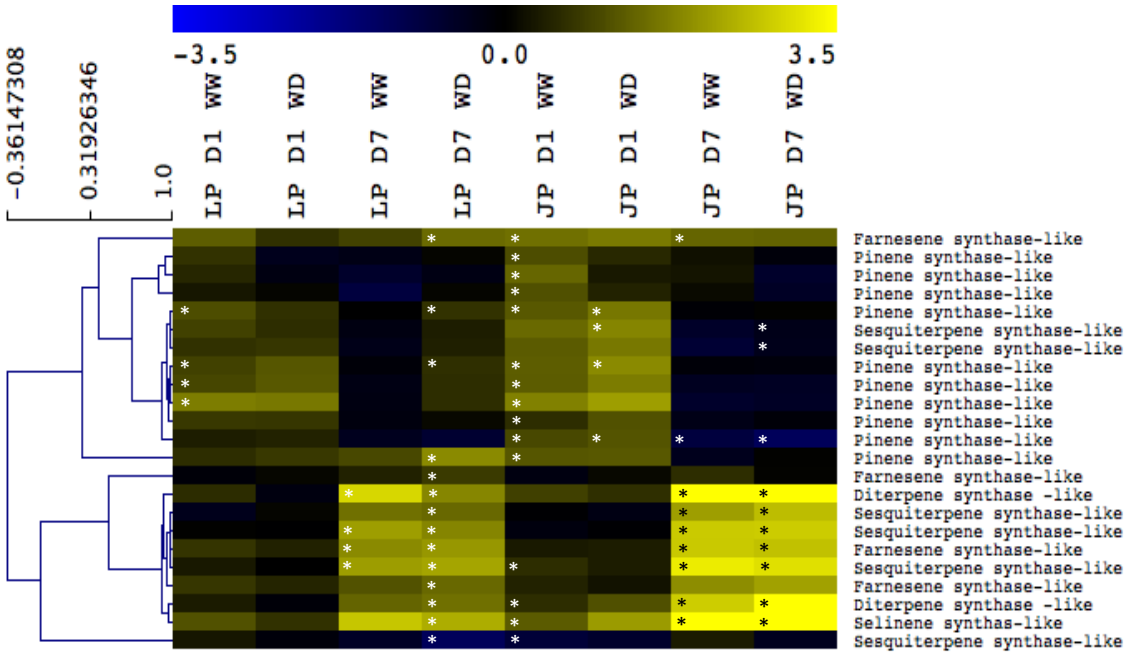


Figure 2.6. Heatmap representation of transcript abundance of putative lodgepole and jack pine terpene synthase genes in *Grosmannia clavigera*-inoculated trees. Data are expressed as $\log_2$ fold-change relative to control uninoculated trees. Yellow: upregulation, blue: downregulation, significant differences in transcript abundance are indicated by an asterisk (*). Genes are clustered hierarchically based on shared patterns of expression.

2.3.7 Activation of JA/ethylene signaling pathway in lodgepole and jack pine

Examination of genes involved in biosynthesis and downstream signaling of defense associated hormones revealed concerted changes in transcript abundance profiles of JA and ethylene pathway sequences in both lodegpole and jack pine seedlings (Fig 2.7).

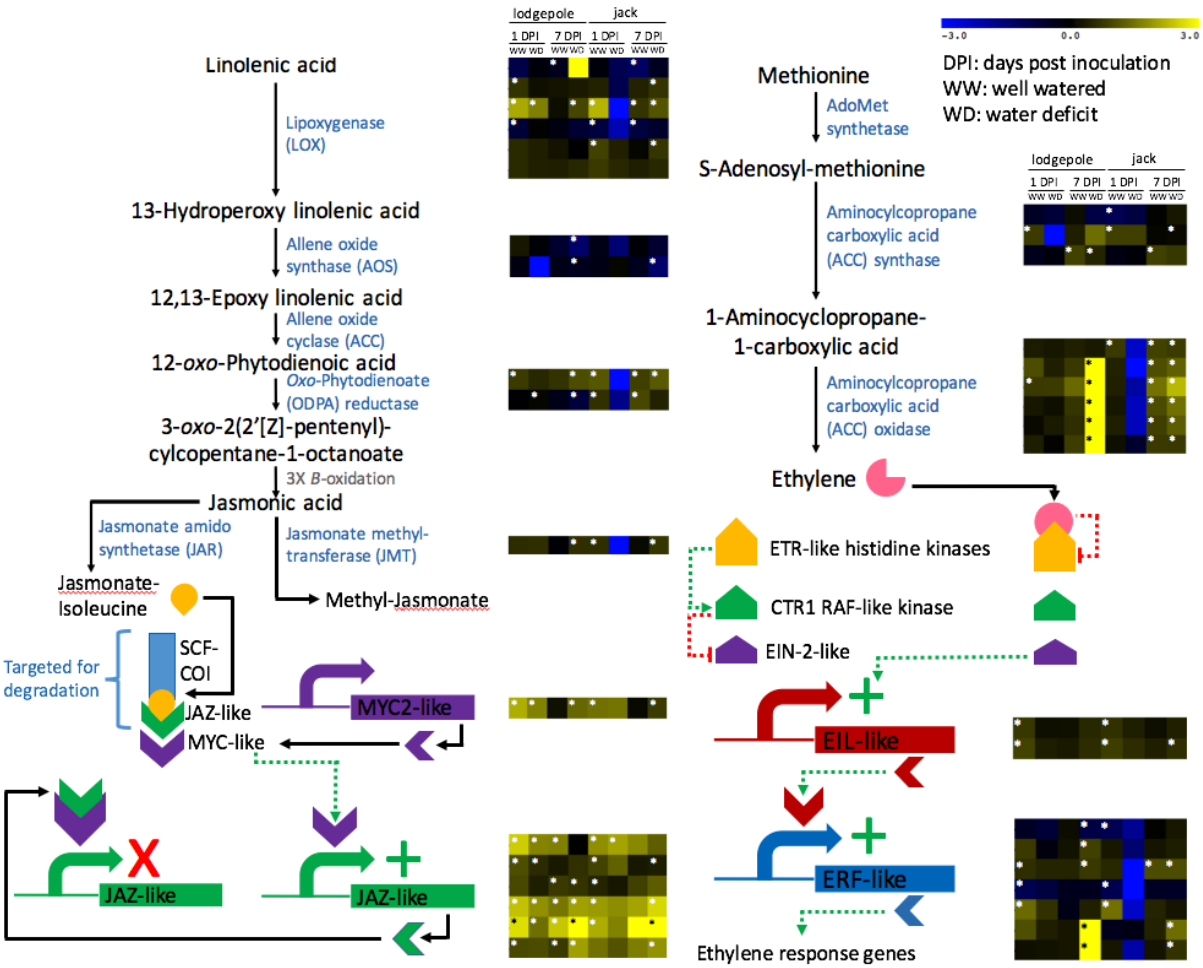


Figure 2.7: Heat map representation of transcript abundance of sequences involved in JA and ethylene biosynthesis and downstream signaling in lodgepole and jack pine inoculated seedlings vs control seedlings, determined by microarray analysis. Data are expressed as log2 fold-change inoculated vs control treatments at 1 and 7 dpi, under well watered (WW) and water deficit (WD) conditions. Yellow: upregulation. Blue: down regulation. Differentially expressed genes: starred (*). In the presence of JA, the SCF-COI protein complex is target for degradation, allowing expression of MYC2-like genes which induce expression of JAZ-like TF. In the presence of ethylene ERT histidine kinases are deactivated, allowing EIN2-like TFs to induce expression of EIL-like TFs which in induce expression of ERF-like TFs, promoting expression of downstream of ethylene responsive genes.

Genes putatively involved in biosynthesis of jasmonic acid such as LOX (spotID:58.4.1, spotID:7.4.9) and ODPA-reductase (spotID:3.9.13) were significantly upregulated at 1 dpi in both lodgepole and jack pine. A subset of these genes was also significantly upregulated under water deficit conditions. Putative members of the JAZ-like transcription factor family (spotID:61.4.16, spotID:61.7.15, spotID:6.22.1, spotID:63.5.4) shared similar but not identical transcript abundance patterns. Different members were significantly upregulated at both time points and both water availability levels in both species (Fig 2.7).

Genes determined to be putatively involved in the biosynthesis of ethylene also display upregulation in both lodgepole and jack pine, particularly at 7 dpi (Fig. 2.7). Fold change increases in transcript abundance for inoculated vs. control samples were particularly high for several ACC oxidase sequences (spotID: 46.23.12, spotID:4.4.6, spotID:28.2.5, spotID:40.14.9, spotID:9.8.15, spotID:56.18.8) and a subset of putative ERF-like sequences (spotID:15.24.6, spotID:25.5.14) in 7dpi lodgepole pine under water deficit conditions. Other putative ERF-like TFs displayed diverse transcript abundance profiles.

2.4 Discussion

2.4.1 Transcriptomes of lodgepole and jack pine are substantially altered in response to *G. clavigera*

Conifers rely on an array of induced and constitutive defenses to defend against pests and pathogens. Fungal invaders are known to induce a wide array of molecular responses in conifers (Eyles *et al.* 2010, Franceschi *et al.* 2005, Koslova & Bohlmann 2012), including (1) chemical defenses, often secondary metabolites, (2) protein defenses, including increased expression of PR proteins, and (3) anatomical defenses, such as cell wall fortification and formation of traumatic

resin ducts (Keeling & Bohlmann 2006, Van Loon *et al.* 2006, Eyles *et al.* 2010). Many sequences associated with this diversified response were evident in the transcriptome-wide changes that we observed in both pine species in response to inoculation with *G. clavigera*, and indicates that both lodgepole and jack pine respond to challenge by *G. clavigera* through activation of multiple defense strategies.

Timing of defense responses are critical in effective pathogen containment, and studies have linked early induction of important PR proteins such as chitinases in pine to resistant phenotypes, and late induction of the same proteins to susceptible phenotypes (Davis *et al.* 2002, Heitala *et al.* 2004). There were a greater number of DE sequences at 7 dpi than at 1 dpi in response to *G. clavigera* inoculation in both lodgepole and jack pine, suggesting that at a global level, there was no overt differences in timing of response to pathogen attack between the co-evolved and naïve species. However, as will be discussed below, there are temporal differences in expression profiles for subsets of genes that have the potential to impact defense capacity of these two species against *G. clavigera*.

2.4.2 JA and ethylene are implicated in the response of both lodgepole and jack pine to *G. clavigera*

Both species responded to inoculation with *G. clavigera* by invoking genes implicated in the JA and ethylene biosynthesis and signaling pathways as early as 1 dpi. Some of these sequences showed significantly increased transcript abundance in response to *G. clavigera* at 1 dpi in both lodgepole and jack pine, suggesting that JA and ethylene play roles in mediating early responses to *G. clavigera* infection in both species. These inferences from gene expression patterns are corroborated by a recent study reporting early *in vivo* increases in levels of JA and JA-Ile in lodgepole and jack pine seedlings inoculated with *G. clavigera* (Arango-Velez *et al.*

2016). In conifers, exogenous application of JA has been reported to induce an array of defense responses such as increased expression of terpene synthases and formation of traumatic resins ducts (Hudgins & Franceschi 2004; Martin *et al.* 2002; Franceschi *et al.* 2002; Gould *et al.* 2009). These reports support our findings that the JA signaling pathway is responsible for activating defense responses to *G. clavigera* in both lodgepole and jack pine. In contrast, much less work has been carried out on the role of ethylene in the defense response of conifers. Formation of tramatic resin ducts and swelling of polyphenolic parenchyma cells was reported following the exogenous application of ethylene (Hudgins & Franceschi 2004). There is indirect evidence that ethylene biosynthesis increases in response to wounding and in response to JA. In Douglas-fir, select ACC synthase genes were expressed at increased levels and ACC synthase proteins accumulated in greater quantities in response to wounding (Ralph *et al.* 2007). Accumulation of ACC oxidase was reported in Douglas-fir treated with both MeJA and wounding, suggesting that in the presence of JA, ethylene biosynthesis is induced and the resulting increase in ethylene promotes downstream defenses such as formation of tramautic resin ducts and accumulation of polyphenolic cells (Hudgins *et al.* 2006). These reports support our model that ethylene works in concert with JA to mediated the inducible defense response of lodgpole and jack pine in response to *G. clavigera.*

In annual plants such as *Arabidopsis*, activation of the JA and ethylene pathways is associated with response to necrotrophic pathogens (Thomma *et al.* 1998; Glazebrook 2005). In contrast, the SA pathway is invoked in response to biotrophic pathogens (Thomma *et al.* 1998; Glazebrook 2005). Our transcriptomic data – in which several sequences associated with the JA and ethylene biosynthetic and signaling pathways are DE, but few if any sequences associated

with the SA biosynthesis and signally pathways are DE – supports the postulation by Arango-Velez et al. (2016) that *G. clavigera* is likely a necrotrophic pathogen.

While there are similarities in the transcriptomic responses of JA and ethylene biosynthesis and signaling genes for both lodgepole and jack pine, there are also some intriguing differences in timing and/or amplitude of increased transcript abundance for some of these genes under conditions of water deficit. It will be important to conduct more in-depth studies investigating the roles for JA and ethylene in response of lodgepole and jack pine to *G. clavigera*, and how water deficit changes JA and ethylene signaling. Similarly, it will be of great interest to determine JA- and ethylene-responsive genes in these species. Defense associated genes such as putative terpene synthases, chitinases and DIR-like genes show similar expression patterns to several JA and ethylene biosynthesis and signaling genes in both species, leading to the hypothesis that they could be activated through the JA/ethylene signaling pathway.

2.4.3 Evidence for core responses and species-specific responses between lodgepole and jack pine to *G. clavigera*

Comparison of genes differentially expressed in response to *G. clavigera* between lodgepole and jack pine seedlings revealed a core set of responses that were invoked in both species, including a number of genes commonly associated with defense responses against diverse pests and pathogens in many species (Ralph *et al.* 2006a; Koslova & Bohlmann 2012). Comparison of transcriptomic profiles also revealed defense response genes that were significantly differentially expressed in one species but not the other, suggesting that even though the global temporal transcriptomic response between the species was similar and there was a shared set of responses, there are also aspects of their responses to *G. clavigera* that differ between lodgepole and jack pine. In some cases, the difference was due to timing of differential

expression of putative orthologues, i.e. a differential temporal response. In other cases, the difference was due to differential expression of genes in one species but not the other, i.e. a unique response.

Mounting an early defense response to pathogen challenge is critical in effective pathogen containment, and differences in timing in the expression of certain defense response genes in one species vs. another may serve as indication of earlier pathogen recognition and response. Analysis of the temporal patterns of DE in lodgepole and jack pine revealed shared defense related regulators expressed earlier in lodgepole and later in jack pine, including a putative ACC oxidase involved in ethylene biosynthesis, and an ERF-like TF involved in ethylene signal transduction. Co-expressed with these potential regulators are an osmotin-like and a DIR-like gene. In contrast, no defense associated transcriptional regulators, nor hormone biosynthesis and signaling genes were differentially expressed earlier in jack pine and later in lodgepole pine. Plants recognize and respond to invaders through generalized and/or specific pathogen recognition mechanisms (Jones & Takemoto 2008). Earlier expression of certain defense associated regulators co-expressed with specific defense response genes such as osmotins, may indicate earlier recognition of *G. clavigera* by lodgepole pine, leading to earlier activation of pathogen specific defense pathways. Osmotins are important anti-microbial proteins which act against fungal invaders by inhibiting hyphal growth by permeabilizing and degrading fungal cell walls (Abad 1996; Osmond 2001; Zearie 2002). Certain ascomycete fungi are highly sensitive to osmotins, and it has been proposed that osmotins may potentially play a role in the specialized, pathogen-specific response in plants (Vernoese *et al.* 2003). DIR proteins are thought to contribute to stereospecific coupling of monolignols to form lignin dimers, also known as lignans (Ralph *et al.* 2006b), as well as other phenoxy radical coupling of other

secondary metabolites such as complex terpenoids (Pickel & Schaller 2013). Many of these compounds, such as lignans, have known roles in plant defense (Pickel & Schaller 2013). In contrast to jack pine, lodgepole pine is considered to have co-evolved with MPB and by extension with MPB fungal associate *G. clavigera,* and early low levels of DE in some defense response genes such as those encoding osmotins and DIR proteins may indicate earlier recognition and response to *G. clavigera* as a result of a co-evolved relationship (Cudmore *et al.* 2010, Safranyik *et al.* 2010).

Enrichment analysis of genes DE exclusively in lodgepole and exclusively in jack pine inoculated vs. control seedlings under well watered conditions revealed the importance of secondary metabolism to the defense response of these sister species. Closer inspection of the genes contained within the secondary metabolism category further revealed differences between the two species: whereas secondary metabolism sequences DE only in lodgepole were predominantly involved in flavonoid biosynthesis, the majority of which were chalcone synthases, and none were involved in the biosynthesis of isoprenoids, secondary metabolism sequences DE only in jack pine were primarily associated with isoprenoid biosynthesis, specifically putative terpene synthases (Table 2.8 & Table 2.10).

Chalcone synthases are often induced in conifers responding to pathogen challenge, and provide the precursor molecules for important anti-microbial compounds such as phytoalexins and phytoanticipins (Nagy *et al.* 2004; Dao *et al.* 2011). Terpene synthases catalyze steps in the complex network of reactions which produce terpenoids, the compounds that form a major component of oleoresins (Byun-McKay *et al.* 2006, Zulak & Bohlmann 2010). While both flavonoid and terpenoid compounds are reported to play a role in conifer defense response, the quantity and qualitative composition of induced oleoresin is reported to effect pest-host

interactions and tree host resistance (Tomlin *et al.* 2000; Raffa & Berryman 1982). We observed notable differences in expression of genes encoding putative terpene synthases between lodgepole and jack pine. Three of the putative terpene synthases DE uniquely in jack pine were α-pinene synthases. This finding is consistent with a previously published report comparing the monoterpene profiles of lodgepole and jack pine seedlings challenged with *G. clavigera,* which demonstrated that jack pine seedlings contained and emitted greater amounts of α-pinene in comparison to lodgepole pine seedlings (Lusebrink *et al.* 2011; Arango-Velez *et al.* 2016). Interestingly, though, these same studies showed that lodgepole pine exhibited the greater induction of total monoterpene levels in response to *G. clavigera* infection, and also greater increases in other specific monoterpenes such as myrcene, β-phellandrene, and terpenolene. There were also species-specific differences in transcript profiles for several putative diterpene and sesquiterpene synthases, including putative farnesene synthases and selinene synthases. Expression of these putative diterpene and sesquiterpene synthases was significantly greater in response to *G. clavigera* in both lodgepole and jack pine at 7 dpi, but jack pine exhibited higher fold-changes in transcript abundance for these genes than lodgepole pine. Along with monoterpenes, diterpene resin acids and sesquiterpenes form important components of oleoresins which accumulate in pine challenged by pests and pathogens. Early work has demonstrated that lodgepole pine resistant to MPB attack accumulated greater quantities of oleoresins in response to inoculations with *G. clavigera* (Raffa & Berryman 1982). Higher concentrations of diterpene resin acids found in wounding induced oleoresin was positively correlated to weevil resistance in Sitka spruce (Tomlin 1996 & 2000). MeJA induced increases in sesquiterpene accumulation along with monoterpene and diterpenes which resulted in increased resistance in Norway spruce to *Ceratocytstis polonica,* a fungus associated with the spruce bark beetle *Ips typograhus* (Zeneli

et al. 2005). The scale of induced total oleoresin accumulation often is a marker of the strength of overall tree host resistance to pathogen challenge, and higher fold changes in transcript abundance of diterpenes and sesquiterpenes in jack pine compared to lodegpole pine is suggestive of species-specific differences in the defense responses to *G. clavigera* between species that may impact the tree's ability to contain the *G. clavigera* infection.

2.4.4 Water deficit augments constitutive defenses and attenuates induced defenses in lodgepole and jack pine

Water deficit had a substantive effect on the transcriptomic responses of both lodgepole and jack pine to *G. clavigera*. A subset of sequences that were significantly DE in lodgepole and/or jack pine under water deficit conditions were not DE under well watered conditions. The number of genes that were DE in inoculated vs. control samples exclusively under water deficit was three fold greater in lodgepole pine than jack pine, with only a small proportion of these DE genes shared between species. The greater impact of water deficit on the transcriptomic response of lodgepole pine to *G. clavigera* may be a function of the stronger isohydric nature of this species relative to jack pine (Arango-Velez *et al.* 2016).

In both species, water deficit induced the expression of multiple genes encoding enzymes involved in synthesis of secondary metabolites implicated in defense (Table 2.15 & 16, Table 2.3). Previously published reports suggest that water limitation in conifers increases expression of constitutive defenses while decreasing induced defenses (Lorio *et al.* 1995, Lombardo *et al.* 2000). Other studies examining lodgepole and jack pine defense response to *G. clavigera* have demonstrated that non-inoculated seedlings under water deficit displayed increased total monoterpene content under control conditions as well as an increase in expression of some terpene synthase genes in comparison to seedlings under well watered conditions (Lusebrink *et*

al. 2011, Arango-Velez *et al.* 2014). In our findings, enrichment analysis of sequences DE only in inoculated vs control trees under water deficit revealed statistical overrepresentation of sequences involved in secondary metabolism for both lodgepole and jack seedlings, including terpene synthases, which are not DE under well watered conditions in either species. The augmentation of these defenses under water deficit conditions occurs to much greater extent in lodgepole pine, supporting our initial hypothesis that water deficit will exert a greater effect on the more drought responsive lodgepole pine seedlings.

Water deficit conditions also delayed differential expression of defense associated regulators and defense response genes in both lodgepole and jack pine. In lodgepole pine, the upregulation of three JAZ-like JA signaling TFs was delayed under water deficit conditions along with the upregulation of key defense genes such as those encoding chitinases and DIR-like proteins. A similar pattern was observed in jack pine seedlings, where water deficit conditions delayed the upregulation of three different JAZ-like TFs along with putative orthologues of the lodgepole pine chitinases and DIR-like genes. These results are consistent with other studies that have demonstrated that lodgepole and jack seedlings inoculated with *G. clavigera* released significantly lower monoterpene emissions while under water deficit, as well as demonstrated reduced transcript abundance of chitinases relative to inoculation under well watered conditions (Lusibrink *et al.* 2011, Arango-Velez *et al.* 2014). Our results demonstrate water deficit conditions delayed expression of JAZ-like defensive transcriptional regulators, and co-expressed defense associated genes in both lodgepole and jack pine. We hypothesize that both species respond to carbon limitation imposed under water deficit conditions by stomatal closure and the resulting decrease in photosynthesis (Arango-Velez *et al.* 2014, Arango-Velez *et al.* 2016) by delaying or limiting expression of inducible defense regulons.

2.4.5 Conclusion

The objective of our study was to examine transcriptome-wide responses of jack pine and lodgepole pine seedlings inoculated with MPB fungal associate *G. clavigera* while under well watered or water deficit conditions. We aimed to identify species-specific differences between the evolutionarily co-evolved lodgepole pine and naïve jack pine as well as to determine the effect of water deficit on their respective responses. Our results support our previous study suggesting that activation of defense response pathways in both species to *G. clavigera* occurs through the JA signaling pathway (Arango-Velez *et al.* 2016), and likely also invokes ethylene signalling. As has been established, secondary metabolism appears to be central to both lodgepole and jack pine defense against *G. clavigera*. Our results suggest that there are qualitative differences in the secondary metabolites induced defense response of lodgepole and jack pine, with a greater emphasis on changes in gene expression associated with flavonoid biosynthesis in lodgepole pine versus isoprenoid biosynthesis in jack pine. This difference may reflect that lodgepole pine typically generates longer lesions in response to *G. clavigera* inoculation than jack pine. We also conclude that expression of a subset of genes that are normally part of the induced defenses becomes attenuated under water deficit conditions, while expression of another subset of genes that contributes to the plant's constitutive defenses under well watered is augmented under water limitation. Finally, we conclude that water deficit impacts defense response to a greater extent in lodgepole pine than in jack pine, and we propose this is due to the stronger isohydric nature of lodgepole pine.

3.0 Chapter 3: Characterization of the chitinase gene family in lodgepole and jack pine

3.1 Introduction

Conifers induce an array of defenses to defend against pests and pathogens. Inducible defenses allow the tree to effectively defend against invader while allocating fewer resources to defense compared to constitutive defenses (Eyles *et al.* 2010). Pathogenesis response (PR) proteins are antimicrobial proteins found in low or undetectably levels in healthy tissue, which accumulate in high concentrations during the onset of pathogen challenge (van Loon *et al.* 2006). Transcription of PR proteins is thought to be induced by R genes, or NB-LRR genes, that detect pathogen-produced elicitors (Valuthakkal *et al.* 2012). Chitinases are PR proteins, belonging to families PR-3, PR-4, PR-8 and PR-11, and are commonly induced in conifers at very high levels in response to pathogen challenge (Valuthakkal *et al.* 2012; Heitela *et al.* 2004; Koslova *et al.* 2014; Islam *et al.* 2010). Within the chitinase family, some chitinases are thought to play an antimicrobial role in response to pathogen challenge, as some chitinases hydrolyze the β-1-4 linked N-acetyl glucosamine units which form chitin, an important component in fungal cell walls. Enzymatic assays have confirmed chitinolytic activity in a subset of chitinases originating from spruce and lodgepole pine (Valuthakkal *et al.* 2012; Koslova *et al.* 2014).

Chitinases are grouped into two separate families of glycosyl hydrolases, GH18 and GH19. Chitinases are further classified in to seven distinct classes based on the presence or absence of highly conserved domains (Neuhaus 1999). Class I, II, IV and VII chitinases are members of GH family 19 and share high sequence similarity with one another (Neuhaus 1999). Class I chitinases contain an N-terminal signal peptide domain, a chitin binding domain (CBD), a

catalytic domain, and ending in a C-terminal vacuolar localization signal. Class II chitinases are highly similar to class I chitinases, but lack a CBD and carry a deletion in the second of three of the loops in the catalytic domain. Class IV chitinases contain a signal peptide and CBD domain, but the catalytic domain is shortened and contains deletions in third of three loops and lack the C-terminal vacuolar signal. Class VII chitinases are highly similar to class IV chitinases, but lack the CBD (Neuhaus 1999). Some GH 19 chitinases have chitinolytic activity, which work endogenously to hydrolyze chitin polymers, and mostly occur in plants (Islam *et al.* 2010). GH 18 chitinases, often referred to as "bacterial", include class III and class V chitinases, and work exogenously on the ends of chitin polymer (van Aalten *et al.* 2001). Less is known about the role of class III and class V chitinase in conifer defense.

The current mountain pine beetle (MPB; *Dendroctonus ponderosae*) outbreak is estimated to have affected 19 million ha. of Canadian forests, resulting in widespread tree mortality and large scale ecological consequences. Within MPB's historic range in south-central British Columbia, lodgepole pine, *Pinus contorta* var. *latifolia* has been a major host for MPB. However, during the course of the outbreak, the beetle has spread eastwards across the Rocky Mountain barrier into Alberta (Safranyik & Carroll 2006). In 2011, MPB was reported to have undergone a tree host species expansion from lodgepole pine to jack pine, *Pinus banksiana,* found in north central Alberta (Cullingham *et al.* 2011). Ecological studies have shown that tree hosts located beyond the historic range of MBP have lower host quality and invest fewer resources to defense than their counterparts that share a co-evolutionary history with MPB (Wu *et al.* 1996; Raffa *et al.*2013). MPB was also reported as having a higher rate of reproductive success in lodgepole pine found outside of MPB's historic range (Cudmore *et al.* 2010). Furthermore, ecological studies suggest that trees subjected to abiotic stresses such as drought

are more susceptible to MPB attack (Breshears *et al.* 2009; McDowell *et al.* 2008; Safranyik *et al.* 2010). Northern Alberta has been subjected extended periods of drought over the past two decades, which have negatively impacted forests, and which may increase risk of MPB infestation in these regions (Chhin *et al.* 2008, Michaelian *et al.* 2011, Hogg & Michaelian 2015). Due to the size and scale of the current MPB outbreak, management practices must be targeted to areas of higher risk. Therefore, efforts to identify resistant and susceptible pine stands have become critical in locating stands at higher risk of infestation. Additionally, understanding the impact of drought on host defense response may also become an important factor in predicting stand susceptibility.

MPB vectors a number of fungal associates, which are thought to play an important role in suppressing tree host defenses and providing nutrition for the beetle (Safranyik & Carroll 2006). The most pathogenic MPB fungal associate is *Grosmannia clavigera*, which grows into tree hosts xylem tissue, blocking water and mineral transport, and contributing to eventual tree host mortality (Solheim & Krokene 1998; Lee *et al.* 2006; Ballard *et al.* 1984). Early response to pathogen challenge is critical in effective containment, and induced chitinase expression may serve as reporters or "sentinels" of early tree host defense response. The timing and magnitude of chitinase expression is correlated with the relative degree of pathogen containment in conifers. In slash pine, *Pinus elliottii*, inoculated with a wound pathogen *Fusarium subglutinans,* responsible for pitch canker disease, class II chitinases were expressed earlier and at lower levels in resistant genotypes in slash pine relative to susceptible genotypes (Davis *et al.* 2002). In Norwary spruce inoculated with the root rot pathogen *Heterobasidion annosum,* class II and IV chitinases were upregulated earlier in resistant clone lines, and later but at greater magnitude than in clone lines determined to be susceptible (Heitala *et al*. 2004).

Chitinases show high levels of allelic variation across plant species, with a rate of non synonymous mutations often located within the catalytic cleft, that exceeds the rate of non synonymous substitutions in other gene families and work on western white pine chitinases has demonstrated that they display high levels of haplotype diversity (Bishop 2000, Liu *et al.* 2013). Work done in western white pine, *Pinus monticola,* identified two unique isozymes of a class IV chitinas present in seedlings resistant to *Cronartium ribicola*, but absent in susceptible seedlings (Liu *et al.* 2011). It has been proposed that chitinases undergo adaptive mutation to overcome the diverse forms of inhibitors produced by some fungal invaders.

We hypothesize that chitinases play and important role in pine defense response to *G. clavigera,* and expression of some chitinases will be highly induced in response to inoculation, but induction will be attenuated under water deficit conditions. We further hypothesize that co-evolution between lodegpole pine and the MPB fungal associate *G. clavigera* has placed selective pressure on lodgepole pine but not jack pine chitinases, and that this difference will be evident in spatially explicit patterns or allelic variation with in chitinase genes across the ranges of lodgepole and jack pine. To address these hypotheses, we aimed to (1) describe the role of chitinase enzymes in lodgepole and jack pine response to MPB fungal associate *G. clavigera,* and (2) determine the extent of allelic variation for a subset of chitinase genes from jack and lodgepole pine sampled from across their ranges. Specifically, we have identified members of the chitinase gene family in lodgepole and jack pine using sequence data from multiple transcriptomic data sets, and characterized these sequences through phylogenetic and other *in silico* analyses. We then identified a subset of *G. clavigera*-responsive lodgepole and jack pine chitinases from a large microarray experiment dataset, and determined the effect of water deficit on expression profiles for these defense-associated chitinases. Finally, we assessed allelic

variation of class I, II, IV and VII chitinases in jack and lodgepole pine seedlings sampled from 11 pine provenances samples from across Canada.

3.2 Materials and Methods

3.2.1 Data mining and sequence extraction

Putative chitinase sequences were mined from in house Illumina, 454 Roche, and Sanger transcriptome data using as queries previously identified chitinases from lodgepole pine, jack pine and white spruce, to query assembled jack pine and lodgepole pine transcriptomes for candidate chitinase genes using tBLASTn (Hall *et al.* 2013; Kolosova *et al.* 2014; González *et al.* 2015) (Appendix 6.1). BLAST hits with a return value e-value< 10E-10 were classified as candidate sequences. Reciprocal BLASTs were performed between species in order to identify potential orthologs shared between lodgepole and jack pine (Ziemann *et al.* 2013). Putative chitinase sequences were also identified using tBLASTn for loblolly pine (*Pinus taeda*) from NCBI dbEST using lodgepole pine, jack pine and white spruce putative chitinase queries.

Redundant gene sequences within each species were removed using USEARCH (Edgar 2010). Sequences with > 96% sequence identity were designated as originating from the same loci, and the longest sequence within a cluster was extracted as the representative sequence of that loci. Open reading frames and resulting amino acid sequences were predicted using ORF predictor (Min *et al.* 2005). Sequences were further filtered using BLASTp to compare predicted amino acid sequences against NCBI nr database, removing sequences that did not return convincing chitinase hits.

Lodgepole and jack pine predicted amino acid chitinase sequences were aligned to a full length representative class I chitinase from tobacco described in Neuhaus *et al.* 1991 using a MAFFT version 7 alignment (gap penalty -3.0) followed by manual adjustment in Mesquite version 3.10 (Katoh *et al.* 2002; Maddison & Maddison 2010). Amino acid sequences were assigned to appropriate biochemical classes based on the presence or absence of highly conserved domains.

3.2.2 Biochemical classification and phylogenetic analysis

Predicted amino acid sequences from lodgepole pine, jack pine, loblolly pine, *Arabidopsis thaliana*, white spruce (*Picea glauca*), and Douglas-fir (*Pseudotsuga menziesii*) shown in Appendix 6.2 were aligned using MAFFT multiple sequence alignment online tool version 7 implementing the L-INS-i method and instructing the program to implement the "leave gappy regions" parameter (Katoh *et al.* 2002). Protest version 2.4 was used to determine the most suitable model of heterogeneity (Gamma) and amino acid substitution model (WAG) based on their Bayesian information criterion score (BIC) (Abascal *et al.* 2005; Yang 1994; Whelan & Goldman 2001). RaxML version 8 was used to construct a maximum likelihood tree was built using these models, with 100 bootstrap iterations (Stamatakis *et al.* 2014). Putative orthologs between jack and lodgepole pine were identified on the basis of proximal positions within the phylogenetic tree. Jack and lodgepole pine chitinases were named based on the pine chitinase naming conventions introduced in Kolosova *et al.* (2006).

3.2.3 Transcript abundance profiling

Expression profiles for members of the lodgepole and jack pine chitinase family were mined from the large microarray dataset described in Chapter 2. Lodgepole and jack pine chitinase sequences above were used to identify chitinase probes on the loblolly pine PtGen microarray used for transcriptome profiling experiments by BLAST against the original loblolly pine sequences representing each probe, as well as against the corresponding lodgepole and jack pine sequences from Hall *et al.* (2013) that had been used to annotate the PtGen array (Lorenz *et al.* 2009). Fold-change expression data between control seedlings and seedlings inoculated with *G. clavigera* across 1, 7 and 28 days post inoculation (dpi) under well watered and water deficit conditions was extracted for each of the chitinases represented on the array. A Log_2 transformation was performed on fold change (FC) data, and used to create a heat map in multi-experiment viewer (MeV) (Saeed *et al.* 2003). Heatmap data were superimposed onto phylogenetic trees.

3.2.4 Gene Cloning

Genomic chitinase sequences for three putative orthologous gene pairs of chitinases (*Pcchia1-1 and Pbchia1-1, Pcchia2-1 and Pbchia2-1, and Pcchia4-1* and *Pbchia4-1)* were cloned from lodgepole pine (Pc) and jack pine (Pb) genomic DNA (gDNA) using primers designed from jack pine CDS (Supplemental table 6.3 & Appendix 6.4). Promoter regions were cloned separately using primers designed against loblolly pine genomic sequence accessed through congenie.org (Zimin *et al.* 2014; Nystedt *et al.* 2013) (Table 6.1). Gene targets were amplified using PCR with standard Taq DNA polymerase (NEB, Ipswich, MA USA) under the following cycling parameters: 95°C 5 min, 95 °C 30 sec, 52 °C 30 sec, 68 °C 1 min 30 sec x 25

cycles. PCR products were purified using the GeneJet PCR purification kit (Thermo Fisher Scientific, Waltham, MA, USA), ligated into the pGEMT-easy vector system (Promega, Madison, WI, USA) and transformed into DH-5 α cells (Thermo Fisher Scientific, Waltham, MA, USA) using heat shock in 42 °C water bath. Inserts were sequenced using the BigDye terminator sequencing system (Thermo Fisher Scientific, Waltham, MA, USA) and sequenced on a 3730 DNA analyzer (Thermo Fisher Scientific, Waltham, MA, USA) (Appendix 6.4).

3.2.5 Seedling germination

Seeds originating from 12 different sampling location representing 11 different provenances were obtained from the National Tree Seed Centre (NTSC), Natural Resources Canada, Canadian Forest Service (Fig 3.1).

Lodgepole pine (*Pinus contorta var. latifolia*) provenances were derived from Minto YK (62.60000°, -136.83330°), Baldy Hughes BC (53.66667°, -122.95000°), Edgewood BC (49.91667°, -118.18330°), Cypress Hills AB (59.63981°,-109.98333°) and Nose Mountain AB (54.63333°, -119.11667°). Jack pine (*Pinus banksiana*) provenances were derived from Stoney Mountain AB (56.26670°, -111.60000°), Creighton SK (54.85000°, -102.41670°), Hudson Bay SK (52.98333°, -102.60000°), seed zone 8 Weagamow lake ON (51.00000°, -90.00000°), seed zone 37 London ON (43.00000°,- 81.00000°), and Despres Lake NB (46.65000°, -65.56667°). One lodgepole-jack pine hybrid, *Pinus contorta x Pinus banksiana,* provenance was included from Blue Ridge AB (54.1000°, -115.53333°).

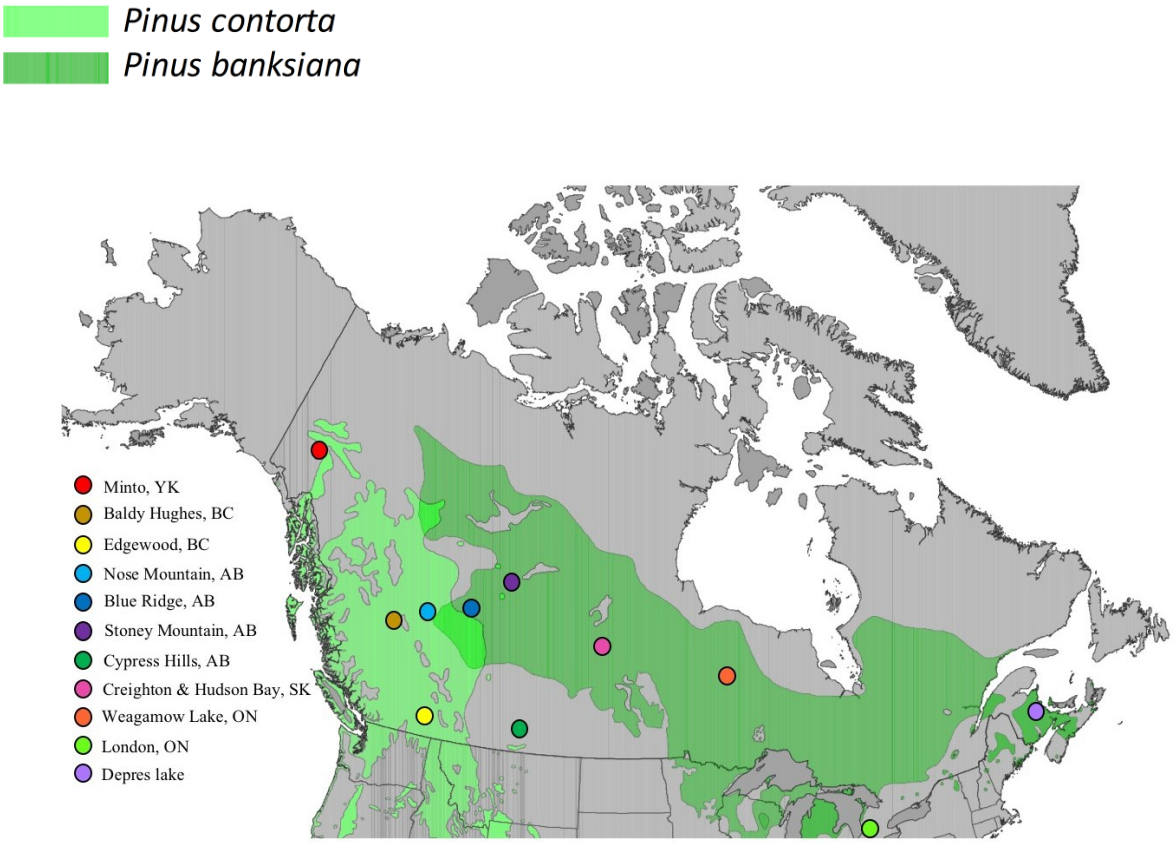


Figure 3.1 Locations of the provenances used for allelic resequencing, superimposed upon the ranges for lodgepole pine and jack pine. Lodgepole pine: light green. Jack pine: dark green. Creighton and Hudson Bay provenances were combined into one location for analyses.

Seeds were hydrated under running water for 30 min, then wetted in a diluted solution of TWEEN-20, and rinsed under running water for an additional 30 mins. Seeds were immersed in a 1% (v/v) sodium hypochlorite (bleach) solution for 5 min, and rinsed under running water. Seeds were stratified at 4 °C for three weeks between two dampened sterile Kimpads inside clamshell containers. After stratification, seeds were sterilized in 1% (v/v) sodium hypochlorite (bleach) solution for 10 min, rinsed with deionized water, then transferred to fresh water-

dampened Kimpads in sterile seed germination trays. Seedlings were germinated at 25 °C with ca. 75% humidity under 12 h light / 12 h dark photoperiod in controlled environment growth chambers. Once cotyledons had elongated but megagametophytes were still attached to the seedlings, megagametophytes and seedlings were separated, flash frozen in liquid nitrogen, and stored at -80°C until processing.

3.2.6 DNA extraction

DNA was extracted from seedling tissue using a CTAB (hexadecyl-trimethyl ammonium bromide) protocol modified from Chang *et al.* (1993) and Roe *et al.* (2010). Seedlings were quickly crushed in 1.5 mL Eppendorf tubes using small pestles, and 400 µL CTAB, supplemented with 25 mAU proteinase K and 10 µg/mL RnaseA, was immediately added. Tissue was incubated for 1 hour at 65 °C, samples were centrifuged for 5 min at 14 000 rpm and supernatant was collected. A phase extraction using 500 µl of chloroform: isoamyl alchohol (24:1) added to the supernatant was performed twice, centrifuging at 14 000 rpm for 5 min each time and collecting the upper phase. 500 µl of ice cold isopropanol was added, samples were incubated at -20°C for 2 hours, and centrifuged at 14 000 rpm for 30 min at 4°C. Supernatant was removed and precipitates were washed with 95% ethanol followed by 70% ethanol and suspended in deionized water then quantified using the Tecan Nanoquant™. Samples were diluted to a standard concentration of 100 ng/µL.

3.2.7 Sanger sequencing

PCR products were generated from lodgepole and jack pine gDNA for the three chitinase genes (*chia1-1, chia2-1, chia4-1*) including promoter sequences using two set of external and internal primers designed against lodgepole and jack pine genomic clones described above (Figure 3.2; Appendix 6.2).

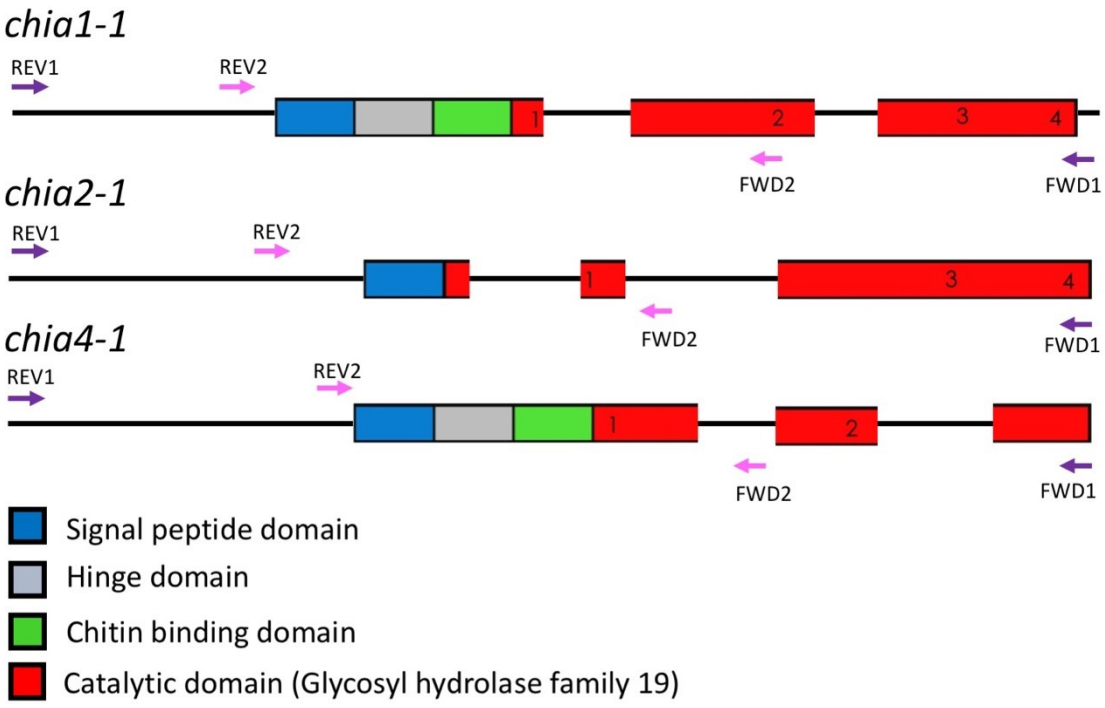


Figure 3.2: Schematic representation of two primer sets used to PCR amplify and Sanger sequence *Chia1-1, Chia2-1* and *Chia4-1* from lodgepole and jack pine seedlings gDNA. Two sets of primers were designed specifically for each gene. The first set of primers, FWD1 and REV1, (indicated in dark purple) were used to amplify an sequence the exterior flanking sequences of each gene and the second set of primers, FWD2 and REV2 (indicated in light purple) were used to amplify and sequence the interior sequences of each gene. See Appendix 6.2 for primer sequences.

PCR reactions were carried out in 96 well plates under the following cycling parameters: 95 °C 30 sec, 51 °C 30 sec, 68 °C 2.5 min x 35 cycles. PCR reactions were cleaned by supplementation with Exonuclease I and shrimp alkaline phosphatase (NEB, Ipswich, MA USA). PCR amplicons were used in forward and reverse cycling reactions using the BigDye^{TM} terminator system (Thermo Fisher Scientific, Waltham, MA, USA) under the following cycling parameters (96 °C 30 sec, 50 °C, 15 sec, 60 °C 2 min x 25 cycles). DNA was precipitated with 1.5 M sodium acetate/250 mM EDTA and 95% ethanol at 4 °C for 15 min. Samples were centrifuged at 2500 g for 30 min at 4 °C, the supernatant was removed and pellets were washed with 70% ethanol. The pellets were dried under vacuum and suspended in deionized water. Resuspended sequencing products were combined with formamide and sequenced on an Applied Biosystems 3730 DNA analyzer (Thermo Fisher Scientific, Waltham, MA, USA).

3.2.8 Data analysis

Sequences were initially trimmed using fastq trimmer as part of the Fastq Toolkit, formatted into fasta files and aligned using MAFFT version 7 (gap opening penalty of -3.0), and assembled by hand into contigs (Gordon *et al.* 2010; Katoh *et al.* 2002) (Appendix 6.7). Contigs were mapped to the reference genomic clones (comprising both the CDS and promoter) in CLC Genomics Server (Qiagen, Redwood City CA) and variants were called using CLC basic variant detection tool, using minimum coverage requirement of ten and minimum count requirement of two. Ambiguous base calls (N) at variant sites were resolved by returning to the original chromatogram output and calling the base with the highest peak of fluorescence.

Sequence motifs and exon-intron boundaries were identified in the genomic reference clone sequences based on pairwise sequence alignment with the coding sequence and resulting

amino acid sequence mined from transcriptomic data using the ebi EMBOSS water local alignment tool gap opening penalty set to -50 and gap extension penalty set -0.05 (Appendix 6.5; Rice *et al.* 2000). SwissProt was used to perform protein homology modeling of allelic variants carrying non-synonymous amino acid substitutions and insertion-deletions (Boekmann *et al.* 2003). Superpose was used to perform superpositioning of 3-D protein models (Maiti *et al.* 2004). SignalP version 4.1 was used to identify location of signal peptide and anlazye allelic variation within the signal peptide (Peterson *et al.* 2011). Allelic richness within each sampling location as well as within each species, and well as pairwise Fst values between sampling locations and between species were calculated across varients identified in each gene separately using FSTAT version 2.9.3 (Goudet 2001).

3.3 Results

3.3.1 In silico characterization of the lodgpole and jack pine chitinase gene families

We identified 42, 40, and 44 putatively unique chitinase expressed genes in the loblolly pine, lodgepole pine, and jack pine transcriptome assemblies, respectively (Appendix 6.2). Based on the presence or absence of defining motifs – signal peptide domain, hinge domain, chitin binding domain, and catalytic domain- we identified eight class I chitinase sequences in loblolly pine, four in lodgepole pine, and three in jack pine. Five chitinase sequences were classified as class II chitinases in loblolly pine, three in lodgepole pine, and four in jack pine. Ten class IV chitinase sequences were identified in loblolly pine, 12 in lodgepole pine, and ten in jack pine. Nine class VII chitinase sequences were identified in loblolly pine, ten in lodgepole pine, and 16 in jack pine. Among the GH 18 family, one class III chitinase sequence was identified in loblolly pine, while nine class V chitinase sequences were identified in loblolly, nine in lodgepole pine,

and seven in jack pine. No lodgepole or jack pine sequences shared enough sequence similarity with previously characterized class III chitinases to be confidently classified as such.

A maximum likelihood tree with 100 bootstraps, constructed from the full length deduced amino acid sequences from the pine chitinases together with previously characterized chitinases from *Arabidposis thaliana*, white spruce and Douglas fir yielded five predominant clades (Figure 3.3).

Cluster one contains, GH 18 chitinases assigned to class V and class III which form a distinct cluster separate from the GH 19 chitinases (Fig 3.3). Cluster two contains both class I and class II chitinases, with conifer and angiosperm chitinases grouping separately from each other (Fig 3.3). Cluster three contains members of class IV and cluster four contains memebers of class VII (Fig 3.3). Finally cluster 5 contains members of class II from lodgepole pine, jack pine and spruce which separate from class II chitinases found in cluster two (Fig 3.3).

3.3.2 Transcript abundance analysis of chitinase gene family

We next extracted transcript abundance profiles for members of the chitinase family from a large microarray dataset comparing *G. clavigera*-inoculated versus control lodgepole and jack pine seedlings grown under well watered or water deficit conditions and sampled at 1, 7 or 28 days post-inoculation (dpi). Heatmaps were generated, with fold change data arranged by treatment along the horizontal axis, and by phylogenetic relationships between sequences along the vertical axis (Fig 3.4). Only lodgepole and jack pine sequences were included in the maximum likelihood phologenetic tree used. Notably in this analysis the group of class II chitinases which formed a separate cluster (cluster 5, Fig 3.3) in the previous analysis using amino acid sequences, grouped together with other class I and class II chitinases.

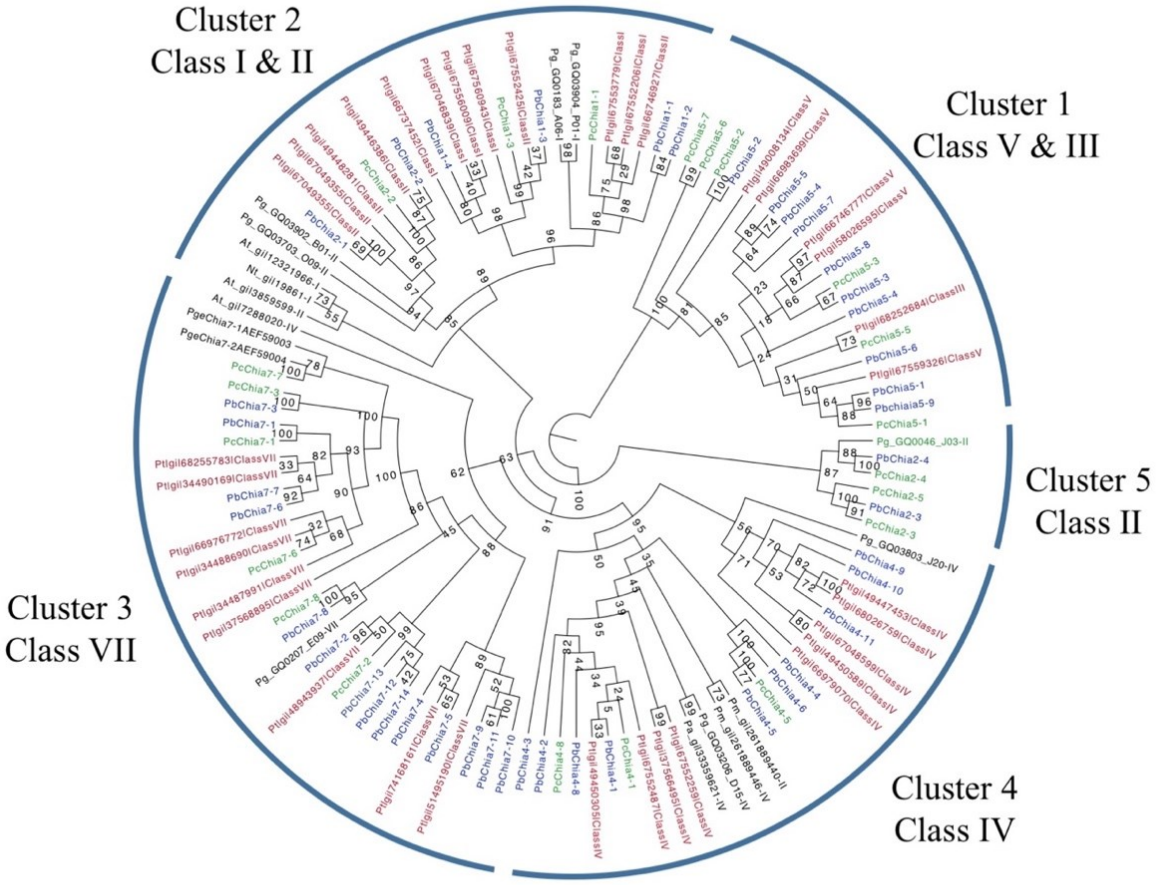


Figure 3.3: Maximum likelihood tree with 100 bootstraps displaying relationships between chitinases identified in this study from lodgepole pine (green), jack pine (blue), and loblolly pine (red) with previously identified chitinases from white spruce, Douglas-fir and *Arabidopsis thaliana* (black). Species are identified by the first two letters of their name: Pc, lodgepole pine (*Pinus contorta*); Pb, jack pine (*Pinus banksiana*); Pt, loblolly pine (*Pinus taeda*); At, *Arabidopsis thaliana*; Nt, *Nicotiana tabacum;* Pg, white spruce (*Picea glauca*), and Pm, Douglas-fir (*Pseudotsuga menziesii*). Species identifiers are followed by their GenBank ID in the case of Arabidopsis, Douglas-fir and loblolly pine, the full length cDNA insert (FLIC) ID for white spruce, and putatively assigned gene names for lodgepole and pine sequences available in supplemental table 1. The inferred structural class for each chitinase is indicated at the end of the sequence identifier.

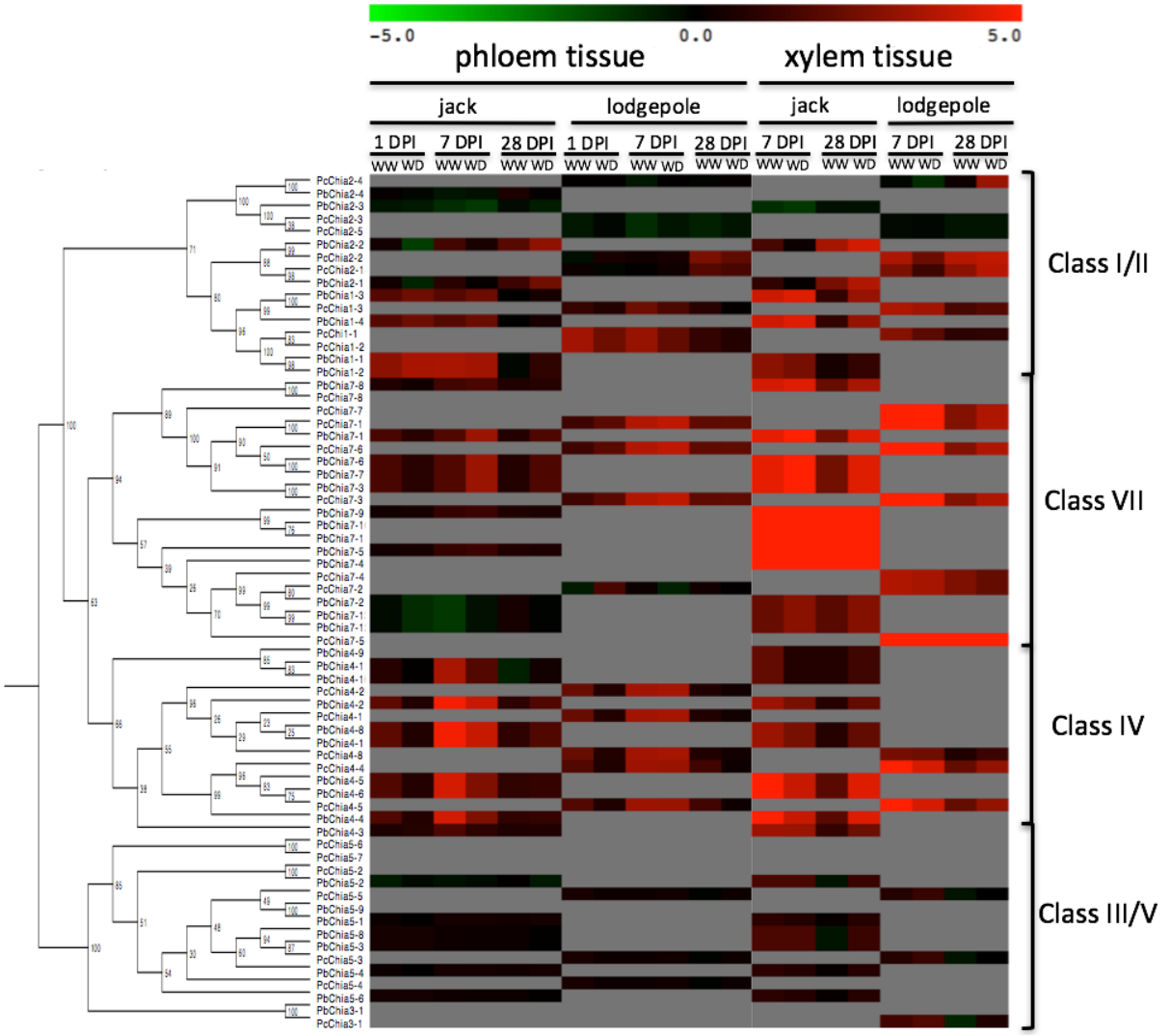


Figure 3.4: Changes in transcript abundance of chitinase genes in *G. clavigera*-inoculated vs. control jack (Pb) and lodgepole (Pc) pine at 1, 7 or 28 dpi, under either well watered or water deficit conditions. Data are represented as $\log_2$ fold-change between *G. clavigera*-inoculated and non-inoculated seedlings, with red representing upregulation, and green representing down regulation. Heatmap expression data is organized on the vertical axis as dedrogram depicting a maximum likelihood phylogeny of lodgepole and jack pine chitinases.

Phylogenetic representation of expression patterns revealed patterns of co-expression between related groups of chitinases (Fig 3.4). Class III and class V chitinase members of GH 18 displayed no significant up or down regulation in response to *G. clavigera* in either xylem or phloem (Fig 3.4). Class I chitinases in lodgepole and jack pine phloem displayed strong and statistically significant upregulation at 1 dpi in and sustained upregulation until 7 dpi and lesser but still significant upregulation at 28 dpi (Fig 3.4). In xylem, class I chitinases displayed a similar pattern of significant at upregulation at 7 dpi, and lesser, but still significant, upregulation at 28 dpi (Fig 3.4). Two distinct expression pattern groups were observed for class II chitinases that correlated with phylogenetic relationships. One group displayed slight non significant down regulation in response to *G. clavigera* in jack pine and lodgepole pine phloem and xylem, while the other displayed upregulation at 7 and 28 dpi in phloem and xylem (Fig 3.4). Class IV chitinases in both jack and lodgepole pine displayed concerted and significant upregulation at 7 dpi in phloem and xylem (Fig 3.4). Similar to class II chitinases, two groups of class VII chitinases could be distinguished, with each group displaying patterns of co-expression unique to these two groups (Fig 3.4). In jack and lodgepole pine phloem, one group of class VII chitinases was non-significantly upregulated at 1 dpi and significantly pregulated at 7 dpi, while the other group displayed slight non significant down regulation (Fig 3.4). Within jack and lodgepole pine xylem, both groups of class VII chitinases were highly up regulated at 7 and 28 dpi (Fig 3.4).

Differences in expression levels between *G. clavigera*-inoculated vs control seedlings under well watered or water deficit conditions were evident across all GH 19 chitinases. Water deficit led to a marked decrease in expression of class IV chitinases at 7 dpi (Fig 3.4). Class I chitinases demonstrated similar patterns of reduced expression at 7 dpi, however at 28 dpi, expression levels were higher under water deficit conditions (Fig 3.4). At 7 and 28 dpi, some

jack pine class VII chitinases displayed increased expression in phloem under water deficit, to a greater extent than their lodgepole pine counterparts in phloem (Fig 3.4). Both jack and lodgepole pine class VII chitinases displayed higher levels of expression under water deficit conditions at 28 dpi in xylem (Fig 3.4). qRT-PCR analysis of select chitinases in phloem tissue confirmed the expression patterns described in microarray data, and is shown in Appendix 6.8.

3.3.3 Allelic variation of chitinase genes

We selected representatives of class I, class IV (*Pcchia1-1 & Pbchia1-1, Pcchia4-1 & Pbchia4-1)* in order to characterize allelic variation in chitinase classes which were highly induced in response to *G. clavigera,* and we selected representative gene of class II chitinases (*Pcchia2-1 & Pbchia2-1)* to characterized allelic variation in a class of chitinases which were not induced in response to inoculation. Allelic re-sequencing of putative orthologous pairs of class I, class II and class IV chitinases in lodgepole and jack seedlings revealed sequence variation within promoter and coding sequences.

CHIA1-1

Sequencing of *chia1-1* in lodgepole and jack pine displayed allelic variation at nine locations in the 888 bp immediately upstream up the transcriptional start site (Table 3.1). Of these nine sites of variation, eight posed no effect on the predicted transcriptional factor binding sites. At one site, Indel -476, insertion of the nucleotide sequence GCTTAC results in a MYB transcription factor biding site (MBS). Within the coding sequence of *chia1-1,* we identified three sites of allelic variation (Table 3.1). The first site of variation was an insertion-deletion found 199 bp after the transcriptional start site, and was located in the hinge region between the

Table 3.1: Variants identified in jack and lodgepole pine promoter sequence of class I chitinase, *chia1-1,* sequenced from gDNA of seedlings from 12 provenances collected across Canada. Allele counts specific to each sample location indicated below total allele count. MBS: MYB transcription factor binding site. Total individuals: total number of individual seedlings sequenced per sample location. SNV: single nucleotide variant. Indel: insertion/deletion

Sequence position		-706		-671		-645		-616		-529		-485		-476		-455		-170	
Type		SNV		SNV		SNV		SNV		SNV		SNV		Indel		SNV		SNV	
Coverage		96		100		91		89		73		85		97		94		187	
Alleles		A	G	A	G	C	T	T	C	C	G	T	A	Insert CGTTAC	Del	C	G	G	C
Functional Change		none		none		None		none		none		none		MBS	Loss of MBS	none		none	
Total allele count		89	7	92	8	89	2	84	5	52	21	78	7	82	15	81	13	67	120
Sample location	Total individuals																		
Minto YK	18	13	0	12	1	12	1	13	0	11	1	10	0	11	0	11	2	10	8
Baldy Hughes BC	13	6	0	9	0	7	1	7	0	3	0	4	0	5	8	19	2	7	6
Edgewood BC	17	9	1	7	1	8	0	8	1	4	1	10	0	10	0	10	0	8	9
Creighton SK	6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
Hudson Bay SK	11	4	1	4	1	4	0	2	0	3	1	3	1	2	1	3	1	1	10
London ON	18	7	3	10	2	9	0	8	3	4	5	6	2	7	2	4	2	4	14
Weagamow Lake ON	18	7	1	7	1	8	0	8	0	5	1	7	1	7	2	7	2	1	17
Nose Mountain AB	16	15	0	16	0	15	0	13	0	8	5	14	1	14	0	9	1	8	6

Sequence position		-706		-671		-645		-616		-529		-485		-476		-455		-170	
Type		SNV		SNV		SNV		SNV		SNV		SNV		Indel		SNV		SNV	
Coverage		96		100		91		89		73		85		97		94		187	
Alleles		A	G	A	G	C	T	T	C	C	G	T	A	Insert CGTTAC	Del	C	G	G	C
Functional Change		none		none		None		none		none		none		MBS	Loss of MBS	none		none	
Total allele count		89	7	92	8	89	2	84	5	52	21	78	7	82	15	81	13	67	120
Blue Ridge, AB	16	9	0	9	0	9	0	9	0	7	1	9	0	9	0	8	0	6	11
Stoney Mountain AB	18	3	1	5	0	5	0	4	1	2	2	5	1	5	1	4	1	4	14
Cypress Hills AB	18	7	0	7	0	7	0	7	0	4	1	5	1	6	1	2	1	15	4
Despres Lake NB	18	8	0	6	2	5	0	5	0	1	3	5	0	6	0	4	1	3	15

signal peptide and the chitin binding domain. Thirty nine individuals contained an insertion of the nucleotide sequence CCACGCCTCCTTCAC which resulted in an insertion of proline-tyrosine-proline-proline-serine-proline amino acid residues. Two individuals carried a shorted insertion at this site of the nucleotide sequence CTTCAC, resulting in an insertion of serine-proline amino acid residues, and 141 individuals carried a complete deletion at this site (Table 3.2).

Protein homology models were built for the two amino acid sequences, one containing the full proline-tyrosine-proline-proline-serine-proline insertion and the other containing the corresponding deletion. Superimposing the two resulting three-dimensional protein models demonstrated that the insertion conferred an extension of the hinge domain but did not alter the structure of the flanking chitin binding domain or catalytic domain (Fig 3.5).

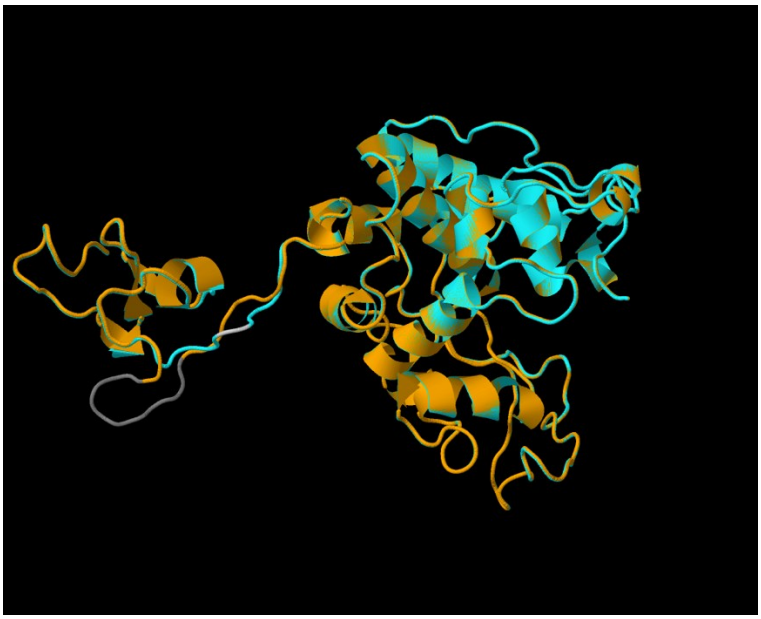


Figure 3.5: Superpositioning of two class one, chia1-1, alleles. Orange: chia-1-1 carrying a proline-tyrosine-proline-proline-serine-proline in the hinge domain. Blue: chia-1-1 carrying a deletion in this position. Hinge domain extends between the chitin binding domain (left) and the catalytic domain (right). Extended hinge domain displayed in grey.

Table 3.2: Variants identified in jack and lodgepole pine coding sequence of class I chitinase, *chia1-1,* sequenced from gDNA of seedlings from 12 provenances collected across Canada. Allele counts specific to each sample location indicated below total allele count. Total individuals: total number of individual seedlings sequenced per sample location. SNV: single nucleotide variant. Indel: insertion/deletion

Sequence position		199			324		357	
Type		Indel			SNV		SNV	
Location		Hinge domain			Catalytic domain		Catalytic domain	
Coverage		182			184		191	
Alleles		Insert CCACGCCTCCTTCAC	Insert CTTCAC	Deletion	C	T	T	C
Functional changes		PTPPSP insertion	SP Insertion	deletion	silent	silent	silent	silent
Total allele count		39	2	141	65	119	65	126
Sample Location	Total individuals							
Minto, YK	18	4	0	14	12	6	12	6
Baldy Hughes, BC	13	2	1	10	4	9	3	10
Edgewood, BC	18	4	0	14	3	14	6	12
Creighton, SK	6	2	0	4	2	4	2	4
Hudson Bay, SK	12	4	1	7	2	9	4	7
London, ON	18	6	0	12	3	15	6	12
Weagamow Lake, ON	18	2	0	13	4	13	0	16
Nose Mountain, AB	16	2	0	13	8	8	5	12
Blue Ridge, AB	16	3	0	13	6	11	3	13
Stoney Mountain, AB	18	4	0	13	3	14	4	15
Cypress Hills, AB	18	1	0	17	17	1	15	5
Despres Lake, NB	18	5	0	11	1	15	5	14

The other two allelic variants found in the coding region of *chia1-1* are SNVs (Single nucleotide variants) located 324 and 357 bp from the start site. Both were located in the catalytic domain, and were synonymous mutations resulting in no amino acid substitutions (Table 3.2)

Variants represented by greater than 150 individuals (i.e. coverage > 150), were used to construct haplotypes observed for each individual. We observed 16 unique haplotypes among the 179 individual seedlings sequenced (Table 3.3)

Table 3.3: Unique haplotypes observed within the class I, *chia1-1,* chitinase gene in 179 individual seedlings. Variants are indicated by their type, and sequence location with respect to the transcriptional start site. SNV: single nucleotide variation, InDel: insertion/deletion. Variants are described in Table 3.1 and 3.2

Haplotype	SNV -170	Indel 199	SNV 324	SNV 357
Hap 1	C	Deletion	T	C
Hap 2	G	Deletion	C	T
Hap 3	C	CCACGCCTCCTTCAC	T	C
Hap 4	G	Deletion	T	C
Hap 5	G	Deletion	C	C
Hap 6	G	Deletion	T	T
Hap 7	C	Deletion	C	C
Hap 8	C	Deletion	C	T
Hap 9	C	Deletion	T	T
Hap 10	C	CCACGCCTCCTTCAC	T	T
Hap 11	C	CCACGCCTCCTTCAC	C	C
Hap 12	C	CCACGCCTCCTTCAC	C	T
Hap 13	G	CCACGCCTCCTTCAC	T	C
Hap 14	C	CTTCAC	T	C
Hap 15	G	CCACGCCTCCTTCAC	C	C
Hap 16	C	CTTCAC	T	T

The most abundant haplotype is haplotype one (Hap 1) which appeared to be predominately found in seedlings from jack pine populations of Stoney Mountain AB, Creighton and Hudson Bay SK, Weagamow Lake ON, London ON and Despres Lake and to a lesser extent

Nose Mountain AB, and Blue Ridge AB (Table3.4: Fig 3.6). Haplotype two (Hap 2) was the second most abundant haplotype and appeared mostly in seedlings from Minto YK, and Cypress Hills (Table3.4: Fig 3.6).

Jack pine sample locations displayed the greatest allelic richness in *chia1-1* across the four variant sites found in *chia1-1* in comparison to lodgepole sample sites and the lodgepole x jack hybrid site (Table 3.5). However, examination of allelic richness with in individual sampling sites revealed lodgepole pine samples found in Baldy Hughes carried the greatest allelic richness in relation to other sampling sites (Appendix 6.6.1). Estimation of pairwise Fst values between lodgepole and jack pine samples revealed very little differentiation between species (Table 3.6). Pairwise Fst values between each sampling location revealed moderate to great differentiation between samples from Minto and Cypress Hills compared to other sampling locations (Appendix 6.6.2).

Table 3.4: Counts of haplotypes identified in jack and lodgepole pine class I chitinase, *chia1-1,* sequenced from gDNA of seedlings from 12 provenances collected across Canada. Haplotypes are described in table 3.3.

Sample locations	Minto YK	Baldy Hughes BC	Edgewood BC	Creighton SK	Hudson Bay SK	London ON	Weagamow Lake ON	Nose Mountain AB	Blue Ridge AB	Stoney Mountain AB	Cypress Hills AB	Despres Lake NB
Total individuals	18	13	17	6	11	18	15	14	16	17	18	16
Haplotype count												
Hap 1	2	3	5	2	5	7	11	4	5	7	0	5
Hap 2	7	1	0	0	0	1	0	2	1	0	11	0
Hap 3	1	0	2	1	1	3	1	0	1	1	0	4
Hap 4	0	3	2	0	0	1	0	2	2	3	0	2
Hap 5	1	1	1	0	0	1	1	3	1	1	2	0
Hap 6	1	1	5	0	1	1	0	0	0	0	0	1
Hap 7	1	0	1	0	1	0	1	0	2	0	1	1
Hap 8	2	0	0	1	0	0	0	2	0	0	3	0
Hap 9	0	1	0	1	0	1	0	0	2	2	0	1
Hap 10	1	0	0	0	1	2	0	0	0	1	0	1
Hap 11	0	1	0	1	0	0	1	0	1	2	0	0
Hap 12	1	0	1	0	1	1	0	0	0	0	0	0
Hap 13	1	0	0	0	0	0	0	1	0	0	1	0
Hap 14	0	1	0	0	0	0	0	0	0	0	0	1
Hap 15	0	1	0	0	0	0	0	0	1	0	0	0
Hap 16	0	0	0	0	1	0	0	0	0	0	0	0

Figure 3.6: Map showing the distribution of unique haplotypes observed in the class I, *chia1-1*, chitinase for each of the 12 sampling locations across Canada. Sampling locations Creighton SK and Hudson Bay SK have been combined to represent on location in this map. Total number of individuals sequenced within each population are indicated as "N=" and are found above each pie chart. Haplotype counts specific to each sampling location described in table 3.4.

Table 3.5: Allelic richness based on four variant loci in jack and lodgepole pine class I chitinase, *chia1-1,* sequenced from gDNA of seedlings from 12 provenances collected across Canada. Seedlings sampled from Minto, Edgewood, Baldy Hughes, and Nose Mountain are designated as lodgepole. Seedlings sampled from Stoney Mountain, Cypress Hills, Hudson Bay, Weagamow lake, London, and Despres Lake are designated as jack pine, and seedlings sampled from Blue Ridge are designated as lodgepole x jack pine hybrids.

	Lodgepole pine	Jack pine	Lodgepole x jack pine hybrid	Overall
SNV -170	2.00	2.00	2.00	2.00
Indel 199	2.41	2.46	2.00	2.34
SNV324	2.00	2.00	2.00	2.00
SNV357	2.00	2.29	2.00	2.17

Table 3.6: Matrix of pairwise Fst values for *chia1-1* among pine seedlings designated as lodgepole pine, jack pine and lodgepole x jack pine hybrids. Fst values were estimated on the basis of four variant loci sequenced from gDNA of seedlings from 12 provenances collected across Canada. Seedlings sampled from Minto, Edgewood, Baldy Hughes, and Nose Mountain were designated as lodgepole. Seedlings sampled from Stoney Mountain, Cypress Hills, Hudson Bay, Weagamow Lake, London, and Despres Lake were designated as jack pine, and seedlings sampled from Blue Ridge were designated as lodgepole x jack pine hybrids. Negative Fst values corrected to be values of 0.

	Lodgepole pine	Jack pine	Lodgepole x jack pine hybrid
Lodgepole pine	0.000	0.045	0.009
Jack pine	0.045	0.000	0.000
Lodgepole x jack pine hybrid	0.009	0.000	0.000

CHIA2-1

Allelic resequencing of the class II chitinase, *chia2-1*, revealed fewer sites of variation within the promoter in comparison to *chia1-1.* We identified two SNVs within the promoter region, including a SNV located 284 bp upstream of the transcriptional start site (Table 3.7). Ninety-seven individuals carried a T at this location, which contributed to a CAT-box core promoter element, and 70 individuals carried a C which resulted in a loss of the CAT-box element (Table 3.7).

We identified an additional SNV located 141 bp upstream of the start site. Ninety eight individuals carried an A at this location, which contributed to a TATA-box core promoter element, and 67 individuals carried a C at this location which resulted in a loss of the TATA-box core promoter element (Table 3.7). An insertion-deletion was identified 67 bp after the transcriptional start site of *chia2-1* within the signal peptide domain. Thirty six individuals carry a TGT insertion at this site resulting in insertion of a cysteine residue, and 129 individuals carry a deletion at this site (Table 3.7). Signal peptide analysis of putative amino acid sequences

containing either and insertion or a deletion using SignalP version 4.1, revealed that the additional cysteine residue at amino acid position 24 occurred before the predicted cleavage site. As such, it is likely cleaved off during post translational modification and therefore does not likely affect protein confirmation. Interestingly, insertion of a cysteine residue at position 24 increased the calculated Y score at this position, which determines the likelihood of a cleavage site. However, the Y score at position 28 was still higher, making it more likely to be the point of cleavage (Appendix 6.9).

Finally, we identified a SNV located 413 bp after the start site, located within loop 1 of the catalytic domain of *chia2-1* (Table 3.7). One hundred and twenty-three individuals carried a T at this location and 55 carried a C. This SNV was a silent mutation (Table 3.7). All four variants identified in *chia2-1* coding region and upstream promoter region had coverage greater than 150 sequences. From this, we identified 12 unique haplotypes from a total of 167 individual seedlings (Table 3.8). Haplotype one (Hap 1) was the most abundant for *chia2-1*, which occured predominantly in western Canada provenances, including lodgepole pine sampled near Minto YK, Baldy Hughes BC, Edgewood BC and Nose Mountain AB, the lodgepole x jack pine hybrids found in Blue Ridge AB, and the lodgepole pine found in Cypress Hill AB (Table 3.9: Fig 3.7). Haplotype 2 (Hap 2) was the second most abundant haplotype and appeared predominantly in the sample locations of eastern Canada, including jack pine sampled near Creighton and Hudson Bay SK, Weagamow Lake ON, London ON and Despres Lake NB (Table 3.9: Fig 3.7). Other abundant haplotypes occurred along similar pattern of east-west distribution. Haplotype three (Hap 3) occurred frequently in Creighton and Hudson Bay SK, Weagamow Lake ON, Despres Lake NB and to a lesser extent (only one occurrence) in London ON and Baldy Hughes BC (Table 3.9: Fig 3.7).

Table 3.7: Variants identified in jack and lodgepole pine promoter and coding sequence of class II chitinase, *chia2-1,* sequenced from gDNA of seedlings from 12 provenances collected across Canada. Allele counts specific to each sample location indicated below total allele count. Total individuals: total number of individual seedlings sequenced per sample location. SNV: single nucleotide variant. Indel: insertion/deletion.

Sequence position		-284		-141		67		413	
Type		SNV		SNV		Insertion-deletion		SNV	
Location		promoter		promoter		N-terminal signal peptide		Loop 1 Catalytic domain	
Coverage		167		165		165		178	
Alleles		T	C	A	C	TGT	Deletion	T	C
Functional Changes		CAT-box	loss of CAT	TATA-box	loss of TATA-box	insertion of C residue	deletion of C residue	silent mutation	silent mutation
Total Allele count		97	70	98	67	36	129	123	55
Sample location	**Total individuals**								
Minto, YK	18	4	10	14	0	2	13	15	3
Baldy Hughes, BC	10	3	6	9	0	1	8	6	4
Edgewood, BC	14	5	8	11	2	2	11	10	4
Creighton, SK	6	6	0	1	5	0	6	4	2
Hudson Bay, SK	11	9	2	4	7	4	7	7	4
London, ON	18	14	3	5	12	8	9	7	11
Weagamow Lake, ON	18	16	1	3	12	7	9	9	10
Nose Mountain, AB	16	3	11	13	1	1	13	15	1
Blue Ridge, AB	16	4	12	13	3	2	14	13	3
Stoney Mountain, AB	18	16	1	1	16	2	15	14	3
Cypress Hills, AB	18	4	14	18	0	0	18	14	4
Despres Lake, NB	15	13	2	6	9	7	6	9	6

Table 3.8: Unique haplotypes observed within the class II, *chia2-1,* chitinase gene in 167 individual seedlings. Varients are indicated by their type, and sequence location with respect to the transcriptional start site. SNV: single nucleotide variation, InDel: insertion/deletion. Variants are described in table 3.5.

Haplotype	SNV -184	SNV -141	Indel 67	SNV 413
Hap 1	C	A	deletion	T
Hap 2	T	C	deletion	T
Hap 3	T	A	TGT	C
Hap 4	T	A	deletion	T
Hap 5	C	A	TGT	C
Hap 6	T	C	deletion	C
Hap 7	T	C	TGT	C
Hap 8	C	A	deletion	C
Hap 9	T	A	deletion	C
Hap 10	C	C	deletion	T
Hap 11	T	A	TGT	T
Hap 12	T	C	TGT	T

Haplotype four (Hap 4) occurred in Minto YK, Edgewood BC, Nose Mountain AB, Blue Ridge AB and Cypress Hills (Table 3.9: Fig 3.7). Sample locations representing jack pine and the lodgepole jack pine hybrid site displayed moderately greater allelic variation in *chia2-1* across these four variant sites in comparison to lodgepole pine sample sites (Table 3.10). However, examination of allelic variation within sampling locations revealed little to no difference in allelic richness between locations (Appendix 6.6.3). Interestingly, in contrast to the variation identified in *chia1-1,* calculation of pairwise Fst values for *chia2-1* between lodgepole and jack pine samples revealed substantial genetic differentiation between species (Table 3.11). This differentiation between species was reflected in pairwise comparisons between sampling locations representing either species (Appendix 6.6.4).

Table 3.9: Counts of haplotypes identified in jack and lodgepole pine class II chitinase, *chia2-1,* sequenced from gDNA of seedlings from 12 provenances collected across Canada. Haplotypes are described in Table 3.8

Sample location	Minto, YK	Baldy Hughes, BC	Edgewood, BC	Creighton, SK	Hudson Bay, SK	London, ON	Weagamow Lake, ON	Nose Mountain, AB	Blue Ridge, AB	Stoney Mountain, AB	Cypress Hills, AB	Despres Lake, NB
Total individuals	14	9	13	6	11	17	17	14	16	17	18	15
Haplotype count												
Hap 1	9	5	5	0	0	0	0	10	9	0	12	0
Hap 2	0	0	2	4	7	6	9	1	1	13	0	5
Hap 3	1	1	0	0	2	1	4	0	0	1	0	3
Hap 4	3	0	3	0	0	1	0	1	2	0	2	0
Hap 5	0	0	2	0	2	3	1	0	2	0	0	2
Hap 6	0	0	0	1	0	2	1	0	1	2	0	3
Hap 7	0	0	0	0	0	4	2	0	0	0	0	0
Hap 8	1	1	1	0	0	0	0	1	0	0	2	0
Hap 9	0	2	0	1	0	0	0	0	0	0	2	0
Hap 10	0	0	0	0	0	0	0	0	1	1	0	0
Hap 11	0	0	0	0	0	0	0	1	0	0	0	1
Hap 12	0	0	0	0	0	0	0	0	0	0	0	1

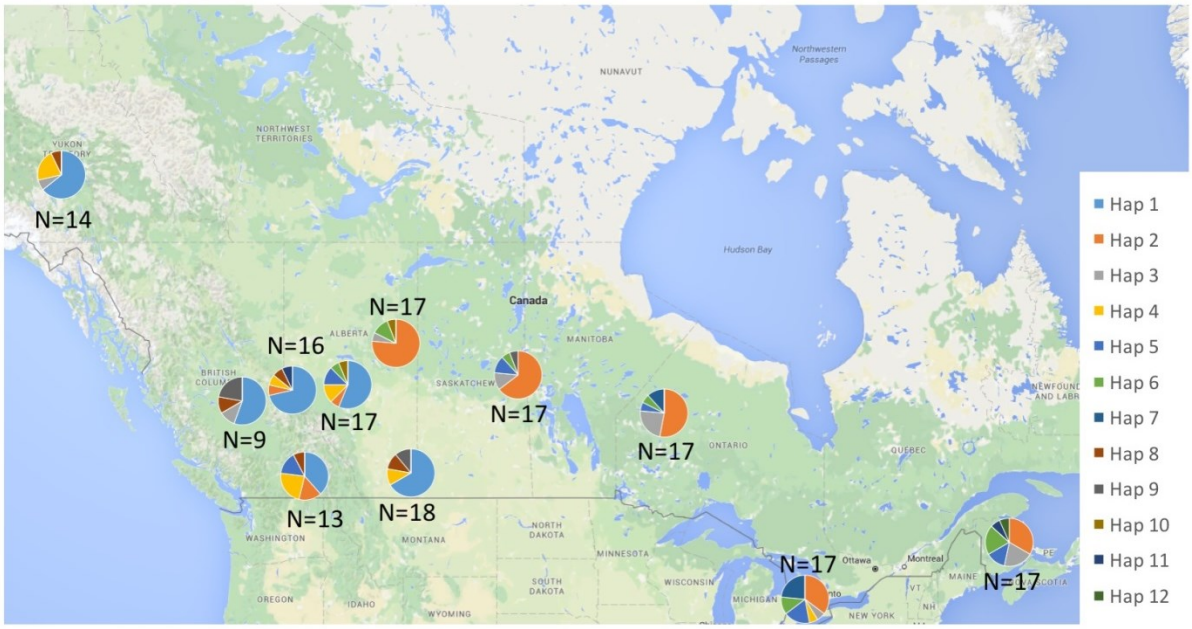


Figure 3.7: Map showing the distribution of unique haplotypes observed in the class II, *chia2-1*, chitinase for each of the 12 sampling locations across Canada. Sampling locations Creighton SK and Hudson Bay SK have been combined to represent on location in this map. Total number of individuals sequenced within each population are indicated as "N=" and are found above each pie chart. Haplotype counts specific to each sampling location described in Table 3.4.

Table 3.10: Allelic richness based on four variant loci in jack and lodgepole pine class II chitinase, *chia2-1,* sequenced from gDNA of seedlings from 12 provenances collected across Canada. Seedlings sampled from Minto, Edgewood, Baldy Hughes, and Nose Mountain are designated as lodgepole. Seedlings sampled from Stoney Mountain, Cypress Hills, Hudson Bay, Weagamow lake, London, and Despres Lake are designated as jack pine, and seedlings sampled from Blue Ridge are designated as lodgepole x jack pine hybrids.

	Lodgepole pine	Jack pine	Lodgepole x jack pine hybrid	Overall
SNV -284	2.00	2.00	2.00	2.00
SNV -141	1.91	2.00	2.00	2.00
InDel 6	1.99	2.00	2.00	2.00
SNV 413	2.00	2.00	2.00	2.00

Table 3.11: Matrix of pairwise Fst values for *chia2-1* among pine seedlings designated as lodgepole pine, jack pine and lodgepole x jack pine hybrids. Fst values estimated on the basis of four variant loci found in *chia2-1,* sequenced gDNA of seedlings from 12 provenances collected across Canada. Seedlings sampled from Minto, Edgewood, Baldy Hughes, and Nose Mountain were designated as lodgepole. Seedlings sampled from Stoney Mountain, Cypress Hills, Hudson Bay, Weagamow Lake, London, and Despres Lake were designated as jack pine, and seedlings sampled from Blue Ridge were designated as lodgepole x jack pine hybrids. Negative Fst values corrected to be values of 0.

	Lodgepole pine	Jack pine	Lodgepole x jack pine hybrid
Lodgepole pine	0.000	0.272	0.000
Jack pine	0.272	0.000	0.215
Lodgepole x jack pine hybrid	0.000	0.215	0.000

CHIA4-1

Allelic resequencing of class IV chitinase, *chia4-1,* revealed variation at three sites along the upstream promoter region and five sites of variation within the coding region across a total of 183 sequenced individuals (Table 3.12).

We identified SNVs 167, 119 and 45 bp upstream of the *chia4-1* transcriptional start site. However only variation at SNV -119 was non-synonymous and could potentially result in functional changes. There were 15 individuals who carried an A at this location, conserving a putative CAT-box core promoter element, and 161 individuals who carried a G at this location which resulted in a loss of the CAT-box core promoter element (Table 3.12). We identified a GTGGTG insertion in the signal peptide domain of *chia4-1* in 173 individuals which resulted in an insertion of two valine amino acid residues at amino acid position 11 and 12 within the signal peptide domain (Table 3.13).

Table 3.12: Variants identified in jack and lodgepole pine promoter sequence of class IV chitinase, *chia4-1,* sequenced from gDNA of seedlings from 12 provenances collected across Canada. Allele counts specific to each sample location indicated below total allele count. Total individuals: total number of individual seedlings sequenced per sample location. SNV: single nucleotide variant. Indel: insertion/deletion

Sequence position		-167		-119		-45	
Type		SNV		SNV		SNV	
Location		Promoter		Promoter		Promoter	
Coverage		175		176		169	
Alleles		G	C	A	G	A	G
Functional changes		None	none	CAA T-box	Loss of CAA T box	none	none
Functional changes							
Total Allele count		59	116	15	161	28	141
Sample location	Total individuals						
Minto, YK	16	13	3	1	15	7	8
Baldy Hughes, BC	13	9	2	2	9	2	7
Edgewood, BC	18	7	8	1	15	1	13
Creighton, SK	6	0	6	0	6	0	6
Hudson Bay, SK	11	1	10	0	11	2	9
London, ON	17	0	17	0	17	0	17
Weagamow Lake, ON	18	0	18	2	16	1	17
Nose Mountain, AB	15	10	5	3	12	3	10
Blue Ridge, AB	16	7	8	2	13	2	13
Stoney Mountain, AB	18	4	14	2	16	5	13
Cypress Hills, AB	18	7	10	1	16	4	13
Despres Lake, NB	17	1	15	1	15	1	15

Table 3.13: Variants identified in jack and lodgepole pine coding sequence of class IV chitinase, *chia4-1,* sequenced from gDNA of seedlings from 12 provenances collected across Canada. Allele counts specific to each sample location indicated below total allele count. Total individuals: total number of individual seedlings sequenced per sample location. SNV: single nucleotide variant. Indel: insertion/deletion

Sequence position		36		56			737		793		867		
Type		Insertion		SNV			SNV		SNV		SNV		
Location		Signal peptide		Signal peptide			Intron		Catalytic domain		Catalytic domain		
Coverage		186		181			174		177		178		
Alleles		Deletion	GTGGTG	T	C	A	T	C	C	A	G	T	C
Functional changes		Del 2 V	Ins 2V		silent		none		glutamine	lysine	silent	silent	silent
Total Allele count		13	173	54	127	47	125	2	65	112	16	160	2
Sample location	**Total individuals**												
Minto, YK	16	3	13	7	8	10	5	1	14	2	4	11	1
Baldy Hughes, BC	13	4	9	7	4	4	9	0	8	5	3	10	0
Edgewood, BC	18	3	16	9	10	5	11	0	10	6	2	14	0
Creighton, SK	6	0	6	0	6	0	5	1	0	6	0	6	0
Hudson Bay, SK	11	0	11	1	10	0	11	0	0	11	0	11	0
London, ON	17	0	18	0	18	1	17	0	0	18	0	18	0
Weagamow Lake, ON	18	0	18	1	17	0	18	0	0	18	0	18	0
Nose Mountain, AB	15	1	15	9	6	9	4	0	10	4	2	12	0
Blue Ridge, AB	16	0	15	5	9	3	13	0	9	6	0	15	1
Stoney Mountain, AB	18	1	18	4	15	3	13	0	4	13	2	15	0
Cypress Hills, AB	18	1	17	10	8	10	5	0	7	10	3	14	0
Despres Lake, NB	17	0	17	1	16	2	14	0	3	13	0	16	0

Thirteen individuals carried a deletion at this site (Table 3.13). Signal peptide analysis of putative amino acid sequences containing either an insertion or a deletion revealed that the additional valine residues occurred before signal peptide the cleavage site, and therefore did not likely affect protein confirmation. (Appendix 6.9). We found an additional SNV 56 bp after the transcriptional start site, found within the signal peptide domain and resulting in no amino acid changes, as well as a SNV 737 bp from the transcriptional start site found within the first intron (Table 3.13). Resequencing revealed a non-synonymous substitution 793 bp from the *chia4-1* transcriptional start site within the catalytic domain (Table 3.13). Sixty-five individuals carried a C at this position resulting in a glutamine residue at position 188 along the peptide, and 112 individuals carried an A at this position resulting in a lysine at amino acid residue 188 (Table 3.13). Protein homology modeling revealed that this substitution occurred outside of the catalytic cleft and a calculated RMSD value of 0 indicated that there were no structural differences between allelic variants (Fig 3.8: Table 6.2).

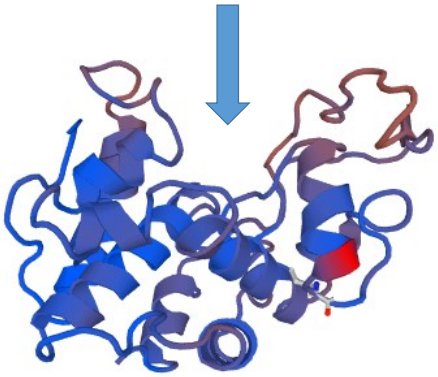


Figure 3.8: Protein homology model of the class IV chitinase, chia4-1. Catalytic cleft indicated by arrow. Site of amino acid substitution glutamine to lysine at amino acid position 188 (Q188K) is highlighted in red.

Finally, we identified an SNV 867 bp from the transcriptional start site of *chia4-1,* located in the catalytic domain which results in no putative amino acid changes (Table 3.13). There was a large diversity of haplotypes within *chia4-1* across sequenced individuals, and using sites of variation with greater than 150 we identified 54 unique haplotypes (Table 3.14). In the jack pine provenances Stoney Mountain AB, Creighton and Hudson Bay SK, Weagamow Lake ON, London ON, and Despres Lake NB, haplotype one (Hap 1) represented the majority of sequences individuals (Table 3.15: Fig 3.9). In the lodgepole and lodgepole x jack pine provenances Minto YK, Baldy Hughes BC, Edgewood BC, Nose Mountain AB, Blue Ridge AB, and Cypress Hills, Hap 1 occurred to a lesser extent compared to eastern sample locations (Table 3.15: Fig 3.9). Interestingly, lodgepole pine provenances appeared to contain much greater haplotype diversity than jack pine. Haplotype two was the second most abundant haplotype, but occurred at a much lower frequency than haplotype one and was found in central provenances of Baldy Hughes BC, Edgewood BC, Nose Mountain AB, Blue Ridge AB and Hudson Bay SK. However, haplotype two carried no functional differences in comparison to haplotype one (Table 3.14). In contrast to haplotype one and two, haplotype three and four both carried an A resulting in a lysine rather than glutamine at residue 188 in the catalytic domain, and occurred in western sampling locations Minto YK, Edgewood BC, Nose Mountain AB, Blue Ridge AB, Stoney Mountain AB and Cypress Hills AB (Table 3.14: Fig 3.9).

Sample locations belonging representing lodgepole pine sampling locations displayed greater allelic variation in *chia4-1* across these eight variant sites in comparison to jack pine and lodgepole x jack pine sample sites (Table 3.16). Examination of allelic variation within sampling locations reveals lodgepole pine sample sites Minto, Edgewood and Nose Mountain carried the greatest amount of allelic richness across these sites (Appendix 6.6.5).

Table 3.14: Unique haplotypes observed within the class IV, *chia4-1,* chitinase gene in 171 individual seedlings. Variants are indicated by their type, and sequence location with respect to the transcriptional start site. SNV: single nucleotide variation, InDel: insertion/deletion. Variants are described in table 3.12 and 3.13

Haplotype	SNV -167	SNV -119	SNV -45	Indel 36	SNV 56	SNV 737	SNV 793	SNV 867
Hap 1	C	G	G	GTGGTG	C	T	A	T
Hap 2	G	G	G	GTGGTG	T	T	A	T
Hap 3	G	G	G	GTGGTG	T	T	C	T
Hap 4	G	G	G	GTGGTG	C	A	C	T
Hap 5	G	G	G	GTGGTG	T	A	A	T
Hap 6	C	G	G	GTGGTG	C	T	C	T
Hap 7	C	G	G	GTGGTG	C	A	C	T
Hap 8	C	G	A	GTGGTG	C	T	A	T
Hap 9	C	G	G	GTGGTG	C	A	A	T
Hap 10	C	G	G	GTGGTG	T	A	C	T
Hap 11	G	G	G	GTGGTG	T	A	C	T
Hap 12	C	A	A	GTGGTG	T	T	A	T
Hap 13	C	G	A	GTGGTG	C	A	C	G
Hap 14	C	G	G	GTGGTG	T	T	A	T
Hap 15	G	A	A	GTGGTG	T	T	C	T
Hap 16	G	G	A	GTGGTG	C	A	C	T
Hap 17	G	G	A	GTGGTG	T	A	C	T
Hap 18	G	G	G	DEL	C	T	A	T
Hap 19	G	G	N/A	GTGGTG	T	A	C	G
Hap 20	C	A	A	GTGGTG	C	A	A	T
Hap 21	C	A	G	GTGGTG	C	A	C	T
Hap 22	C	A	G	GTGGTG	C	T	A	T

Haplotype	SNV -167	SNV -119	SNV -45	Indel 36	SNV 56	SNV 737	SNV 793	SNV 867
Hap 23	C	G	A	DEL	C	A	C	G
Hap 24	C	G	A	DEL	T	A	C	T
Hap 25	C	G	A	GTGGTG	C	A	C	G
Hap 26	C	G	A	GTGGTG	C	T	C	G
Hap 27	C	G	A	GTGGTG	T	T	C	T
Hap 28	C	G	G	DEL	C	A	C	T
Hap 29	C	G	G	DEL	C	T	C	T
Hap 30	C	G	G	GTGGTG	C	A	C	G
Hap 31	C	G	G	GTGGTG	T	C	A	T
Hap 32	G	A	A	DEL	C	A	C	G
Hap 33	G	A	A	GTGGTG	C	A	C	T
Hap 34	G	A	A	GTGGTG	T	A	C	G
Hap 35	G	A	A	GTGGTG	T	A	C	T
Hap 36	G	A	A	GTGGTG	T	T	C	G
Hap 37	G	A	N/A	DEL	T	T	C	T
Hap 38	G	G	A	DEL	C	N/A	C	T
Hap 39	G	G	A	DEL	T	A	C	G
Hap 40	G	G	A	GTGGTG	C	A	C	G
Hap 41	G	G	A	GTGGTG	T	C	C	T
Hap 42	G	G	A	GTGGTG	T	T	C	T
Hap 43	G	G	G	DEL	C	A	A	T
Hap 44	G	G	G	DEL	C	A	C	T
Hap 45	G	G	G	DEL	N/A	A	A	T
Hap 46	G	G	G	GTGGTG	C	A	A	G
Hap 47	G	G	G	GTGGTG	C	C	C	T
Hap 48	G	G	G	GTGGTG	C	T	A	T

Haplotype	SNV -167	SNV -119	SNV -45	Indel 36	SNV 56	SNV 737	SNV 793	SNV 867
Hap 49	G	G	G	GTGGTG	T	A	C	G
Hap 50	G	G	G	GTGGTG	T	C	C	T
Hap 51	G	G	G	GTGGTG	T	T	A	C
Hap 52	G	G	N/A	GTGGTG	C	T	C	T
Hap 53	N/A	G	G	GTGGTG	C	T	C	C
Hap 54	T	G	T	GTGGTG	T	T	C	G

Table 3.15: Counts of haplotypes identified in jack and lodgepole pine class IV chitinase, *chia4-1,* sequenced from gDNA of seedlings from 12 provenances collected across Canada. Haplotypes are described in table 3.14

Sample Location	Minto YK	Baldy Hughes BC	Edgewood BC	Creighton SK	Husdon Bay SK	London ON	Weagamow Lake ON	Nose Mountain AB	Blue Ridge AB	Stoney Mountain AB	Cypress Hills AB	Despres Lake NB
Total individuals	17	10	16	6	11	17	18	13	14	17	17	15
Haplotype count												
Hap 1	0	0	3	5	8	16	16	0	5	11	3	13
Hap 2	0	1	2	0	1	0	0	1	1	0	0	0
Hap 3	1	0	2	0	0	0	0	1	1	0	1	0
Hap 4	3	0	0	0	0	0	0	0	1	1	0	0
Hap 5	0	0	0	0	0	0	0	2	0	0	2	0
Hap 6	0	0	1	0	0	0	0	1	1	0	0	1
Hap 7	0	0	1	0	0	0	0	0	0	0	0	0
Hap 8	0	0	1	0	1	0	0	0	0	0	0	0
Hap 9	0	0	0	0	1	1	0	0	1	0	0	0
Hap 10	0	0	0	1	0	0	0	0	1	2	1	0
Hap 11	1	1	0	0	0	0	0	2	0	0	1	0
Hap 12	0	0	0	0	0	0	1	0	0	0	0	0
Hap 13	0	1	0	0	0	0	0	0	0	0	0	0
Hap 14	0	0	1	0	0	0	0	0	0	0	1	0
Hap 15	0	0	0	0	0	0	0	0	2	0	0	0
Hap 16	1	0	0	0	0	0	0	1	0	0	0	0
Hap 17	0	1	0	0	0	0	0	0	0	0	1	0
Hap 18	1	1	0	0	0	0	0	0	0	0	0	0
Hap 19	1	1	0	0	0	0	0	0	0	0	0	0
Hap 20	0	0	0	0	0	0	0	0	0	0	1	0
Hap 21	0	0	0	0	0	0	0	1	0	0	0	0

Sample Location	Minto YK	Baldy Hughes BC	Edgewood BC	Creighton SK	Husdon Bay SK	London ON	Weagamow Lake ON	Nose Mountain AB	Blue Ridge AB	Stoney Mountain AB	Cypress Hills AB	Despres Lake NB
Total individuals	17	10	16	6	11	17	18	13	14	17	17	15
Hap 22	0	0	0	0	0	0	1	0	0	0	0	0
Hap 23	1	0	0	0	0	0	0	0	0	0	0	0
Hap 24	0	0	0	0	0	0	0	1	0	0	0	0
Hap 25	0	0	0	0	0	0	0	0	0	0	1	0
Hap 26	0	0	0	0	0	0	0	1	0	0	0	0
Hap 27	1	0	0	0	0	0	0	0	0	0	0	0
Hap 28	0	0	1	0	0	0	0	0	0	0	0	0
Hap 29	0	1	0	0	0	0	0	0	0	0	0	0
Hap 30	0	0	0	0	0	0	0	0	0	0	1	0
Hap 31	0	0	0	0	0	0	0	0	0	0	1	0
Hap 32	1	0	1	0	0	0	0	0	0	0	0	0
Hap 33	1	0	0	0	0	0	0	0	0	0	0	0
Hap 34	0	0	0	0	0	0	0	0	0	1	0	0
Hap 35	0	0	0	0	0	0	0	0	0	0	0	1
Hap 36	0	0	0	0	0	0	0	0	0	1	0	0
Hap 37	0	1	0	0	0	0	0	0	0	0	0	0
Hap 38	0	0	0	0	0	0	0	0	0	1	0	0
Hap 39	1	0	0	0	0	0	0	0	0	0	0	0
Hap 40	1	0	0	0	0	0	0	0	0	0	0	0
Hap 41	0	0	0	0	0	0	0	0	0	0	1	0
Hap 42	1	0	0	0	0	0	0	0	0	0	0	0
Hap 43	0	0	0	0	0	0	0	0	0	0	1	0
Hap 44	0	0	1	0	0	0	0	0	0	0	0	0
Hap 45	0	1	0	0	0	0	0	0	0	0	0	0
Hap 46	0	0	0	0	0	0	0	1	0	0	0	0

Sample Location	Minto YK	Baldy Hughes BC	Edgewood BC	Creighton SK	Husdon Bay SK	London ON	Weagamow Lake ON	Nose Mountain AB	Blue Ridge AB	Stoney Mountain AB	Cypress Hills AB	Despres Lake NB
Total individuals	17	10	16	6	11	17	18	13	14	17	17	15
Hap 47	1	0	0	0	0	0	0	0	0	0	0	0
Hap 48	0	1	0	0	0	0	0	0	0	0	0	0
Hap 49	0	0	0	0	0	0	0	0	0	0	1	0
Hap 50	0	0	0	0	0	0	0	1	0	0	0	0
Hap 51	1	0	0	0	0	0	0	0	0	0	0	0
Hap 52	0	0	1	0	0	0	0	0	0	0	0	0
Hap 53	0	0	0	0	0	0	0	0	1	0	0	0
Hap 54	0	0	1	0	0	0	0	0	0	0	0	0

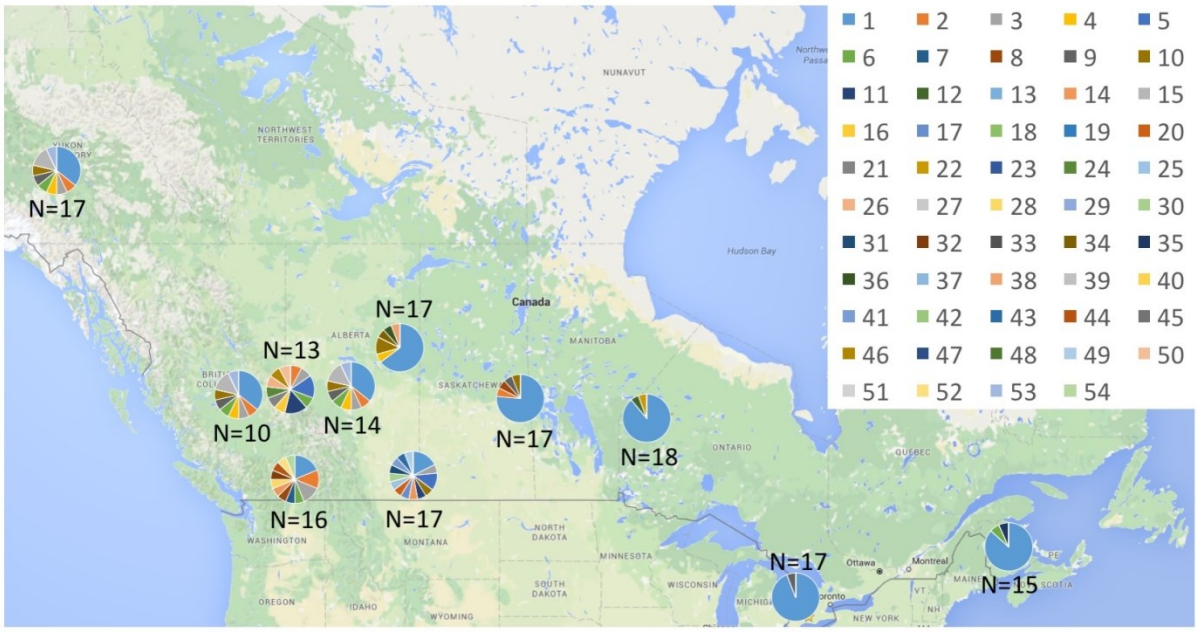


Figure 3.9: Map showing the distribution of unique haplotypes observed in the class IV, *chia4-1*, chitinase for each of the 12 sampling locations across Canada. Sampling locations Creighton SK and Hudson Bay SK have been combined to represent on location in this map. Total number of individuals sequenced within each population are indicated as "N=" and are found above each pie chart. Haplotype counts specific to each sampling location described in table 3.14.

Calculation of pairwise Fst values for *chia4-1* between lodgepole and jack pine samples revealed substantial genetic differentiation between pure species, as well as between jack pine and lodgepole x jack pine hybrid samples (Table 3.17). However, there was little differentiation between lodgepole samples and lodgepole x jack pine hybrid samples (Table 3.17). This genetic differentiation between species was reflected in pairwise Fst calculations between sampling locations (Appendix 6.6.6).

Table 3.16: Allelic richness based on four variant loci in jack and lodgepole pine class IV chitinase, *chia4-1,* sequenced from gDNA of seedlings from 12 provenances collected across Canada. Seedlings sampled from Minto, Edgewood, Baldy Hughes, and Nose Mountain were designated as lodgepole. Seedlings sampled from Stoney Mountain, Cypress Hills, Hudson Bay, Weagamow Lake, London, and Despres Lake were designated as jack pine, and seedlings sampled from Blue Ridge were designated as lodgepole x jack pine hybrids.

	Lodgepole pine	Jack pine	Lodgepole x jack pine hybrid	Overall
SNV -167	2.43	1.98	2.00	2.15
SNV -119	1.98	1.84	2.00	1.93
SNV -45	2.47	1.98	2.00	2.15
Indel 36	2.00	1.43	1.00	1.90
SNV 56	2.00	2.00	2.00	2.00
SNV 737	2.67	2.59	2.00	2.57
SNV 793	2.00	2.24	2.00	2.15
SNV 867	2.42	1.78	1.99	2.22

Table 3.17: Matrix of pairwise Fst values for *chia4-1* among pine seedlings designated as lodgepole pine, jack pine and lodgepole x jack pine hybrids. Fst values estimated on the basis of four variant loci found in *chia4-1,* sequenced gDNA of seedlings from 12 provenances collected across Canada. Seedlings sampled from Minto, Edgewood, Baldy Hughes, and Nose Mountain were designated as lodgepole. Seedlings sampled from Stoney Mountain, Cypress Hills, Hudson Bay, Weagamow Lake, London, and Despres Lake were designated as jack pine, and seedlings sampled from Blue Ridge were designated as lodgepole x jack pine hybrids. Negative Fst values corrected to be values of 0.

	Lodgepole pine	Jack pine	Lodgepole x jack pine hybrid
Lodgepole pine	0.00	0.303	0.043
Jack pine	0.303	0.000	0.161
Lodgepole x jack pine hybrid	0.043	0.161	0.000

3.4 Discussion

3.4.1 Phylogenetic analysis of lodgepole and jack pine chitinase family

Pine chitinases grouped into five distinct clades, which correspond strongly to their biochemical classifications. We named each of these clades as clusters, for consistency with the naming convention of Galindo-Gonzalez *et al.* (2015). Clusters one, two, three and four are consistent with four of the five clusters described in a phylogenetic analysis of white spruce chitinases (Galindo-Gonzalez *et al.* 2015). However, the phylogenetic analysis conducted by Galindo-Gonzalez *et al.* 2015 identified a cluster composed of class III chitinases, and we were unable to identify class III chitinases within the jack and lodgepole transcriptomes with confidence. Rather the fifth cluster we identified was composed of class II chitinases which formed their own cluster (cluster five) distinct from the class II chitinases found in cluster two. It has been suggested that class II chitinases may be further divided into subclass class IIa, termed pathogenesis chitinases, which are more closely related to class I chitinases compared to class IIb chitinases (Araki & Torikata 1995; Neuhaus 1999). It has been proposed that class IIa and class IIb were derived from class I chitinases following two independent deletion events of the CBD from a class I chitinase (Neuhaus 1999). We suggest that the chitinases which form cluster five are less related to class I chitinases than those class II chitinases found in cluster 2, and these two lineages of class II chitinases arose from a separate event leading to the deletion of the CBD.

Our findings add further support to the proposition that class I and class II chitinases likely share a common ancestor and class IV and class VII chitinases likely share a different common ancestor (Galindo-Gonzalez *et al.* 2015; Araki & Torikata 1995; Neuhaus 1999).

Placement of class IV and class VII chitinases together, and class I and class IIa chitinases together suggests that shared similarities in the catalytic domains are driving the placement of these sequences, rather than presence or absence of the CBD. It has been proposed that the CBD has been lost many times throughout the evolutionary history of chitinases within the GH 19 family (Neuhaus 1999). Our results are in line with this theory, suggesting – as did Galindo-Gonzalez et al. (2015) – that class II chitinases arose following a deletion events of the CBD in class I chitinases, and similarly class VII chitinase arose following a deletion event of the CBD in a class IV chitinases.

Galindo-Gonzalez *et al.* (2015) suggest that the loss of the CBD in class VII and class II chitinases occurred after the gymnosperm-angiosperm split, as evidenced by class IV conifer chitinases being more closely related to class VII conifer chitinases than class IV angiosperm chitinases, and class I conifer chitinases being more closely related to class II chitinases. Our results also show *Arabidopsis* class I and II chitinases clustering more closely to one another than to conifer chitinase members of the equivalent class.

3.4.2 Expression analysis of chitinase family in response to *G. clavigera*

Analysis of microarray expression data revealed that many chitinase members of GH family 19 in pine are induced in response to challenge with *G. clavigera.* This is consistent with large increases in pathogen-induced chitinase transcript abundance reported for other conifers (Davis *et al.* 2004; Heitala *et al.* 2004; Koslova *et al.* 2014). Our results showed that different classes (clusters) of GH 19 chitinases displayed distinct transcript abundance profiles, differing in the timing and magnitude of their induction, but that closely related chitinases within classes (clusters) display similarities in expression profiles. These results suggest that there is a

correlation between polygenetic relationship, biochemical classification and expression patterns among the chitinase gene family.

Our results are consistent with qRT-PCR analysis of two class VII chitinases in lodgepole pine inoculated with *G. clavigera* (Koslova *et al.* 2014). It was reported that one class VII chitinase was upregulated in response to inoculation while the other displayed no change in expression pattern, and the authors suggest that some chitinase may play a role in defense while others play a role in development (Koslova *et al.* 2014). Plant chitinases have been reported to play roles in a variety of developmental processes including somatic embryogenesis (Wiweger *et al.* 2003), seed senescence and germination (Hanfrey *et al.* 1996; Wu *et al.* 2001) and cell wall biosynthesis and assembly (Zhang *et al.* 2004: Sánchez-Rodríguez *et al.* 2012: Wu *et al.* 2012). Conifer chitinases have been reported to play a role in a number of metabolic and developmental processes, such as somatic embryogenesis (Johnsen *et al.* 2005), the transition from growth to dormancy (Galindo-Gonzalez *et al.* 2015), and cold hardiness through antifreeze activity (Zamani *et al.* 2003; Jarząbek *et al.* 2009). It is possible that class II and class VII pine chitinases which are not upregulated in response to *G. clavigera* do not play roles in defense in pine. It is also possible that they do play a defensive role, but not in response specifically to *G. clavigera.*

We identified lodgepole and jack pine chitinases *PcChia2-2* and *PbChia2-2* respectively which both display low levels of down regulation in response to *G. clavigera.* These chitinases are highly similar to a class II chitinase, *Pschi4*, whose expression is upregulated in slash pine seedlings when treated with SA but not upregulated when treated with JA (Davis *et al.* 2004). In Chapter 2, we noted a coordinated upregulation of gene involved in the JA and ethylene biosynthesis and signaling pathways, and little upregulation of genes involved in SA biosynthesis or signaling within both jack and lodgepole pine seedlings inoculated with *G. clavigera.* Furthermore, significant increases in JA have been reported in *G. clavigera*-inoculated lodgepole

and jack pine trees under well watered and water deficit conditions, where as in the same study SA was only upregulated in jack pine trees inoculated with *G. clavigera* under water deficit conditions *(*Arango-Velez *et al.* 2016) *pcchia2-2* and *pbchia2-2* are likely upregulated through activation of SA pathway, not through activation of JA pathway, and are therefore not induced in response to *G. clavigera*. It is possible that the same pattern exists amongst class VII chitinases observed to have low levels down regulation in response to *G. clavigera*. Taken together, these results suggest that the certain members of the chitinase family are specialized to respond specifically challenge by biotrophic pathogens through SA signaling. Our results demonstrated that water deficit influences the induced expression of chitinases genes in pine by either attenuating or augmenting expression in response to *G. clavigera*. This is consistent with qRT-PCR analysis of class I and class I chitinases in jack x lodgepole hybrids which displayed contrasting pattern of expressions when under water deficit (Appendix 6.7).

3.4.3 Functional consequences of allelic variation found within chitinase genes

For all three resequenced chitinases, very little variation identified within promoter sequences resulted in changes to putative transcription factor binding sites, and only *chia1-1* displayed alteration of a putative MYB transcription factor binding site (MBS). MYBs are known to play a role in conifer defense response through regulation of the phenylpropanoid pathway (Liu *et al.* 2015 Patzlaff *et al.* 2003a; Patzlaff 2003b Bedon 2010). MYB TFs have recently been reported to regulate transcription of chitinases in *Brassica juncea* in response to pathogen challenge (Gao *et al.* 2016). While variation at the MBS site in the *chia1-1* promoter may influence expression in response to pathogen challenges, it is important to note that two other putative MBS are located within the upstream promoter region and neither are altered by any of

the variation identified in this study. Further study is required to confirm any interaction between a defense associated MYB TF and the *chia1-1* promoter, as well as any potential changes to transcriptional regulation conferred by conservation or loss of an MBS.

We also identified allelic variation in *chia2-1* and *chia4-1* CAT-box core promoter elements. The CAT-box is present in the promoters of most eukaryotic genes, and is bound by a core binding element or nuclear factors which facilitate transcription by RNA polymerases (Laloum *et al.* 2013). In *chia 2-1,* we further identified alteration of a TATA box, another core promoter element found in most eukaryotic genes which promotes recruitment of RNA polymerases and facilitates transcription (Smale & Kadonaga 2003). Other CAT and TATA boxes were identified at multiple alternative locations in the promoters of *chia2-1* and *chia4-1* which were unaffected by variation identified in this study. Further study is required to determine if any differences in expression patterns exist as a result of this variation. Taken together, our results suggest that variation identified within the promoter regions of *chia1-1, chia2-1, and chia4-1* likely does not result in any differences in expression patterns across the individuals we sequenced.

In both *chia2-1,* and *chia4-1*, we identified insertion-deletions within the signal peptide domain. Signal peptides are responsible for targeting proteins to the secretory pathway in eukaryotes and prokaryotes (Gierasch1989 & Rapoport 1992). Plant chitinases exhibit diagnostic signal peptides which target them to the secretory pathway (Neilson *et al.* 1997). The variation identified in the signal peptides of *chia4-1* and *chia2-1* did not change the location of predicted signal peptide and both occurred prior to the cleavage site. These results suggest that in both cases they do not play a role in altering the protein structure and function, and because the signal peptide remains detectable in both cases, it is unlikely that these variants effect targeting of protein to secretory pathway.

Chia1-1 contained an insertion within the hinge domain located between the chitin binding domain and the catalytic domain. Protein homology modeling and superposition of the two amino acid sequences revealed and extension of the hinge loop between the chitin bind domain and the catalytic domain. However, the extended hinge domain did not appear to affect the folding of the chitin binding or catalytic domains. Hinge regions are typically proline rich and hyper variable regions, therefore the variation we observed in the hinge region of *chia1-1* is not unexpected (Bishop 2000). Further study is required to determine if extension of the hinge domain confers any changes to the function of *chia1-1*.

Chia4-1 contains a non-synonymous substitution of a polar glutamine to a charged lysine at amino acid residue 188 in the catalytic domain. Protein homology modeling revealed this substitution occurs outside of the catalytic cleft and results in no conformational changes to the protein. Analysis of chitinases found in white spruce demonstrated low conservation of identical or similar amino acid residues at this position, and in the two separate class IV white spruce chitinases reported one carried a glutamine at this position and the other a lysine (Galindo-Gonzalez *et al.* 2015). Taken together, it is unlikely this non-synonymous substitution results in changes to the function of *chia4-1* in lodgepole or jack pine, however further study is required to confirm this.

3.4.5 Geographic distribution of chitinase haplotypes reveals specifies specific differences

Analysis of chitinase haplotypes revealed a difference in distribution between eastern and western lodgepole and jack pine sample locations. Often the most abundant haplotypes were found either predominantly in eastern populations of Stoney Mountain AB, Creighton and Hudson Bay SK, Weagamow Lake ON, London ON, and Despres Lake NB, or alternatively were

found predominantly in western populations of Minto YK, Baldy Hughes BC, Edgewood BC, and Nose Mountain AB. In the case of *chia2-1,* the east-west partition is particularly well illustrated, and with the exception of Cypress Hills, is indicative of haplotypes specific to lodgepole pine, found in the western sample locations, or jack pine, found in the eastern sample locations.

The occurrence of haplotypes specific to lodgepole or jack pine is well illustrated by the haplotypes observed across the three sample locations (Nose Mountain AB, Blue Ridge AB and Stoney Mountain AB) which represent the hybridization zone between lodgepole and jack pine previously characterized in Cullingham *et al.* 2012. Samples collected from Nose Mountain AB represent lodgepole pine, samples collected from Blue Ridge represent lodgepole x jack pine hybrids and samples collected from Stoney Mountain represent jack pine. Accordingly, a gradient of lodgepole to jack pine specific haplotypes can be observed across these sample locations, with Nose Mountain containing predominantly lodgepole pine specific haplotypes, Stoney Mountain containing predominantly jack pine specific haplotypes, and Blue Ridge containing intermediate counts of haplotypes specific to both species.

As expected, pairwise Fst calculations using the variants identified in *chia2-1 and chia4-1* revealed substantial genetic differenation between species. Samples collected from the lodgepole x jack pine hybrids in Blue Ridge displayed less genetic differentiation in *chia2-1* and *chia4-1* from lodgepole samples compared to jack pine samples. This is consistent with earlier studies which suggest that hybrid ancestry is biased toward lodgepole pine (Cullingham *et al.* 2012). However, Fst calculations using the variants identified in *chia1-1* little genetic differentiation between species suggesting that *chia1-1* many be more conserved in comparison to *chia2-1* and *chia4-1.* Additionally, in contrast to *chia2-1* and *chia4-1,* lodgepole x jack pine hybrid samples from Blue Ridge showed less genetic differentiation in *chia1-1* from jack pine samples than from

lodgepole pine samples. Taken together, this is compelling evidence that different selective pressures are acting on *chia1-1* in comparison to *chia2-1* and *chia4-1,* and that the chitinase gene family is playing diverse roles in deference and development.

In the case of *chia4-1,* much greater proportion of the haplotype diversity was represented in western lodgepole pine samples found in MPB historic range. These samples also displayed much greater allelic richness in comparison to the eastern jack pine samples. Chitinases undergo greater rates of non-synonymous substitution at their active site, and it has been suggested that this a strategy to overcome inhibition by carbohydrate and protein inhibitors released by an invading pathogen (Bishop *et al.* 2000). Higher levels of haplotype diversity, and allelic richness found in lodgepole pine located in MPB historic range may suggest greater diversifying selective pressure on chitinases within these populations. However, the variants we observed did not occur within the catalytic cleft and would need to be compared to neutral markers in order to confirm that they are the result of selective pressure.

3.5 Conclusion

The objectives of our study were to characterize the chitinases gene family in lodgepole and jack pine, to investigate their role in response to *G. clavigera,* and to examine their allelic variation across Canada. Using transcriptome assemblies generated from loblolly, lodgepole, and jack pine, we identified 42 putative unique expressed chitinase genes in loblolly pine, 44 in jack pine and 40 in lodgepole pine. Our data are consistent with the theory that class II chitinases likely arose from class I chitinases following a deletion of the CBD and that class VII chitinases arose from class IV chitinases following a deletion of the CBD. A number of lodgepole and jack pine chitinases appear to be involved in defense against the putative necrotroph *G. clavigera*, with closely related genes showing co-expression patterns. Similar to the changes in global gene

expression described in Chapter Two, inoculation under water deficit attenuated expression of some chitinases while increasing expression of others. Allelic resequencing revealed species specific differences in chitinase haplotypes between lodegpole and jack pine species, and in the case of *chia4-1* it revealed greater haplotype diversity and allelic richness in lodgepole pine samples collected from MPB historic range which is suggestive of greater selective pressure being placed on defensive enzymes such as chitinases in those populations.

4.0 Conclusions

The current MPB outbreak has resulted in wide spread tree mortality, and its continued eastward spread across the Rocky Mountains into Alberta remains a significant threat to Canadian forests. In order to effectively combat an outbreak of this magnitude, management practices must be targeted to areas of higher risk. This has made efforts to understand the underlying mechanisms which contribute to tree host susceptibility critical in predicting future MPB spread risk. Differences in the response of co-evolved and naïve pine attacked by MPB have been studied mainly in field studies. These ecological studies have shown that trees outside of the traditional range of MPB have better host quality, resulting in higher beetle reproductive success, and invest fewer resources in inducible defenses (Cudmore *et al.* 2010; Raffa *et al.* 2013). Both qualitative and quantitative differences in monoterpene emission profiles and differences in lesion lengths have been reported between co-evolved lodgepole and naïve jack pine seedlings inoculated with *G. clavigera,* and these responses are impacted by water deficit (Lusibrink *et al.* 2011; Arango-Velez *et al.* 2016). Yet, little work has been done to elucidate the underlying genetic and molecular mechanisms responsible for the observed differences in host quality and defense response between co-evolved and naïve pine hosts, and even less work has been done to examine the effect of water deficit on these underlying molecular mechanisms.

In this study, we investigated the transcriptome-wide responses of lodgepole and jack pine seedlings to inoculation with the MPB fungal associate *G. clavigera,* under either well watered or water deficit conditions. Secondary metabolite production is an important component of plant defense against pests and pathogens. We identified differences in expression profiles for genes encoding secondary metabolites in lodgepole and jack pine seedlings in response to *G. clavigera.* Lodgepole pine exhibited greater changes in expression of genes involved in flavonoid

biosynthesis, such as dihydroflavonol reductases and chalcone synthases, whereas jack pine exhibited greater changes in expression of genes involved in isoprenoid biosynthesis, in particular putative terpene synthases. These were mainly genes encoding putative diterpene and sesquiterpene synthases, suggesting that the di- and sesquiterpene synthases may represent important differences between these two species. However, most studies to date have concentrated on monoterpenes. For example, lodgepole seedlings were reported to contain greater levels of total monoterpenes in comparison to jack pine seedlings (Arango *et al.* 2016). The same study demonstrates that lodgepole pine seedlings inoculated with *G. clavigera,* developed longer lesions and experience greater loss of stem hydraulic conductivity, and the authors suggest that *G. clavigera* is able to better colonize lodgepole pine. Given that the amount of total oleoresin which accumulates in response to pathogen challenge is often a marker of the strength of overall tree host resistance to pathogen challenge, the higher fold change in expression of sesquiterpene synthases and diterpene synthases that we observed in jack pine compared to lodgepole pine may contribute to jack pine's ability to better contain *G. clavigera*. However, Arango-Velez *et al.* (2016) propose that lodgepole pine mounted a more rapid response to *G. clavigera* compared to jack pine. We identified early upregulation of ERF-like and JAZ-like TF's, along with several osmotin-like genes in lodgepole pine. Given that these regulators are expressed later in jack pine, we suggest they are important in both species defense responses to *G. clavigera.* We propose that early expression of these regulators and defense response proteins in lodgepole pine seedlings indicates earlier recognition of *G. clavigera* effector molecules by NB-LRR proteins which leads to earlier induction of pathogen specific JA and ethylene signaling pathways, and a more rapid response to inoculation on the part of lodgepole pine.

Gene expression profiles suggested that activation of defense response pathways in response to *G. clavigera* occurs through the signaling action of JA and ethylene rather than

through the SA pathway. This gene expression data is supported by hormone profiling carried out by Arango-Velez *et al.* (2016), who demonstrated increased JA levels in both lodgepole and jack pine inoculated with *G. clavigera*. We also observed that water-deficit conditions influence defense responses by attenuating expression of some inducible defense-related genes, while simultaneously increasing expression of some defense related genes which, under well-watered conditions, played a constitutive role in response to inoculation. While upwards and downwards shifts in gene expression are expected in any system responding to external stress, we observed that water deficit specifically influences the expression of genes known to play a role in tree host defense.

We propose that water deficit modulates defensive signalling pathways through the action of ABA (Fig. 4.1), a well characterized stress hormone that shows dramatically increased levels in lodgepole and jack pine trees subjected to water deficit (Arango-Velez *et al.* 2016). The JA signaling pathway comprises two branches termed the ERF branch and the MYC branch (Pieterse *et al.* 2012; Broekgaarden *et al.* 2015). The ERF branch is believed to promote expression of necrotrophic pathogen response genes, and the MYC branch is believed to promote expression of wounding response genes. The MYC branch is favoured in the presence of ABA (Pieterse *et al.* 2012; Broekgaarden *et al.* 2015) (Figure 5.1). Increased levels of ABA were previously reported in phloem tissue of lodgepole and jack pine seedlings inoculated with *G. clavigera* under water deficit compared to those inoculated under well-watered conditions (Arango-Velez *et al.* 2016). We propose that in well watered lodgepole and jack pine seedlings, the ERF branch of the JA signalling pathway is activated to a greater degree than in seedlings under water deficit, due to lower levels of ABA in well watered tissues. Alternatively, under water deficit conditions, increased levels of ABA promote the MYC branch of the JA signalling pathway.

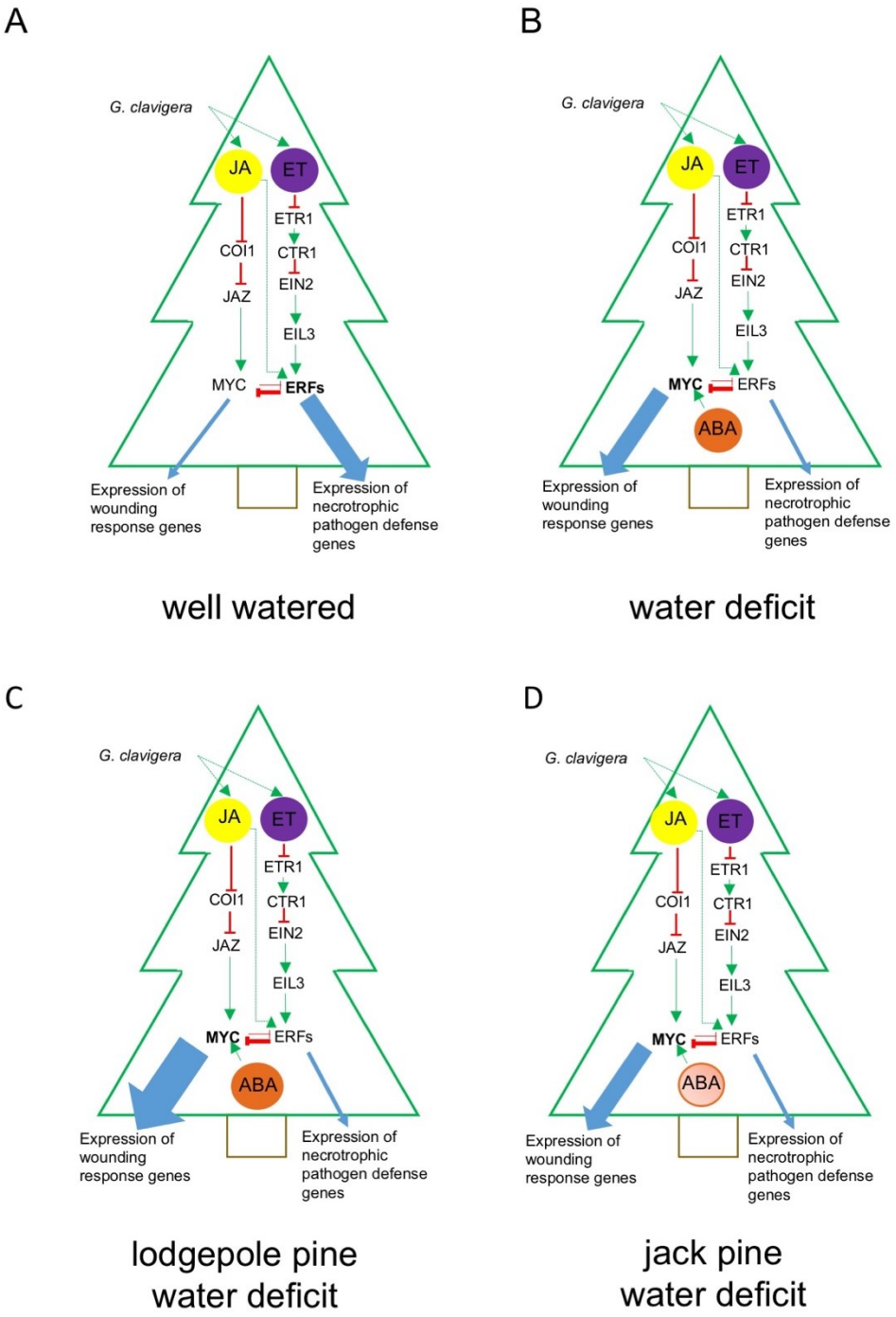


Figure 4.1: Conceptual model of defense signalling pathways in lodgepole and jack pine bark in response to inoculation with *G. clavigera.* Panel A: Inoculation of a lodgepole or jack pine host under well watered Panel B: Inoculation of a lodgepole or jack pine host under water deficit conditions. When inoculated, jasmonic acid (JA) and ethylene (ET) accumulate in response to

necrotrophic pathogen challenge (Pieterse *et al.* 2009). In the presence of JA, COI1 is targeted for degradation, allowing JAZ-like TFs to induce expression of MYC-like TFs which in turn induced expression wounding responsive genes (Chini *et al.* 2007; Thines *et al.* 2007; Dombrecht *et al.* 2007; Broekgaarden *et al.* 2015). In the presence of ethylene (ET), ERT-like histidine kinases are deactivated, which deactivates CRT-like histidine kinases and allows EIN2-like TF's to induce expression of EIL3-like TFs which in turn induces expression of ERF-like TFs (Binder 2008; Qiao *et al.* 2009; Solano *et al*. 1998). MYCs and ERFs work antagonistically to inhibit one another's expression (Zhang *et al.* 2014; Song *et al.* 2014). Under well watered conditions (panel A) JA and ET work synergistically to induce expression of ERFs, which activates expression of genes involved in necrotrophic pathogen defense response, and inhibiting the action of MYC TFs. Under water deficit conditions (panel B) we hypothesize that abscisic acid (ABA) accumulates in the phloem tissue, and promotes expression of MYCs which inhibits the action of ERF TFs reducing expression of genes involved in necrotrophic pathogen defense response and increasing expression of genes involved in response to wounding (Vos *et al.* 2013). Panel C & D: Inoculation of lodgepole and jack pine under water deficit conditions. lodgepole pine (panel C) is the more isohydric species, and is likely more sensitive to ABA than jack pine (panel D). This results in greater promotion of MYC TFs and downstream wounding response genes in lodgepole compared to jack pine, as well as greater inhibition of the ERF JA signalling branch and downstream necrotrophic pathogen response genes in lodgepole compared to jack pine.

We observed attenuation of some inducible defense genes under water deficit conditions, and we suggest this is due to inhibition of the ERF JA signalling branch in the presence of ABA, resulting in reduced expression of necrotrophic pathogen response genes under water deficit conditions. We also observed increased expression of some defense related genes which were constitutively expressed under well-watered conditions, and we suggest this is due to promotion of the MYC JA signalling branch in the presence of ABA, resulting in increased expression of wounding responsive genes under water conditions.

To add further support to our model, microarray analysis revealed that expression of a putative MYC gene increases under water deficit in both lodgepole and jack pine, and expression under water deficit increased earlier in lodgepole seedlings at 1 dpi, but later in jack pine at 7 dpi. We propose that accumulation of ABA in the more isohydric lodgepole pine results in earlier upregulation of a MYC-like genes, and therefore impacts expression of defense related genes to greater extent than jack pine. This results in greater downstream increases of wounding responsive genes and greater inhibition of necrotrophic pathogen responsive genes observed in lodgepole compared to jack pine. However, further qRT-PCR expression analysis of MYC-like genes in jack and lodgepole pine after inoculation with *G. clavigera* is be required to confirm this pattern of expression. Additionally, qRT-PCR analysis of of MYC-like genes in jack and lodgepole inoculated with *G. clavigera* and subjected to exogenous application of ABA would be required to confirm a causal link between increased expression of MYC in the presence of increased ABA.

Using transcriptome assemblies, we identified members of the chitinases gene family in loblolly, lodgepole and jack pine, and investigated their corresponding expression patterns in response to *G. clavigera*. We observed that biochemical classification, based on the presence or absence of conserved amino acid sequence motifs, corresponded to phylogenetic relationships, and we determined that the phylogenetic relationships among members of pine chitinase gene families is similar to the relationships described for the white spruce and Norway spruce chitinase gene families (Galindo-Gonzalez *et al.* 2015). We also observed that the expression patterns of pine chitinases reflect phylogenetic relationships. Interestingly, under water deficit conditions, expression patterns among the chitinase gene family mirrored global patterns described by the microarray data. Class I and class IV chitinases were highly induced in response to inoculation but this response was attenuated under water deficit conditions, whereas for some class II and

class VII chitinases, expression was induced only in seedlings inoculated under water deficit conditions. Given that different classes of chitinases play different biological roles in the plant (Veluthakkal *et al.* 2012), we propose that chitinases are specialized to play different role in defense response, and that some members of the chitinase gene family form part of the necrotrophic pathogen defense response, while others form part of the wounding defense response. We propose that diversification of the pine chitinase gene family evolved as a mechanism to better tailor defense responses against a wide range of pests and pathogens, which long-lived pine encounter over their lifetimes. Further investigation into the enzymatic functions of different pine chitinase members will be required in order to test the hypothesis that they play different roles in defense response.

Finally, we examined the allelic variation of three putative orthologous pairs of chitinase genes in individuals from both lodgepole and jack pine sampled from across Canada. We identified a several changes in the upstream promoter regions of these genes, including alterations of a MYB binding site, along with CAT and TATA core promoter elements. However we were not able to sequence farther than 500-800 bp upstream of the transcriptional start site, and allelic variation in TF binding sites may exist further upstream. Expression analysis of chitinase genes in response to *G. clavigera* across populations would be required to confirm if differences in transcriptional regulation for a unique chitinase gene(s) exists between populations, and promoter deletion analysis would be required to confirm if allelic variation in the promoter is responsible for this difference.

We identified allelic variation in the signal peptide, hinge and catalytic domains of class I, class II and class IV chitinase genes, and frequency of haplotypes differed within and between lodgepole and jack pine populations. In particular, we observed a spatially explicit pattern of occurrence for the non-synonymous substitution observed in the catalytic domain of the class IV

chitinases across lodgepole and jack pine ranges. In order to determine if any of the alleles observed in eastern lodgepole pine populations within MPB's traditional range are a result of selective pressure, the alleles we reported will need to be compared to frequencies of neutral markers. For all three putative orthologous chitinase pairs, we observed a distribution of haplotypes suggestive of differentiation between western-lodgepole and eastern-jack pine populations. Calculation of Fst values across variant sites in all three orthologous chitinase pairs displayed high levels of population differentiation between lodgeople and jack pine samples. The greater haplotype variation found within the class IV chitinase across lodgepole pine populations could be evidence of greater selective pressure on these populations. Lodgepole samples displayed substantially higher levels of allelic richness across variant sites within the class IV chitinase compared to jack pine samples. Microarray expression and qRT-PCR data demonstrated that class IV chitinases display the highest fold change in expression in response to *G. clavigera,* and likely play a more prominent role in defense response compared to other classes of chitinases. Therefore, we would expect greater selective pressure would be placed on class IV chitinases in comparison to other classes.

Differences between naïve and co-evolved pine hosts to MPB and its fungal associates have been described at the ecological and physiological level (Cudmore *et al.* 2010; Raffa *et al.* 2013; Arango *et al.* 2016). The significant impact of water deficit on pine host defense has also been described at the physiological level (Lusibrink *et al.* 2011; Arango-Velez *et al.* 2014). The results of this study have revealed that there are substantial differences in the transcriptomic response of lodgepole and jack pine to MPB fungal associate *G. clavigera* and these differences occur among genes that are an important part of defense response. Our results also reveal that water deficit greatly impacts defense response of both species, however to a greater degree in lodgepole pine. We have provided evidence of the molecular and genetic mechanisms which

underlie earlier observations made at at the ecological and physiological level. Furthermore, our results describe the roles played by different classes of the pine chitinase gene family in response to *G. clavigera,* and provide evidence for greater diversification of defense related gene families in long lived species such as pine.

5.0 References

Abad L.R., D'Urzo M.P., Liu D., Narasimhan M.L., Reuveni M., Zhu J.K., Niu X., Singh N.K., Hasegawa P.M. & Bressen R.A. (1996) Antifungal activity of tobacco osmotin has specificity and involves plasma membrane permeabilization. *Plant Science* **118**,11–23.

Abascal F., Zardoya R., & Posada D. (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics 21*, 2104-2105.

Adams H. D., Guardiola-Claramonte M., Barron-Gafford G. A., Villegas J. C., Breshears D. D., Zou C. B. & Huxman T. E. (2009) Temperature sensitivity of drought-induced tree mortality portends increased regional die-off under global-change-type drought. *Proceedings of the National Academy of Sciences USA* **106,** 7063-7066.

Adomas A., Heller G., Li G., Olson Å., Chu T. M., Osborne J. & Dean R. A. (2007) Transcript profiling of a conifer pathosystem: response of *Pinus sylvestris* root tissues to pathogen (*Heterobasidion annosum*) invasion. *Tree Physiology* **27,** 1441-1458.

Araki T. & Torikata T. (1995) Structural classification of plant chitinases: two subclasses in class I and class II chitinases. *Bioscience, Biotechnology & Biochemistry* **59**, 336-338

Arango-Velez A., El Kayal W., Copeland C. C., Zaharia L. I., Lusebrink I. & Cooke, J. E. (2016). Differences in defense responses of *Pinus contorta* and *Pinus banksiana* to the mountain beetle fungal associate *Grosmannia clavigera* are affected by water deficit. *Plant Cell & Environment* **39,** 726-744.

Arango-Velez A., Galindo-González L.M., Meents M.J., El Kayal W., Cooke B.J., Linsky J., Lusebrink I. & Cooke J.E.K. (2014) Influence of water deficit on the molecular responses of *Pinus contorta* × *Pinus banksiana* mature trees to infection by the mountain beetle fungal associate, *Grosmannia clavigera*. *Tree Physiology* **34**, 1220–1239.

Bale J. S. (2002). Insects and low temperatures: from molecular biology to distributions and abundance. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **357**, 849-862.

Ballard, R. G., Walsh, M. A., & Cole, W. E. (1982) Blue-stain fungi in xylem of lodgepole pine: a light-microscope study on extent of hyphal distribution. *Canadian Journal of Botany*, **60,** 2334-2341.

Beckman C. H. (2000) Phenolic-storing cells: keys to programmed cell death and periderm formation in wilt disease resistance and in general defense responses in plants? *Physiological and Molecular Plant Pathology* **57**, 101-110.

Bedon F., Bomal C., Caron S., Levasseur C., Boyle B., Mansfield S. D. & MacKay J. (2010) Subgroup 4 R2R3-MYBs in conifer trees: gene family expansion and contribution to the isoprenoid-and flavonoid-oriented responses. *Journal of Experimental Botany*, **67**, 3827-3864.

Benjamini Y. & Hochberg Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* **57**, 289–300.

Bentz B. J., & Mullins D. (1999) Ecology of mountain pine beetle (Coleoptera: Scolytidae) cold hardening in the intermountain west. *Environmental Entomology* **28***,* 577-587.

Bentz B. J., Logan J. A. & Vandygriff J. C. (2001) Latitudinal variation in *Dendroctonus ponderosae* (Coleoptera: Scolytidae) development time and adult size. *The Canadian Entomologist* **133,** 375-387.

Berryman A. A. (1982) Mountain pine beetle outbreaks in Rocky Mountain lodgepole pine forests. *Journal of Forestry* **80**, 410-419.

Binder B. M. (2008) The ethylene receptors: complex perception for a simple gas. *Plant Science* **175,** 8-17.

Bishop J. G., Dean A. M. & Mitchell-Olds T. (2000) Rapid evolution in plant chitinases: molecular targets of selection in plant-pathogen coevolution. *Proceedings of the National Academy of Sciences USA* **97***,* 5322-5327.

Boeckmann B., Bairoch A., Apweiler R., Blatter M. C., Estreicher A., Gasteiger E., & Pilbout S. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research* **31**, 365-370.

Bonnett T. R., Robert J. A., Pitt C., Fraser J. D., Keeling C. I., Bohlmann J. & Huber D. P. (2012) Global and comparative proteomic profiling of overwintering and developing mountain pine beetle, *Dendroctonus ponderosae* (Coleoptera: Curculionidae), larvae. *Insect Biochemistry and Molecular Biology* **42,** 890-901.

Boone C.K., Aukema B.H., Bohlmann J., Carroll A.L. & Raffa K.F. (2011) Efficacy of tree defense physiology varies with bark beetle population density: a basis for positive feedback in eruptive species. *Canadian Journal of Forest Research* **41**, 1174–1188.

Borden, J. H., Pureswaran, D. S., & Lafontaine, J. P. (2008). Synergistic blends of monoterpenes for aggregation pheromones of the mountain pine beetle (Coleoptera: Curculionidae). *Journal of Economic Entomology*, *101*(4), 1266-1275.

Breshears D. D., Myers O. B., Meyer C. W., Barnes F. J., Zou C. B., Allen C. D. & Pockman W. T. (2008) Tree die-off in response to global change-type drought: mortality insights from a decade of plant water potential measurements. *Frontiers in Ecology and the Environment* **7,** 185-189.

Burke J. L. & Carroll A. L. (2016) The influence of variation in host tree monoterpene composition on secondary attraction by an invasive bark beetle: Implications for range expansion and potential host shift by the mountain pine beetle. *Forest Ecology and Management* **359**, 59-64.

Byun-McKay A., Godard K.A., Toudefallah M., Martin D.M., Alfaro R., King J. & Bohlmann J. (2006) Wound-induced terpene synthase gene expression in Sitka spruce that exhibit resistance or susceptibility to attack by the white weevil. *Plant Physiology* **140**, 1009–1021.

Carlson M. R., Murphy J. C., Berger V. G. & Ryrie L. F. (1999) Genetics of elevational adaptations of lodgepole pine in the interior. *Journal of Sustainable Forestry* **10**, 35-44.

Chang S., Puryear J. & Cairney J. (1993) A simple and efficient method for isolating RNA from pine trees. *Plant molecular biology reporter* **11**, 113-116.

Chhin S., Hogg E.H., Lieffers V.J. & Huang S. (2008) Potential effects of climate change on the growth of lodgepole pine across diameter size classes and ecological regions. *Forest Ecology and Management* **256**, 1692–1703.

Chini A., Fonseca S., Fernandez G., Adie B., Chico J. M., Lorenzo O. & Micol J. L. (2007) The JAZ family of repressors is the missing link in jasmonate signalling. *Nature* **448,** 666-671.

CLC Genomics Workbench v3. 6.(2010) *Now new version can be available at http://www. clcbio. com/products/clcgenomicsworkbench*.

Cletus J., Balasubramanian V., Vashisht D. & Sakthivel N. (2013) Transgenic expression of plant chitinases to enhance disease resistance. *Biotechnology Letters* **35**, 1719-1732.

Critchfield W. B., & Little E. L. (1966) *Geographic distribution of the pines of the world*. US Department of Agriculture, Forest Service. (No. 991)

Cudmore T. J., Björklund N., Carroll A. L. & Staffan Lindgren, B. (2010) Climate change and range expansion of an aggressive bark beetle: evidence of higher beetle reproduction in naive host tree populations. *Journal of Applied Ecology **47*,**1036-1043.

Cui H., Tsuda K., & Parker J. E. (2015) Effector-triggered immunity: from pathogen perception to robust defense. *Annual Review of Plant Biology* **66,** 487-511.

Cullingham C. I., James P., Cooke J. E. & Coltman D. W. (2012) Characterizing the physical and genetic structure of the lodgepole pine× jack pine hybrid zone: mosaic structure and differential introgression. *Evolutionary Applications* **5**, 879-891.

Cullingham C.I., Cooke J.E.K., Dang S., Davis C.S., Cooke B.J. & Coltman D.W. (2011) Mountain pine beetle host-range expansion threatens the boreal forest. *Molecular Ecology* **20**, 2157-2171.

Dangl J. L., Horvath D. M. & Staskawicz B. J. (2013). Pivoting the plant immune system from dissection to deployment. *Science* **341**, 746-751.

Dao T. T. H., Linthorst H. J. M. & Verpoorte R. (2011) Chalcone synthase and its functions in plant resistance. *Phytochemistry Reviews **10***, 397-412.

Davis J.M., Wu H.G., Cooke J.E.K., Reed J.M., Luce K.S. & Michler C.H. (2002) Pathogen challenge, salicylic acid, and jasmonic acid regulate expression of chitinase gene homologs in pine. *Molecular Plant-Microbe Interactions* **15**, 380–387.

Deslandes, L., & Rivas, S. (2012) Catch me if you can: bacterial effectors and plant targets. *Trends in Plant Science* **17**, 644-655.

DiGuistini S., Wang Y., Liao N. Y., Taylor G., Tanguay P., Feau N. & Tsui, C. K. (2011) Genome and transcriptome analyses of the mountain pine beetle-fungal symbiont *Grosmannia clavigera*, a lodgepole pine pathogen. *Proceedings of the National Academy of Sciences USA* **108**, 2504-2509.

Dombrecht B., Xue G. P., Sprague S. J., Kirkegaard J. A., Ross J. J., Reid J. B. & Kazan K. (2007) MYC2 differentially modulates diverse jasmonate-dependent functions in Arabidopsis. *The Plant Cell* **19**, 2225-2245.

Edgar R. C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461.

El Kayal W., Allen C.C.G., Ju C.J., Adams E., King-Jones S., Zaharia L.I., Abrams S.R. & Cooke J.E.K. (2011) Molecular events of apical bud formation in white spruce, *Picea glauca. Plant Cell & Environment* **34,** 480-500.

Eyles A., Bonello P., Ganley E. & Mohammed C. (2010) Induced resistance to pests and pathogens in trees. *New Phytologist* **185,** 893–908

Fossdal C. G., Nagy N. E., Johnsen Ø. & Dalen L. S. (2007). Local and systemic stress responses in Norway spruce: similarities in gene expression between a compatible pathogen interaction and drought stress. *Physiological and Molecular Plant Pathology*, **70,** 161-173.

Fossdal C. G., Yaqoob N., Krokene P., Kvaalen H., Solheim H., & Yakovlev I. A. (2012). Local and systemic changes in expression of resistance genes, nb-lrr genes and their putative microRNAs in Norway spruce after wounding and inoculation with the pathogen *Ceratocystis polonica*. *BMC Plant Biology* **12**,105.

Fossdal, C. G., Sharma, P., & Lönneborg, A. (2001). Isolation of the first putative peroxidase cDNA from a conifer and the local and systemic accumulation of related proteins upon pathogen infection. *Plant Molecular Biology* **47,** 423-435.

Franceschi V. R., Krokene P., Christiansen E. & Krekling T. (2005) Anatomical and chemical defenses of conifer bark against bark beetles and other pests. *New Phytologist* **167**, 353-376.

Franceschi V., Krekling T. & Christiansen E. (2002) Application of methyl jasmonate on *Picea abies* (Pinaceae) stems induces defense-related responses in phloem and xylem. *American Journal of Botany* **89**, 578–586.

Gao Y., Jia S., Wang C., Wang F., Wang F. & Zhao K. (2016) BjMYB1, a transcription factor implicated in plant defence through activating BjCHI1 chitinase expression by binding to a W-box-like element. *Journal of Experimental Botany* **67,** 4647-4658.

Gierasch L. M. (1989) Signal sequences. *Biochemistry* **28**, 923-930.

Glazebrook J. (2005) Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annual Review of Phytopathology **43*,** 205-227.

Galindo-González L. M., El Kayal W., Morris J. S., & Cooke, J. E. (2015) Diverse chitinases are invoked during the activity-dormancy transition in spruce. *Tree Genetics & Genomes* **11**, 1-21.

Gordon A. & Hannon G. J. (2010) Fastx-toolkit. *FASTQ/A short-reads preprocessing tools (unpublished) http://hannonlab. cshl. edu/fastx_toolkit*.

Goudet J. (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9. 3).

Gould N., Reglinski T., Northcott G. L., Spiers M. & Taylor J. T. (2009) Physiological and biochemical responses in *Pinus radiata* seedlings associated with methyl jasmonate-induced resistance to *Diplodia pinea*. *Physiological and Molecular Plant Pathology **74***, 121-128.

Government of Alberta. Mountain pine beetle in Alberta. http://aep.alberta.ca/lands-forests/mountain-pine-beetle/documents/BeetleBattleHeightens-Aug11-2016.pdf (Aug 11 2016, date last accessed).

Government of British Columbia. Facts about B.C.'s mountain pine beetle. https://www.for.gov.bc.ca/hfp/mountain_pine_beetle/Updated-Beetle-Facts_April2013.pdf (Aug 11 2016, date last accessed).

Hall D. E., Yuen M. M., Jancsik S., Quesada A. L., Dullat H. K., Li M. & Chan, S. K. (2013) Transcriptome resources and functional characterization of monoterpene synthases for two host species of the mountain pine beetle, lodgepole pine (Pinus contorta) and jack pine (Pinus banksiana). *BMC Plant Biology* **13**, 1471-2229.

Hanfrey C., Fife M. & Buchanan-Wollaston V. (1996) Leaf senescence in Brassica napus: expression of genes encoding pathogenesis-related proteins. *Plant Molecular Biology* **30**,597-609.

Harrington T. C. (1993) Biology and taxonomy of fungi associated with bark beetles. *Beetle-pathogen interactions in conifer forests* (eds T.D. Schowatter & G.M. Filip) **pp.** 37-58. Academic Press Inc, London.

Hietala A.M., Kvaalen H., Schmidt A., Jøhnk N., Solheim H. & Fossdal C.G. (2004) Temporal and spatial profiles of chitinase expression by Norway spruce in response to bark

colonization by *Heterobasidion annosum*. *Applied Environmental Microbiology* **70**, 3948–3953.

Hogg, E.H., Michaelian M. (2015). Factors affecting fall down rates of dead aspen (*Populus tremuloides*) biomass following severe drought in west-central Canada. *Global Change Biology*, **21**, 1968-1979.

Hudgins J. W., Christiansen E. & Franceschi V. R. (2004) Induction of anatomically based defense responses in stems of diverse conifers by methyl jasmonate: a phylogenetic perspective. *Tree Physiology* **24**, 251-264.

Hudgins J. W., Ralph S. G., Franceschi V. R. & Bohlmann J. (2006) Ethylene in induced conifer defense: cDNA cloning, protein expression, and cellular and subcellular localization of 1-aminocyclopropane-1-carboxylate oxidase in resin duct and phenolic parenchyma cells. *Planta* **224**, 865-877.

Hudgins J.W. & Franceschi V.R. (2004) Methyl jasmonate-induced ethylene production is responsible for conifer phloem defense responses and reprogramming of stem cambial zone for traumatic resin duct formation. *Plant Physiology* **135**, 2134–2149.

Hudgins, J. W., Christiansen, E., & Franceschi, V. R. (2003). Methyl jasmonate induces changes mimicking anatomical defenses in diverse members of the Pinaceae. *Tree Physiology*, *23*(6), 361-371.

Hynum B. G. & Berryman A. A. (1980) *Dendroctonus ponderosae* (Coleoptera: Scolytidae): pre-aggregation landing and gallery initiation on lodgepole pine. *The Canadian Entomologist* **112,** 185-191.

IPCC (2013) Climate Change 2013: The Physical Science Basis. Working Group I: Contribution to the Intergovernmental Panel on Climate Change Fifth Assessment Report. *Cambridge University Press*, Cambridge, UK.

Islam M. A., Sturrock R. N., & Ekramoddoullah, A. K. M. (2011). Conifer chitinases. *The American Journal of Plant Science and Biogechnology* **5,** 22-33.

Islam M. A., Sturrock R. N., Williams H. L. & Ekramoddoullah A. K. M. (2010) Identification, characterization, and expression analyses of class II and IV chitinase genes from Douglas-fir seedlings infected by *Phellinus sulphurascens*. *Phytopathology* **100(4),** 356-366

Jarząbek M., Pukacki P. M. & Nuc K. (2009) Cold-regulated proteins with potent antifreeze and cryoprotective activities in spruces (Picea spp.). *Cryobiology* **58**, 268-274.

Johnsen Ø., Fossdal C. G., Nagy N., Mølmann J., Dahlen O. L. A. & Skrøppa T. (2005) Climatic adaptation in Picea abies progenies is affected by the temperature during zygotic embryogenesis and seed maturation. *Plant, Cell & Environment* **28**, 1090-1102.

Jones D. A., & Takemoto D. (2004). Plant innate immunity–direct and indirect recognition of general and specific pathogen-associated molecules. *Current Opinion in Immunology* **16,** 48-62.

Jones J. D. & Dangl J. L. (2006) The plant immune system. *Nature* **444,** 323-329.

Karasov T. L., Horton, M. W. & Bergelson J. (2014) Genomic variability as a driver of plant–pathogen coevolution? *Current Opinion in Plant Biology* **18,** 24-30.

Katoh K., Misawa K., Kuma K. I., & Miyata T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059-3066.

Keeling C. I. & Bohlmann J. (2006a) Diterpene resin acids in conifers. *Phytochemistry* ***67***, 2415-2423.

Keeling C.I. & Bohlmann J. (2006b) Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defense of conifers against insects and pathogens. *New Phytologist* **170**, 657–675.

Kenkel N., Hendrie M. L. & Bella I. E. (1997) A long-term study of *Pinus banksiana* population dynamics. *Journal of Vegetation Science* **8,** 241-254.

Kessler A., Halitschke R. & Baldwin I.T. (2004) Silencing the jasmonate cascade: induced plant defenses and insect populations. *Science* **305**,665–668

Khadempour L., LeMay V., Jack D., Bohlmann J. & Breuil C. (2012) The relative abundance of mountain pine beetle fungal associates through the beetle life cycle in pine trees. *Microbial Ecology* **64,** 909-917.

Kim M. K., Jeon J. H., Fujita M., Davin L. B. & Lewis N. G. (2002) The western red cedar (*Thuja plicata*) 8-8′ DIRIGENT family displays diverse expression patterns and conserved monolignol coupling specificity. *Plant Molecular Biology* **49,** 199-214.

Kolb T.E., Homberg K.M., Wagner M.R. & Stone J.E. (1998) Regulation of ponderosa foliar physiology and insect resistance mechanisms by by basal area treatments. *Tree Physiology* **18**,375–381.

Kolosova N. & Bohlmann J. (2012) Conifer defense against insects and fungal pathogens. In *Growth and defence in plants.* (eds R. Matyssek & H. Schnyder), pp. 85-109. Springer, Berlin Heidelberg, Germany.

Kolosova N., Breuil C. & Bohlmann J. (2014). Cloning and characterization of chitinases from interior spruce and lodgepole pine. *Phytochemistry* **101,** 32-39.

Kozlowski G., & Métraux J. P. (1998) Infection of Norway spruce (*Picea abies* (L.) Karst.) seedlings with *Pythium irregulare* Buism. and *Pythium ultimum* Trow.: histological and biochemical responses. *European Journal of Plant Pathology* **104,** 225-234.

Kozlowski G., Buchala A. & Métraux J. P. (1999) Methyl jasmonate protects Norway spruce [*Picea abies* (L.) Karst.] seedlings against *Pythium ultimum* Trow. *Physiological and Molecular Plant Pathology* **55**, 53-58.

Krekling T., Franceschi V. R., Berryman A. A. & Christiansen E. (2000) The structure and development of polyphenolic parenchyma cells in Norway spruce (*Picea abies*) bark. *Flora (Jena)* **195,** 354-369.

Krokene P., Nagy N.E. & Solheim H. (2008) Methyl jasmonate and oxalic acid treatment of Norway spruce: anatomically based defense responses and increased resistance against fungal infection. *Tree Physiology* **28**, 29–35.

Laloum T., De Mita S., Gamas P., Baudin M., & Niebel A. (2013) CCAAT-box binding transcription factors in plants: Y so many?. *Trends in Plant Science* **18**, 157-166.

Lee S., Kim J. J. & Breuil C. (2006) Diversity of fungi associated with mountain pine beetle, *Dendroctonus ponderosae*, and infested lodgepole pines in British Columbia. Natural Resources Canada, Canadian Forest Service no. 06.

Lee S., Kim J. J., Fung S. & Breuil C. (2003) A PCR-RFLP marker distinguishing *Ophiostoma clavigerum* from morphologically similar *Leptographium* species associated with bark beetles. *Canadian Journal of Botany* **81,** 1104-1112.

Little E.L. Jr. (1971) Atlas of the United States Trees, Conifers and Important Hardwoods, Volume 1. *US Department of Agriculture Miscellaneous Publication* No. 1146, Washington, D.C.

Liu J. J., Ekramoddoullah A. K. & Yu X. (2003) Differential expression of multiple PR10 proteins in western white pine following wounding, fungal infection and cold-hardening. *Physiologia Plantarum* **119**, 544-553.

Liu J. J. & Ekramoddoullah A. K. M. (2004) Isolation, genetic variation and expression of TIR-NBS-LRR resistance gene analogs from western white pine (*Pinus monticola* Dougl. ex. D. Don.). *Molecular Genetics and Genomics* **270,** 432-441.

Liu J. J., Ekramoddoullah A. K. & Zamani A. (2005) A class IV chitinase is up-regulated by fungal infection and abiotic stresses and associated with slow-canker-growth resistance to *Cronartium ribicola* in western white pine (*Pinus monticola*). *Phytopathology* **95,** 284-291.

Liu J. J. & Ekramoddoullah A. K. (2007) The CC-NBS-LRR subfamily in *Pinus monticola*: targeted identification, gene expression, and genetic linkage with resistance to *Cronartium ribicola*. *Phytopathology* **97,** 728-736.

Liu J. J., Zamani A. & Ekramoddoullah A. K. (2010) Expression profiling of a complex thaumatin-like protein family in western white pine. *Planta* **231,** 637-651.

Liu J. J., Sniezko R. A. & Ekramoddoullah A. K. (2011) Association of a novel *Pinus monticola* chitinase gene (PmCh4B) with quantitative resistance to *Cronartium ribicola*. *Phytopathology* **101,** 904-911.

Liu J. J., Chan D., Sturrock R. & Sniezko R. A. (2014) Genetic variation and population differentiation of the endochitinase gene family in *Pinus monticola*. *Plant Systematics and Evolution* **300,** 1313-1322.

Logan, J. A., & Bentz, B. J. (1999) Model analysis of mountain pine beetle (Coleoptera: Scolytidae) seasonality. *Environmental Entomology* **28,** 924-934.

Lombardero M.H., Ayres M.P., Lorio P.L. Jr. & Ruel J.J. (2000) Environmental effects on constitutive and inducible resin defences of *Pinus taeda*. *Ecology Letters* **3,** 329–339.

Lorenz, W. W., Yu, Y. S., Simões, M., & Dean, J. F. (2009). Processing the loblolly pine PtGen2 cDNA microarray. *Journal of Visualized Experiments*, **25**, 1182-1182.

Lorio P.L., Stephen F.M. & Paine T.D. (1995) Environment and ontogeny modify loblolly response to induced acute water deficits and bark beetle attack. *Forest Ecology and Management* **73**, 97-110.

Lu, M., Wingfield, M. J., Gillette, N. E., Mori, S. R., & Sun, J. H. (2010). Complex interactions among host pines and fungi vectored by an invasive bark beetle. *New Phytologist* **187**, 859-866.

Lusebrink I., Evenden M.I., Blanchet F.G., Cooke J.E.K. & Erbilgin N. (2011) Effect of water stress and fungal inoculation on monoterpene emission from an historical and a new host of the mountain pine beetle. *Journal of Chemical Ecology* **37**, 1013–1026.

Maddison W. P. & Maddison D. R. (2010) Mesquite: a modular system for evolutionary analysis. Version 2.6. 2009. *URL http://mesquiteproject. org*

Maiti R., Van Domselaar G. H., Zhang H. & Wishart D. S. (2004) SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Research* **32**, 590-594.

Martin D., Tholl D. & Gershenzon J. (2002) Methyl jasmonate induces traumatic resin ducts, terpenoid resin biosynthesis, and terpenoid accumulation in developing xylem of Norway spruce stems. *Plant Physiology* **129**, 1003–1018.

McDonald B. A. & Linde C. (2002) Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology* **40**, 349-379.

McDowell N., Pockman W. T., Allen C. D., Breshears D. D., Cobb N., Kolb T. & Yepez E. A. (2008) Mechanisms of plant survival and mortality during drought: why do some plants survive while others succumb to drought? *New Phytologist* **178,** 719-739.

Michaelian M., Hogg E.H., Hall R.J. & Arsenault E. (2011) Massive mortality of aspen following severe drought along the southern edge of the Canadian boreal forest. *Global Change Biology* **17**, 2084–2094

Miller B., Madilao L. L., Ralph S. & Bohlmann J. (2005) Insect-induced conifer defense. White pine weevil and methyl jasmonate induce traumatic resinosis, de novo formed volatile emissions, and accumulation of terpenoid synthase and putative octadecanoid pathway transcripts in Sitka spruce. *Plant Physiology* **137,** 369-382.

Min X.J., Butler G., Storms R. and Tsang A. (2005) OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Residues* Web Server Issue W677-W680

Nagy N. E., Fossdal C. G., Krokene P., Krekling T., Lönneborg A. & Solheim H. (2004) Induced responses to pathogen infection in Norway spruce phloem: changes in polyphenolic parenchyma cells, chalcone synthase transcript levels and peroxidase activity. *Tree Physiology* **24**, 505-515.

Natural Resources Canada, Canadian Forest Service. (2015) The State of Canada's Forests. Annual Report 2015. Candian Forest Service Publication, Ottawa, Ontario, Canada. ISSN 1488-2736.

Neuhaus J.M. (1999) Plant chitinases (PR-3, PR-4, PR-8, PR-11). In *Pathogenesis-Related Proteins in Plants* (eds Datta S.K., Muthukrishnan S.), pp. 77–105. CRC Press, Boca Raton, Florida, USA.

Nielsen H., Engelbrecht J., Brunak S. & von Heijne G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* **10**, 1-6.

Nystedt B., Street N. R., Wetterbom A., Zuccolo A., Lin Y. C., Scofield D. G. & Vicedomini R. (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature* **497,** 579-584.

Osmond R.I., Hrmova M., Fontaine F., Imberty A. & Fincher G.B. (2001) Binding interactions between barley thaumatin-like proteins an (1, 3)-β-d-glucans. *European Journal of Biochemistry* **15**,4190–419.

Palaniswamy S. K., James S., Sun H., Lamb R. S., Davuluri R. V., & Grotewold E. (2006). AGRIS and AtRegNet. a platform to link cis-regulatory elements and transcription factors into regulatory networks. *Plant Physiology* **140**, 818-829.

Pasonen H. L., Seppänen S. K., Degefu Y., Rytkönen A., von Weissenberg K. & Pappinen A. (2004) Field performance of chitinase transgenic silver birches (Betula pendula): resistance to fungal diseases. *Theoretical and Applied Genetics* **109,** 562-570.

Patzlaff A, McInnis S, Courtenay A, et al. 2003b. Characterisation of a pine MYB that regulates lignification. *The Plant Journal* **36**, 743–754.

Patzlaff A, Newman LJ, Dubos C, Whetten RW, Smith C, McInnis S, Bevan MW, Sederoff RR, Campbell MM. 2003a. Characterisation of PtMYB1, an R2R3-MYB from pine xylem. *Plant Molecular Biology* **53**, 597–608.

Pavy N., Boyle B., Nelson C., Paule C., Giguère I., Caron S.& Cooke J. (2008) Identification of conserved core xylem gene sets: conifer cDNA microarray development, transcript profiling and computational analyses. *New Phytologist*, **180,** 766-786.

Petersen T. N., Brunak S., von Heijne G. & Nielsen H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785-786.

Pickel B. & Schaller A. (2013) Dirigent proteins: molecular characteristics and potential biotechnological applications. *Applied Microbiology and Biotechnology* **97**, 8427-8438.

Pierce Jr H. D., Conn J. E., Oehlschlager A. C. & Borden J. H. (1987) Monoterpene metabolism in female mountain pine beetles, *Dendroctonus ponderosae* Hopkins, attacking ponderosa pine. *Journal of Chemical Ecology* **13,** 1455-1480.

Pieterse C. M., Leon-Reyes A., Van der Ent S., & Van Wees S. C. (2009) Networking by small-molecule hormones in plant immunity. *Nature Chemical Biology* **5,** 308-316.

Piggott N., Ekramoddoullah A. K., Liu J. J. & Yu X. (2004) Gene cloning of a thaumatin-like (PR-5) protein of western white pine (Pinus monticola D. Don) and expression studies of members of the PR-5 group. *Physiological and Molecular Plant Pathology* **64,** 1-8.

Popp M. P., Johnson J. D. & Lesney M. S. (1995) Changes in ethylene production and monoterpene concentration in slash pine and loblolly pine following inoculation with bark beetle vectored fungi. *Tree Physiology* **15,** 807-812.

Qiao H., Chang K. N., Yazaki J. & Ecker J. R. (2009) Interplay between ethylene, ETP1/ETP2 F-box proteins, and degradation of EIN2 triggers ethylene responses in Arabidopsis. *Genes & Development* **23,** 512-521.

Raffa K. F. & Berryman A. A. (1982). Physiological differences between lodgepole pines resistant and susceptible to the mountain pine beetle and associated microorganisms. *Environmental Entomology* **11**, 486-492.

Raffa K. F., Aukema B. H., Bentz B. J., Carroll A. L., Hicke J. A., Turner M. G. & Romme W. H. (2008) Cross-scale drivers of natural disturbances prone to anthropogenic amplification: the dynamics of bark beetle eruptions. *Bioscience* **58,** 501-517.

Raffa K. F., Powell E. N. & Townsend P. A. (2013) Temperature-driven range expansion of an irruptive insect heightened by weakly coevolved plant defenses. *Proceedings of the National Academy of Sciences USA **110***, 2193-2198.

Raffa K.F., Aukema B.H., Erbilgin N., Klepzig K.D. & Wallin K.F. (2005) Interactions among conifer terpenoids and bark beetles across multiple levels of scale: an attempt to

understand links between population patterns and physiological processes. *Recent Advances Phytochemistry* **39**, 79–118

Raffaele S., Farrer R. A., Cano L. M., Studholme D. J., MacLean D., Thines M. & Meyers B. C. (2010) Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science* **330**, 1540-1543.

Ralph S. G., Hudgins J. W., Jancsik S., Franceschi V. R. & Bohlmann J. (2007) Aminocyclopropane carboxylic acid synthase is a regulated step in ethylene-dependent induced conifer defense. Full-length cDNA cloning of a multigene family, differential constitutive, and wound-and insect-induced expression, and cellular and subcellular localization in spruce and Douglas fir. *Plant Physiology* **143**, 410-424.

Ralph S. G., Yueh H., Friedmann M., Aeschliman D., Zeznik J. A., Nelson C. C. & Marra M. A. (2006a) Conifer defence against insects: microarray gene expression profiling of Sitka spruce (*Picea sitchensis*) induced by mechanical wounding or feeding by spruce budworms (*Choristoneura occidentalis*) or white pine weevils (*Pissodes strobi*) reveals large-scale changes of the host transcriptome. *Plant, Cell & Environment* **29(8),** 1545-1570.

Ralph S., Park J. Y., Bohlmann J. & Mansfield, S. D. (2006b) Dirigent proteins in conifer defense: gene discovery, phylogeny, and differential wound-and insect-induced expression of a family of DIR and DIR-like genes in spruce (*Picea* spp.). *Plant Molecular Biology* **60**, 21-40.

Rapoport T. A. (1992) Transport of proteins across the endoplasmic reticulum membrane. *Science* **258,** 931-931.

Rice P., Longden I. & Bleasby A. (2000) EMBOSS: the European molecular biology open software suite. *Trends in Genetics* **16,** 276-277.

Richardson, D. M. (2000). *Ecology and biogeography of Pinus*. Cambridge University Press.

Roe A. D., Rice A. V., Bromilow S. E., Cooke J. E. & Sperling F. A. (2010) Multilocus species identification and fungal DNA barcoding: insights from blue stain fungal symbionts of the mountain pine beetle. *Molecular Ecology Resources* **10,** 946-959.

Rombel I. T., Sykes K. F., Rayner S. & Johnston S. A. (2002) ORF-FINDER: a vector for high-throughput gene identification. *Gene* **282**, 33-41.

Saeed A.I., Sharov V., White J., Li J., Liang W., Bhagabati N., Braisted J., Klapa M., Currier T., Thiagarajan M., Sturn A., Snuffin M., Rezantsev A., Popov D., Ryltsov A., Kostukovich E., Borisovsky I., Liu Z., Vinsavich A., Trush V. & Quackenbush J. (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**,374-8.

Safranyik L. & Carroll, A. L. (2006). The biology and epidemiology of the mountain pine beetle in lodgepole pine forests. In: *The Mountain Pine Beetle: A Synthesis of its Biology,*

Management, and Impacts on Lodgepole Pine (eds L. Safranyik & B. Wilson), pp. 3-66. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC., Canada. ISBN 0-662-42623-1. pp. 304.

Safranyik L., Carroll A. L., Régnière J. Langor D. W., Riel W. G., Shore T. L., Peter B., Cooke B.J., Nealis V.G. & Taylor S.W. (2010) Potential for range expansion of mountain pine beetle into the boreal forest of North America. *The Canadian Entomologist* **142**, 415–442.

Sampedro L., Moreira X. & Zas R. (2011). Resistance and response of *Pinus pinaster* seedlings to *Hylobius abietis* after induction with methyl jasmonate. *Plant Ecology* **212,** 397–401.

Sánchez-Rodríguez C., Bauer S., Hématy K., Saxe F., Ibáñez A. B., Vodermaier V. & Neumetzler L. (2012) Chitinase-like1/pom-pom1 and its homolog CTL2 are glucan-interacting proteins important for cellulose biosynthesis in Arabidopsis. *The Plant Cell* **24**, 589-607.

Seager, R., Ting, M., Held, I., Kushnir, Y., Lu, J., Vecchi, G., ... & Li, C. (2007). Model projections of an imminent transition to a more arid climate in southwestern North America. *Science* **316,** 1181-1184.

Shore T. L., Safranyik L., Hawkes B. C. & Taylor S. W. (2006) Effects of the mountain pine beetle on lodgepole pine stand structure and dynamics. *In: The Mountain Pine Beetle: A Synthesis of its Biology, Management, and Impacts on Lodgepole Pine (eds L. Safranyik & B. Wilson),* pp. 95-116. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC., Canada. ISBN 0-662-42623-1. pp. 304.

Solheim, H., & Krokene, P. (1998) Growth and virulence of mountain pine beetle associated blue-stain fungi, Ophiostoma clavigerum and Ophiostoma montium. *Canadian Journal of Botany* **76**, 561-566.

Song S., Huang H., Gao H., Wang J., Wu D., Liu X., Yang S., Zhai Q., Li C., Qi T. (2014) Interaction between MYC2 and ETHYLENE INSENSITIVE3modulates antagonism between jasmonate and ethylene signaling in Arabidopsis. *Plant Cell* **26**, 263–279

Six D. L. & Klepzig K. D. (2004) Dendroctonus bark beetles as model systems for studies on symbiosis. *Symbiosis **37,*** 207-232.

Six D.L. & Paine T.D. (1998) Effects of mycangial fungi and host tree species on progeny survival and emergence of *Dendroctonus ponderosae* (Coleoptera: Scolytidae). *Environmental Entomology* **27,**1393–1401

Smale S. T. & Kadonaga J. T. (2003) The RNA polymerase II core promoter. *Annual Review of Biochemistry*, ***72,*** 449-479.

Smyth G.K. (2005) LIMMA: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (eds R. Gentleman, V. Carey, S. Dudoit, R. Irizarry Irizarry &W. Huber), pp. 397–420. Springer, New York.

Solano R., Stepanova A., Chao Q. & Ecker J. R. (1998) Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. *Genes & Development* **12,** 3703-3714.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312-1313.

Taira T., Hayashi H., Tajiri Y., Onaga S., Uechi G. I., Iwasaki H. & Fukamizo T. (2009) A plant class V chitinase from a cycad (Cycas revoluta): biochemical characterization, cDNA isolation, and posttranslational modification. *Glycobiology* **19,** 1452-1461.

Taylor S. W., Carroll A. L., Alfaro R. I. & Safranyik L. (2006) Forest, climate and mountain pine beetle outbreak dynamics in western Canada. *In: The Mountain Pine Beetle: A Synthesis of its Biology, Management, and Impacts on Lodgepole Pine (*eds L. Safranyik & B. Wilson), pp. 67-94. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC., Canada. ISBN 0-662-42623-1. pp. 304.

Thines B., Katsir L., Melotto M., Niu Y., Mandaokar A., Liu G. & Browse J. (2007). JAZ repressor proteins are targets of the SCFCOI1 complex during jasmonate signalling. *Nature* **448,** 661-665.

Thomma B. P., Eggermont K., Penninckx I. A., Mauch-Mani B., Vogelsang R., Cammue B. P. & Broekaert W. F. (1998) Separate jasmonate-dependent and salicylate-dependent defense-response pathways in Arabidopsis are essential for resistance to distinct microbial pathogens. *Proceedings of the National Academy of Sciences USA* **95**, 15107-15111.

Tomlin E. S., Antonejevic E., Alfaro R. I., & Borden J. H. (2000) Changes in volatile terpene and diterpene resin acid composition of resistant and susceptible white spruce leaders exposed to simulated white pine weevil damage. *Tree Physiology* **20,** 1087-1095.

Tomlin E. S., Borden J. H. & Pierce Jr H. D. (1996). Relationship between cortical resin acids and resistance of Sitka spruce to the white pine weevil. *Canadian Journal of Botany **74*,** 599-606.

Tsui C.K., Roe A.D., El-Kassaby Y.A., Rice A.V., Alamouti S.M., Sperling F.A. & Hamelin R.C. (2012) Population structure and migration pattern of a conifer pathogen, *Grosmannia clavigera*, as influenced by its symbiont, the mountain pine beetle. *Molecular Ecology* **21**, 71-86.

Usadel B., Poree F., Nagel A., Lohse M., Czedik-Eysenberg A. N. & Stitt M. (2009) A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize. *Plant, Cell & Environment* **32,** 1211-1229.

Van Aalten D. M. F., Komander D., Synstad B., Gåseidnes S., Peter M. G. & Eijsink V. G. H. (2001) Structural insights into the catalytic mechanism of a family 18 exo-chitinase. *Proceedings of the National Academy of Sciences* **98**, 8979-8984.

Van Loon L.C., Rep M. & Pieterse C.M.J. (2006) Significance of inducible defense-related proteins in infected plants. *Annual Review of Phytopathology* **44**,135–162

Veluthakkal R., Sundari B. & Dasgupta M. G. (2012) Tree chitinases–stress-and developmental-driven gene regulation. *Forest Pathology* **42**, 271-278.

Veronese P., Ruiz M. T., Coca M. A., Hernandez-Lopez A., Lee H., Ibeas J. I. & Narasimhan M. L. (2003) In defense against pathogens. Both plant sentinels and foot soldiers need to know the enemy. *Plant Physiology* **131,** 1580-1590.

Vos I.A., Verhage A., Schuurink R.C., Watt L.G., Pieterse C.M.J., Van Wees S.C.M. (2013) Onset of herbivore-induced resistance in systemic tissue primed for jasmonate-dependent defenses is activated by abscisic acid. *Frontiers in Plant Science* **4,** 539.

Wallin K.F. & Raffa K.F. (2002) Density-mediated responses of bark beetles to host allelochemicals: a link between individual behavior and population dynamics. *Ecological Entomology* **27,** 484–492.

Whelan S., & Goldman, N. (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution* **18,** 691-699.

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.

Wood D. L. (1982) The role of pheromones, kairomones, and allomones in the host selection and colonization behavior of bark beetles. *Annual Review of Entomology* **27,** 411-446.

Wu B., Zhang B., Dai Y., Zhang L., Shang-Guan K., Peng Y. & Zhu Z. (2012) Brittle Culm15 encodes a membrane-associated chitinase-like protein required for cellulose biosynthesis in rice. *Plant Physiology* **159,** 1440-1452.

Wu C. T., Leubner-Metzger G., Meins F. & Bradford K. J. (2001) Class I β-1, 3-glucanase and chitinase are expressed in the micropylar endosperm of tomato seeds prior to radicle emergence. *Plant Physiology* **126**,1299-1313.

Wu H.X., Ying C.C. & Muir J.A. (1996) Effect of geographic variation and jack pine introgression on disease and insect resistance in lodgepole pine. *Canadian Journal of Forest Research* **26**, 711-726.

Yang S., Zhang X., Yue J. X., Tian D. & Chen J. Q. (2008) Recent duplications dominate NBS-encoding gene expansion in two woody species. *Molecular Genetics and Genomics*, **280**, 187-198.

Yang Z. (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution* **39**,306–314

Zamani A., Sturrock R., Ekramoddoullah A. K. M., Wiseman S. B. & Griffith M. (2003) Endochitinase activity in the apoplastic fluid of Phellinus weirii-infected Douglas-fir and its association with over wintering and antifreeze activity. *Forest Pathology* **33**, 299-316.

Zeneli G., Krokene P., Christiansen E., Krekling T. & Gershenzon J. (2006) Methyl jasmonate treatment of mature Norway spruce (*Picea abies*) trees increases the accumulation of terpenoid resin components and protects against infection by *Ceratocystis polonica*, a bark beetle-associated fungus. *Tree Physiology* **26**, 977-988.

Zhang D., HrmovaM., Wan C. H., Wu C., Balzen J., Cai W. & Haigler C. H. (2004) Members of a new group of chitinase-like genes are expressed preferentially in cotton cells with secondary walls. *Plant molecular biology* **54,** 353-372.

Zhang X, Zhu Z, An F, Hao D, Li P, Song J, Yi C, Guo H (2014) Jasmonate- activated MYC2 represses ETHYLENE INSENSITIVE3 activity to an- tagonize ethylene-promoted apical hook formation in Arabidopsis. *Plant Cell* **26**,1105–1117.

Ziemann M., Kamboj A. & Bhave M. (2013) OrthoRBH: A streamlined pipeline for mining large gene family sequences in related species. *Bioinformatics* **9**, 267.

Zimin A., Stevens K. A., Crepeau M. W., Holtz-Morris A., Koriabine M., Marçais G. & Neale D. B. (2014) Sequencing and assembly of the 22-Gb loblolly pine genome. *Genetics* **196**, 875-890.

Zipfel C. (2014). Plant pattern-recognition receptors. *Trends in Immunology* **35,** 345-351.

Zulak K. G., Lippert D. N., Kuzyk M. A., Domanski D., Chou T., Borchers C. H. & Bohlmann J. (2009). Targeted proteomics using selected reaction monitoring reveals the induction of specific terpene synthases in a multi-level study of methyl jasmonate-treated Norway spruce (*Picea abies*). *The Plant Journal* **60,** 1015-1030.

Zulak K.G., & Bohlmann J. (2010) Terpenoid biosynthesis and specialized vascular cells of conifer defense. *Journal of Integrative Plant Biology* **52**,86–97.

6.0 Appendices

Appendix 6.1

Nucleotide sequences of previsously identified conifer chitinases including sequences from a lodgepole and jack pine from a sanger generated clone library designated as PCO for lodgepole and PBA for jack pine, as well as odgepole, and white spruce sequences from NCBI EST data base.

>PCO0111_M10
AATTCGGCCATTACGGCCGGGGAACTGAAAGTTGCAATTAATGCAGTGCTGATATAGAAACGAGTGTCTGCAATATGA
AATCGATGAAGTTCTGCGCGATGGCGATCGCCTTGCTTACAATGGCGACGATGAATATGTATTTTGTATCTGCTGAGC
AATGCGGACAGCAAGCAGGCGGAGCTCTTTGTCCCGGTGGCTTGTGCTGCAGCAAATGGGGATGGTGTGGCAACACGG
ACGCCCATTGCGGGCAGGATTGCCAGAGCCAATGCAGCGGATCGACTCCCACTCCCGGTGGACAGGGAGTTGCATCTA
TCATCACTGAAAGCATTTTCAATGAGTTNTNAAGCACAGAAACGACGCCGGNNGCAAGNCCNAGCGNNTTCTACACNT
ANTNTCCCTTCATTGCAGCTGCCAATGCTTTTCCTTCCTTCGGCACCACCGGCGATGTCGCTACTCGGAAAAGAGAGC
TCGCTGCTTTCTTTGGCCAAACCTCCCACGAAACCACAGGAGGATGGGCGACAGCCCCAGACGGCGCGTACGCGTGGG
GTTATTGTTTCAAAGAGGAGCAAGGCAATCCTCCCGCCGAGTACTGCCAGGCAACCTCCCAGTGGTCCTGTGCATCTG
GAAAGAGATACTACGGACGAGGGCCCGTTCAATTGTCATGGAATTACAACTATGGACCGGCCGGGAAGGCAATCGGAT
TCGATGGCATAAACAACCCCGACATTGTTGCTAGCGATGCCACGGTCTCTTTCAAGACCGCAATCTGGTTNTGGATGA
CCGCCCAATCTCCGAAACCTTCTTGCCACGATGTCATGACCGGGAAATGGACTCCGTCCGGCAGCGACAGCGCCGCTG
GGAGAGCTGCGGGATATGGAGCAGTTACCAACATCATCAACGGCGGGCTGGAGTGCGGGAAAGGCAGTGACTCGAGGC
AGCAGGATCGCATCGGCTTCTACAAAAGATACAGTGACATTCTTGGGGTGAGCTACGGATCAAACCTGGATTGCAACA
ACCAGAGGCCTTTCGGCGCTGCAGTTCAATCTGAACCTCGTCTTATCAAAACCGTGGTCTGAACACTTCTCATAAATC
CGAGATTTCGGATTAGCAGCAGTTCGATCTCCATCTGTGTTCCAAGTTGCCATAATAAGATATAATAAAATCCNNNNN
NNNNNNNNNAAAAAAAAA

>PCO024_O12
ATATGCCCGGGAATTCGGCCATTACGGCCGGGTGGAGAGTAG
TAGAGGTAACTCGGTAATGGTGCTGGTGGTGGTGCTAGTCCTGCTACTGG
CGGGTGTCAGTGTTAATGCTCAAAACTGTGGCTGTGCCAGCGGACTGTGT
TGCAGCAAGTTTGGATATTGCGGGACCACCTCTGCTTACTGCGGCGGATGGCTGCCAGAGCGGTCCTTGTACCAGTTC
AGGAGGAGGATCTCCCAAGTGGCGGAGGTGGAAGCGTGGGTACCATAATTTCCCAGAGTTTCTTCAATGGCCTCGCGG
GCGGAGCTGCCAGCTCCTGCGAGGGCAAGGGATTCTACACTTACGATGCTTTCATCGCAGCGGCCAATGCGTATTCTG
GATTTGGCACCACGGGATCCGCAGACGTCACAAAGAGAGAACTCGCGGCCTTCTTCGCTAATGTTATGCACGAAACTG
GAGGGATGTGCTACATCAATGAGAGAACCCCTCCAATGATNTACTGTATGAGTTCAGCCACGTGGCCGTGTGCGAGCG
GTAAGGACTACCATGGACGTGGCCCTCTCCAGTTGAGCTGGAATTACAATTATGGAGCGGCTGGGAAGAACATTGGGT
TCGACGGGGTGAACAACCCAGAGAAGGTGGGGCAAGACTCCACCATTTCGTTCAAGACGGCGGTCTGGTTCTGGATGG
ACAACAGCAACTGCCACACCGCCATAACGTCAGGGCAGGGCTTCGGTGGCACAATCAAAGCCATCAACAGCCAGGAAT
GCAACGGTGGAAACAGTGGGGAAGTGAATAGCAGAGTGAACTACTACAAGAATATCTGTAGCCAGTTGGGTGTGGATC
CTGGAGCCAACGTCTCGTGCTAAATGTCTCTTTTTCTTCATGGCTACTAAATATATATGAATAAGACATCAGTGATCA
TTGATGTCCNTNNCNNNNNNNTAAATAAAAATTGCTCTGCACATCNNNNNNNNNNNNNNNAAAAAAAAAGGCCGCCT
CGGCCCTCGAGAAGCTTTCTAGACC

>PBA019_O12
GGGGGCAAATAATATC
TGTGAGTGGAGAGTCCTAAGAAACTGTTGGCAATGGGGAGTAGTAGCGGT

AACTCGGCAATGGTGGTGGTGGTGCTAGTCCTGGCTACTGGTGGGCGTTAGTGTTACTGCTCAAAACTGTGGCTGTGC
CAGCGACCTGTGTTGCAGCATGTATGGATACTGCGGGACCACCTCTGCTTACTGCGGCGATGGCTGCCAGAGCGGTCC
TTGTACCAGTTCAGGAGGATCTNTAAGTGGCGGAGGTGGAAGCGTGGGTACCATAATTTCCCAGAGTTTCTTCGATGG
CCTCAAGGGCGGAGCTGCCAGCTCCTGCGAGGGCAAGGGATTCTACACTTACAATGCTTTCATCGCAGCGGCCAATGC
ATATTCTGGATTTGGCACCACCGGATCCGCCGACGTCACAAAGAGAGAACTCGCGGCTTTCTTCGCTAATGTTATGCA
CGAAACTGGAGGGATGTGCTACATCAATGAGAGAAACCCTCCCATGAACTACTGTATGGGTTCAGCCACGTCGCCGTG
TGCGAGCGGTAAGAGCTACTATGGACGTGGCCCTCTCCAGTTGAGCTGGAATTACAATTATGGAGCGGCTGGGCAGAG
CATTGGGTTCGACGGGGTGAACAACCCAGAGAAGGTGGGGCAAGACTCCACCATTTCGTTCAAGACGGCCGTTTGGTT
CTGGATGAAGAATAGCAACTGTCACTCCGCCATAACGTCAGGGCAGGGCTTCGGTGGCACAATCAAAGCCATCAACAG
CCAGGAATGCAACGGCGGAAACAGTGGAGAAGTGAATAGCAGAGTGAACTACTACAAGAATATCTGTAGCCAGTTGGG
TGTGGATCCTGGAGCCAATGTTTCGTGTTAAATCTGTATTTTTTTCCTGCTTGCTAAATATATGAATAAGACATCAAT
GATCGTCGATGTCCATGGCCACTTACTAAATAAAAATTGACTTGCACATCTTTTTCCAAAAAAAAAAAAAAAAAAAG
GCCGCCTCGGCCCTCGAGAAGNNNNTAGNNCATTCG

>JPPBA0221_H13
GGGGACTGTGTTTGAATGTTTCACTGGCGTCGAAGAAGATGAAGGTCTGCGACAAGGGATGGGAATGCAAGGGGACTT
ATTGCTGCAACCAGACTATATCGGAAATTTTCACAGTGGACAATTTCGAAGAACTATTCTCCAAGAGGAATACCCCGG
TGGCTCATGCGGTGGGTTTCTGGGATTACTATTCCTTCATCAACGCGGCTGCACAGTTCGAGGGCATAGGATTTGGTA
CTACTGGAGGACAGGTGATGCAGCAGAAGGAGCTTGCAGCCTTCCTTGGCAATGTTGCAGCTGAAACTTCATGTGGTT
ACAATGTTGCTACTGGTGGTCCAACTGCTTGGGGTTTGTGCTACAAAGAAGAAATGAGTCCGGACCAGCTTTACTGTG
ATCAAAATTTGCTATACCCTTGCGCTCCAGGAGCCAGTTACCATGGACGTGGTGCCCTGCCTATTTACTGGAACTTTA
ACTACGGACCGATTGGTGTAGCACTGAAACTTGATTTGCTTACTAGTCCAGACATGGTGTCAAACAATGCCACGATAG
GATTTCTAACTGCAATGTGGAGGTGGATGAATCCCATCAAGCCAAAACAGCCCTCTGCTCATGATGTGTTTGTGGGTA
ACTGGAAACCAACCAAAAATGACACTGAATCATATAGGCTCCCTGGTTTTGGAATGGTCATCAATGTTCTAAATGGGG
GTTTAGAGTGTGGGGAAAGGGGATATTGATGCAATGAACAACAGGATTTCACATTACCTGTATTTCCTTGATTTG

>JPcontig7707
TTTTTTTTTTTTTTTTAACGAAATGTACCAGCATGCACGAGGTGCTGCAGCGCAACTTTTATTCACCAGTTCGCCATA
GACTTACTCACATTTTCTACTCGCGATTATTTAGTTATATATTTGGCGGCCATGAAAATAGATATTTAGCAAGAGATG
TTGGTTCCCGGATCCACCCCCAACTGGCTACATAATTTCTTGTAGTAATTCACTCTGCTATTCACCTGCGCACTATGT
CCACCGTTGCACTCCCCGCCGTTGATGGCTTTGATTGTCGCACCAAAACCCTGCCCGGAGGTTATGGCAGAATGGCAG
TTGCTGTTCTTCATCCAGAACCACACAGCCGTCTTGAACGAAATGGTGGCGTCTTGCCCCACTTTCTCTGGGTTGTTC
AGGCCGTCGAAGCCAATGTCCTGCCCCGCCGCTCCATAATTGTAATTCCAGGTTAGTTGGATGGGGCCACGCCCATGG
TAGCTCTTCCCGCTCGCACACGGCCACGTGCTGGAACTCTGACAGAAGTTTGACAAAGGATTTATCTCGTTGATGTAG
CACAGTCCTCCAGTTTCGTGCATCACATTAGCGAAGAAGGCAGCGAGCTCTCTCTTCTGGTTGTCGCTGGCACCGGTG
GTGCCAAATCCAGAATACGCATTTGCAGCTGTGATGAAAGCGTTGTAAGTGTAGAAACCCTTGCCCTCGCACGAGCTA
CCAGCAGCATTGGCTAGGCCACTGAAGAAATTCTGAGAAATTATAGTTCCCACGTTTCCACCTCCACCACTTGGAGAT
CCTCCACCTGAGCCGTAACAGGGGCCGCTCTGGCAGCCATTGCCGCAATAAGCAGAGGTGGTGCCGCAGTATCCCCAC
TTACTGCAACACAGTCCGCTGGCGCAGCCACAGTTTTGAGCATAAACAACGTTAACGCTCACGACCACCACTAACAGC
AGCACCATACTCATCCTCAGCACCATATGCTCATAAGGGCTAACAATCACCGATCTACCAGTACGAGTCCCCATTACT
AATAATCGCTCAGGACTCCCACATTCACTCTGAATATCCCC

>LPContig6903
GGGGAGAGCAAAGAAGTACACTCTCTGTGCTAGTGACAGTAGTGATAGACATTAGCCAATATGAGAAATTACCTCCGT
GTTATGACACCGAGCGTTATTATTCTATTGCTCGTTCTTGCCTTCGATGTTGTTTCCATCTGCCGAGGAGATGTTGGG
GATTTTGCCACCCAAGATTTCTTCAATGGAATTTTGTCTGGTGCTTCTGATAGTTGCGCAGGAAAGACCTTCTATACA
TACAATAACTTCATGGACGCAGCCACTGCATTCTCTGGCTTTGGCACAACAGGCGCCGACGTTGACCACAAGAGAGAA
ATCGCTGCTTTCTTCGCCAATATTGCTCACGAGACTTCCAGACTGTGCTACGTTGAGCAAATAGAGAAGTCTGACTAC
TGCGATTCCACCAACAAGAAATACCAGTGTGTGGCGGGAAAGCAGTACTACGGGCGTGGCCCCTTGCAACTCACATGG
AATTATAATTATGATGCAGCTGGTGATTACCTGGGCTTTGATGGCCTGAATCACCCTGAAATTGTAGCCCAGAATGGC
TCAATCTCTTGGAAGACTGCTGTCTGGTTTTGGATGAAGCATAGTAATTGCCACTCTGCTATTACATCTGGACAAGGA
TTCAGGGCAACCATCAAAGCAATCAGTGGGGATGAATGCAATGGTGGAGACTCCGACGCAATTGATGAACGTGTCAAT
TATTATACCAACTATTGCAACGAGTTTGGTGTAGATCCAGGAAGTAATCTTAGCTGTTAAAAGTATATTAATATGCTG
TGACCTCTTGATCTCTGTCAAAGGATCCAGAATAGTTCAAAATACTATATCGATAGAGTCTGATTACCACGGTTAGCA
TTAGCAGGAAACGGATCTATGTGACATGTAAAACGAGACAACATTCAACATGTAAGCACATCCATGCCTTTTGCATGC
CAATGAACTGTGCAAATACAGCAAATTAAGATTAGATTATCTCAGATTTTAAATTTTTGCGGCCATTAATAAGGTAGT
TCATTTTTTCGTTCTTTTTCATCCTGAAAAATAACTATAGTTGCTGTGCTGGTTCGAAAAGCCCAATGGTAAATATGG

TGAATGGGATGAATCCAGAAATTTAATGTTTACCGCACAAGATTAGATTAATTAGGTGTACCTAATATTAGTGCATTT
GGTGCCACATTGATTTTATATTTCAGGGATGATTTATACTTTTTCTAGCATTAAATAAGATGCTGCTTTTTGGTTAAA
AAAAAAAAAAAAGAAAAAAAAAAAAAAAA

>JPContig8129
TTTTTTTTTTTTTTTTACCTTAATTTGCTGTATTTGCACAGTTTATTGGCATGCAAAAGGCATGGATGTGCTTACATG
TTGAAAGAAAGAGAATGTTGTCTTGTTTTACATGTCACATAGATCCGTTTCCTGCTAATGCTAACCGTGGTAATCAGA
CTCTATCGATATAGTATTTTGAACTATTCTGGATCCTCTGACAGAGATCAAGAGGTCACAGCATATTAATATACTTTT
AACAGCTAAGATTACTTCCTGGATCTACACCAAACTCGTTGCAATAGTTGGTATAATAATTGACACGTTCATCAACTG
AGTCGGAGTCTCCACCATTGCATTCATCCCCACTGATTGCTTTGATGGTTGCCCTGAATCCTTGTCCAGATGTAATAG
CAGAGTGGCAATTACTGTGCTTCATCCAAAACCAGACAGCAGCCTTCCACGAGATTGAGCCATTCTGGGCTACAATTT
CAGGGTGATTCAGGCCATCAAAGCCCAGGTAATCACCAGCTGCATCATAATTATAATTCCATGTGAGTTGCAAGGGGC
CACGCCCGTAGTACTGCTTTCCCGCCACACACTGGTATTTCTTGTTGGTGGAATCGCAGTAGTCAGACTTCTCTATTT
GCTCAACGTAGCACAGTCTGGAAGTCTCGCGAGCAATATTGGCGAAGAAAGCAGCGATTTCTCTCTTGTGGTCAACGT
CGGCGCCTGTTGTGCCAAAGCCAGAGAATGCAGTGGCTGCGTCCATGAAGTTATTGTATGTATAGAAGGTCTTTCCTG
CGCAACTATCAGAAGCACCAGACAAAATTCCATTGAAGAAATCTTGGGTGGCGAAATCCCCAACATCTCCTCGGCAGA
TGGAAACAACATCGAAGGCAAGAACGAGCAATAGAATAATAACGCTCGGTGTCATAACACGGAGGTAATTTCTCATAT
TGGCTAATGTCTATCACTACTGTCACTAGCACAGAGAATGTACTTCTTTGCTCTCCCCC

>LPContig2984
TTTTTTTTTTTTTTTTGAGCTCTAAAGAAAAACTTTTGTTCCATTTCCACTGTTTCTAGGGCGATGTTTACACGTCAA
TCTCACCTTTCATCTTCGGTGCTTAAGACTTGCTTTAATCAATTTCCATCTAATAACTAAAACACAATTTGTTTACAT
TTTTTATCGCAACAATATAGAATGAATGGCTGTAGCCGATATGGATTTCTAAGTAATTCATTATTCACAAAAATTCAG
AACAATTTGACATCCAGACTGACTTTGCATGAGCAGCTTCTATTCAAGCCTCTCTTAAATCATTAATAACTGAAGTCC
ATCTTCAATATGGATATAGAGTACTGATGCCATAAGTGATATTCACGATCTTCAGTCTTGCCTTTTTTCTCCAGCAGC
ATTGGATACAGGTGGAAGCGAAACATAATGTAGAATCTTCTGTACGACAAGAACATCAAGTTGAAACCTCTCTGCAAT
ACCTTTAGCATCCATTGACTTCTCTTGATTTTCTGCTTTGCCTTCATGAAGGAGCAACATTTGGTGAAGTTGGGCAAC
AGTTAATGTTCCAGGAGCCAAGATTTTCTCCCCATTTGGACCTGTATTAACACTAGTATTTCGAAGCTTTGGCAATGG
TCTATTATATTTTGGCACAACTGCAGCATCACCCATTTCTGCTGGCCCTCCGGGCTTAACCTTGACACGTCCAAACAT
CTCATTCAGCATTACATCATAACTTGGGTCTCTCTGTTCCAAAACATTAGAATCAGGAGCCTGATTAGGATCTGCCTT
GGGATCCCAATTCGGGTTATTCACTGGTATGGAATCTTGGGACTTGGATTGTGATGGAGGTCTGCTGAATAGCTCGGG
CTTGTGATCTTTAACAGAAGATGATTTTGTTGGTGCCGTGGGATTCGCACTTCTGAGCCACCCTGTTGCTCTACGTAG
AGTCTGCCCCATGTTTTCTTCAGTTAAATAAGATCAATCCAGAAGCCTTGAGCTCTTCATGTGTGTAAATACAGGCCT
TACACCTGCTACGTCTCGTCTCTCGACACATTTGGTTGTGACTGTTTGATTTGAAGCTCAAACAACTGTAAACTGTAA
AGTATTACGAAAATAGACACCCCCC

>LP-Contig3900
TTTTTTTTTTTTTTTTATTTTATTTTAATTTTTTTTCAAAAGAATGATACATGTTTATACAATGATATATGTTTATAC
AGATATTTCAGGCCTCGGCAATAGCTGGCTTGCACAAAATAACAAATAACATATTTTCATCATGTGATCCACAATAAT
TAAATTAAGTCGTTGCACAAATAGCGTCTTAGAAAGCGCTGCTGGTATATTTTACATAACCGGCAACAAACCTTATCA
CGATCACCTACAAATAGGAAGAGAGAGCCATTCTATTATTTCATCACTGAACGGAACCAAACTTCCTAACTGTGAAAT
AAAGAGCCCAAACAATACACATAATGGTCAAAGGCATGCTCCTTTATTTACCCTACTATCAATTGGATCAAGGATGGC
CTTGGATAGATAAAACTACAGACTGCCAGCTAACGAGAGGTGGTAGAAGTTGATACAGGGTTAAGAGGAACTTGCTGG
CCACAATCAAGATTATCACCAGCTTGTTCTCGTCCAACGCCCAGCAAATCAAGGAAATACAGGTAATGTGAAATCCTG
TTGTTCATTGCATCAATATCCCCTTTCCCACACTCTAAACCCCCATTTAGAACATTGATGACCATTCCAAAACCAGGG
AGCCTATATGATTCAGTGTCATTTTTGGTTGGTTTCCAGTTACCCACAAACACATCATGAGCAGAGGGCTGTTTTGGC
TTGATGGGATTCATCCACCTCCACATTGCAGTTAGAAATCCTATCGTGGCATTGTTTGACACCATGTCTGGACTAGTA
AGCAAATCAAGTTTCAGTGCTTCACCAATCGGTCCGTAGTTAAAGTTCCAGTAAATAGGCAGGGCACCACGTCCATGG
TAACTGGCTCCTGGAGCGCAAGGGTATAGCAAATTTTGATCACAGTAAAGCTGGTCCGGACTCATTTCTTCTTTGTAG
CACAAACCCCAAGCAGTTGGACCACCAGTAGCAACATTGTAACCACATGAAGTTTCAGCTGCAACATTGCCAAGGAAG
GCTGCAAGCTCCTTCTGCTGCATCACCTGTCCTCCAGTAGTACCAAATCCTATGCCCTCGAACTGTGCAGCCGCGTTG
ATGAAGGAATAGTAATCCCAGAAACCCACCGCATGAGCCACCGGGGTATTCCTCTTGGAGAAGAGTTCTTCGAAATTG
TCCACTGTGAAAATTTCCGATATAGTCTGGTTGCAGCAATAAGCCCCCTTGCATTCCCATCCCTTGTCGCATATCTTC
ATCTTCTTCGACGCCAGTGAAACATTCAAACACAGTGCCAATGCCAACAAAACAACCAAACAAATCCTTCCCATCGCC
ATGACAGGCAAGATCTCCGCCAGATCTAGGAAGATTTTGAAGCTTAATCCCAATATTCGGGCCCC

>1.GQ0183_A06.2 clone=GQ0183_A06 cluster=GQ0183_A06 status=complete
gb=BT102361 date=2009-11-25
GTCAGATCTGTCTAATAACATACATACAGATACTGCAGTTGAGGGAGATGCCCTTATCCAGCACAAATCTCAGCTTTG
GATTTCATCGAGAAGCTGAAAAAATCCTCACATCTAAATGCGGCTCAACTTTCATCAAGCTTCTTCATATTCAAGGGG
TTCCTCTCAGAGGGAATAAATGTAGTCGATGATATTATTCTTGGCGAAGCAAAGACAATTGCTTGTCTTTATTATTCT
CTCAGATATTTTGGAAAGATTCGAAACCCTTTATCCCAAAGGAGAGGAGTTGGTCCAAATATTAAAGGGGAAGCAATA
GAATTGATAGTACATGGGTCCATCATTGAAATTTAAAAATTTTCAAGGATATTCAAGGGATCAAAGATGTTTGATTGA
ATGAAAATAAGGATCCATTGGATATATCTGCATCAAAATTGGACTTAATGTATAATTATATCTGCATCAAAATTGGAT
TTTAGGTAGAAATTAAAAAGGAATACTTGAGTTAGTTATTGCTTATTGTTTAGCATGGGAGTTGAGATACTTGAGTTA
GTTATTGTGTAGCATGGGAGTTGAGATACTTGAGTTAGTTATTGTGTAGCATGGATTAGCAATGTATGAGTCCTTATC
GGGACTAAATGTAATATTGGGAACAGAGTTGGTAATGTTTATTTGGCTGAAGCTCATGCATGAGCCTGAGGCTTTGGA
TTCTCTTGTTGAGACTTTTGATTAANAAAAAAAAAAAAAAAAAAAAAATTAATGCAGTGCTAAGACGTTAACGAGTGT
CCGAGAGATGAAATCGATTCGGTTCTCCGCGATGGCGATCGCCTTGGTTACAATGGGGACAATGAATTTGTATTTTGC
GTCTGCAGAACAATGCGGACGGCAAGCAAGCGGAGCGCTCTGTCCCGGTGGCTTGTGCTGCAGCAAATGGGGATGGTG
CGGCAACACAGAGGCCCATTGCGGGCAGGATTGCCAAAGTCAATGCGGAGGATCGACTCCCACGCCTCCCTCACCCAC
TCCCGGTGGACAGGGAGTCTCATCTGTCATCACTGAAAGCATTTTCAACGATTTATTGAAGCACAGAAACGACGCCGC
TTGCAAGGCCAAAGGATTCTACTCCTACGCCGCCTTCATTGCGGCTGCCAATGCTTTCCCCGCTTTCGGCACCACAGG
CGATCTCTCTACTCGGAAAAGAGAGCTCGCTGCTTTCTTTGGCCAAACCTCCCACGAAACCACAGGCGGATGGCCGGC
GGCCCCAGACGGCCCGTACGCGTGGGGTTATTGTTTCAAAGAGGAGCAAGGCAATCCTCCCGGCGAGTACTGCCAGGC
ATCTTCCCAATGGCCCTGCGCATCTGGAAAGAGATACTACGGACGAGGGCCCGTTCAAATATCATGGAACTACAACTA
TGGACCGGCCGGGAGGGCAATTGGGTTCGATGGCATAAACAACCCGGACATTGTTGCGAACGATGCCACGGTCTCTTT
TAAGACGGCAATGTGGTTTTGGATGACCGCCCAGTCTCCAAAGCCTTCTTGCCACGACGTGATGACCGGGAGATGGAG
TCCGTCTGGCAGAGACAGCGCCCCCGGGAGAGCGGCGGGATATGGAGTGGTAACAAATATCTTCAACGGCGGGTTGGA
GTGCGGGAAAGGCAGCGACTCGAGGCAGGTGGACCGCATCGGGTTTTACAAAAGATATTGCGACATTTTCGGGGTGAG
ATACGGATCAAACCTGGAGTGCAACACCCAGAAGCCTTTTGGATTTGCGGGTCAATCTCAACCTCACCTTATCAAAAC
CGTGGTTTAAATACTGTTTATAAATCTGAGATTTGTGGGGGCGCTTGCGTTGTTAGGGTTAAAGCAGTTGGATCTCCA
TCTGTTTGCCCTAATAAAATAGAATAAAGTCCCTATAAATTAACTCTCTAAAAAA

>2.GQ03811_E13.2 clone=GQ03811_E13 cluster=GQ03811_E13 status=complete
gb=BT117024 date=2009-11-25
CAGAACACCGCACTTTTCTTTGTGTCTGTGCAAGTGATAAGAATAGTACTGATAGACAATAGTCGCCAATATGGCAAC
TCATTTCCGTGTGAACGTTATTTTTCTATGGCTTGCTTTTGCCCTTTCTGCTCTTTCCATCTGCCGTGGAGCAGTTTC
GGATATTGCTACCCAAGATTTCTTCAATGGAATTTTATCTGCTGCTACTGATGGCTGCGCAGGAAAGACCTTCTATAC
ATACAGTGATTTCATTACTGCGGCCAATGGATTCTCCGGCTTTGGCACAACTGGGACCTCCGATGACAACAAGAGAGA
AATTGCTGCTTTCTTCGCCAACGTTGCTCACGAGACTACCAATCTGTGCTACGTGGAGGAGATCGCCAAGAGCGACTA
CTGCTCCAGCAATACACAATATCCATGCGCATCCGGACAGCAATACTACGGCCGCGGCCCCTTGCAACTCACAGGAAA
TGCAAACTATGGTGCAGCTGGAACTTATCTGAGCGCGGATTTGCTGAATAACCCTGGATTGGTGGCTCAGGACGACCT
AACCTCGTGGAzGGCTGCGCTCTGGTTTTGGAATGTGAACAGCAATTGCCACACTGCTATTACGTCCGGTCAGGGATT
TGGGGCAACCATTCAAGCAATCAACGGAGCGATAGAATGCAACGGTGGAAACACCGACCAAGTTAATGACCGCATCAG
TCGCTACACCAACTATTGCAGTCAGTTTGGTGTAGACCCAGGAAGCAACCTTAGCTGTTAACTTAACGGTTAACAATA
ATTAGCCAGTCCCTTTTGATCGTAGCTATCACATCACGTCTCAGTAATAACGAATATTAAATATTTTCTCTCACTCAA
TAATAGTGATAGAGGCTAATAATTACCGTTACTGGTATTTAGCAGAACAGAGTAGGGTTCGACTCATGTCGTATGTAC
GGTGACTCAAAATCCTGCGACTCTAAATAAAAGGATAAATTTCGCATATTCCCACCGCCAAAAAA

>3.GQ03206_D15.1 clone=GQ03206_D15 cluster=GQ03206_D15 status=complete
gb=BT109235 date=2009-03-31
AATCAACAAGTCAAATAATCAGAGCCAGTGGGTAGTCTTAAGGAATTATTGGCAATGGGGAGTAGTAGTAGTAATAAA
TCGGTAATGGCGCTGGTGCTAGTGCTGTTACTGGTGGGCGTCAGTGTTAATGCTCAAAACTGTGGCTGTGCTAGCGGA
GTGTGTTGCAGTCAGTATGGATACTGCGGGACAACCTCTGCTTACTGCGGGAAGGGCTGTAAGAGCGGTCCCTGTTAC
AGTTCAGGAGGAGGATCTCCGAGTGCCGGGGGAGGAAGCGTGGGGGGCATAATTTCCCAGAGTTTCTTCAATGGCCTT
GCGGGTGGAGCTGGCAGCTCCTGCGAGGGCAAGGGATTCTACACTTATAACGCTTTCATCGCAGCGGCCAATGCGTTT
TCTGGATTTGGCACCACCGGATCCAACGACGTCAAAAAGAGAGAACTGGCCGCCTTCTTCGCTAATGTAATGCACGAA
ACTGGAGGGCTGTGTTACATCAACGAGATAAACCCTAAAATAATCTACTGTCAAAGTTCAAGCACGTGGCCGTGTACG
AGCGGTAAGAGCTACCATGGACGTGGGCCTCTCCAGTTGAGCTGGAATTACAATTATGGAGCGGCTGGGAAGAGCATT
GGGTTCGACGGGCTGAACAACCCAGAGAAGGTGGGGCAAGACTCCACCATCTCGTTCAAGACGGCCGTGTGGTTCTGG
ATGAAGAATAGCAACTGCCACTCCGCCATAACCTCAGGGCAGGGCTTTGGTGGCACAATCAAAGCCATCAACAACATG
GAATGCAACGGCGGAAACAGTGGCGAAGTGAGCAGCAGAGTGAACTACTACAAGAAAATATGTAGCCAGTTGGGTGTC

GACCCTGGAGTCAATGTCTCCTGCTAAATTTGTCTTCTGTTCATGGCTGCTAAGTATATCAATAAGATGTCAATGATC
ATTGATGTCCATGGCTTCTAGTAAATAAAAATTGCACTGCACCTCTCGTGGCGTGCATCTAGATTCTCTCAAAAAA

>4.GQ03803_J20.1 clone=GQ03803_J20 cluster=GQ03803_J20 status=complete
gb=BT116631 date=2009-03-31
ACCAACACATCAAATATTCAGAGTGAATGTGAGAGTGCTGAGCAATTATTAGCAATGGCGGGTAGTAGTGGTAAATTT
GCTAGTCCTCGTGGCCGTGTGGTGGTGAGGATGAGTCTGGTGCTACTGCTGGTGGTGGGCGTGAGCGTTAACGTTGTT
AACGCCCAAAACTGTGGCTGTGCCAGCGGACTGTGTTGTAGTAAGTGGGGATACTGCGGCACCACCTCTGCTTATTGC
GGCAATGGCTGCCAGAGCGGACCCTGTTCTGGCGGAGGATCTCCAAGTGGCGGAGGTGGAAACGTGGGGACTATAATT
TCTCAGAATTTCTTCAATAGCCTTGCGAGCGGAGCTGGTAGCTCCTGCAAGGGCAAGGGATTCTACACATACAACGCT
TTCATCGCAGCTGCCAATGCGTATTCTGGATTTGGCACCACCGGTTCCAACGAGGTACAGAAGAGAGAGCTTGCTGCC
TTCTTCGCAAATGTAATGCACGAAACTGGAGGACTGTGCTACATCAACGAGATAAGTCCTTCATCAAACTACTGTCAG
AGTTCAAGCACTTGGCCATGTACGAGCGGGAAGAGCTACCATGGGCGTGGGCCCCTCCAAATAAGTTGGAATTACAAT
TATGGAGCGGCGGGGCAGAGCATTGGGTTCGACGGGCTGAATAACCCAGAGAAGGTTGGGCAAGACGCCACCATCTCG
TTCAAGACAGCGGTGTGGTTCTGGATGAAGAACAGCAACTGCCATTCTGCCATAACCTCCGGGCAGGGCTTCGGTGGG
ACAATCAAAGCCATTAACAGCGGGGAATGCAACGGTGGAAACAGTGGCCAAGTGAATAGCAGAGTTACTTACTACAAG
AAATTCTGTAGCCAGTTGGGGGTGGATACCGGAACCAACGTCTCCTGCTAAATATTTCTCTTCATGGCTGCCAAATAT
ATAAATAAATAAATGCAAGAAGAAAATGTGACTGTCTATATAACAGTCTGTTATATAACTAGCACAGTCGCTAACTCG
TGGCTGAGAACGCCGAAAGATAGTGGCTGATAACGCCTGTCCTCAATACATTGAGAATGTGTGTCTCTAATATAAAAC
GTTTGTGGTAATCTGTTTGTGGTAATAAAAAA

>5.GQ0183_A06.2 clone=GQ0183_A06 cluster=GQ0183_A06 status=complete
gb=BT102361 date=2009-11-25
GTCAGATCTGTCTAATAACATACATACAGATACTGCAGTTGAGGGAGATGCCCTTATCCAGCACAAATCTCAGCTTTT
GGATTTCATCGAGAAGCTGAAAAAATCCTCACATCTAAATGCGGCTCAACTTTCATCAAGCTTCTTCATATTCAAGGG
GTTCCTCTCAGAGGGAATAAATGTAGTCGATGATATTATTCTTGGCGAAGCAAAGACAATTGCTTGTCTTTATTATTC
TCTCAGATATTTTGGAAAGATTCGAAACCCTTTATCCCAAAGGAGAGGAGTTGGTCCAAATATTAAAGGGGAAGCAAT
AGAATTGATAGTACATGGGTCCATCATTGAAATTTAAAAATTTTCAAGGATATTCAAGGGATCAAAGATGTTTGATTG
AATGAAAATAAGGATCCATTGGATATATCTGCATCAAAATTGGACTTAATGTATAATTATATCTGCATCAAAATTGGA
TTTTAGGTAGAAATTAAAAAGGAATACTTGAGTTAGTTATTGCTTATTGTTTAGCATGGGAGTTGAGATACTTGAGTT
AGTTATTGTGTAGCATGGGAGTTGAGATACTTGAGTTAGTTATTGTGTAGCATGGATTAGCAATGTATGAGTCCTTAT
CGGGACTAAATGTAATATTGGGAACAGAGTTGGTAATGTTTATTTGGCTGAAGCTCATGCATGAGCCTGAGGCTTTGG
ATTCTCTTGTTGAGACTTTTGATTAANAAAAAAAAAAAAAAAAAAAAATTAATGCAGTGCTAAGACGTTAACGAGTG
TCCGAGAGATGAAATCGATTCGGTTCTCCGCGATGGCGATCGCCTTGGTTACAATGGGGACAATGAATTTGTATTTTG
CGTCTGCAGAACAATGCGGACGGCAAGCAAGCGGAGCGCTCTGTCCCGGTGGCTTGTGCTGCAGCAAATGGGGATGGT
GCGGCAACACAGAGGCCCATTGCGGGCAGGATTGCCAAAGTCAATGCGGAGGATCGACTCCCACGCCTCCCTCACCCA
CTCCCGGTGGACAGGGAGTCTCATCTGTCATCACTGAAAGCATTTTCAACGATTTATTGAAGCACAGAAACGACGCCG
CTTGCAAGGCCAAAGGATTCTACTCCTACGCCGCCTTCATTGCGGCTGCCAATGCTTTCCCCGCTTTCGGCACCACAG
GCGATCTCTCTACTCGGAAAAGAGAGCTCGCTGCTTTCTTTGGCCAAACCTCCCACGAAACCACAGGCGGATGGCCGG
CGGCCCCAGACGGCCCGTACGCGTGGGGTTATTGTTTCAAAGAGGAGCAAGGCAATCCTCCCGGCGAGTACTGCCAGG
CATCTTCCCAATGGCCCTGCGCATCTGGAAAGAGATACTACGGACGAGGGCCCGTTCAAATATCATGGAACTACAACT
ATGGACCGGCCGGGAGGGCAATTGGGTTCGATGGCATAAACAACCCGGACATTGTTGCGAACGATGCCACGGTCTCTT
TTAAGACGGCAATGTGGTTTTGGATGACCGCCCAGTCTCCAAAGCCTTCTTGCCACGACGTGATGACCGGGAGATGGA
GTCCGTCTGGCAGAGACAGCGCCCCCGGGAGAGCGGCGGGATATGGAGTGGTAACAAATATCTTCAACGGCGGGTTGG
AGTGCGGGAAAGGCAGCGACTCGAGGCAGGTGGACCGCATCGGGTTTTACAAAAGATATTGCGACATTTTCGGGGTGA
GATACGGATCAAACCTGGAGTGCAACACCCAGAAGCCTTTTGGATTTGCGGGTCAATCTCAACCTCACCTTATCAAAA
CCGTGGTTTAAATACTGTTTATAAATCTGAGATTTGTGGGGGCGCTTGCGTTGTTAGGGTTAAAGCAGTTGGATCTCC
ATCTGTTTGCCCTAATAAAATAGAATAAAGTCCCTATAAATTAACTCTCTAAAAAA

>6.GQ01310_K18.3 clone=GQ01310_K18 cluster=GQ01310_K18 status=complete
gb=BT101680 date=2009-11-25
CAACTATAAAGCTGCCGGGGATGCCTTAGGATACGATCTCATAAACAACCCGGAGCTTGTGGTAACCGATGCCACGGT
CTCTTTTAAGACGGCGGTGTGGTTCTGGATGACCCCGCAGTCTCCCAAGCCTTCCTGCCACGACGTCATATTGGGACG
ATGGAGTCCATCCGACACCGACACTGCCGCCGGCAGAGTTCCGGGATATGGAATGGTGACCAACATCATAAATGGCGG
GGTGGAGTGTGGTCAAGGCACATCAAGCGCCACGCAGCAGGGCCGCATCGGCTTCTACCAGACATTCTGCAACAAGCT
GGGCGTCGACTCTGGATCCAACCTCGACTGCAACAACCAGAAGCATTTCGGAAACTAATCTCAAAAATACCAACATTT
GTCGTGCTATCGATGGTATGTCTAATAAATAAATAATCTGAAAATTGTCATAACTGATGGGTCCGTTTGCCGGAGGGT

TCCCATCCCTGACTTGCGCAATAATTCAATAAATTTCAAAATCTGAAGAATAAATCTTCTCACTATTTCTGTAACTGG
TATAATTAGCATGATGATTTTCTCGATATTTCAAATATTATAGATGTTTTCGAAATGATGGAGTTTTGTTATTTTAAT
CCAATTGATGATTTAAGGTGAATTTAACAAGGCAGCGAGGCTGCGCGTACTGCCTCCTCAGTCTGCCATGAAAATATT
GCCGAATTCCATGGATACAAGCTTCTTCAAGTCCTCAATTGCTGAGCTACTGAATTGAATTGTTCAGTATGATCCTTC
AATATTTAAAGAGCAGAAACTCCTCTTCCTCTGTACTGAGATAAATTCGTAGATACGTATGCATACAGTGTGAGGAGT
CGTGTGGTCAATATCCGATTAGTTGGCGCCCCGAGGGTGCCTTTCTGTCTCCGACTATTCTTGAGACCCAGAGCGAAT
ATTGTGTTAGGTGGTGTTGAGATTCGAACCCACCCCC

>7.GQ03811_E13.2 clone=GQ03811_E13 cluster=GQ03811_E13 status=complete
gb=BT117024 date=2009-11-25
CAGAACACCGCACTTTTCTTTGTGTCTGTGCAAGTGATAAGAATAGTACTGATAGACAATAGTCGCCAATATGGCAAC
TCATTTCCGTGTGAACGTTATTTTTCTATGGCTTGCTTTTGCCCTTTCTGCTCTTTCCATCTGCCGTGGAGCAGTTTC
GGATATTGCTACCCAAGATTTCTTCAATGGAATTTTATCTGCTGCTACTGATGGCTGCGCAGGAAAGACCTTCTATAC
ATACAGTGATTTCATTACTGCGGCCAATGGATTCTCCGGCTTTGGCACAACTGGGACCTCCGATGACAACAAGAGAGA
AATTGCTGCTTTCTTCGCCAACGTTGCTCACGAGACTACCAATCTGTGCTACGTGGAGGAGATCGCCAAGAGCGACTA
CTGCTCCAGCAATACACAATATCCATGCGCATCCGGACAGCAATACTACGGCCGCGGCCCCTTGCAACTCACAGGAAA
TGCAAACTATGGTGCAGCTGGAACTTATCTGAGCGCGGATTTGCTGAATAACCCTGGATTGGTGGCTCAGGACGACCT
AACCTCGTGGAAGGCTGCGCTCTGGTTTTGGAATGTGAACAGCAATTGCCACACTGCTATTACGTCCGGTCAGGGATT
TGGGGCAACCATTCAAGCAATCAACGGAGCGATAGAATGCAACGGTGGAAACACCGACCAAGTTAATGACCGCATCAG
TCGCTACACCAACTATTGCAGTCAGTTTGGTGTAGACCCAGGAAGCAACCTTAGCTGTTAACTTAACGGTTAACAATA
ATTAGCCAGTCCCTTTTGATCGTAGCTATCACATCACGTCTCAGTAATAACGAATATTAAATATTTTCTCTCACTCAA
TAATAGTGATAGAGGCTAATAATTACCGTTACTGGTATTTAGCAGAACAGAGTAGGGTTCGACTCATGTCGTATGTAC
GGTGACTCAAAATCCTGCGACTCTAAATAAAAGGATAAATTTCGCATATTCCCACCGCCAAAAAA

>8.GQ0046_J03.1 clone=GQ0046_J03 cluster=GQ0046_J03 status=complete
gb=BT100827 date=2009-11-25 (A13clone??)
CCCTATCTCAATTCCCATTGTATTTCAGGTGAGAGATTGCAATTCGCAGAGGAAGATGGAAACGAAGCAGATTATAGG
GTCGGTCAACAGAATGCTCGTTGTGGTTGGTATTTGTTTAATGATCTCGTCATGGCTTTGCTGTATCGAACCTGCAGC
AGCGGCAGACGAAGATCAGGCGATGAAGTCGAAGAAGATAGCATGTATAAAGGGAGCGGAGTGCAAAAATAAGACCAT
TTCGGAGCTATTCACAGTCGACCAGTTCGAATCTCTCTTCTCTCACCGCAATGCTCCCATGGCGCATGCTCAGGGATT
CTGGGACTACCACTCTTTCATCACCGCGGCGGCCCACTTCGAACCCAAAGGATTTGGTGCCACCGGCGGAGATCTGGT
CCAGAAAAAGGAGCTCGCTGCTTTCTTCGCCCACGTTGCCACTGAAACTTCATGTGAATCGTTGATGGCCCAGAGTTC
CACCGCAACTACAGATTCCCCTACCAAGTGGGGATTATGCTACAAAGAGGAACTCAGCCCAGACAGCACTTACTGTGA
ATCAAGCCTAGTCTACCCGTGTGCTCCTGGAGTCAGCTACCATGGACGGGGAGCTCTGCCAGTATACTGGAACTACAA
TTATGGACAATTAGGGCAGGCTCTGAAGGTGGATTTGTTGCATCATGCCGAGGATCTGTCGCAGAATGCAACATTGGC
GTTCCAAGCGGCCATGTGGAGATGGATGAATCCAATTAAAGTGAAGCAGCCTTCGGCTCATCAAGTGATGGTGGGGAA
GTGGGTTCCCACGAAGAACGACACGAACTCGTTGAGACATCCTGGCTTCGGCATGACCATCAACATTCTCAAGGGAGA
TGCCGAATGTGGAGCCGGCTCCGATGATAAGCAGATGAACAAGAGAATCGCGCATTACCTCTACTTTCTTGACCAGTT
GGACGTCGGTCGAGACAATGCTGGCGACAATCTCGACTGCTCCGACCAAAAAGTTCTAAATCCTTCTTCCGCGTCCAC
TTAGCTAGGGTTCCATTTCTTCAAATAAATGAAGATAAATTTATGGTATTTTTTCTCAGATTTCTCCTGGTATTTCCT
ATATTTCTCCTGGATCGTAATCTGTCTACAATGAAAGCTTTATTTCCTGGTTGGTGCCTGTTTTCTGATCAGAAATCA
GATATTGCACAACAATGTTGATATACAGACGGTATTATTGAGTTTATAATGAAAGCTTTATTTCCTGGTTGGTGCCTG
TTTTCTGATCAGAAATCAGATATTGCACAACAATGTTGATATACAGACGGTATTATGAGTTT

>gi|333601359|gb|HM219843.1| Picea engelmannii x Picea glauca class I
chitinase (Chia1-1) mRNA, complete cds
ATGAAGTCGATTAGGTTCTCCGCGATGGCGATCGCCTTGGTTACAATGGGGACAATGAATTTGTATTTTG
CGTCTGCAGAACAATGCGGACGGCAAGCAAGCGGAGCGCTCTGTCCCGGTGGCTTGTGCTGCAGCAAATG
GGGATGGTGCGGCAACACAGAGGCCCATTGCGGGCAGGATTGCCAAAGTCAATGCGGAGGATCGACTCCC
ACGCCTCCCTCACCCACTCCCGGTGGACAGGGAGTCTCATCTGTCATCACTGAAAGCATTTTCAACGATT
TATTGAAGCACAGAAACGACGCCGCTTGCAAGGCCAAAGGATTCTACTCCTACGCCGCCTTCATTGCGGC
TGCCAATGCTTTCCCCGCTTTCGGCACCACAGGCGATCTCTCTACTCGGAAAAGAGAGCTCGCTGCTTTC
TTTGGCCAAACCTCCCACGAAACCACAGGCGGATGGCCGGCGGCCCCAGACGGCCCGTACGCGTGGGGTT
ATTGTTTCAAAGAGGAGCAAGGCAATCCTCCCGGCGAGTACTGCCAGGCATCTTCCCAATGGCCCTGCGC
ATCTGGAAAGAGATACTACGGACGAGGGCCCGTTCAAATATCATGGAACTACAACTATGGACCGGCCGGG
AGGGCAATTGGGTTCGACGGCATAAACAACCCGGACATTGTTGCGAACGATGCCACGGTCTCTTTTAAGA

CGGCAGTGTGGTTCTGGATGACCGCCCAGTCTCCAAAGCCTTCCTGCCACGACGTGATGACCGGGAGATG
GAGTCCGTCCGGCAGCGACAGCGCCGCCGGGAGAGCGGCGGGATATGGAGTGGTAACAAATATCATCAAC
GGCGGGTTGGAGTGCGGGAAAGGCAGCGACTCGAGGCAGGTGGACCGCATCGGGTTCTACAAAAGATACT
GCGACATTCTCGGCGTGAGCTACGGATCAAACCTGGACTGCAACAGCCAGAAGCCTTTCGGATTTGCGGC
TCAATCTCATCCTCGCCTTATCAAAACCGTGGTTTAA

>gi|333601363|gb|HM219845.1| Picea engelmannii x Picea glauca class II chitinase (Chia2-1) mRNA, complete cds
ATGGAAACTAGTGTCCGACTAATGAAATCGATGAGGTTCTCCGTGATGGTGATGGCCTTGGTTACAATGG
CGACGATGAATTTCTATTTTGTGTCTGCAGAGCAATGCGGACGGCAAGCAAGCGGAGTGCTCTGTACCGG
CGGCGTACAGGGAGTCTCATCTGTCATCACTGAAAGCATTTTCAATAATTTATTGAAGCACAGAAACGAC
GCCGGTTGCAAGGCCAAAGGATTCTACTCCTACGACGCCTTCATTGCGGCTGCCAATGCTTTCCCTGGTT
TCGGCACCACCGGCGATCTCACTAGTCAGAAAAGAGAGCTCGCTGCTTTCTTTGGCCAAACCTCCCACGA
AACCACAGGCGGATGGCCGACGGCCCCAGACGGCCCGTACGCGTGGGGTTATTGTTTCAAAGAGGAGCAA
GGCAATCCTCCCGGCGAGTACTGCCAGGCGTCCTCCCAATGGCCCTGTGCATCTGGGAAGAGATACTACG
GACGAGGGCCCGTTCAAATATCATGGAACTACAACTATGGACCGGCCGGGAAGGCAGTTGGATTCGACGG
CATAAACAACCCGGACATTGTTGCTAACGATGCCACGGTCTCTTTTAAGACGGCGGTGTGGTTCTGGATG
ACCGAACAGTCTCCAAAGCCTTCCTGCCACAACGTGATGGCCGGGGGATGGGGTCCGTCCGGCAGCGACA
CCGCCGCCGGGAGAGCGGCGGGATATGGAGTGGTAACCAACATCATCAACGGCGGGTTGGAGTGCGGGAA
AGGCAGCGACTCAAGGCAGGAGGACCGCATCGGGTTCTACAAAAGATACTGCGACATTCTTGGCGTGAGC
TACGGATCAAACCTGGACTGCAACAGCCAGAAGCCTTTCGGCTTTGCGGCTCAATCTCAACCTCGCCTTA
TCAAAACCGTGGTTTAA

>gi|333601365|gb|HM219846.1| Picea engelmannii x Picea glauca class IV chitinase (Chia4-1) mRNA, complete cds
ATGGCGCTGGTGCTAGTGCTTTTACTGGTGGGCGTCAGTGTTAATGCTCAAAACTGTGGCTGTGCTAGCG
GAGTGTGTTGCAGTCAGTATGGATACTGCGGGACAACCTCTGCTTACTGCGGGAAGGGCTGTAAGAGCGG
TCCCTGTTACAGTTCAGGAGGAGGATCTCCGAGTGCCGGGGGAGGAAGCGTGGGGGGCATAATTTCCCAG
AGTTTCTTCAATGGCCTTGCGGGTGGAGCTGGCAGCTCCTGCGAGGGCAAGGGATTCTACACTTACAACG
CTTTCATCGCAGCGGCCAATGCGTTTTCTGGATTTGGCACCACCGGATCCAACGACGTCAAAAAGAGAGA
ACTGGCTGCCTTCTTCGCTAATGTAATGCACGAAACTGGAGGGTTGTGCTACATCAACGAGAAGAACCCT
CCAATGAAGTACTGTCAGAGTTCAAGCACGTGGCCGTGTACGAGCGGTAAGAGCTACCATGGACGTGGGC
CTCTCCAGTTGAGCTGGAATTACAATTATGGAGCGGCTGGGAAGAGCATTGGGTTCGACGGGCTGAACAA
CCCAGAGAAGGTGGGGCAAGACTCCACCATCTCGTTCAAGACGGCCGTGTGGTTCTGGATGAAGAATAGC
AACTGCCACTCCGCCATAACCTCAGGGCAGGGCTTTGGTGGCACAATCAAAGCCATCAACAGCATGGAAT
GCAACGGCGGAAACAGTGGCGAAGTGAGCAGCAGAGTGAACTACTACAAGAAAATATGTAGCCAGTTGGG
TGTCGACCCTGGAGCCAATGTCTCCTGCTAA

>gi|333601367|gb|HM219847.1| Picea engelmannii x Picea glauca class VII chitinase (Chia7-1) mRNA, complete cds
ATGGCAACTCATTTCCGTGTCAACGTTATTTTTCTATGGCTTGCTTTTGCCCTTTCTGCTCTTTCCATCT
GCCGTGGAGCTGTTTCGGATATTGCTACCCAAGATTTCTTCAATGGAATTTTATCTGCTGCTACTGATGG
CTGCGCAGGAAAGACCTTCTATACATACACTGACTTCATTAATGCGGCCAATTCATTTTCTAGCTTTGGC
ACAACTGGGACCTCCGATGACAACAAGAGAGAAATTGCTGCTTTCTTCGCCAACGTTGCTCACGAGACTA
CCAATCTGTGCTACGTGGAGGAGATAGACAAGAGCGACTACTGTGATTCCAGCAATACACAATATCCATG
CGCATCCGGACAGCAATACTACGGCCGTGGCCCCTTGCAACTCACAGGAAATGCAAACTATGGTGCAGCT
GGAACTTATCTGAGCGCGGATTTGCTGAATAACCCTGGATTGGTGGCTCAGGACGACCTCACCTCGTGGA
AGACTGCGCTCTGGTTTTGGAATGTGAACAGCAATTGCCACACCGCTATTACTTCCGGTCAGGGATTCGG
GGCAACCATTCAGGCAATCAACGGAGCGATAGAATGCAACGGTGGAAACCCCGACGAAGTTAATGACCGC
ATCAGTCACTATACCAACTATTGCAGTCAGTTTGGTGTAGACCCAGGAAGCAACCTTAGCTGTTAA

>gi|333601369|gb|HM219848.1| Picea engelmannii x Picea glauca class VII chitinase (Chia7-2) mRNA, complete cds
ATGGCAACTCATTTCCGTGTGAACGTTATTTTTCTATGGCTTGCTTTTGCCCTTTCTGCTCTTTCCATCT
GCCGTGGAGCTGTTTCGGATATTGCTACCCAAGATTTCTTCAATGGAATTTTATCTGCTGCTACTGATGG
CTGCGCAGGAAAGACCTTCTATACATACACTGACTTCATTAATGCGGCCATTTCATTTTCTAGCTTTGGC
ACAACTGGGACCTCCGATGACAACAAGAGAGAAATTGCTGCTTTCTTCGCCAACGTTGCTCACGAGACTA

CCAATCTCTGCTACGTGGAGGAGATCGCCAAGAGCGACTACTGCTCCAGCAATACACAATATCCATGCGC
ATCCGGACAGCAATACTACGGCCGTGGACCCTTGCAACTCACAGGAAATGGAAACTATGGTGCAGCTGGG
GATTACCTGGGCGTCGATCTGCTGAATAACCCAGGATTGGTGGCTCAGGACGACCTAACCTCGTGGAAGA
CTGCGCTCTGGTTTTGGAATGTGAACAGCAATTGCCACACCGCTATTACGTCCGGTCAGGGATTCGGGGC
AACCATTCAAGCAATCAACGGAGCGATAGAATGCAACGGTGGAAACACCGACCAAGTTAATGACCGCATC
AGTCGCTATACCAACTATTGCAGTCAGTTTGGTGTAGACCCAGGAAGCAACCTTAGCTGTTAA

>gi|333601371|gb|HM219849.1| Pinus contorta class I chitinase (Chia1-1) mRNA, complete cds
ATGAAATCGATGAAGTTCTCCGCGATGGCGATCGCCTTGCTTACAATGGCGACGATGAATTTGTATTTTG
TATCTGCTGAGCAATGCGGACAGCAAGCAGGCGGAGCTCTTTGTCCCGGTGGCTTGTGCTGCAGCAAATG
GGGATGGTGTGGCAACACGGACGCCCATTGCGGGCAGGATTGCCAGAGCCAATGCGGAGGATCGACTCCC
ACGCCTCCTTCACCCACTCCCGGTGGACAGGGAGTTGCATCTATCATCACTGAAAGCATTTTCAATGAGT
TATTGAAGCACAGAAACGACGCCGGTTGCAAGGCCAGCGGATTCTACACCTACTCTGCCTTCATTGCAGC
TGCCAATGCTTTTCCTTCCTTCGGCACCACCGGCGATGTCGCTACTCGGAAAAGAGAGCTCGCTGCTTTC
TTTGGCCAAACCTCCCACGAAACCACAGGAGGATGGGCGACAGCCCCAGACGGCGCGTACGCGTGGGGTT
ATTGTTTCAAAGAGGAGCAAGGCAATCCTCCCGCCGAGTACTGCCAGGCAACCTCCCAGTGGCCCTGTGC
ATCTGGAAAGAGATACTACGGACGAGGGCCCGTTCAATTGTCATGGAATTACAACTATGGACCGGCCGGG
AAGGCAATCGGATTCGATGGCATAAACAACCCCGACATTGTTGCTAGCGATGCCACCGTCTCTTTCAAGA
CCGCAATCTGGTTCTGGATGACCGCCCAATCTCCGAAACCTTCTTGCCACGATGTCATGACCGGGAAATG
GACTCCGTCCGGCAGCGACAGCGCCGCTGGGAGAGCTGCGGGATATGGAGCAGTTACCAACATCATCAAC
GGCGGGCTGGAGTGCGGGAAAGGCAGTGACTCGAGGCAGCAGGATCGCATCGGCTTCTACAAAAGATACA
GTGACATTCTTGGGGTGAGCTATGGATCAAACCTGGATTGCAACAACCAGAGGCCTTTCGGCGCTGCAGT
TCAATCTGAAGCTCGTCTTATCAAAACCGTGGTTTGA

>gi|333601373|gb|HM219850.1| Pinus contorta class I chitinase (Chia4-1) mRNA, complete cds
ATGGGGAGTAGTAGCGGTAACTGGGTAATGGCGGTGCTAGTCCTGCTACTGGTGAGCGTCAGTGTTAATG
CTCAAAACTGTGGCTGTGCCAGCGGACTGTGTTGCAGCAAGTATGGATACTGTGGAACCACCTCTGCTTA
CTGCGGCGCTGGCTGCAAGAGCGGTCCATGTTCCAGTTCAGGGGGAGGATCTCCAAGTGGCGGAGGTGGA
AGCGTGGGAACCATAATTTCCCAGAGTTTCTTCAATGGCCTCGCGGGCGGAGCTGCCAGCTCCTGCGAGG
GCAAGGGATTCTACACTTACAATGCTTTCATCGCAGCGGCCAATGCGTATTCTGGATTTGGCACCACCGG
ATCTGCCGACGTCAGAAAGAGAGAACTCGCCGCCTTCTTCGCTAATGTTATGCACGAAACTGAAGGGATG
TGCTACATCAATGAGATAAACCCTCAAAGTAACTACTGTAATAGTTCAGCCACGTGGCCATGTGCGAGTG
GTAAGAGCTACCATGGACGGGGCCCTCTCCAGTTGAGCTGGAATTACAATTATGGGGCGGCTGGGCAGAC
CATTGGGTTCGATGGGGTGAACAACCCAGAGAAGGTGGGGCAAGACCCCACCATTTCGTTCAAGACGGCT
GTCTGGTTCTGGATGAAGAACAGCAACTGTCACTCCGCCATAACGTCAGGGCAGGGCTTCGGTGGCACAA
TCAAAGCCATCAACAGCCAGGAATGCAACGGTGGAAAGACTGGGGAGGTGAATAACAGAGTGAACTACTA
CAAGAATATCTGTAGCCAGTTGGGTGTGGATCCTGGAGCCAATGTTTCGTGTTAA

>gi|333601375|gb|HM219851.1| Pinus contorta class VII chitinase (Chia7-1) mRNA, complete cds
ATGGCAACTCATTTCACTGTGCGCGCTGTTATTCTATGGTTTGTTTTTGCCCTTTCCGCTCTTTACATCT
GCCGCGGAGCTGTTTCGGATATTGCTACCCAAGATTTCTTCAATGGAATTTTATCTGCTGCTACTGATGG
CTGCGCAGGAAAGACCTTCTATACATACAGCGACTTCATTAATGCGGCCAATTCATTTTCTAGTTTTGGC
ACGACCGGAACATCCGATGACAACAAGAGAGAAATTGCTGCTTTCTTCGCCAATGTTGCTCACGAGACTA
CCAATCTGTGCTACGTGGAGGAGATAGCCAAGAGCGCCTACTGTGATTCCACCAATACACAATATCCATG
CGCATCTGGACAGCAATACTACGGCCGTGGCCCCTTGCAACTCACAGGAAATGCAAACTATGGTGCAGCT
GGAGCTTATCTGGCAGTGGATCTGCGGAATAACCCTGGATTGGTGGCTCAGGACGACCTAACCTCGTGGA
AGACAGCTCTGTGGTTTTGGAATGTGAACAGCAACTGCCACACCGCTATTACATCCGGTCAGGGATTCGG
GGCAACTATTCAAGCAATCAATGGAGCAGTAGAATGCAACGGTGGAAACACCGCCGAAGTTAACGACCGC
GTCAGTCGTTACACCACCTATTGCAGTCAGCTTGGTGTAGACCCAGGAAGCAATCTTACCTGTTAA

>gi|333601377|gb|HM219852.1| Pinus contorta class VII chitinase (Chia7-2) mRNA, complete cds
ATGAGAAATTACCTCCGTGTTATGACACTGAGCGCTATTATTCTATGGCTCGTTCTTGCCTTCGATCTTG
TTTCCATGTGCCGAGGAGATGTTGGGGATTTTGCTACTCAAGATTTCTTTAATGGAATTTTGTCTGGTGC

TTCTGATAGTTGCACAGGAAAGACCTTCTATACATACAATAACTTCATGGACGCAGCCACTGCATTCTCT
GGCTTTGGCACAACGGGCCCCGACGTTGACCACAAGAGAGAAATCGCTGCTTTCTTCGCCAATGTTGCTC
ACGAGACTTCCAGACTGTGCTACGTTGAGCAAATAGAGAAGTCTGACTACTGCGATTCCACCAACCAGAA
ATACCAGTGTGTGGCGGGAAAGCAGTACTATGGGCGTGGCCCCTTGCAACTCACATGGAATTATAATTAT
GGTGCAGCTGGTGATTACCTGGGCTTTGATGGCCTGAATCACCCTGAAATTGTAGCCCAGAATGGCTCAA
TCTCGTGGAAGACTGCTGTCTGGTTTTGGATGAAGCATAGTAATTGCCACTCTGCTATTACATCTGGACA
AGGATTCAGGGCAACGATCAAAGCAATCAGCGGGGATGAATGCAATGGTGGAGACTCCAACGCAGTTGAT
GAACGTGTCAATTATTATACCAACTATTGCAATGAGTTTGGTGTAGATCCAGGAAATAACCTTAGCTGTT
AA

Appendix 6.2

>PcChi1-1
MKSMKFCAMAIALLTMATMNMYFVSAEQCGQQAGGALCPGGLCCSKWGWCGNTDAHCGQD
CQSQCSGSTPTPGGQGVASIITESIFNELLKHRNDAGCKASGFYTYSAFIAAANAFPSFG
TTGDVATRKRELAAFFGQTSHETTGGWATAPDGAYAWGYCFKEEQGNPPAEYCQATSQWS
CASGKRYYGRGPVQLSWNYNYGPAGKAIGFDGINNPDIVASDATVSFKTAIWFWMTAQSP
KPSCHDVMTGKWTPSGSDSAAGRAAGYGAVTNIINGGLECGKGSDSRQQDRIGFYKRYSD
ILGVSYGSNLDCNNQRPFGAAVQSEPRLIKTVV
>PcChia1-2
TTGDVATRKRELAAFFGQTSHETTGGWATAPDGAYAWGYCFKEEQGNPPAEYCQATSQWP
CASGKRYYGRGPVQLSWNYNYGPAGKAIGFDGINNPDIVASDATVSFKTAIWFWMTAQSP
KPSCHDVMTGKWTPSGSDSAAGRAAGYGAVTNIINGGLECGKGSDSRQQDRIGFYKRYSD
ILGVSYGSNLDCNNQRPFGAAVQSEPRLIKTVV
>PcChia1-3
MEGTIGKMSSILLATTLAIFTAMTIVITGLPSVSADDCGKNAGGALCPGGLCCSKYGYCG
NTQAHCGKDCQSQCGGGGSTPTPTPPTPTPTTPTPTPSGQGVASIITEDLFNQLLKYKDD
SRCKANGFYSYDAFINVANAYPGFGTAGDVTSNKRELAAFFGQTSHETTGGWPAAPDGPY
AWGYCLKEEIGTPPPVYCNETAQWPCGSGKSYYGRGPIQISWNYNYGPAGNALGFDGINN
PDIVASDATVSFKTAIWFWMTAQSPKPSCHDVMTGKWTPSGSDSAAGRVPGYGAVTNIIN
GGLECGKGTVSRQQDRIGFYQRYCDILGVSYGANLDCNNQTPFGS
>PcChia1-4
MKAFWTVTIFILATSLAISAEQCGRQAGNAICPNGLCCSQFGFCGSTPDYCTKNCQSQCS
TGGGGGATPAPSGGGGGDVGSIITEGIFNEMLKYRNDGRCHANGFYTYSAFISAAKSFGG
FGTTGNDVQRKQELAAFLAQTSHETTGGWPSAPDGPYAWGYCFLKENNQGTFCTSNAYPC
APGKQYYGRGPIQLTHNYNYAQAGKAIGVDLINNPDLVATDATISFKTAIWFWMTPQANK
PRSEE
>PcChia2-1
MARKMSMKLLLALAAVAIMSTLCYVSEQQGVGSIITEDVFNEFLKHRNDAACQARDFYTY
SAFIAATNSFSDFGNNGDLESRKRELAAFFGQTSQETTGGWATAPDGPYAWGYCFKEENS
GDKYHGRGPIQLTGDYNYKAAGDALGYDLINNPDLLVTDATVSFKTAVWFWMTAQAPKPS
CHDVILGRWSPSNDDTAAGRVPGYGLLTNIINGGMECGTGTISDRQQGRIGFYQRYCSLL
GV
>PcChia2-2
MAYTNMGRRMSIMRLLLALTAVAIMSSLCCYVSAQQGVASIISEDVFNQFLKHRNDAACP
AKGFYTYTAFIAAANSFPDFGNNGDLQSRKRELAAFFGQTSQETTGGWATAPDGPYAWGY
CFKDQVNSGGDRYHGRGPIQLTGDYNYKAAGDALGYDLINNPDLVVNDATISFKTAVWFW
MTAQSPKPSCHDVILGRWSPSATDTAAGRVAGYGMVTDIINGGPECGTGTISDVQKGRIG
FYQRYCNMLGVAVGSNLDCKNQKPFGT
>PcChia2-3
MAMGRICLVVLLALALCLNVSLASKKMKICDKGWECKGAYCCNQTISEIFTVDNFEELFS
KRNTPVAHAVGFWDYYSFINAAAQFEGIGFGTTGGQVMQQKELAAFLGNVAAETSCGYNV
ATGGPTAWGLCYKEEMSPDQLYCDQNLLYPCAPGASYHGRGALPIYWNFNYGPIGEALKL
DLLTSPDMVSNNATIGFLTAMWRWMNPIKPKQPSAHDVFVGNWKPTKNDTESYRLPGFGM
VINVLNGGLECGKGDIDAMNNRISHYLYFLDLLGVGREQAGDNLDCGQQVPLNPVSTSTT
SR
>PcChia2-4

METKQVNRILVVAGICLMMSSWFCCIEAAAAAAAADENEETKFKKILCAKATDCKNKTIS
ELFTVDQFESLFSHRNAPLAHAQGFWDYHSFITAAAHYEPKGFGTTGGDIVQKRELAAFF
AHIATETSCESLMAQAASTAPSDSPTKWGLCYKEELSPDSTYCESSLVYPCAPGVSYHGR
GALPVYWNYNYGQLGQALKVDLLHHAEYLSQNATLAFAAAIWRWMTPMKVKQPSAHQVMV
GKWVPTKNDTEALRLPGFGMTINILKADAECGTDSDDKQMNTRIAHYLDFLDHMDVGREN
AGDNVDCSEQKVLNPSSSAST
>PcChia2-5
MAMGRICLVVLLALALCLNVSLASKKMKVCDKGWECKGTYCCNQTISEIFTVDNFEELFS
KRNTPVAHAVGFWDYYSFINAAAQFEGIGFGTTGGQVMQQKELAAFLGNVAAETSCGYNV
ATGGPTAWGLCYKEEMSPDQLYCDQNLLYPCAPGASYHGRGALPIYWNFNYGPIGEALKL
DLLTSPDMVSNNATIGFLTAMWRWMNPIKPKQPSAHDVFVGNWKPTKNDTESYRLPGFGM
VINVLNGGLECGKGDIDAMNNRISHYLYFLDLLGVGREQAGDNLDCGQQVPLNPVSTSTT
SR
>PcChia2-6
MEAKIKMGLFIALVMVTLASVNGDSSKEILVKTVKGKKVCSQGWECKQWSKYCCNQTISD
LFQVYQFEDLFAKRNTPVAHAVGFWDYQSFITAASIYEPLGFGTTGGKLMQMLEIAAFLG
HIGSKTSCGYGVATGGPLAWGLCYNREMSPSQSYCDDSYKYTYPCAPGAEYYGRGALPLY
WNFNYGAAGEALKVDLLNHPEYIEQNATLAFQAAIWRWM
>PcChia4-1
MGSSSGNSVMVVVVLVLLLVSVSVNAQNCGCASGLCCSKYGYCGTTSAYCGAGCKSGPCS
SSGGGSPSGGGGSVGTIISQSFFNGLAGGAASSCEGKGFYTYNAFIAAANAYSGFGTTGS
ADVRKRELAAFFANVMHETEGMCYINEINPQSNYCNSSATWPCASGKSYHGRGPLQLSWN
YNYGAAGQTIGFDGVNNPEKVGQDPTISFKTAVWFWMKNSNCHSAITSGQGFGGTIKAIN
SQECNGGKTGEVNNRVNYYKNICSQLGVDPGANVSC
>PcChia4-9
MHETGGLCYINEQNPASIYCDATSTSWPCASGKSYHGRGPLQLSWNYNYGVAGQNIGFDG
LNNPDVVGQDATISFKTAVWFWMQNCHSAITSGQGFGATIKAINSMECNGGNTPAVNSRV
SYYNTICSQLGVDPGANVSC
>PcChia4-10
CNCSPDLCCSQYGYCGTGDDYCGKGCQGGPCYNAAPSGGGTGGGSVADIVTDGFFNGIIS
QADGSCTGKNFYTRSAFLEATGSYTGFGSGSGDALKREIAAFFAHVTHETGHFCYTEEID
GRSKDYCDETKTQYPCAPNKYYFGRGPLQLTWNYNYAEAGKSVVFDGLQNPDIVASDAVV
SFKTALWFWTNNVESVMGQGFGATIRAINSIECNGGAPDAVNARISYYTSYCSQLGVAPG
ENLSC
>PcChia4-11
MLGCQEGPCDAPPPQNSVVVEDIVTDNFFNGIIDQAEESCAGKGFYSRSAFLEALKFYSG
FGTVGSEDDSKREIAAFFAHATHETGHFCYIEEINGPSKDYCDENNTQYPCAPNKGYYGR
GPIQLS
>PcChia4-12
MHETGGLCYINEQNPASIYCDATSTSWPCASGKSYHGRGPLQLSWNYNYGVAGQNIGFDG
LNNPDVVGQDATISFKTAVWFWMQNCHSAITSGQGFGATIKAINSMECNGGNTPAVNSRV
SYYNTICSQLGVDPGANVSC
>PcChia4-2
GCQSGPCTSSGGGSPSGGGGSVGTIISQSFFNGLAGGAASSCEGKGFYTYDAFIAAANAY
SGFGTTGSADVTKRELAAFFANVMHETGGMCYINERTPPMIYCMSSATWPCASGKDYHGR
GPLQLSWNYNYGAAGKNIGFDGVNNPEKVGQDSTISFKTAVWFWMDNSNCHTAITSGQGF
GGTIKAINSQECNGGNSGEVNSRVNYYKNICSQLGVDPGANVSC

>PcChia4-3
MHETGGLCYINEVNPGSIYCDSSSNTWPCASGKSYHGRGPLQLSWNYNYGAAGEDIGFDG
LNHPEEVALDAIISFRTGVWFWMKNSNCHSAITSGQGFGATIKAINSMECNGGNTAAVNS
RINYYTKFCS
>PcChia4-4
CGDGCQSGPCTSSGGGSPSTGGGNVGTIISQSVFDGLAGGAASSCEGKGFYTYDAFIAAA
NAYSGFGTTGASDDQKRELAAFFANVMHETGGLCYINEQNPPSIYCDATSTSWPCASGKS
YHGRGPLQLSWNYNYGVAGQNIGFDGLNNPDVVGQDATISFKTAVWFWMQNCHSAITSGQ
GFGATIKAINSMECNGGNTPAVNSRVSYYNTICSQLGVDPGANVSC
>PcChia4-5
MGVRCGTSVMALVLVLLLVGVSVNAQNCGCASGLCCSKYGYCGTTSDYCGDGCQSGPCTS
SGGGSPSTGGGNVGTIISQSVFDGLAGGAPSSCEGKGFYTYDAFIAAANAYSGFGTTGAS
DDQKRELAAFFANVMHETGGLCYINEQNPASIYCDATSTSWPCASGKSYHGRGPLQLSWN
YNYGVAGQNIGFDGLNNPDVVGQDATISFKTAVWFWMQNCHSAITSGQGFGATIKAINSM
>PcChia4-6
MHETGGLCYINEINPLSNFCQSSSTWPCASGKSYHGRGPIQLTWNYNYGAAGQDIGFDGL
NNPEKVGQDATISFKTA
>PcChia4-7
SGGGSPSGGGGSVGTIISQSFFNGLAGGAASSCEGKGFYTYNAFIAAANAYSGFGTTGSA
DVRKRELAAFFANVMHETGGMCYINERNPPMNYCMSSATWPCASGKSYHGRGPLQLSWNY
NYGAAGKSIGFDGVNNPEKVGQDPTISFKTAVWFWMKNSNCHSAITSGQGFGGTIKAINS
QECNGGNSGEVNSRVKYYKNICSQLGVDPGANVSC
>PcChia4-8
MGGSSGNSVMVLVMVLVLLLAGVSVNAQNCGCASGLCCSKYGYCGSTPAYCGRGCRSGPC
SSSGGGSPSGGGGSVGTIISQSFFNGLAGGAASSCEGKGFYTYDAFIAAANAYSGFGTTG
SADVTKRELAAFLANVMHETGGMCYINERNPPMNYCMSSATWPCASGKSYHGRGPLQLSW
NYNYGAAGQTIGFDGVNNPEKVGQDPTISFKTAVWFWMKNSNCHSAITSGQGFGGTIKAI
NSQECNGGNSGEVNSRVKYYKNICSQLGVDPGANVSC
>PcChia5-1
MGMPLYGRSWILKSLDETGIGAPAIAAGPKQTLSSETGVMFFSEIIDFISQKNATEVFDK
ETVSAYAYSSDMMWVGYDNKQSVAAKVSFAKKKHLLGYFFWAISQDSNWMLSAQASVSWN
QAR
>PcChia5-2
MAALHRLLNITGVAFTFLCVISPHMASASKGGYWPAYSYSYFPPSQINASLYTHLYYAFV
DVDNQTFQVGVSVENQQSIQQFTAQVQTNNPSVKTLLSIGGGGDDTIHTKFAKMAADASS
RKAFIDSSIALARNYSFHGLDLDWEYPQDTT
>PcChia5-3
MAQAPKNLCIILFFITSSIIYSSSSSNNTTSNNGVKAAYWPSDSSLSPSSISTDLFTHIF
YAFADLNDQTFQVQLPATADPAEFTSTLLQKNPSLKTLISIGGGGSNATAFALMASNASH
RKVFINSAIAVARKYGFHGLDLDWEFPRDAAEMEYLGTLFDDWRHAIDLEATASGRPRLL
LTAAVYFAQYFFVWGKKRAYPVTSIARNLDWVNVMCYDYHGSWDISATGAHAALYDPTSN
ISTSFGIGSWLHSGVPPNKVAMGMPLYGRSWILKSLDETEIGAPAVAAGPKQTLSNEKGV
MFFSEIRELINQKNATEVFDKETVSAYSYSSDLLWVGYDNQESVATKVSFAKEMHLLGYF
FWAIGQDNNWMLSAQASDSWN
>PcChia5-4
MASNASDRKVFINSAIAIARKYGFHGLDLDWEFPKNSVEMENLGTLFSEWRHAIDLEATA
SGRPRILLTAAVYFAQYFFLSSEKRAYPATSIAQNLDWVNVMCYDYHGSWNISATGAHAA

LYDPSSNISTSFGIGSWLHSGVSPNNIVMGMPLYGRSWILKSLDETGIGAPAMAAGPKQT
LSNEKGVMFFSEIRDLINQKNATEVFDKDTVSAYSYSSGMIWVAYDNPDSVATKVSFAKE
RRLLGYFFWGYQSR
>PcChia5-5
MAQTQKYLLFILLSMTTSLVYSSSSQSASHGGGVKAAYWPSYAVSYSPPSSINAALFTHI
FYAFANLNDQTFQVELPATAKPAEFTLTLRKKNPFLKTLISIGGGGSNSTAFALMASNAS
RRKVFINSAIALARKYGFHGLDLDWEFPKDATEMQNLGSLFAEWRHAIELEARAWPRCPR
LLLTAAVYFAQYFLLAADKRAYPATSITQNLDWVNVMCYDYHGSWDISATGAHAALYDPT
SNISTSFGIGSWLHSGVPPNKVAMGMPLYGRSWILKSLDDTEIGAPAVAAGPKQTLSNEK
GVMFFSEIRELINQKNATEVFDKETVSAYSYSSDLLWVGYDNQESVATKVSFAKEMHLLG
YFFWAIGQDNNWMLSAQASDSWN
>PcChia5-6
MALKTLAITLLVFLVQLLQFSVPAQSQSSVNAGYWFLDSGLAASDINSTLFTHLFCAFAD
LDPNTKQVTISSSNNASFAQFTQTVRLKNPSVITLLSIGGGNSNDADFAAIASNSTSRKS
FIDSSLKVAGRLVCYRSLA
>PcChia5-7
MALKTLAITLLVFLVQLLQFSVPAQSQSSVNAGYWFLDSGLAASDINSTLFTHLFCAFAD
LDPNTKQVTISSSNNASFAQFTQTVRLKNPSVITLLSIGGGNSNDADFAAIASNSTSRKS
FIDSSLKVARSYNFLGLDLDYEYPLSATDMVNLGTLLSEWRTAAATEATN
>PcChia5-8
MENLGTLFSEWRHAIDLEATASGRPRILLTAAVYFAQYFFLSSEKRAYPATSIAQNLDWV
NVMCYD
>PcChia5-9
SVPCIRSCAGQCPIGGVLHDVPEEVAQQPFLFRKHRLPGYRLVIIVANPVESRVTVSRDG
GVIKHLRGCILCNEVPNIRVCNGAIFYTYATSCGSYSHLPGALEFPRVAVEGHSH
>PcChia7-1
RAKHRTCLCVCTSDNSSTATQLPPIWQLISLCALLFYGLFLPFPLLSICRGAVSDIATQD
FFNGILSAATDGCAGKTFYTYSDFINAANSFSSFGTTGTSDDNKREIAAFFANVAHETTN
LCYVEEIVKSAYCDSTNTQYPCASGQQYYGRGPLQLTGNANYGAAGAYLALDLLNNPGLV
AQDDLTSWKTALWFWNVNSNCHTAITSGQGFGATIQAINGAVECNGGNTAEVNDRVSRYT
TYCSQLGVDPGSNLTC
>PcChia7-9
NAFSGFGTTGTSDDEKREIAAFFANVAHETIGFCYIDEIQKGDYCDSNSTQYPCAPGKQY
YGRGPIQLSWNFNYGPAGNYLGVDLLNHPEIVAQDALISWKSALWYWNVNSDCHSAITSG
KGFGATIQAINGAIECNGGNTDEVNDRISYYTKYCSQFGVDPGSNLSC
>PcChia7-10
MHSLASAQLGLPMMRREKLPLSSPTLLTKLSKGDYCDSNSTQYPCAPGKQYYGRGPIQLS
WNFNYGPAGNYLGVDLLNHPEIVAQDALISWKSALWYWNVNSDCHSAITSGKGFGATIQA
INGAIECNGGNTDEVNDRISYYTKYCSQFGVDPGSNLSC
>PcChia7-11
MHSLASAQLGLPMMRREKLPLSSPTLLTKLSKGDYCDSNSTQYPCAPGKQYYGRGPIQLS
WNFNYGPAGNYLGVDLLNHPEIVAQDALISWKSALWYWNVNSDCHSAITSGKGFGATIQA
INGAIECNGGNTDEVNDRISYYTKYCSQFGVDPGSNLSC
>PcChia7-2
MRNYLRVMTPSVIILWLVLAFDLVSICRGDVGDFATQDFFNGILSGASDSCAGKNFYTYN
NFMDAATAFSGFGTTGPDVDHKREIAAFFANVAHETSRLCYVEQIEKSDYCDSTNEKYQC
VAGKQYYGRGPLQLTWNYNYDAAGDYLGFDGLNHPEIVAQNGSISWKTAVWFWMKHSNCH

SAITSGQGFRATIKAISGDECNGGDSNAVDERVNYYTNYCNEFGVDPGNNLSC
>PcChia7-3
MAAQLPVSAIILWFVFALSALSICRGAVSDIATEDFFNGILSAATDGCAGKTFYTYSDFI
NAANSFSSFGTTGTSDDNKREIAAFFGNVAHETTNLCFVEENAKSDNCDSSNTQYPCASG
QQYYGRGPLQLTGNGNYGAAGSYLGVDLLNNPGLVAQDDLTSWKTALWFWNVNSNCHTAI
TSGQGFGATIQAINGAKECNGVNTAEVNDRVSRYTTYCSQLVVDPGSNLSC
>PcChia7-4
MATHLRSVLILWLALALYVVSICRGDVGDFATQDFFNGILSGASDSCAGETFYTYNNFMD
AATSFSGFGTTGPDVDHKREIAAFFANVAHETSRLCYVEQIEKSDYCDSTNQKYQCVAGK
QYYGRGPLQLTWNYNYDAAGDYLGFDGLNHPEIVAQNGSISWKTAVWFWMLDSNCHSAIT
SGQG
>PcChia7-5
MATHLRSVLILWLALALYVVSICRGDVGDIVTRDFFNGILSGAADSCAGKNFYTYDDFIT
AANAFSGFGTTGTSDDNKRELAAFFANVAHETGSLCYIEEINKSDYCDSTNTQYPCVAGK
QYYGRGPLQLTWNYNYDAAGDYLGFDGLNHPEIVAQNGSISWKTAVWFWMLDSNCHSAIT
SGQG
>PcChia7-6
MATHFTVRAVILWFVFALSALSICRGAVSDIATQDFFNGILSAAADGCAGKTFYTYSDFI
NAANSFSSFGTTGTSDDNKREIAAFFANVAHETTDLCYVEEIAKSAYCDSTNTQYPCASG
QQYYGRGPLQLTGNANYGAGGAYLALDLLNNPGLVAQDDLTSWKTALWFWNVNSNCHTAI
TSGQGFGATIQAINGAVECNGGNTDEVNDRVSRYTTYCSQLGVDPGSNLTC
>PcChia7-7
MATHFRVNVIFLWLAFALSALSICRGAVSDIATQDFFNGILSAATDGCAGKTFYTYTDFI
NAAISFSSFGTTGTSDDNKREIAAFFANVAHETTNLCYVEEIAKSDYCSSNTQYPCASGQ
QYYGRGPLQLTGNGNYGAAGDYLGVDLLNNPGLVAQDDLTSWKTALWFWNVNSNCHTAIT
SGQGFGATIQAINGAIECNGGNTDQVNDRISRYTNYCSQFGVDPGSNLSC
>PcChia7-8
MASHLPVTTVSLFIIGLALAFFTCRGDVENIATEDFFNGIISGVSDTCEAKNFYTYSDFI
TAANVFSGFGTTGTSDDEKREIAAFFANVADETIGFCYINEIQKGDYCDSNSTQYPCAPG
KQYYGRGPIQLSWNFNYGPAGNYLGVDLLNHPEIVAQDALISWKSALWYWNVNSDCHSSI
TSGKGFGATIQAINGAIECNGGNTDEVNDRISYYTKYCSQFGVDPGSNLSC
>PbChia1-1
MKSMKFSAMAIALLTMATMNLYFVSAEQCGQQAGGALCPGGLCCSKWGWCGNTDAHCGQD
CQSQCGGSTPTPPSPTPGGQGVASIITESIFNELLKHRNDAGCKASGFYTYSAFIAAANA
FPSFGTTGDVATRKRELAAFFGQTSHETTGGWATAPDGAYAWGYCFKEEQGNPPAEYCQA
TSQWPCASGKRYYGRGPVQLSWNYNYGPAGKAIGFDGINNPDIVASDATVSFKTAIWFWM
TAQSPKPSCHDVMTGKWTPSGSDSAAGRAAGYGAVTNIINGGLECGKGSDSRQQDRIGFY
KRYSDILGVSYGSNLDCNNQRPFGAAVQSEPRLIKTVV
>PbChia1-2
MKSMKFSAMAIALLTMATMNLYFVSAEQCGQQAGGALCPGGLCCSKWGWCGNTDAHCGQD
CQSQCGGSTPTPGGQGVASIITESIFNELLKHRNDAGCKASGFYTYSAFIAAANAFPSFG
TTGDVATRKRELAAFFGQTSHETTGGWATAPDGAYAWGYCFKEEQGNPPAEYCQATSQWP
CASGKRYYGRGPVQLSWNYNYGPAGKAIGFDGINNPDIVASDATVSFKTAIWFWMTAQSP
KPSCHDVMTGKWTPSGSDSAAGRAAGYGAVTNIINGGLECGKGSDSRQQDRIGFYKRYSD
ILGVSYGSNLDCNNQRPFGAAVQSEPRLIKTVV
>PbChia1-3
MSSILLATTLAIFTAMAIVITGLPSVSADDCGKNAGGALCPGGLCCSKYGYCGNTQAHCG

KDCQSQCGGGGSTPTPTPPTPTPTTPTPTPSGQGVASIITEDLFNQLLKYKDDSRCKANG
FYSYDAFINVANAYPGFGTAGDVTSNKRELAAFFGQTSHETTGGWPAAPDGPYAWGYCLK
EEIGTPPPVYCNETAQWPCGSGKSYYGRGPIQISWNYNYGPAGKAIGFDGINNPDIVASD
ATVSFKTAIWFWMTAQSPKPSCHDVMTGRWSPSSSDSAAGRAPGYGAVTNIINGGLECGK
GSYPKQEDRIGFYKRYCDILGVSYGSNLDCNNQTPFGS
>PbChia1-4
MAIMITGLPSVSAEDCGQQAGGAVCPRGLCCSKWGWCGNTEAHCGLGCQSQCGGGGSTPT
PTPPTPTPTNPTPTPSGQGVASIITEDLFNQLLKYKDDSRCKANGFYSYAAFIAAANAFP
GFGTAGDVTTNKRELAAFLGQTSHETTGGWQTAPDGPYAWGYCFKEEQGSPPPVYCQASS
QWPCVSGKSYYGRGPIQLTWNYNYGPAGKAIGFDGINNPDIVASDATVSFKTAIWFWMTA
QSPKPSCHDVMTGKWTPSGSDSAAGRVPGYGAVTNIVNGGLECGKGTDSRQQDRIGFYQR
YCDILGVSYGANLDCNNQTPFGS
>PbChia2-1
MARKMSMKLLLALAAVAIMSTLCYVSGQQGVGSIITEDVFNEFLKHRNDAACQARDFYTY
SAFIAATNSFSDFGNNGDLESRKRELAAFFGQTSQETTGGWATAPDGPYAWGYCFKEENS
GDKYHGRGPIQLTGDYNYKAAGDALGYDLINNPDLLVTDATVSFKTAVWFWMTAQAPKPS
CHDVILGRWSPSNDDTAAGRVPGYGLLTNIINGGEECGTGTISDRQQGRIGFYQRYCSLL
GVDTGSNLDCQNQKHF
>PbChia2-2
MAYTNMGRRMSIMRLLLALTAVAIMSSLCCYVSAQQGVASIISEDVFNQFLKHRNDAACS
AKGFYTYTAFIAAANSFPDFGNNGDLESRKRELAAFFGQTSQETTGGWATAPDGPYAWGY
CFKDQVNSGGDRYHGRGPIQLTGDYNYKAAGDALGYDLINNPDLVVNDATISFKTAVWFW
MTAQSPKPSCHDVILGRWSPSATDTAAGRVAGYGMVTDIINGGPECGTGTISDVQKGRIG
FYQRYCNMLGVDVGSNLDCKNQKPFGT
>PbChia2-3
MAMGRICLVVLLALALCLNVSLASKKMKICDKGWECKGAYCCNQTISEIFTVDNFEELFS
KRNTPVAHAVGFWDYYSFINAAAQFEGIGFGTTGGQVMQQKELAAFLGNVAAETSCGYNV
ATGGPTAWGLCYKEEMSPDQLYCDQNLLYPCAPGASYHGRGALPIYWNFNYGPIGEALKL
DLLTSPDMVSNNATIGFLTAMWRWMNPIKPKQPSAHDVFVGNWKPTKNDTESYRLPGFGM
VINVLNGGLECGKGDIDAMNNRISHYLYFLDLLGVGREQAGDNLDCGQQVPLNPISTSTT
SR
>PbChia2-4
METKQVNRILVVAGICLMMSSWFCCIEAAAAAAAADENEETKFKKILCAKATDCKNKTIS
ELFTVDQFESLFSHRNAPLAHAQGFWDYHSFITAAAHYEPKGFGTTGGDIVQKRELAAFF
AHIATETSCESLMAQAASTAPSDSPTKWGLCYKEELSPDSTYCESSLVYPCAPGVSYHGR
GALPVYWNYNYGQLGQALKVDLLHHAEYLSENATLAFAAAIWRWMTPMKVKQPSAHQVMV
GKWVPTKNDTEALRLPGFGMTINILKADAECGTDSDDKQMNTRIAHYLDFLDHMDVGREN
AGDNVDCSEQKVLNPSSSAST
>PbChia4-1
MGSSSGNSVMAVLVLLLVSVSANAQNCGCASGLCCSKYGYCGTTSAYCGAGCKSGPCSSS
GGGSPSGGGGSVGTIISQSFFNGLAGGAASSCEGKGFYTYNAFIAAANAYSGFGTTGSAD
VRKRELAAFFANVMHETGGMCYINERNPPMNYCMSSATWPCASGKSYHGRGPLQLSWNYN
YGAAGKSIGFDGVNNPEKVGQDPTISFKTAVWFWMKNSNCHSAITSGQGFGGTIKAINSQ
ECNGGNSGEVNSRVKYYKNICSQLGVDPGANVSC
>PbChia4-10
MGTRTGRSVIASPYEHMVLRMSMVLLLVVVVSVNVVYAQNCGCASGLCCSKWGYCGTTSA
YCGNGCQSGPCYGSGGGSPSGGGGNVGTIISQNFFSGLANAAGSSCEGKGFYTYNAFITA

ANAYSGFGTTGASDNQKRELAAFFANVMHETGGLCYINEINPLSNFCQSSSTWPCASGKS
YHGRGPIQLTWNYNYGAAGQDIGFDGLNNPEKVGQDATISFKTAVWFWMKNSNCHSAITS
GQGFGGTIKAINSRECNGGNSAQVNSRVNYYKKLCSQLGVDPGTNISC
>PbChia4-11
MGTRTGRSVIVSPYEHMVLRMSMVLLLVVVVSVNVVYAQNCGCASGLCCSKWGYCGTTSA
YCGNGCQSGPCYGSGGGSPSGGGGNVGTIISQNFFSGLANAAGSSCEGKGFYTYNAFITA
ANAYSGFGTTGASDNQKRELAAFFANVMHETGGLCYINEINPLSNFCQSSSTWPCASGKS
YHGRGPIQLTWNYNYGAAGQDIGFDGLNNPEKVGQDATISFKTAVWFWMKNSNCHSAITS
GQGFGATIKAINGGECNGGHSAQVNSRVNYYKKLCSQLGVDPGTNISC
>PbChia4-2
MGSSSGNSAMVVVVLVLLLVGVSVTAQNCGCASDLCCSKYGYCGTTSAYCGDGCQSGPCT
SSGGSLSGGGGSVGTIISQSFFDGLKGGAASSCEGKGFYTYNAFIAAANAYSGFGTTGSA
DVTKRELAAFFANVMHETGGMCYINERNPPMNYCMGSATSPCASGKSYYGRGPLQLSWNY
NYGAAGQSIGFDGVNNPEKVGQDSTISFKTAVWFWMKNSNCHSAITSGQGFGGTIKAINS
QECNGGNSGEVNSRVNYYKNICSQLGVDPGANVSC
>PbChia4-3
MATVLVLVQLLLGVIVNVVNAESVANIVSRSFFDSILNAPTLANCSCAIGLCCSKWGYCG
NTASYCGDGCQSGSCYNQSGGSSTGRHGGAGAIISSSFFDGLAGGAASSCEGKGFYTYKA
FVAASSAYSVFGTTGSSDDQKRELAAFFANVMHETGGLCYINEVNPGSIYCDSSSNTWPC
ASGKSYHGRGPLQLSWNYNYGAAGEDIGFDGLNHPEEVALDAIISFRTGVWFWMKNSNCH
SAITSGQGFGATIKAINSMECNGGNTAAVNSRINYYTKFCSELGVDPGTKLSC
>PbChia4-4
MGVRCGTSVMALVLVLLLVGVSVNAQNCGCASGLCCSKYGYCGTTSDYCGDGCQSGPCTS
SGGGSPSTGGGNVGTIISQSVFDGLAGGAASSCEGKGFYTYDAFIAAANAYSEFGTTGAS
DDQKRELAAFFANVMHETGGLCYINEQNPPSIYCDATSTSWPCASGKSYHGRGPLQLSWN
YNYGAAGQNIGFDGLNNPDVVGQDATISFKTAVWFWMQNCHSAITSGQGFAATIKAINSM
ECNGGNTPAVNSRVNYYNTICSQLGVDPGSNVSC
>PbChia4-5
MGVRCGTSVMALVLVLLLVGVSVNAQNCGCASGLCCSKYGYCGTTSDYCGDGCQSGPCTS
SGGGSPSTGGGNVGTIISQSVFDGLAGGAPSSCEGKGFYTYDAFIAAANAYSGFGTTGAS
DDQKRELAAFFANVMHETGGLCYINEQNPASIYCDATSTSWPCASGKSYHGRGPLQLSWN
YNYGVAGQNIGFDGLNNPDVVGQDATISFKTAVWFWMQNCHSAITSGQGFGATIKAINSM
ECNGGNTPAVNSRVSYYNTICSQLGVDPGANVSC
>PbChia4-6
MGVRCGTSVMALVLVLLLVGVSVNAQNCGCASGLCCSKYGYCGTTSDYCGDGCQSGPCTS
SGGGSPSTGGGNVGTIISQSVFDGLAGGAPSSCEGKGFYTYDAFIAAANAYSGFGTTGAS
DNQKRELAAFFANVMHETGGLCYINEQNPASIYCDATSTSWPCASGKSYHGRGPLQLSWN
YNYGVAGQNIGFDGLNNPDVVGQDATISFKTAVWFWMQNCHSAITSG
>PbChia4-7
MGVRCGTSVMALVLVLLLVGVSVNAQNCGCASGLCCSKYGYCGTTSDYCGDGCQSGPCTS
SGGGSPSTGGGNVGTIISQSVFDGLAGGAPSSCEGKGFYTYDAFIAAANAYSGFGTTGAS
DNQKRELAAFFANVMHETGGLCYINEINPSSNYCQSSSTWPCASGKSYHGRGPIQLSWNY
NYGAAGQDIGFDGLNNPEKVGQDATISFKTA
>PbChia4-8
MGVRCGTSVMALVLVLLLVGVSVNAQNCGCASGLCCSKYGYCGTTSAYCGAGCKSGPCSS
SGGGSPSGGGGSVGTIISQSFFNGLAGGVASSCEGKGFYTYNAFIAAANAYSGFGTTGSA
DVRKRELAAFFANVMHETGGMCYINERNPQINYCMSSTTWPCASGKSYHGRGPLQLSWNY

NYGPAGKSIGFDGVNNPEKVGQDPTISFKTAVWFWMKNSNCHSAITSGQGFGGTIKAINS
QECNGGNSGEVNSRVNYYKNICSQLGVDPGANVSC
>PbChia4-9
MVLRMSMVLLLVVVVSVNVVYAQNCGCASGLCCSKWGYCGTTSAYCGNGCQSGPCYGSGG
GSPSGGGGNVGTIISQNFFSGLANAAGSSCEGKGFYTYNAFIAAANAYSGFGTTGASDNQ
KRELAAFFANVMHETGGLCYINEINPSSNYCQSSSTWPCASGKSYHGRGPIQLSWNYNYG
AAGQDIGFDGLNNPEKVGQDATISFKTAVWFWMKNSNCHSAITSGQGFGGTIKAINSGEC
NGGNSAQVNRRVNYYKKFCSQLGVDPGTNISC
>PbChia5-1
MTTSLVYSSSSQSASHGGGVKAAYWPSYTVSYSPPSSINAALFTHIFYAFANLNDQTFQV
ELPATAKPAEFTLTLRKKNPFLKTLISIGGGGSNSTAFALMASNAFRRKVFINSAITLAR
KYGFHGLDLDWEFPKDATEMRNLGSLFAEWRHAIELEARAWPRRPRLLLTAAVYFAQYFL
LAADKRAYPATSITQNLDWVNVMCFDYHGSWDTSATGAHAALYDPSSNISNSFGIGSWLH
SGVPTNKVVMGMPLYGRSWILKSLDETGIGAPAIAAGPKQTLSSETGVMFFSEIIDFISQ
KNATEVFDKETVSAYAYSSDMMWVGYDNKQSVAAKVSFAKNKHLLGYFFWAISQDSNWML
SAQASVSWNQAR
>PbChia5-2
MAALHRLLNITGVAFTFLCVISPHIASASKGGYWPAYSYSYFPPSQINASLYTHLYYAFV
DVDNQTFQVGVSVENQQSIQQFTAQVQTNNPSVKTLLSIGGGGDDSIHTKFAKMAADASS
RKAFIDSSIALARNYSFHGLDLDWEYPQDTTEMNNLGQLLSEWRSAVQAEATATGNDPLL
LTAAVYFSHSFFVSDIVREYPSEMDKLDWLNIMAYDFHGSWEPKQTGAPAALYDPDSNIS
TSFGVQSWLDAGLPSSKVVMGMPLYGYSWKLQSAGEVGIGAAASGTGIKDGSVTYSDIRD
FIAQNAATEVFDNTTVSAYCYSGLDWIGYDNDQSVARKAVFAKQEGLLGYFFWNVVQDTA
DWALSRTASDTWDG
>PbChia5-3
MASNASHRKVFINSSIAVARKYGFHGLDLDWEFPQDAAEMEYLGTLFDDWRHAIDLEATA
SGRPRLLLTAAVYFAQYFFVWGKKRAYPVTSIARNLDWVNVMCYDYHGSWDISATGAHAA
LYDPTSNISTSFGIGSWLHSGVPPNKVAMGMPLYGRSWILKSLDETEIGAPAVAAGPKQT
LSNEKGVMFFSEIRELINQKNATEVFDKETVSAYSYSSDLLWVGYDNQESVATKVSFAKE
MHLLGYFFWAIGQDNNWMLSAQASDSWN
>PbChia5-4
MAQPQKYLLFILLSMTTSLVYSSSSQSASHGGGVKAAYWPSYTVSYSPPSSINAALFTHI
FYAFANLNDQTFQVELPATAKPAEFTLTLRKKNPFLKTLISIGGGGSNSTAFALMASNAF
RRKVFINSAITLARKYGFHGLDLDWEFPQDAAEMEYLGTLFDDWRHDIDLEATASGRPRL
LLTAAVYFAQYFFVWGEKRAYPVTSIARNLDWVNV
>PbChia5-5
MKKPSLKTLLSIGGGGSSPTVFALMANNASDRKVFINSAIAIARKYGFHGLDLDWEFPKN
SVEMENLGTLFSEWRHAIDLEATASGRPRILLTAAVYFAQYFFLSSEKRAYPATSIAQNL
DWVNVMCYDYHGSWNISATGAHAALYDPSSNISTSFGIGSWLHSGVSPNNIVMGMPLYGR
SWILKSLDETGIGAPAMAAGPKQTLSNEKGVMFFSEIRDLINQKNATEVFDKDTVSAYSY
SSDMIWVAYDNPDSVATKVSFAKER
>PbChia5-6
MGMPLYGCTWILKSLNETGIGAPAVAAGPKLTLSNETGVMFFSDIRNFITQKNVTVVFDN
ETVSAYAYSSDMMWVGYDNPDSVAIKVSFAKERRLLGYFFWAVSQDSNWMLSTRALETWN
QVQ
>PbChia5-7
INTDLFTHLLYASVYLNNHTFQVDVPPTENPLQFTSTLLKKKPSLKTLISISGDTTAFAI

MASNASHRKAFINSTFALARKYGFHG
>PbChia5-8
MAQAPKNLCIILFFITSSIIYSSSSSNNTTSNNGVKAAYWPSDSSLSPSSISTDLFTHIF
YAFADLNDQTFQVQLPATADPAEFTSTLLQKNPSLKTLISIGGGGSNATAFALMASNTSY
RKVFINSTIALARKYGFHGLDLDWEFPQDAAEMEYLGTLFDDWRHDIDLEATASGRPRLL
LTAAVYFAQYFFVWGEKRAYPVTSIARNLDWVNVMCYDYHGSWDISATGAHAALYDPTSN
ISTSFGIGSWLHSGVPPNKVAMGMPLYGRSWILKSLDETEIGAPAVAAGPKQTLSNEKGV
MFFSEIRELINQKNATEVFDKETVSAYSYSSDLLWVGYDNQESVATKVSFAKEMHLLGYF
FWAIGQDNNWMLSAQASDSWN
>PbChia5-9
MAQPQKYLLFILLSMTTSLVYSSSSQSASHGGGVKAAYWPSYTVSYSPPSSINAALFTHI
FYAFANLNDQTFQVELPATAKPAEFTLTLRKKNPFLKTLISIGGGGSNSTAFALMASNAF
RRKVFINSAITLARKYGFHGLDLDWEFPKDATEMRNLGSLFAEWRHAIELEARAWPRRPR
LLLTAAVYFAQYFLLAADKRAYPATSITQNLDWVNVMCFDYHGSWDTSATGAHAALYDPS
SNISNSFGIGSWLHSGVPTNKVVMGMPLYGRSWILKSLDETGIGAPAIAAGPKQTLSSET
GVMFFSEIIDFISQKNATEVFDKETVSAYAYSSDMMWVGYDNKQSVAAKVSFAKNKHLLG
YFFWAISQDSNWMLSAQASVSWNQAR
>PbChia7-1
RAKHRTCLCVCTSDNSSTATQLPPIWQLISLCALLFYGLFLPFPLLSICRGAVSDIATQD
FFNGILSAATDGCAGKTFYTYSDFINAANSFSSFGTTGTSDDNKREIAAFFANVAHETTN
LCYVEEIVKSAYCDSTNTQYPCASGQQYYGRGPLQLTGNANYGAAGAYLALDLLNNPGLV
AQDDLTSWKTALWFWNVNSNCHTAITSGQGFGATIQAINGAVECNGGNTAEVNDRVSRYT
TYCSQLGVDPGSNLTC
>PbChia7-10
MATFLRSTLILWLALALYVVSFCQGDVGDIVTQDFFNAILSGADGSCAGQNFYTYNDFIN
AANAFSGFGTTGTSDDNKRELAAFFANVAHETGSLCYIEEISKSDYCDSTNTQYPCATGK
QYYGRGPIQLTWNYNYGAAGNYVGFDGLNNPDIVAQDDSISWKTAVWFWMLDSNCHSAIT
SGQGFGATIQAINSGECNGGNSAAVTSRVNYYNNYCSQFGVDPGSDVSC
>PbChia7-11
MATFLRSTLILWLALALYVVSFCQGDVGDIVTQDFFNAILSGADGSCAGQNFYTYNDFIN
AANAFSGFGTTGTSDDNKRELAAFFANVAHETGSLCYIEEISKSDYCDSTNTQYPCATGK
QYYGRGPIQLTWNYNYGAAGNYVGFDGLNNPDIVAQDDSISWKTAVWFWMLDSNCHSAIT
SGQGFGATIQAINSGECNGGNTAAVTSRVNYYNNYCSQFGVDPGSDVSC
>PbChia7-12
MRNYLRVMTPSVIILLLVLAFDVVSICRGDVGDFATQDFFNGILSGASDSCAGKTFYTYN
NFMDAATAFSGFGTTGADVDHKREIAAFFANIARETSRLCYVEQIEKSDYCDSTNKKYQC
VAGKQYYGRGPLQLTWNYNYDAAGDYLGFDGLNHPEIVAQNGSISWKAAVWFWMKHSNCH
SAITSGQGFRATIKAISGDECNGGDSDSVDERVNYYTNYCNEFGVDPGSNLSC
>PbChia7-13
MRNYLRVMTPSVIILLLVLAFDVVSICRGDVGDFATQDFFNGILSGASDSCAGKTFYTYN
NFMDAATAFSGFGTTGADVDHKREIAAFFANIAHETSRLCYVEQIEKSDYCDSTNKKYQC
VAGKQYYGRGPLQLTWNYNYDAAGDYLGFDGLNHPEIVAQNGSISWKTAVWFWMKHSNCH
SAITSGQGFRATIKAISGDECNGGDSDAVDERVNYYTNYCNEFGVDPGSNLSC
>PbChia7-14
MLFPSAEEMLGILPPKISSMEFCLVLLIVAQERPSIHTITSWTQPLHSLALAQQAPTLTT
REIAAFFANIAHETSRLCYVEQIEKSDYCDSTNKKYQCVAGKQYYGRGPLQLTWNYNYDA
AGHYLGFDGLNHPEIVAQNGSISWKTAVWFWMKHSNCHSAITSGQGFRATIKAISGDECN

GGDSDSVDERVNYYTNYCNEFGVDPGSNLSC
>PbChia7-15
MDAATAFSGFGTTGADVDHKREIAAFFANIAHETSRLCYVEQIEKSDYCDSTNKKYQCVA
GKQYYGRGPLQLTWNYNYNAAGDYLGFDGLNHPEIVAQNGSISWKTAVWFWMKHSNCHSA
ITSGQGFRATIKAISGDECNGGDSDAVDERVNYYTNYCNEFGVDPGSNLSC
>PbChia7-16
SDIATQDFFNGILSAAADSCAGKNFYAYSDFINAANSFSSFGTTGTSDDNKREIAAFFAN
VAHETGNLCYVEEINKSDYCDSTNTQYPCAAGKQYYGRGPLQLSWNYNYGAAGDYLGQDL
LNNPDTVAQDDLTSWKTALWFWNVNSNCHTAITSGQGFGATIQAINGAIECNGGNTDQVN
DRISRYTNYCSQFNVDPGSNLSC
>PbChia7-2
MRNYLRVMTLSAIILWLVLAFDLVSMCRGDVGDFATQDFFNGILSGASDSCTGKTFYTYN
NFMDAATAFSGFGTTGPDVDHKREIAAFFANVAHETSRLCYVEQIEKSDYCDSTNQKYQC
VAGKQYYGRGPLQLTWNYNYGAAGDYLGFDGLNHPEIVAQNGSISWKTAVWFWMKHSNCH
SAITSGQGFMATIKAISGDECNGGDSNAVDERVNYYTNYCNEFGVDPGNNLSC
>PbChia7-3
MAAQLPVSAIILWFVFALSALSICRGAVSDIATEDFFNGILSAATDGCAGKTFYTYSDFI
NAANSFSSFGTTGTSDDNKREIAAFFGNVAHETTNLCFVEENAKSDNCDSSNTQYPCASG
QQYYGRGPLQLTGNGNYGAAGSYLGVDLLNNPGLVAQDDLTSWKTALWFWNVNSNCHTAI
TSGQGFGATIQAINGAKECNGGNTAEVNDRVSRYTTYCSQLVVDPGSNLSC
>PbChia7-4
MATHLRSVLILWLALALYVVSICRGDVGDIVTRDFFNGILSGAADSCAGKNFYTYDDFIT
AANAFSGFGTTGTSDDNKRELAAFFANVAHETGSLCYIEEINKSDYCDSTNTQYPCVAGK
QYYGRGPLQLTWNYNYGPAGNYVGFDGLNNPDIVAQDDSISWKTAVWFWMLDSNCHSAIT
SGQGFGATIQAINSGECNGGNSAAVADRVKYYNNYCGQFGVDPGSNVSC
>PbChia7-5
MATHLRSVLILWLALALYVVSICRGDVGDIVTRDFFNGILSGAADSCAGKNFYTYDDFIT
AANAFSGFGTTGTSDDNKRELAAFFANVAHETGSLCYIEEINKSDYCDSTNTQYPCVAGK
QYYGRGPLQLTWNYNYGAAGNYLGFDGLNNPDIVAQDDSISWKTAVWFWMLDSNCHSAIT
SGQGFGATIQAINSGECNGGNTAAVTDRVNYYNNYCGQFGVDPGSNVSC
>PbChia7-6
MATHFTVRAVILWFVFALSALSICRGAVSDIATQNFFNGILSAATDGCAGKTFYKYSDFI
NAANSFSSFGTTGTSDDNKREIAAFFANVAHETTNLCYVEEIVKSAYCDSTNTQYPCASG
QQYYGRGPLQLTGNANYGAAGAYLAVDLRNNPGLVAQDDLTSWKTALWFWNVNSNCHTAI
TSGQGFGATIQAINGAVECNGGNTAEVNDRVSRYTTYCSQLGVDPGSNLTC
>PbChia7-7
MATHFTVRAVILWFVFALSALSICRGAVSDIATQNFFNGILSAATDGCAGKTFYKYSDFI
NAANSFSSFGTTGTSDDNKREIAAFFANVAHETTNLCYVEEIVKSAYCDSTNTQYPCASG
QQYYGRGPLQLTGNANYGAAGAYLAVDLRNNPGLVAQDDLTSWKTALWFWNVNSNCHTAI
TSGQGFGATIQAINGAVECNGGNTAEVNDRVSRYTTYCSQLGVDPGSNLTC
>PbChia7-8
MAIHLPVTTVSLFIIGLALAPFTCRGDVENIATEDFFNGLISGVSDTCEAKDFYTYSDFI
TAANAFSGFGTTGTSDDEKREIAAFFANVAHETIGFCYIDEIQKGDYCDSNSTQYPCAPG
KQYYGRGPIQLSWNFNYGPAGNYLGVDLLNHPEIVAQDALISWKSALWYWNVNSDCHSAI
TSGKGFGATIQAINGAIECNGGNTDEVNDRISYYTKYCSQFGVDPGSNLSC
>PbChia7-9
MATFLRSTLILWLALALYVVSFCQGDVGDIVTQDFFNAILSGADGSCAGQNFYTYNDFIN

AANAFSGFGTTGTSDDNKRELAAFFANVAHETGSLRYIEEISKSDYCDSTNTQYPCATGK
QYYGRGPIQLTWNYNYGAAGNYVGFDGLNNPDIVAQDDSISWKTAVWFWMLDSNCHSAIT
SGQGFGATIQAINSGECNGGNTAAVTSRVNYYNNYCGQFGVDPGSDVSR
>Pt|gi|67046839|gb|DR069154|ClassI
MSSILLATTLAIFTAMAIMITGLPSVSAEDCGQQAGGAVCPRGLCCSKWGWCGNTEAHCG
LGCQSQCGGGGSTPTPTPPTPTPTNPTPTPSGQGVASIITEDLFNQLLKYKDDSRCKANG
FYSYAAFIAAANAFPGFGTAGDVTTNKRELAAFLGQTSHETTGGWQTAPDGPYAWGYCFK
EEQGTPPPVYCQASSQWPCVSGKSYYGRGPIQLTWNYNYGPAGKAIGFDGINNPDIVASD
ATVSFKTAIWFWMTAQSPKPSCHDVMTGKWTPSGSDSAAG
>Pt|gi|66737452|gb|DR015082.1|ClassI
IAAANAFPGFGTAGDVTTNXRELAAFLGQTSHETTGGWQTAPDGPYAWGYCFKEEQGSPP
PVYCQASSQWPCVSGKSYYGRGPIQLTWNYNYGPAGKAIGFDGINNPDIVASDATVSFKT
AIWFWMTAQSPKPSCHDVMTGKWTPSGSDSAAGRVPGYGAVTNIINGGLECGKGTVSRQQ
DRIGFYQRYCDILGVSYGANLDCNNQTPFGS
>Pt|gi|67556009|gb|DR096654.1|ClassI
YKDDSRCKANGFYSYAAFIAAANAFPGFGTAGDVTTNKRELAAFLGQTSHETTGGWQSAP
DGPYAWGYCFKEEQGTPPPVYCQASSQWPCVSGKSYYGRGPIQLTWNYNYGPAGKAIGFD
GIGNPDIVASDATVSFKTAIWFWMTAQSPKPSCHDVMTGKWTPSSSDSAAGRVPGYGVVT
NIVNGGLECGKGSDSRQEDRIGFYKRYCDILGVSYGANLDCNNQTPFGS
>Pt|gi|67556254|gb|DR096899.1|ClassI
MTGKWTPSGSDSAAGRVPGYGAVTNIINGGLECGKGTVSRQQDRIGFYQRYCDILGVSYG
ANLDCNNQTPFGS
>Pt|gi|67552206|gb|DR093065.1|ClassI
PSPTPGGQGVASIITESIFNELLKHRNDAGCKASGFYTYSAFIAAANAFPSFGTTGDVAT
RKRELAAFFGQTSHETTGGWATAPDGAYAWGYCFKEEQGNPPAEYCQATSQWPCASGKRY
YGRVPVQLSWNYNYGPAGKAIGFDGINNPDIVASDATVSFKTAIWFWMTAQSPKPSCHDV
MTGKWTPSGSDSAAGRAAGYGAVTNIINGGLECGKGSDSRQQDRIGFYKRYSDILGVSYG
SNLDCNNQRPFGAAVQSEPRLIKTVV
>Pt|gi|67553779|gb|DR094424.1|ClassI
MPPSPTPGGQGVASIITESIFNELLKHRNDAGCKASGFYTYSAFIAAANAFPSFGTTGDV
ATRKRELAAFFGQTSHETTGGWATAPDGAYAWGYCFKEEQGNPPAEYCQATSQWPCASGK
RYYGRGPVQLSWNYNYGPAGKAIGFDGINNPDIVASDATVSFKTAIWFWMTAQSPKPSCH
DVMTGKWTPSGSDTAAGRAAGYGAVTNIINGGLECGKGSDSRQQDRIGFYKRYSDILGVS
YGSNLDCNNQRPFGAAVQSE
>Pt|gi|67560943|gb|DR101588.1|ClassI
SGQGVASIITEDLFNQLLKYIDDSRCKANGFYSYDAFINVANAYPGFGTAGDVTSNKREL
AAFFGQTSHETTGGWPAAPDGPYAWGYCLKEELGTPPPVYCNETAQWPCGSGKSYYGRGP
IQISWNYNYGPAGNAIGFDGINNPDIVASDATVSFKTAIWFWMTAQSPKPSCHDVMTGKW
TPSGSDSAAGRVPGYGAVTNIVNGGLECGKGTDSRQQDRIGFYQRYCDILGVSYGANLDC
NDQTPFGY
>Pt|gi|49446386|gb|CO365069.1|ClassII
MSSLCCYVSAQQGVASIISEDVFNQFLKHRNDAACSAKGFYTYTAFIAAANSFPDFGNNG
DLESRKRELAAFFGQTSQETTGGWATAPDGPYAWGYCFKDQVNSGGDRYHGRGPIQLTGD
YNYKAAGDALGYDLINNPDLVVNDATISFKTAVWFWMTAQSPKPSCHDVILGRWSPSATD
TAAGRVAGYGMVTDIINGGPECGTGTISDVQKGRIGFYQRYCNMLGVDVGSNLDCKNQKP
FGT
>Pt|gi|49448281|gb|CO366964.1|ClassII

MRLLLAITAVAIMSSLCCYVSAQQGVASIISEDVFNQFLKHRNDAACPAKGFYTYTAFIA
AANSFPDFGNNGDLESRKRELAAFFGQTSQETTGGWATAPDGPYAWGYCFKDQVKSGDRY
HGRGPIQLTGDYNYKAAGDALGYDLINNPDLVVNDATISFKTAVWFWMTAQSPKPSCHDV
ILGRWSPSATDTAAGRVAGYGMVTDIINGGPECGTGTISDVQQGRIGFYQRYCNMLGVDV
GSNLDCKNQKPFGT
>Pt|gi|67049355|gb|DR071605.1}ClassII
MSMKLLLALAAVAIMSTLCYVSGQQGVGSIITEDVFNEFLKHRNDAACQARDFYTYSAFI
AATNSFSDFGNNGDLESRKRELAAFFGQTSQETTGGWATAPDGPYAWGYCFKEENSGDKY
HGRGPIQLTGDYNYKAAGDALGYDLINNPDLLVTDATVSFNTAVWFWMTAQAPKPSCHDV
ILGRWSPSNDDTAAGRVPGYGLLTNIINGGVECG
>Pt|gi|67049355|gb|DR071605.1|ClassII
MARKMSMKLLLALAAVAIMSTLCYVSGQQGVGSIITEDVFNEFLKHRNDAACQARDFYTY
SAFIAATNSFSDFGNNGDLESRKRELAAFFGQTSQETTGGWATAPDGPYAWGYCFKEENS
GDKYHGRGPIQLTGDYNYKAAGDALGYDLINNPDLLVTDATVSFKTAVWFWMTAQAPKPS
CHDVILGRWSPSNDDTAAGRVPGYGLLTNIINGGVECGTGTISDRQQGRIGFYQRYCSLL
GVDTGSNLDCQNQKHF
>Pt|gi|34361138|gb|CF402721.1|ClassII
MTAQSPKPSCHDVILGRWSPSATDTAAGRVAGYGMVTDIINGEPECGTGTISDVQQGRIG
FYQRYCNMLGVDVGSNLDCKNQKPFGT
>Pt|gi|67552425|gb|DR093284.1|ClassII
FGFYSYDAFINVANAYPGFGTAGDVTSNKRELAAFFGQTSHETTGGWPAAPDGPYAWGYC
LKEEIGTPPPVYCNETAQWPCGSGKSYYGRGPIQISWNYNYGPAGNALGFDGINNPDIVA
SDATVSFKTAIWFWMTAQSPKPSCHDVMTGKWTPSGSDSAAGRVPGYGVVTNIVNGGLEC
GKGSDSRQEDRIGFYKRYCDILGVSYGANLDCNNQTPFGS
>Pt|gi|66746927|gb|DR024557.1|ClassII
MGDSPRRRVRVGLLFQRGARQSSRRVLPGTSQWPCASGKRYYGRGPVQLSWNYNYGPAGK
AIGFDGINNPDIVASDATVSFKTAIWFWMTAQSPKPSCHDVMTGKWTPSGSDSAAGRAAG
YGAVTNIINGGLECGKGSDSRQQDRIGFYKRYSDILGVSYGSNLDCNN
>Pt|gi|37561417|gb|CF664174.1|ClassIIfragemennt
MVTDIINGGPECGTGTISDVQQGRIGFYQRYCNMLGVDVGSNLDCKNRKPFGTYSVRYQI
ITSTTTTVSYSMLWSAQSLPSTDR
>Pt|gi|49450589|gb|CO369272.1|ClassIV
MGTRTGRSVIASPYQHMVLRMSMVLLLVVVVSVNVVYAQNCGCASGLCCSKWGYCGTTSA
YCGNGCQSGPCYGSGGGSPSGGGGNVGTIISQNFFSGLANAAGSSCEGKGFYTYNAFITA
AKAYSGFGTTGASDNQKRELAAFFANVMHETGGLCYINEINPSSNYCQSSSTWPCASGKS
YHGRGPIQLSWNYNYGAAGQDIGFDGLNNPEKVGQDATISFKTAVWFWMKNSNCHSAITS
GQGFGGTIKAINSGECNGGNSAQVNSRSDYYKKFCSQ
>Pt|gi|49447453|gb|CO366136.1|ClassIV
MGTRTGRSVIGSPYGHLVLRMSMVLLLVVVVSVNVVYAQNCGCASGLCCSKWGYCGTTSA
YCGNGCQSGPCYGSGGGSPSGGGGNVGTIISQNFFNGLANAAGSSCEGKGFYTYNAFITA
ANAYSGFGTTGSNDVQKRELAGFFANVMHETGGLCDINERDPLSNYCQSSSTWPCASGKS
YHGRGPIQLTWNYNYGAAGQDIGFDGLNNPEKVGQDATISFKTAVWFWMKNSNCHSAITS
GQGFGATIKAINGGECNGGHSAQVNSRVNYYKKLCSQLGGGS
>Pt|gi|68026759|gb|DR166981.1|ClassIV
MGTRTGRSVIGSPYGHLVLRMSMVLLLVVVVSVNVVYAQNCGCASGLCCSKWGYCGTTSA
YCGNGCQSGPCYGSGGGSPSGGGGNVGTIISQNFFNGLANAAGSSCEGKGFYTYNAFITA

ANAYSGFGTTGSNDVQKRELAGFFANVMHETGGLCDINERDPLSNYCQSSSTWPCASGKS
YHGRGPIQLTWNYNYGAAGQDIGFDGLNNPEKVGQDATISFKTAVWFWMK
>Pt|gi|67552487|gb|DR093346.1|ClassIV
MVVVVLVLLLVSVSVNAQNCGCASGLCCSKYGYCGTTSAYCGAGCKSGPCSSSGGGSPSG
GGGSVGTIISQSFFNGLAGGAASSCEGKGFYTYNAFIAAANAYSGFGTTGSADVRKRELA
AFFANVMHETGGMCYINERNPPMNYCMSSATWPCASGKSYHGRGPLQLSWNYNYGAAGKS
IGFDGVNNPEKVGQDPTISFKTAVWFWMKNSNCHSAITSGQGFGGTIKAINSQECNGGNS
GEVNSRVNYYKNICSQLGVDPGANVSC
>Pt|gi|67552259|gb|DR093118.1|ClassIV
MESSRGNSVMVLVVVLVLLLAGVSVNAQNCGCASGLCCSKFGYCGTTSAYCGAGCQSGPC
TSSGGGSPSGGGGSVGTIISQSFFNGLAGGAASSCEGKGFYTYDAFIAAANAYSGFGTTG
SADVTKRELAAFFANVMHETGGMCYINERTPPMIYCMSSATWPCASGKSYHGRGPLQLSW
NYNYGAAGKNIGFDGVNNPEKVGQDSTISFKTAVWFWMDNSNCHTAITSGQGFGGTIKAI
NSQECNGGNSGEVNSRVNYYKNICSQLGVDPGANVSC
>Pt|gi|49450305|gb|CO368988.1|ClassIV
MVVVVLVLLLVSVSVNAQNCGCASGLCCSKYGYCGTTSAYCGAGCKSGPCSSSGGGSPSG
GGGSVGTIISQSFFNGLAGGAASSCEGKGFYTYNAFIAAANAYSGFGTTGSADVRKRELA
AFFANVMHETGGMCYINERNPPMNYCMSSATWPCASGKSYHGRGPLQLSWNYNYGAAGQS
IGFDGVNNPEKVGQDPTISFKTAVWFWMKNSNCHSAITSGQGFGGTIKAINSQECNGGNS
GEVNSRVKYYKNICSQLGVDPGANVSC
>Pt|gi|37566495|gb|CF669102.1|ClassIV
MVLVVVLVLLLAGVSVNAQNCGCASGLCCSKFGYCGTTSAYCGDGCQSGPCSSSGGGSPS
GGGGSVGTIISQSFFNGLAGGAASSCEGKGFYTYDAFIAAANAYSGFGTTGSADVTKREL
AAFFANVMHETGGMCYINERTPPMIYCMSSATWPCASGKSYHGRGPLQLSWNYNYGAAGK
NIGFDGVNNPEKVGQDSTISFKTAVWFWMDNSNCHTAITSGRGFGGTIKAINSQECNGGN
SGEVNSRVNYYKNICSQLGVDPGANVSC
>Pt|gi|67048599|gb|DR070857.1|ClassIV
SAYCGNGCQSGPCYGSGGGSPSGGGGNVGTIISQNFFSGLANAAGSSCEGKGFYTYNAFI
TAAKAYSGFGTTGASDNQKRELAAFFANVMHETGGLCYINEINPSSNYCQSSSTWPCASG
KSYHGRGPIQLSWNYNYGAAGQDIGFDGLNNPEKVGQDATISFKTAVWFWMKNSNCHSAI
TSGQGFGGTIKAINSGECNGGNSAQVNSRVNYYKKFCSQLGVDPGTNISC
>Pt|gi|66979070|gb|DR055503.1|ClassIV
MHETGGLCYINEINHSSNYCQSSSTWPCASGKSYHGRGPIELSWNYNYGAAGQDIGFDGL
NNPEKVGQDATISFKTAVWFWMKNSNCHSAITSGQGFGGTIEAINGGECNGGNSGQVNSR
VNYYQNFCSQLGVDPGTNISC
>Pt|gi|5044079|gb|AI725260.1| ClassIV
MHETGGLCSINEINPSSNYCKSSSTWPCASGKSYHGRGPIQLTWNYNYGAAGQDIGFDGL
NNPEKVGQDATISFKTAVWSG
>Pt|gi|34490169|gb|CF472797.1|ClassVII
MATHFTVRAVILWFVFALSALSICRGAVPDIATQDFFNGILSAATDGCAGKTFYTYSDFI
NAANSFSSFGTTGTSDDNKREIAAFFANVAHETTNLCYVEEIVKSAYCDSTNTQYPCASG
QQYYGRGPLQLTGNANYGAAGAYLAVDLRNNPGLVAQDDLTSWKTALWFWNVNSNCHTAI
TSGQGFGATIQAINGAVECNGGNTAEVKDRVSRYTTYCTQLGVDPGSNLTC
>Pt|gi|66976772|gb|DR053205.1|ClassVII
MATHFTVRAVILWFVFALSALSICRGAVSDIATQDFFNGILSAAADGCAKKTFYTYSDFI
NAANSFSSFGTTGTSDDNKREIAAFFANVAHETTNLCYVEEIAKSAYCDSTNTQYPCASG

QQYYGRGPLQLTGNANYGAAGAYLALDLLNNPGLVAQDDLTSWKTALWFWNVNSNCHTAI
TSGQGFGATIQAINGAVE
>Pt|gi|68255783|gb|DR388185.1|ClassVII
LPFPLFHLPRSFPDIATQNFFNGILSAATDGCAGKTFYTYSDFINAANSFSSFGTTGTSD
DNKREIAAFFANVAHETTNLCYVEEIVKSAYCDSTNTQYPCASGQQYYGRGPLQLTGNAN
YGAAGAYLAVDLRNNPGLVAQDDLTSWKTALWFWNVNSNCHTAITSGQGFGATIQAINGA
VECNGGNTAEVNDRVRRYTTYCSQLGVDPGSNLTC
>Pt|gi|37568895|gb|CF671502.1|ClassVII
MATHLRVSAVILWFAFALSALSICRGAVSDIATQDFFNGILSAAADSCAGKTFYAYSDFI
NAANSFSSFGTTGTSDDNKREIAAFFANVAHETGNLCYVEEINKSDYCDSTNTQYPCAAG
KQYYGRGPLQLSWNYNYGAAGDYLGQDLLNNPDTVAQDDLTSWKTALWFWNVNSNCHTAI
TSGQGFGATIQAINGAIECNGGNTDQVNDRISYYTNYCSEFNVDPGSNLSC
>Pt|gi|34488690|gb|CF471318.1|ClassVII
MATHFTVRAVILWFVFALSALSICRGAVSDIATQDFFNGIISSAADSCAGKTFYTHSDFI
NAANSFSSFGTTGTSDDNKREIAAFFANVAHETTDLCYVEEVAKSAFCDSTNTQYPCASG
QQYYGRGPLQLTGNANYGAGGAYLAVDLLNNPGLVAQDDLTSWKTALWFWNVNSNCHTAI
TSGQGFGATIQAINGAVECNGGNTDEVNDRVSHYTTYCSQLG
>Pt|gi|34487991|gb|CF470619.1|ClassVII
AVSDIATQDFFNGILSAAADGCAGKTFYTYSDFINAANSFSSFGTTGTSDDNKREIAAFF
ANVAHETTNLCYVEEIAKSAYCDSTNTQYPCASGQQYYGRGPLQLTGNANYGAAGAYLAV
DLLNNPGLVAQDDLTSWKTALWFWNVNSNCHTAITSGQGFGATIQAINGAVECNGGNTDE
VNDRVSRYTTYCSQLGVDPGSNLTC
>gi|74168161|gb|DT638089.1|ClassVII
MATHLRSVLILWLALALYVVSICRGDVGDIVTRDFFNGILSGAADSCAGKNFYTYDDFIT
AANAFSGFGTTGTSDDNKRELAAFFANVAHETGSLCYIEEINKSDYCDSTNTQYPCVAGK
QYYGRGPLQLTWNYNYGAAGNYLGFDGLNNPDIVAQDDSISWKTAVWFWMLDSNCHSAIT
SGQGFGATIQAINSGECNGGNSAAVTDRVNYYNNYCGQFGVDPGSNVSC
>Pt|gi|51495190|gb|CV032367.1|ClassVII
MATHLRSVLILWLALALYVVSICRGDVGDIVTQDFFNGILSGAADSCAGKNFYTYDDFIT
AANAFSGFGTTGTSDDNKRELAAFFANVAHETGSLCYIEEINKSDFCDSANTQYPCATGK
QYYGRGPLQLTWNFNYGPAGNYVGFDGLNNPDIVAQDDSISWKTAVWFWMLDSNCHSAIT
SGQGFGATIQAINSGECNGGNSPAVTDR
>Pt|gi|48943937|gb|CO171065.1|ClassVII
MRNYLRVMTLSAIILWLVLAFDLVSMCRGDVGDFATQDFFNGILSGASDSCTGKTFYTYN
NFMDAATAFSGFGTTGPDVDHKREIAAFFANVAHETSRLCYVEQIEKSDYCDSTNQKYQV
CGGKAVLWAWPLATHMEL
>Pt|gi|67047413|gb|DR069716.1|GH18
MFFSEIIDFISQTNATEVFDKETVSAYAYSSDMMWVGYDNKQSVAAKVSFAKNKHLLGYF
FWAISQDSNWMLSAQASESWNQAR
>Pt|gi|67960181|gb|DR162791.1|GH18
MFFSEIIDFISQTNATEVFDKETVSAYAYSSDMMWVGYDNKQSVAAKVSFAKNKHLLGYF
FWAISQDSNWMLSAQASESWNQAR
>Pt|gi|68252684|gb|DR385086.1|GH18
MCYDYHGSWDISATGAHAALYDPTSNISTSFGIGSWLHSGVPPNKVAMGMPLYGRSWILK
SLDDTEIGAPAVAAGPKQTLSNEKGVMFFSEIRELINQKNATEVFDKETVSAYSYSSDLL
WVGYDNQESVATKVSFAKEMHLLGYFFWAIGQDNNWMLSAQASDSWN
>Pt|gi|67557727|gb|DR098372.1|GH18

MGMPLYGRSWILKSLDETGIGAPAIAAGPKQTLSSETGVMFFSEIIDFISQKNATEVFDK
ETVSAYAYSSDMMWVGYDNKQSVAAKVSFAKNKHLLGYFFWAISQDSNWMLSAQASESWN
QAR
>Pt|gi|66746777|gb|DR024407.1|ClassV
MASNTSYRKVFINSTIALARKYGFHGLDLDWEFPQDAAEMEYLGTLFDDWRHAIDLEATA
SGRPPLLLTAAVYFAQYFFVWGEKRAYPVTSIARNLDWVNVMCYDYHGSWDISATGAHAA
LYDPTSNISTSFGIGSWLHSGVPPNKVAMGMPLYGRSW
>Pt|gi|58026595|gb|CX713336.1|ClassV
MAQAPKNLCIILFFITSSIIYSSSSSNNTTSNNGVKAAYWPSDSSLSPSSISTDLFTHIL
YAFADLNDQTFQVQLPATADPAEFTSTLLQKNPSLKTLISIGGGGSNATAFALMASNTSY
RKVFINSTIALARKYGFHGLDLDWEFPQDAAEMEYLGTLFDDWRHAIDLEATASGRPPLL
LTAAVYFAQYFFVWGEKRAYPVTSIARNLDWVNVMCYDYHGSWDISATGAHAALYDPTSN
ISTSFGIGSWLHSGVPPNKVAMGMPLYGRSWILKS
>gi|66983699|gb|DR060132.1|ClassV
GSGPKKLLLILLFITCSLIYSSSSPLVDSINNGGGVKAAYWPSHRVSYSPPSSINTDLFT
HIFYAFAHLNNQTFEVEVPSTDDPAEFTSTLLMKKPSLKTLLSIGGGGSSPTAFALMASN
ASDRKVFINSTIAIARKYGFHGLDLDWEFPKNSVEMENLGALFSEWRHAIDLEATASGRP
RILLTAAVYFAQYFFLSSEKRAYPATSIAQNLDWVNVMCYDYHGSWNISATGAHAALYDP
SSNISTSFGIGSWLHSGVSPNNIVMGMPLYGRSWILK
>Pt|gi|67559326|gb|DR099971.1|ClassV
GGVKAAYWPSYTISYSPPSSINAALFTHIFYAFANLNDQTFQVELPATAKPAEFTLTLRK
KNPFLKTLISIGGGGSNSTAFALMASNALRRKVFINSAIALARKYGFHGLDLDWEFPKDA
TEMQNLGSLFAEWRHAIELEARAWPRRPRLLLTAAVYFAQYFLLAADKRAYPATSITQNL
DWVNVMCFDYHGSWDTSATGAHAALYDPSSNISNSFGIGSWLHSGV
>Pt|gi|49008134|gb|CO196959.1|ClassV
MAQPQKYLLIILLSMATSLVYCSSSQPAYHRGHHPHKKHGGVNAAYWPSYTVSFSPPSSI
NAALFTHIFYAFADLNDQTFKVELPATTDPAADPAEFTPTLRKKNPSIKTLISIGGGGSN
ATAFALMASNASHRKVFINSAIAVARKYGFHGLDLDWEFPQNVTEMKNLGILFAEWRHAI
ELEAKASRRPRLLLTAAVYFAEYFLLAADQKAYPATSIARNLDWVNAMCFDYHGSWDISA
TGAHAALYDPS
>Nt_gi|19861-I
MRLCKFTALSSLLFSLLLLSASAEQCGSQAGGARCPSGLCCSKFGWCGNTNDYCGPGNCQ
SQCPGGPTPTPPTPPGGGDLGSIISSSMFDQMLKHRNDNACQGKGFYSYNAFINAARSFP
GFGTSGDTTARKREIAAFFAQTSHETTGGWATAPDGPYAWGYCWLREQGSPGDYCTPSGQ
WPCAPGRKYFGRGPIQISHNYNYGPCGRAIGVDLLNNPDLVATDPVISFKSALWFWMTPQ
SPKPSCHDVIIGRWQPSAGDRAANRLPGFGVITNIINGGLECGRGTDSRVQDRIGFYRRY
CSILGVSPGDNLDCGNQRSFGNGLLVDTM
>Pg_GQ03904_P01-I
FIEAQKRRRCKAKGFYSYAAFIAAADAFPGFGTTGDLTAQKRELAAFFGQTSHETTGGWP
TAPDGPYAWGYCLKEEQGDPPGEYCQASSEWPCASGKKYYGRGPVQISWNYNYGPAGKAV
GFDGINNPDIIANDATVSFKTAVWFWMTEQSPKPSCHNVMAGGWGPSGSDTAAGRAAGYG
VVTNIINGGLECGKGSDSRQEDRIGFYKRYCDILGVSYGSNLDCNSQKPFGFAAQSQPRL
IKTVV

>Pg_GQ0183_A06-I
LTSVREMKSIRFSAMAIALVTMGTMNLYFASAEQCGRQASGALCPGGLCCSKWGWCGNTE
AHCGQDCQSQCGGSTPTPPSPTPGGQGVSSVITESIFNDLLKHRNDAACKAKGFYSYAAF

IAAANAFPAFGTTGDLSTRKRELAAFFGQTSHETTGGWPAAPDGPYAWGYCFKEEQGNPP
GEYCQASSQWPCASGKRYYGRGPVQISWNYNYGPAGRAIGFDGINNPDIVANDATVSFKT
AMWFWMTAQSPKPSCHDVMTGRWSPSGRDSAPGRAAGYGVVTNIFNGGLECGKGSDSRQV
DRIGFYKRYCDIFGVRYGSNLECNTQKPFGFAGQSQPHLIKTVV

>Pg_GQ03902_B01-II
RGTRICIGSLKMASATIGRMKSMRVLSVLAAVAMMGTLCCHVSAQQGVASIISEDVFNQF
LKHRNDDACPAKGFYTYSAFIAAANSFPDFGNNGDLETSKRELAAFFGQTSQETTGGWAT
APDGPYAWGYCFKEENSADRYHGRGPIQLTGDYNYKAAGDALGYDLINNPELVVTDATVS
FKTALWFWMTPQSPKPSCHDVILGRWSPSDTDTAAGRVPGYGMVTNIINGGLECGQGTSS
AAQQGRIGFYQTYCNKLGVDSGSNLDCNNQKPFGS

>Pg_GQ03703_O09-II
MRVLLALTVVAMMNTFCYVSAQQGVAFIITEDVFNRFLKHRNDGACAAKGFYTYNSFILA
ANRFPSFGNVGNLETRKRELAAFFAQTSHETTGGWETAPDGPYAWGYCFKEEQDADPEEM
YYGRGPIQLTGKSNYEAAGKALGYDLLNNPGIVASNPTISFKTAVWFWMTTQSPKPSPHT
VMIGKWSPSGSDTAAGRVPGYGVVTNIINGGVECGRGSDSNQEDRIGFYKAYCDILGASY
GSNVDCNTQRSFGG

>Pg_GQ0046_J03-II
METKQIIGSVNRMLVVVGICLMISSWLCCIEPAAAADEDQAMKSKKIACIKGAECKNKTI
SELFTVDQFESLFSHRNAPMAHAQGFWDYHSFITAAAHFEPKGFGATGGDLVQKKELAAF
FAHVATETSCESLMAQSSTATTDSPTKWGLCYKEELSPDSTYCESSLVYPCAPGVSYHGR
GALPVYWNYNYGQLGQALKVDLLHHAEDLSQNATLAFQAAMWRWMNPIKVKQPSAHQVMV
GKWVPTKNDTNSLRHPGFGMTINILKGDAECGAGSDDKQMNKRIAHYLYFLDQLDVGRDN
AGDNLDCSDQKVLNPSSAST

>Pg_GQ0207_E09-VII
TILANMAIHLRVMTASVFVLWLALALSICRGDVGDIATEDFFNGILSGAADTCAGKKFYT
YSDFITAANGFSGFGTTGTSDDEKREIAAFFANVAHETIGFCYIEEIQKDDYCDSNNTQY
PCAAGKQYYGRGPIQLSWNFNYGPAGSYLGVDLLNNPETVAQDALIAWKTALWYWNVNSD
CHSAITSGQGFGATIQAINGALECNGGNTDEVNDRVSYYTKYCSQFGVDPGSNLSC

>Pg_GQ03803_J20-IV
MAGSSGKFASPRGRVVVRMSLVLLLVVGVSVNVVNAQNCGCASGLCCSKWGYCGTTSAYC
GNGCQSGPCSGGGSPSGGGGNVGTIISQNFFNSLASGAGSSCKGKGFYTYNAFIAAANAY
SGFGTTGSNEVQKRELAAFFANVMHETGGLCYINEISPSSNYCQSSSTWPCTSGKSYHGR
GPLQISWNYNYGAAGQSIGFDGLNNPEKVGQDATISFKTAVWFWMKNSNCHSAITSGQGF
GGTIKAINSGECNGGNSGQVNSRVTYYKKFCSQLGVDTGTNVSC

>Pg_GQ03206_D15-IV
MGSSSSNKSVMALVLVLLLVGVSVNAQNCGCASGVCCSQYGYCGTTSAYCGKGCKSGPCY
SSGGGSPSAGGGSVGGIISQSFFNGLAGGAGSSCEGKGFYTYNAFIAAANAFSGFGTTGS
NDVKKRELAAFFANVMHETGGLCYINEINPKIIYCQSSSTWPCTSGKSYHGRGPLQLSWN
YNYGAAGKSIGFDGLNNPEKVGQDSTISFKTAVWFWMKNSNCHSAITSGQGFGGTIKAIN
NMECNGGNSGEVSSRVNYYKKICSQLGVDPGVNVSC

>At_gi|7288020-IV
MLTPTISKSISLVTILLVLQAFSNTTKAQNCGCSSELCCSQFGFCGNTSDYCGVGCQQGP
CFAPPPANGVSVAEIVTQEFFNGIISQAASSCAGNRFYSRGAFLEALDSYSRFGRVGSTD
DSRREIAAFFAHVTHETGHFCYIEEIDGASKDYCDENATQYPCNPNKGYYGRGPIQLSWN
FNYGPAGTAIGFDGLNAPETVATDPVISFKTALWYWTNRVQPVISQGFGATIRAINGALE
CDGANTATVQARVRYYTDYCRQLGVDPGNNLTC

>At_gi|2342435-III
MTNMTLRKHVIYFLFFISCSLSKPSDASRGGIAIYWGQNGNEGNLSATCATGRYAYVNVA
FLVKFGNGQTPELNLAGHCNPAANTCTHFGSQVKDCQSRGIKVMLSLGGGIGNYSIGSRE
DAKVIADYLWNNFLGGKSSSRPLGDAVLDGIDFNIELGSPQHWDDLARTLSKFSHRGRKI
YLTGAPQCPFPDRLMGSALNTKRFDYVWIQFYNNPPCSYSSGNTQNLFDSWNKWTTSIAA
QKFFLGLPAAPEAAGSGYIPPDVLTSQILPTLKKSRKYGGVMLWSKFWDDKNGYSSSILA
SV

>At_gi|3859599-II
MEKQISLLLCLLLFIFSISSSLHETEARKHNKYKPAPIMSLVPRTLYDQIFIHKDNNACP
AKGFYPYEAFVEATRSFPKFGSVGNFWTRRREVAAFLAQISHETTGGWATAPDGPYAWGL
CFKEEVSPQSNYCDASNKDWPCVSGKSYKGRGPIQLSWNYNYGQAGRALGFDGLQNPELV
ANNSVLAFKTALWFWMTEQTPKPSCHNVMVNRYRPTKADRAANRTVGYGLVTNIINGGLE
CGIPGDGRVTDRVGYFQRYAQLFKVTTGPNLDCENQRPFS

>At_gi|12321966-I
MKTNLFLFLIFSLLLSLSSAEQCGRQAGGALCPNGLCCSEFGWCGNTEPYCKQPGCQSQC
TPGGTPPGPTGDLSGIISSSQFDDMLKHRNDAACPARGFYTYNAFITAAKSFPGFGTTGD
TATRKKEVAAFFGQTSHETTGGWATAPDGPYSWGYCFKQEQNPASDYCEPSATWPCASGK
RYYGRGPMQLSWNYNYGLCGRAIGVDLLNNPDLVANDAVIAFKAAIWFWMTAQPPKPSCH
AVIAGQWQPSDADRAAGRLPGYGVITNIINGGLECGRGQDGRVADRIGFYQRYCNIFGVN
PGGNLDCYNQRSFVNGLLEAAI

>At_gi|3250684-V
MYTEVVKASYWFPDGASSPTTGSVVPQSSAVLIDSTLFTHLFCAFADLDPQTNSVVVSGA
HEQEFSNFTKIVKKKNPHVQTLLSIGGRNADKSAFASMASNPTSRKSFIWSAISSARYYR
FDGLDLVWKYPKDDVEMRNFGQLLEQWREAIEDDAERTERMPLLLTAAVYYSPVYDSVSY
PIREIKKKLDWVNLIAYDFYSSSTTIGPPAALFDPSNPKGPCGDYGLKEWIKAGLPAKKA
VLGFPYVGWTWSLGSGNDAATSRVATSAEGSINYDQIKRLIVDHKARPVFDSTVVGDYCF
AGTSLIGYDDHQSVVAKVKYAKQKGLLGYFSWHVGADDNFGLSRAGSSFCTYESLFFSLF
TFS

>PgeChia7-2AEF59004
MATHFRVNVIFLWLAFALSALSICRGAVSDIATQDFFNGILSAATDGCAGKTFYTYTDFI
NAAISFSSFGTTGTSDDNKREIAAFFANVAHETTNLCYVEEIAKSDYCSSNTQYPCASGQ
QYYGRGPLQLTGNGNYGAAGDYLGVDLLNNPGLVAQDDLTSWKTALWFWNVNSNCHTAIT
SGQGFGATIQAINGAIECNGGNTDQVNDRISRYTNYCSQFGVDPGSNLSC

>PgeChia7-1AEF59003

MATHFRVNVIFLWLAFALSALSICRGAVSDIATQDFFNGILSAATDGCAGKTFYTYTDFI
NAANSFSSFGTTGTSDDNKREIAAFFANVAHETTNLCYVEEIDKSDYCDSSNTQYPCASG
QQYYGRGPLQLTGNANYGAAGTYLSADLLNNPGLVAQDDLTSWKTALWFWNVNSNCHTAI
TSGQGFGATIQAINGAIECNGGNPDEVNDRISHYTNYCSQFGVDPGSNLSC

>Pm_gi|261889440-II
MGRTGGEKWVMALVLVLLLVGVGVNAQNCGCASGLCCSKFGYCGTTSAYCGTGCQSGPCS
SSGGGSPSGGGGSVGTIISESVFNGLAGGAASSCEGKGFYTYNAFIAAASAYSGFGTTGS
SDVQKRELAAFFANVMHESGGLCYINEINPPIIYCQSSSTWPCTSGKSYHGRGPLQLSWN
YNYGAAGQSIGFDGLNNPEKVGQDATISFKTAVVFWMKNSNCHSAITGGQGFGATIKAIN
SGECNGGNSGEVSSRVNYYKKICSQLGVDPGANVSC

>Pm_gi|261889446-IV
MGSSTAAENWVMSLVVALLLAGVSVNARKCGCASGLCCSQYGYCGTTSAYCGKGCQSGPC
TSSGGGSPSGGGGSVGTIISQSIFNGLAGGAASSCEGKGFYTYTAFIKAASAYSGFGTTG
SNDVKKRELAAFFANVMHETGGLCYINERNPPIIYCQSSSTWPCASGKSYHGRGPLQLSW
NYNYGAAGKNIGFDGLNNPEKVGQDATISFKTAVWFWMNNSNCHSAITGGQGFGATIKAI
NSGECNGGNSGEVSSRVNYYRKICSQLGVDPGANVSC

>Pa_gi|33359621-IV
MGSIIIDKSVMALVLVLLLVGVSVNAQNCGCATGVCCSQYGYCGTTSAYCGKGCKSGPCY
SSGGGSPSAGGGSVGGIISQSFFNGLAGGAASSCEGKGFYTYNAFIAAANAYSGFGTTGS
NDVKKRELAAFFANVMHETGGLCYINEKNPPINYCQSSSTWPCTSGKSYHGRGPLQLSWN
YNYGAAGKSIGFDGLNNPEKVGQDSTISFKTAVWFWMKNSNCHSAITSGQGFGGTIKAIN
SMECNGGNSGEVSSRVNYYKKICSQLGVDPGANVSC

6.3 Appendix

Table 6.2.1: Primer sequences used to amiplify chitinase (*chia1-1, chia2-1, chia4-1*) genes for Sanger sequencing. Primers FWD1 and REV1 PCR amplify and seqeunce outer gene sequences and, primers FWD2 and REV2 use to PCR amplify and sequence inner sequences.

Primers	***chia1-1***	***chia2-1***	***chia4-1***
FWD1	GAGGGATAATTTGGAAACGCTAAA	GCTGTGCCTACCCTTAGATTT	AGACAGATTTAACACGAAACATTGG
FWD 2	GTCGTCATCTCCATGTTCGTAG	CAGCTTTATAGTTGTAGTCCCTATCA	CAGTAGTTCATTGGAGGGTTTCT
REV 1	ATAGAAACGACCGACCATGAC	ACGGTAATGTAGCGTAGTCT	ATCTCTGTGAGTGGGAGTCTTA
REV 1	CTCAGATTTATGAGAAGTGTTCAAACC	TAGCTAGACGGTAGACTCTGTG	GAAGCAGGAGTCCTCATTGTTA

6.4 Appendix

Chitinase nucleotide sequences representing the uptstream promoter and coding regions of *Pcchia1-1, Pbchia1-1, Pcchia2-1, Pbchia2-1, Pcchia4-1* and *Pbchia4-1* genes cloned from lodgepole (Pc) and jack pine (Pb) gDNA.

>Pbchia1-1
GAGGGATAATTTGGAAACGCTAAAAGTGACACGTGATTTAAGTTTTAATAAGTGATTTAAGCAGA
GATTAGTTGTAATGTTATTTCTTATTGGGAAATTGGAAGGGCATTTATTTTTTTACCAATTTTTG
GCACATTATTCGAAATGAGATTTTTATGAAGACAAAATGATTTAAGTAGTGGTATTATCTAAATA
AACACAGTTAATACACATTTCGTTTAAAAATAAAAAATAAAGCCATTACATATTATGAAGTTAAA
CTAATAAACTACTATTCAAATATAAAATAAAATTAATAAACTATTATTTAAATATAATAACCTAG
AACTACTTCAAGCTCAAGGATCTATATAAAACAGTAATCAAATTTCCACTTTGAGAAAAAAGGGA
TGAATTTGAAATACCAGGCTTAGTTAGACTACGACTTTGGGGGCAAAGTCGTCCATGTGTCGGAT
GCGGACATCGAGGACCACATGGGCGACTGGACTGTCGTCATCTCCATGTAATCATTTTGGTCTGG
GCGGCGGCCTCATTAATGACAATTAAGAACATGTCATAATTATTTTAATTTCAACTTTATCCATG
AAAGGTCTTTATGGTGTTAAGAGATAAGTCTCTCATGATCACAAATAAAATATAACCTGGGTTTC
ATTGAATTGAAAATTTGGCCGTCAGAGTCAAGATGAGCGAACACGGAAGTGGATAATATTTCATA
TACCTTGAAGAGACGAGTGAACACGGAAGCGGATAATATTTCATATAGCTTCGACAGACGTGAAC
GTGAAGGAGACCGGAAATACGGCTATAAAAGACGGTCTGCGGTCTCCAAAATTTAACTGAAAGTT
GCAATTAATGCAGTGCTGATATAAGAAACGAGTGTCTGCAATATGAAATCGATGAAGTTTTCCGC
GATGGCGATCGCCTTGATTACAATGGCGACGATGAATTTGTATTTTGTATCTGCTGAGCAATGCG
GACAGCAAGCAGGCGGAGCTCTTTGTCCCGGTGGCTTGTGCTGCAGCAAATGGGGATGGTGCGGA
CGCCCATTGCGGGCAGGATTGCCAGAGCCAATGCGGAGGATCGACTCCCACGCCTCCTTCACCCA
CTCCCGGTGGACAGGGAGTTGCATCTATCATCACTGTAAGTATTTTCAATGAGTTATTGAAGCAT
AGAAACGACGCCGGCTGCAAGGCCAGCGGATTCTACACCTACTCTGCCTTCATTGCAGCTGCCAA
TGCTTTTCCTTCCTTCGGCACCACCGGCGATGTCGCTACTCGGAAAAGAGAGCTCGCTGCTTTCT
TTGGTCAAACCTCCCACGAAACCACAGGTCATATTTTCATCTCTATTACTTGTGATTTACCGCTG
TGATCTTACTTGGCTTTACATCTCATTGGGTAAGAAATATTGGTAAACAGGAGGATGGGCAACAG
CCCCAGACGGCGCGTACGCGTGGGGTTATTGTTTGAAAGAGGAGCAAGGCAATCCTGCCGCCGAG
TCCTGCCAGGCAACCTCCCAGTGGCCCTGTGCATCTGGAAAGAGATACTACGGACGAGGGCCCGT
TCAATTGTCATGGTCGGTCGTTTCTATATTTATGCAGTTTTAATTTCTTTACTTATTTTCTGCAC
AGACGCACCCAAAAAGAATATCTGAATGAAATTTAAGACTAATACGCCGTTTNNTCAATTTGCTT
TGGGTTTTTCAGGAATTACAACTATGGACCGGCCGGGAAGGCAATCGGATTCGATGGCATAAACA
ACCCCGACATTGTTGCTAGCGATGCCACGGTGTCTTTCAAGACCGCAATNTGGTTCTGGATGACC
GCCCAATCTCCGAAACCTTCTTGCCACGATGTCATGACCGGGAAATGGACTCCGTCCGGCAGCGA
CAGCGCCGCTGGGAGAGCTGCGGGATATGGAGCAGTTACCAACATCATCAACGGCGGGCTGGAGT
GCGGGAAAGGCAGTGACTCGAGGCAGCAGGATCGCATCGGCTTCTACAAAAGATACAGTGACATT
CTTGGGGTGAGCTACGGATCAAACNTGGATTGCAACAACCAGAGGCCTTTCGGCGCTGCAGTTCA
ATCTGAACCTCGTCTTATCAAAACCGTGGTTTGAACACTTCTCATAAATCTGAGA
>Pbchia2-1
TGCTGTGCCTACCCTTAGATTTAAGGCGTATATTATCTAGCGATGGATTTAAGCTGTACATTATC
CAACGATAGATATATACTGCCCATCTTATTATTAATGGCTAGATAAACTTGTGAATATGCGTTGT
CCACCTAATTTATCAAATCATTGAGATATAATAGATGCGAGCCGCAAGCTCCACCATAAAGTTAC
CTTTAAATCTTTGCTACTTAGAATTGATCTTGGAGATATACCAGAACAATGAATAATGAAATATA

CAGTGAGAGGTAAAGATTTACACCACTGTATCATCCCATTAAAGACAATGCAGGCTATCTATTTG
TTGCTTGACCGCATTGTGCCGCGACAATGAGCACCTCTTGCAACTTGAATGGAATGAACTTCCTC
TCCGTGTGTATGGTATCTTTACATAACCGTAATGTAGCGTAGTATGATATTCTTACATAACGGTA
ATGTAGCGTAGTCTAAACGTTTAAAAAGCCAGGAGTTTATTCCCCTAAAAACAATGAAGTTGCCC
TAAGTTTGAAAAGACCAGAATGGCCGTTAGAGGGGCGGACTCCTCAAACCGCTGAGTTGTTAAAA
TCATAGGATCTTAGTATACCCGACTTACAGGTGGTTAGATATACTCTTGATATTTTAGATCAAAT
ATGAGATTTGATCCACGGTCTCCTGGATTAAGCTTGATAGTTATGGTTTCACCATATAATTAAAA
AAAAAAAATTGCTTTCCAAAAAGATTANCTTCCATATTGTGGCATGTGTGGAAAGGCCAAAGCCA
AAGTCTGCCCTTCAAAGGAAGAGTGGAAAATACTCCACGCGAGGAAGAAGCGAAGAAAGATCTGA
AGTCAAGATTATGAAATACGGGAGAGAAGCGGCTATAGTCGCCACGGAAGTCGTTTTACTTTTAA
CCTTAACCCATCTTACACGGTATGTACCTGCTTTCGCAAAAATGGTGGCCGGGTTGCAGGTTGGA
CCATGAGAACACTATATCACCGTGAACGCCACGGAATGGTACAAAAGTTTCAACAAAGCGCCGCC
CCCCACCTCCCATGGCACTGTATCTGACTCTTGCGGAAATACTGCTATAAAAGGAGGGGCTTGCA
GCCTAGGATCATCACCAGAATTTGCTTTGGCAGTTTAAAGACAGCATATACGACTATGGCGAGAA
AGATGTCGATGAANTTGNTGTTGGCCCTCGCTGCAGTGGCCATAATGAGTACTTTGTGTTATGTT
TCTGGACAACAGGGAGTCGGCTCCATCATAACTGAAGATGTTTTCAATGAGTTTCTCAAGCACCG
AAATGACGCCGCATGCCAGGCGAGAGACTTCTACACCTACAGCGCCTTCATTGCGGCCACTAATA
GCTTCTCAGACTTCGGCAACAACGGCGATCTAGAGAGCCGCAAAAGAGAGCTCGCGGCTTTCTTT
GGTCAAACGTCGCAGGAAACCACAGGTATTATTAGTTTAGCCTCCTCTAACTCTTCTGTCTCTCT
GCTATTCCTTATTGTTATTAATGGCATTAAGTTAATTGAGTTTGTACAGGCGGGTGGGCCACGGC
CCCCAGACGGCCCCATATGCGTGGGGTTACTGCTTCAAAGAGGAAAATAGCACAGACAAATACCA
CGGACGGGGACCTATTCAGTTAACAGGGTAATATTTCTATATCTAAGTTTTGTTAATCCATTTCG
AATTAATTGTAGCAATAGATATGGANAAAATCGAATGAATTTCAAGCNTAATTCACTTACCGCTG
TTTGATTAGGGACTACAACTATAAAGCTGCGGGCGATGCCTTAGGGTACGATCTCATAAACAATC
CGGATCTCTTGGTAACTGACGCCACGGTCTCGTTTAAGACGGCGGTGTGGTTCTGGATGACCGCG
CAGGCTCCCAAGCCTTCGTGCCACGACGTGATATTGGGAAGATGGAGTCCATCAAACGACGATAC
TGCCGCCGGCAGAGTGCCGGGATNNNGATTGTTGACGAACATCATTAACGGCGGGAGGGAATGCG
GCACAGGCACAATAAGCGACCCTCAGAAGGGCCGAATCGGGTTCTACCAGAGATACTGCAGCTTG
CTGGGCGTGGACACTGGATCCAACCTCGACTGCCAAAACCAGAAGCATTTCTGAACTCAAATCTC
TACTGTAGCAAGAGGCCACCACATCGATTACTACTTCGAGTTGAATGTTATGGTCAGCACAGAGT
CTACCGTCTAGCTAA
>Pbchia4-1
TAAAATGGAAAAAGAAATGAAAAAGAAAAGAAGAAGCAGGAGTCCTCATTGTTAAACAATTACGC
ATGTTAAATACACATTTCTGCTACGAAAGTCATGGTATCCAAGTGGCATTTTTATCTCGTGACTT
CTCACAACCACTAAATCAATGAAAGTCGGTGAAACATTATCTTTATCGTGGAACACTGGTATGGT
GCAGCTCTGTGCAAATGGGTGCGTTCAACCATCCTCAGTGTTCCGCCGCGTGTTAAGCTTTTAGC
AATTAATCGAAATCTATTCTAGGGATGGTATACGAGCGTGGAAGACCGGATGGAGCAGGTAAGTC
GAAGGCGGCGCACTATCAAATAAAACTCGTTACGTTCAACGATCCGGACCGCAGCCGCCGCGTAT
TCAGCGGATGCAATTGTTCCTAAAAATCAAAGTCTATAATAGTCTTGCCGACCCAAAGAAGAAAA
TGAGGAGCAGAGTGGGTCGGGGTTTCTATATAAATAGAGAACGAACTTGAATCAACAAATCAAAT
AATCTCTGTGAGTGGGAGTCTTAAGGAATTATTCGCAATGGGGGGTAGTAGCGGTAAGTCGGTAA
TGGTGCTGGTGATGGTGCTAGTCCTGCTACTGGCGGGTGTCAGTGTTAATGCTCAAAACTGTGGC
TGTGCCAGCGGACTGTGTTGCAGCAAGTATGGATACTGCGGGAGCACCCCGGCTTACTGCGGCCG
TGGCTGCAGGAGCGGTCCTTGTAGCAGTTCAGGAGGAGGATCTCCAAGTGGCGGAGGTGGAAGCG
GTGGGTACCATAATTTCCCAGAGTTTCTTCAATGGCCTCGCGGGCGGAGCTGCCAGCTCCTGCGA
GGCCAAGGGATTCTACACTTACGATGCTTTCATCGCAGCGGCCAATGCGTATTATGGATTTGGCA
CCACGGGATCCGCAGACGTCACAAAGAGAGAACTCGCGGCCTTCTTGGCTAATGTTATGCACGAA

ACTGGAGGTCTATATTCTGCTCTATCATAATTATTTACCTAGTAATCCTTTTTCTTTATGAGAAA
AAGCTTTAATTGCCATTTACCTTCCAATCGATCTCTAATTTGGTTTCAATTTTTTCACTGATAAT
TGATTAATTTCTGTGTCATGTCAGGAATGTGCTACATCAATGAGAGAAACCCTCCAATGAACTAC
TGTATGAGTTCAGCCACGTGGCCGTGTGCGAGTGGTAAGAGCTACCATGGTCGTGGCCCTCTCCA
GTTGAGCTGGTATGTATGCGAAAACTCTACACCCACTAATCACATCTCATCTCCTCTGTAATGAT
AAAATAGATTGAGACTAACAATTAGATTGTTACAGGAATTACAATTATGGAGCGGCTGGGAAGAG
CATTGGGTTCGACGGGGTGAACAACCCAGAGAAGGTGGGGCAAGACCCCACCATTTCGTTCAAGA
CGGCTGTCTGGTTCTGGATGAAGAACAGCAACTGTCACTCCGCCATAACGTCAGGGCAGGGCTTC
GGTGGCACAATCAAAGCCATCAACAGCCAGGAATGCAACGGTGGAAACAGTGGGGAGGTGAATAG
CAGAGTGAAGTACTACAAGAATATCTGTAGCCAGTTGGGTGTGGATCCTGGAGCCAATGTTTCGT
GTTAAATCTGTC

>Pcchia1-1
GAGGGATAATTTGGAAACGCTAAAAGTGACACGTGATTTAAGTTTTAATAAGTGATTTAAGCAGA
GATTAGTTGTAATGTTATTTCTTATTGGTAAATTGGAAGGGCATTTATTTTTTTACTAGTTTTTG
GCACATTATTCGAAGATTTTTATGAAGACAAAATGATTTAAGTAGTGATATTATCTGAATAAACA
CAGTTAATACACATTTCATTTAAAAATAAAAAATAAAGTCATTCCATATTATGAAGTTAAACCAA
TAAACTATTATTCAAATATAAAATAAAATTAATGAACTATTATTTAAATATAATAACCTAGACTA
CTTCAAGCTCAAGGATCTATATAAAACAGTAATCAAATTTCCACTTTGAGAAAAAAGGGATGAAT
TTGAAACTCCNGGCTTCGTTTACAGTTAGACTACGACTTTGGGGCCAAAGTCGTCCATGTGTCGG
ATGCGGACATCGAGGACCACATGGGACTGGACTGTCGTCATCTCCATGTTCGTAGAAATCATTTT
GGTCTGGGCGGCGGCCTCATTAATGACAATTAAGAACATGTCATAATTATTTTAATTTCAACTTT
ATCCATTAAAGGTCTTTATGGTGTTAAGTGATAAGTCTCTCATGATCACAAATAAAATATAACCT
GGGTTTCATTGAATTGAAAATTTGGCCGTCAGAGTCAAGATGAGCGAACACGGAAGTGGATAATA
TTTCATATAGGTTGAAGAGACGAGTGAACACGGAAGCGGATAATATTTCATATAGCTTCAAGAGA
CGTGAACGTGAAGGACACCAGAAATATGGCTATAAAAGACGACCTACGGTCTCCAAAATTTAACT
GAAAGTTGCAATTAATGCAGTGCTGATATAGAAACGAGTGTCTGCAATATGAAATCGATGAAGTT
CTCCGCGATGGCGATCGCCTTGCTTACAATGGCGACGATGAATTTGTATTTTGTATCTGCTGAGC
AATGCGGACAGCAAGCAGGCGGAGCTCTTTGTCCCGGTGGGTTGTGCTGCAGCAAATGGGGATGG
TGTGGCAACACGGACGCCCATTGCGGGCAGGATTGCCAGAGCCAATGCAGCGGATCGACTCCCAC
TCCCGGTGGACAGGGAGTTGCATCTATCATCAGTGAAAGCATTTTCAATGAGTTATTGAAGCACA
GAAACGACGCCGGTTGCAAGGCCAGCGGATTCTACACCTACTCTGCCTTCATTGCAGCTGCCAAT
GCTTTTCCCTCCTTCGGCACCACCGGCGATGTCGCTACTCGGAAAAGAGAGCTCGCTGCTTTCTT
TGGCCAAACCTCCCACGAAACCACAGGTCATATTTTCATCTCTATTACTTGTGATTTACCTCTGT
GATCTTACTTGGATTTACATCTCATTGCGTAAGAAATATTGGTAAACAGGCGGATGGGCGACAGC
CCCAGACGGCGCGTACGCGTGGGGTTATTGTTTCAAAGAGGAGCAAGGCAATCCTCCCGCCGAGT
ACTGCCAGGCAACCTCCCAANGGCCCTGTGCATCTGGAAAGAGATACTACGNNNNAGGGCCCGTT
CAATTGTCATGNTNGGTCGTTTCTATATTTTTGCAGTTTTAATTTCTTTAATTATATTCTGCACA
GACGCACCCAAAAAGAATATCTGAATGAAATTTAAGACTAATACGCCGTTTGCTCAATTTGCTTT
GGGTTTTTCAGGAATTACAACTATGGACCGGCCGGGAAGGCAATCGGATTCGATGGCATAAACAA
CCCCGACATTGTTGCTAGCGATGCCACGGTCTCTTTCAAGACCGCAATCTGGTTCTGGATGACCG
CCCAATCTCCGAAACCTTCTTGCCACGATGTCATGACCGGGAAATGGACTCCGTCCGGCAGCGAC
AGCGCCGCTGGGAGAGCTGCGGGATATGGAGCAGTTACCAACATCATCAACGGCGGGCTGGAGTG
CGGGAAAGGCAGTGACCCGAGGCAGCAGGATCGCATCGGCTTCTACAAAAGATACAGTGACATTC
TTGGGGTGAGCTACGGATCAAACCTGGATCGCAACAACCAGAGGCCTTTCGGCGCTGCAGTTCAA
TCTGAACCTCGTCTTATCAAAACCGTGGTTTGAACACTTCTCATAAATCTGAGA

>Pcchia2-1
TGCTGTGCCTACCCTTAGATTTAAGGCGTATATTATCTAGCGATGGATTTAAGCTGTACATTATC
CAACGATAGATATATACTGCCCATCTTATTATTAATGGCTAGATAAACTTGTGAATATGCGTTGT
CCACCTAATTTATCAAATCATTGAGATATAATAGATGCGAGCCGCAAGCTCCACCATAAAGTTAC
CTTTAAATCTTTGCTACTTAGAATTGATCTTGGAGATATACCAGAACAATGAATAATGAAATATA
CAGTGAGAGGTAAAGATTTACACCACTGTATCATCCCATTAAAGACAATGCAGGCTATCTATTTG
TTGCTTGACCGCATTGTGCCGCGACAATGAGCACCTCTTGCAACTTGAATGGAATGAACTTCCTC
TCCGTGTGTATGGTATCTTTACATAACCGTAATGTAGCGTAGTATGATATTCTTACATAACGGTA
ATGTAGCGTAGTCTAAACGTTTAAAAAGCCAGGAGTTTATTCCCCTAAAAACAATGAAGTTGCCC
TAAGTTTGAAAAGACCAGAATGGCCGTTAGAGGGGCGGACTCCTCAAACCGCTGAGTTGTTAAAA
TCATAGGATCTTAGTATACCCGACTTACAGGTGGTTAGATATACTCTTGATATTTTAGATCAAAT
ATGAGATTTGATCCACGGTCTCCTGGATTAAGCTTGATAGTTATGGTTTCACCATATAATTAAAA
AAAAAAAATTGCTTTCCAAAAAGATTANCTTCCATATTGTGGCATGTGTGGAAAGGCCAAAGCCA
AAGTCTGCCCTTCAAAGGAAGAGTGGAAAATACTCCACGCGAGGAAGAAGCGAAGAAAGATCTGA
AGTCAAGATTATGAAATACGGGAGAGAAGCGGCTATAGTCGCCACGGAAGTCGTTTTACTTTTAA
CCTTAACCCATCTTACACGGTATGTACCTGCTTTCGCAAAAATGGTGGCCGGGTTGCAGGTTGGA
CCATGAGAACACTATATCACCGTGAACGCCACGGAATGGTACAAAAGTTTCAACAAAGCGCCGCC
CCCCACCTCCCATGGCACTGTATCTGACTCTTGCGGAAATACTGCTATAAAAGGAGGGGCTTGCA
GCCTAGGATCATCACCAGAATTTGCTTTGGCAGTTTAAAGACAGCATACACGACTATGGCGAGAA
AGATGTCGATGAAGTTGTTGTTGGCCCTCGCTGCAGTGGCCATAATGAGTACTTTGTGTTATGTT
TCTGGACAACAGGGAGTCGGCTCCATCATAACTGAAGATGTTTTCAATGAGTTTCTCAAGCACCG
AAATGACGCCGCATGCCAGGCGAGAGACTTCTACACCTACAGCGCCTTCATTGCGGCCACTAATA
GTTTCTCAGACTTCGGCAACAACGGCGATCTAGAGAGCCGCAAAAGAGAGCTCGCAGCTTTCTTT
GGTCAAACGTCGCAGGAAACCACAGGTATTATTAGTTTAGCCTCCTCTAACTCTTCTGTCTCTCT
GCTATTCCTTATTGTTATTAATGGCATTAAGCTAATTGAGTTTGTACAGGCGGGTGGGCCACGGC
CCCAGACGGTCCATATGCGTGGGGTTACTGCTTCAAAGAGGAGAATAGCGGAGACAAATACCACG
GACGAGGACCTATTCAGCTAACAGGGTAATATTTCTATATCTAAGTTTTGTTAATCCACTTCGAA
TTAATTGTAGCAATAGATATGGAAAAAATTGAATGAATTTCAAGCCTAATACACTTACCGCTGCT
GTTTGATAGGGACTACAACTATAAAGCTGCGGGGGATGCCTTANGGTACGATCTCATAAACAATC
CGGATCTCTTGGTGACTGACGCCACGGTCTCGTTTAAGACGGCGGTGTGGTTCTGGATGANCGCG
CAGGCTCCCAAGCCTTCGTGCCACGACGTGATATTGGGAAGATGGAGTCCATCAAACGACGATAC
TGCCGCCGGCAGANTGCCGGGATATGGATTGTTGACGAACATCATTAACGGCGGGGAGGAATGCG
GCACAGGCACATAAGCGACCGTCAGCAGGGCCGAATCGGGTTCTACCAAAGATACTGCAGCTTGC
TGGGCGTGGACACTGGATCCAACCTCGACTGCCAAAACCAGAAGCATTTCTGAACTCAAATCTCT
ACTGTAGCAAGAGGCCACCACATCGATTACTACTTCGAGTTGAATGTTATGGTCAGCACAGAGTC
TACCGTCTAGCTAA
>Pcchia4-1
TAGACAGATTTAACACGAAACATTGGCTCCAGGAATTAACAGCGTTGAATTCTATCCGCAAGAAT
CAATATTTTACCCAGATATTTAAAATTAAATTATTTATGGATAAATGATAATCAGGGAATGACTA
TCAAGACACCATAATGTATATTTCAATTTTCAAATATCGAATATAAAGATTTTTCTCATAAACAC
AATTATTTTTCAAATAAACCAGATAAATCATTATGATAGAGCAGAATAGAGACCTCCAGTTTCGT
GCATAACATTAGCGAAGAAGGCGGCGAGTTCTCTCTTTCTGACGTCGGCAGATCCGGTGGTGCCA
AATCCAGAATATGCATTGGCCGCTGCGATGAAAGCATTGTAAGTGTAGAATCCCTTGCCCTCGCA
GGAGCTGGCAGCTCCGCCCGCGAGGCCATTGAAGAAACTCTGGGAAATTATGGTACCCACGCTTC
CACCTCCGCCACTTGGAGATCCTCCCCCTGAACTGGAACATGGACCGCTCTTGCAGCCGGCGCCG
CAGTAAGCAGAGCTGGTTCCACAGTATCCATACTTGCTGCAACACAGTCCGCTGGCACAGCCACA
GTTTTGAGCATTAACACTGACGCTCACCAGTAGCAGGACTAGCACCGCCATTACCGAGTTACCGC

TACTACTCCCCATTGCCAACAGTTTCTTAGGACTCTCCACTCACAGATATTATTTGATTTGTTGA
TTCAAGTTCGTTCTCTATTTATACAGAAACCCCGACCCACTCTGCTCCTCATTTTCTTCTTTGGG
TCGGCAAGACCATTATAGATTTTAATTTTTAGAAACAATTGCAGCCAGCCAATACGCGGCGGCTG
CGGTCCGGTCTTCCACGCTCTTATCCTATCCCTAGAATAGACTTCGATTTTATGAGAGAATTGCA
AAAAGCTGAACACGCGGCGGAACGCTCAGGATGGTTGAACGCACCCATTTTCACAGAGCTTCTCC
ATGCCAATGTTCCACGATAAAGATAATGTTTCCAAATAAAAAAAACTCACACGTGAATTCCCACA
AGCGACGTCCAGTATCAACGTCCAGTGTCAACATTGTGAACGTGCATGGGCATTTGAACAATCAT
AACACGGCACCAATTACAGTATCCCAATGGGCTAAAAAGTCCTCAGATTTTAACGAGAAATTACA
ACATATCAATACGCGGCCGGACGCGGTTCGCTCCGGATCGTTGAGCAGACCCATTTTTGGTTTGT
AGTGCGCCACCTTGCGCCAATGTCGAACAATAATAATAGTTTATTCATTTATTTAGCGGTTGTGA
GAAGTCACGAGATAACATCAGTGGCTTGGAGATAAGATCCGAAGATGGGTATATTACAAGTGGTC
GTATTTTCATGTGTTAGGGTATGACATTAAAGGGGAGGANATTGTTTTCTTGATGGCAAAGAAAT
ACTTCCTACTTTGTTTGTACTCATTAATATCTATATTTGTAGTTTCNATTTTGTTTGTACTCATT
AATATCTATATTTGTAGTTTCTATTTTGTTTGTACTCATTAATATCTATATTTGTAGAAAAAAGT
TATTGTCCTGGCTGAAAGCAGTTAGCGTATAACGTACTTAAGTTGGGGTTATTTTCAATTAATGA
AATGTTATTAAAATTTTTAGAAAATATATAAAGTGGGCATTAGAACTTATAAATAAATAATTTAT
GGTTAGGGTATAATATACTCAAAAATATGTTGACCAACC

6.5 Appendix

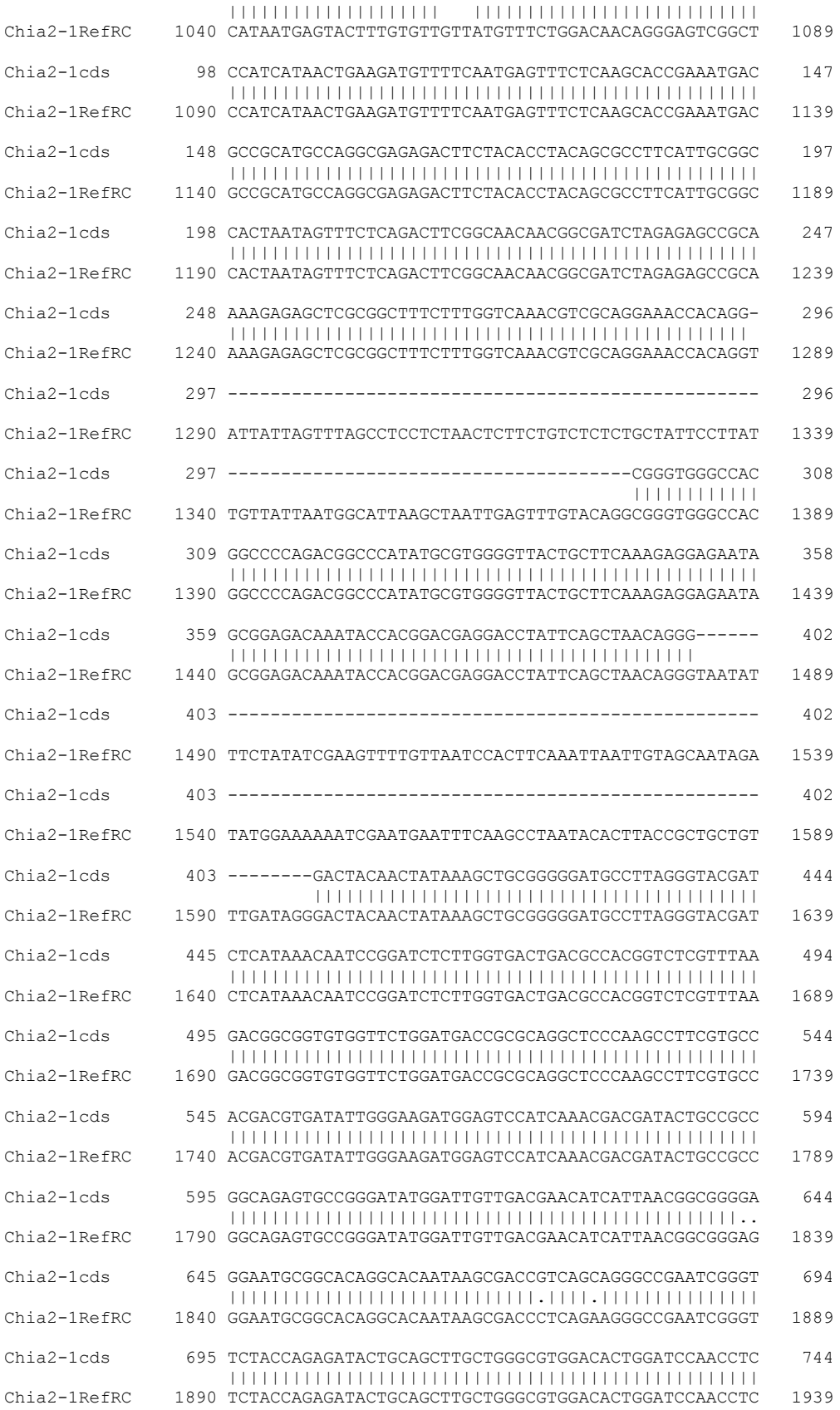
Alignment of Chia1-1 consensus sequence to the coding region (cds).

```
Chia1-1cds         1 ATGAAATCGATGAAGTTCTCCGCGATGGCGATCGCCTTGCTTACAATGGC     50
                     ||||||||||||||||.||||||||||||||||||.||||||||||
Chia1-1Refere    888 ATGAAATCGATGAAGTTTTCCGCGATGGCGATCGCCTTGATTACAATGGC    937

Chia1-1cds        51 GACGATGAATTTGTATTTTGTATCTGCTGAGCAATGCGGACAGCAAGCAG    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Chia1-1Refere    938 GACGATGAATTTGTATTTTGTATCTGCTGAGCAATGCGGACAGCAAGCAG    987

Chia1-1cds       101 GCGGAGCTCTTTGTCCCGGTGGCTTGTGCTGCAGCAAATGGGGATGGTGT    150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Chia1-1Refere    988 GCGGAGCTCTTTGTCCCGGTGGCTTGTGCTGCAGCAAATGGGGATGGTGT   1036

Chia1-1cds       151 GGCAACACGGACGCCCATTGCGGGCAGGATTGCCAGAGCCAATGCGGAGG    200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Chia1-1Refere   1037 GGCAACACGGACGCCCATTGCGGGCAGGATTGCCAGAGCCAATGCGGAGG   1079

Chia1-1cds       201 ATCGACTCCCACGCCTCCTTCACCCACTCCCGGTGGACAGGGAGTTGCAT    250
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Chia1-1Refere   1080 ATCGACTCCCACGCCTCCTTCACCCACTCCCGGTGGACAGGGAGTTGCAT   1129

Chia1-1cds       251 CTATCATCACTGAAAGCATTTTCAATGAGTTATTGAAGCACAGAAACGAC    300
                     ||||||||||||.|||.||||||||||||||||||||.|||||||||||
Chia1-1Refere   1130 CTATCATCACTGTAAGTATTTTCAATGAGTTATTGAAGCATAGAAACGAC   1179

Chia1-1cds       301 GCCGGTTGCAAGGCCAGCGGATTCTACACCTACTCTGCCTTCATTGCAGC    350
                     |||||.||||||||||||||||||||||||||||||||||||||||||||
Chia1-1Refere   1180 GCCGGCTGCAAGGCCAGCGGATTCTACACCTACTCTGCCTTCATTGCAGC   1229

Chia1-1cds       351 TGCCAATGCTTTTCCCTCCTTCGGCACCACCGGCGATGTCGCTACTCGGA    400
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Chia1-1Refere   1230 TGCCAATGCTTTTCCCTCCTTCGGCACCACCGGCGATGTCGCTACTCGGA   1279

Chia1-1cds       401 AAAGAGAGCTCGCTGCTTTCTTTGGCCAAACCTCCCACGAAACC------    444
                     ||||||||||||||||||||||||||.||||||||||||||||
Chia1-1Refere   1280 AAAGAGAGCTCGCTGCTTTCTTTGGTCAAACCTCCCACGAAACCACAGGT   1329

Chia1-1cds       445 --------------------------------------------------    444

Chia1-1Refere   1330 CATATTTTCATCTCTATTACTTGTGATTTACCGCTGTGATCTTACTTGGC   1379

Chia1-1cds       445 --------------------------------ACAGGAGGATGGGCGACA    462
                                                     ||||||||||||||.|||
Chia1-1Refere   1380 TTTACATCTCATTGGGTAAGAAATATTGGTAAACAGGAGGATGGGCAACA   1429

Chia1-1cds       463 GCCCCAGACGGCGCGTACGCGTGGGGTTATTGTTTCAAAGAGGAGCAAGG    512
                     |||||||||||||||||||||||||||||||||||.|||||||||||||||
Chia1-1Refere   1430 GCCCCAGACGGCGCGTACGCGTGGGGTTATTGTTTGAAAGAGGAGCAAGG   1479

Chia1-1cds       513 CAATCCTCCCGCCGAGTACTGCCAGGCAACCTCCCAGTGGCCCTGTGCAT    562
                     |||||||.|||||||||.||||||||||||||||||||||||||||||||||
Chia1-1Refere   1480 CAATCCTGCCGCCGAGTCCTGCCAGGCAACCTCCCAGTGGCCCTGTGCAT   1529

Chia1-1cds       563 CTGGAAAGAGATACTACGGACGAGGGCCCGTTCAATTGTCAT--------    604
                     |||||||||||||||||||||||||||||||||||||||||||||
```

```
Chia1-1Refere   1530 CTGGAAAGAGATACTACGGACGAGGGCCCGTTCAATTGTCATGGTCGGTC   1579

Chia1-1cds       605 ------------------------------------------------    604

Chia1-1Refere   1580 GTTTCTATATTTATGCAGTTTTAATTTCTTTACTTATTTTCTGCACAGAC   1629

Chia1-1cds       605 ------------------------------------------------    604

Chia1-1Refere   1630 GCACCCAAAAAGAATATCTGAATGAAATTTAAGACTAATACGCCGTTTNN   1679

Chia1-1cds       605 ----------------------GGAATTACAACTATGGACCGGCCGGGAA    632
                                           ||||||||||||||||||||||||||||
Chia1-1Refere   1680 TCAATTTGCTTTGGGTTTTTCAGGAATTACAACTATGGACCGGCCGGGAA   1729

Chia1-1cds       633 GGCAATCGGATTCGATGGCATAAACAACCCCGACATTGTTGCTAGCGATG    682
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Chia1-1Refere   1730 GGCAATCGGATTCGATGGCATAAACAACCCCGACATTGTTGCTAGCGATG   1779

Chia1-1cds       683 CCACCGTCTCTTTCAAGACCGCAATCTGGTTCTGGATGACCGCCCAATCT    732
                     ||||.||.||||||||||||||.||||||||||||||||||||||||||
Chia1-1Refere   1780 CCACGGTGTCTTTCAAGACCGCAATNTGGTTCTGGATGACCGCCCAATCT   1829

Chia1-1cds       733 CCGAAACCTTCTTGCCACGATGTCATGACCGGGAAATGGACTCCGTCCGG    782
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Chia1-1Refere   1830 CCGAAACCTTCTTGCCACGATGTCATGACCGGGAAATGGACTCCGTCCGG   1879

Chia1-1cds       783 CAGCGACAGCGCCGCTGGGAGAGCTGCGGGATATGGAGCAGTTACCAACA    832
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Chia1-1Refere   1880 CAGCGACAGCGCCGCTGGGAGAGCTGCGGGATATGGAGCAGTTACCAACA   1929

Chia1-1cds       833 TCATCAACGGCGGGCTGGAGTGCGGGAAAGGCAGTGACTCGAGGCAGCAG    882
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Chia1-1Refere   1930 TCATCAACGGCGGGCTGGAGTGCGGGAAAGGCAGTGACTCGAGGCAGCAG   1979

Chia1-1cds       883 GATCGCATCGGCTTCTACAAAAGATACAGTGACATTCTTGGGGTGAGCTA    932
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Chia1-1Refere   1980 GATCGCATCGGCTTCTACAAAAGATACAGTGACATTCTTGGGGTGAGCTA   2029

Chia1-1cds       933 CGGATCAAACCTGGATTGCAACAACCAGAGGCCTTTCGGCGCTGCAGTTC    982
                     ||||||||||.|||||||||||||||||||||||||||||||||||||||
Chia1-1Refere   2030 CGGATCAAACNTGGATTGCAACAACCAGAGGCCTTTCGGCGCTGCAGTTC   2079

Chia1-1cds       983 AATCTGAACCTCGTCTTATCAAAACCGTGGTTTGAACACTTCTCATAAAT   1032
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Chia1-1Refere   2080 AATCTGAACCTCGTCTTATCAAAACCGTGGTTTGAACACTTCTCATAAAT   2129

Chia1-1cds      1033 CTGAGA   1038
                     ||||||
Chia1-1Refere   2130 CTGAGA   2135
```

Alignment of Chia2-1 consensus sequence to the coding region (cds) and deduced amino acid sequence.

```
Chia2-1cds         1 ATGGCGAGAAAGATGTCGATGAAGTTGTTGTTGGCCCTCGCTGCAGTGGC     50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Chia2-1RefRC     990 ATGGCGAGAAAGATGTCGATGAAGTTGTTGTTGGCCCTCGCTGCAGTGGC   1039

Chia2-1cds        51 CATAATGAGTACTTTGTGTT---ATGTTTCTGGACAACAGGGAGTCGGCT     97
```


Chia2-1RefRC 1040 CATAATGAGTACTTTGTGTTGTTATGTTTCTGGACAACAGGGAGTCGGCT 1089
Chia2-1cds 98 CCATCATAACTGAAGATGTTTTCAATGAGTTTCTCAAGCACCGAAATGAC 147
Chia2-1RefRC 1090 CCATCATAACTGAAGATGTTTTCAATGAGTTTCTCAAGCACCGAAATGAC 1139
Chia2-1cds 148 GCCGCATGCCAGGCGAGAGACTTCTACACCTACAGCGCCTTCATTGCGGC 197
Chia2-1RefRC 1140 GCCGCATGCCAGGCGAGAGACTTCTACACCTACAGCGCCTTCATTGCGGC 1189
Chia2-1cds 198 CACTAATAGTTTCTCAGACTTCGGCAACAACGGCGATCTAGAGAGCCGCA 247
Chia2-1RefRC 1190 CACTAATAGTTTCTCAGACTTCGGCAACAACGGCGATCTAGAGAGCCGCA 1239
Chia2-1cds 248 AAAGAGAGCTCGCGGCTTTCTTTGGTCAAACGTCGCAGGAAACCACAGG- 296
Chia2-1RefRC 1240 AAAGAGAGCTCGCGGCTTTCTTTGGTCAAACGTCGCAGGAAACCACAGGT 1289
Chia2-1cds 297 ---------- 296
Chia2-1RefRC 1290 ATTATTAGTTTAGCCTCCTCTAACTCTTCTGTCTCTCTGCTATTCCTTAT 1339
Chia2-1cds 297 ----------CGGGTGGGCCAC 308
Chia2-1RefRC 1340 TGTTATTAATGGCATTAAGCTAATTGAGTTTGTACAGGCGGGTGGGCCAC 1389
Chia2-1cds 309 GGCCCCAGACGGCCCATATGCGTGGGGTTACTGCTTCAAAGAGGAGAATA 358
Chia2-1RefRC 1390 GGCCCCAGACGGCCCATATGCGTGGGGTTACTGCTTCAAAGAGGAGAATA 1439
Chia2-1cds 359 GCGGAGACAAATACCACGGACGAGGACCTATTCAGCTAACAGGG------ 402
Chia2-1RefRC 1440 GCGGAGACAAATACCACGGACGAGGACCTATTCAGCTAACAGGGTAATAT 1489
Chia2-1cds 403 ---------- 402
Chia2-1RefRC 1490 TTCTATATCGAAGTTTTGTTAATCCACTTCAAATTAATTGTAGCAATAGA 1539
Chia2-1cds 403 ---------- 402
Chia2-1RefRC 1540 TATGGAAAAAATCGAATGAATTTCAAGCCTAATACACTTACCGCTGCTGT 1589
Chia2-1cds 403 --------GACTACAACTATAAAGCTGCGGGGGATGCCTTAGGGTACGAT 444
Chia2-1RefRC 1590 TTGATAGGGACTACAACTATAAAGCTGCGGGGGATGCCTTAGGGTACGAT 1639
Chia2-1cds 445 CTCATAAACAATCCGGATCTCTTGGTGACTGACGCCACGGTCTCGTTTAA 494
Chia2-1RefRC 1640 CTCATAAACAATCCGGATCTCTTGGTGACTGACGCCACGGTCTCGTTTAA 1689
Chia2-1cds 495 GACGGCGGTGTGGTTCTGGATGACCGCGCAGGCTCCCAAGCCTTCGTGCC 544
Chia2-1RefRC 1690 GACGGCGGTGTGGTTCTGGATGACCGCGCAGGCTCCCAAGCCTTCGTGCC 1739
Chia2-1cds 545 ACGACGTGATATTGGGAAGATGGAGTCCATCAAACGACGATACTGCCGCC 594
Chia2-1RefRC 1740 ACGACGTGATATTGGGAAGATGGAGTCCATCAAACGACGATACTGCCGCC 1789
Chia2-1cds 595 GGCAGAGTGCCGGGATATGGATTGTTGACGAACATCATTAACGGCGGGGA 644
Chia2-1RefRC 1790 GGCAGAGTGCCGGGATATGGATTGTTGACGAACATCATTAACGGCGGGAG 1839
Chia2-1cds 645 GGAATGCGGCACAGGCACAATAAGCGACCGTCAGCAGGGCCGAATCGGGT 694
Chia2-1RefRC 1840 GGAATGCGGCACAGGCACAATAAGCGACCCTCAGAAGGGCCGAATCGGGT 1889
Chia2-1cds 695 TCTACCAGAGATACTGCAGCTTGCTGGGCGTGGACACTGGATCCAACCTC 744
Chia2-1RefRC 1890 TCTACCAGAGATACTGCAGCTTGCTGGGCGTGGACACTGGATCCAACCTC 1939

```
Chia2-1cds       745 GACTGCCAAAACCAGAAGCATTTCTGA    771
                     ||||||||||||||||||||||||||
Chia2-1RefRC    1940 GACTGCCAAAACCAGAAGCATTTCTGA   1966
```

Alignment of Chia4-1 consensus sequence to the coding region (cds) and deduced amino acid sequence.

```
Chia4-1cds         1 ATGGGGAGTAGTAGCGGTAACTCGGTAATGGTGGTGGTGGTGCTAGTCCT     50
                     ||||||||||||||||||||||||||||      ||||||||||||||||
Chia4-1refere    479 ATGGGGAGTAGTAGCGGTAACTCGGTAA------TGGTGGTGCTAGTCCT    522

Chia4-1cds        51 GCTACTGGTGAGCGTCAGTGTTAATGCTCAAAACTGTGGCTGTGCCAGCG    100
                     ||||||||||||.||||||||||||||||||||||||||||||||||||||
Chia4-1refere    523 GCTACTGGTGAGTGTCAGTGTTAATGCTCAAAACTGTGGCTGTGCCAGCG    572

Chia4-1cds       101 GACTGTGTTGCAGCAAGTATGGATACTGTGGAACCACCTCTGCTTACTGC    150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Chia4-1refere    573 GACTGTGTTGCAGCAAGTATGGATACTGTGGAACCACCTCTGCTTACTGC    622

Chia4-1cds       151 GGCGCTGGCTGCAAGAGCGGTCCATGTTCCAGTTCAGGGGGAGGATCTCC    200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Chia4-1refere    623 GGCGCTGGCTGCAAGAGCGGTCCATGTTCCAGTTCAGGGGGAGGATCTCC    672

Chia4-1cds       201 AAGTGGCGGAGGTGGAAGCGTGGGTACCATAATTTCCCAGAGTTTCTTCA    250
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Chia4-1refere    673 AAGTGGCGGAGGTGGAAGCGTGGGTACCATAATTTCCCAGAGTTTCTTCA    722

Chia4-1cds       251 ATGGCCTCGCGGGCGGAGCTGCCAGCTCCTGCGAGGGCAAGGGATTCTAC    300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Chia4-1refere    723 ATGGCCTCGCGGGCGGAGCTGCCAGCTCCTGCGAGGGCAAGGGATTCTAC    772

Chia4-1cds       301 ACTTACAATGCTTTCATCGCAGCGGCCAATGCGTATTCTGGATTTGGCAC    350
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Chia4-1refere    773 ACTTACAATGCTTTCATCGCAGCGGCCAATGCGTATTCTGGATTTGGCAC    822

Chia4-1cds       351 CACCGGATCTGCCGACGTCAGAAAGAGAGAACTCGCCGCCTTCTTCGCTA    400
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Chia4-1refere    823 CACCGGATCTGCCGACGTCAGAAAGAGAGAACTCGCCGCCTTCTTCGCTA    872

Chia4-1cds       401 ATGTTATGCACGAAACTGAAGGG--------------------------    423
                     ||||||||||||||||||.|||.
Chia4-1refere    873 ATGTTATGCACGAAACTGGAGGTCTATATTCTGCTCTATCATAATTATTT    922

Chia4-1cds       424 -------------------------------------------------    423

Chia4-1refere    923 ACCTAGTAATCCCTTTTCTTTATGAGAAAAAGCTTTAATTGCCATTTACC    972

Chia4-1cds       424 -------------------------------------------------    423

Chia4-1refere    973 TTCCAATCGATCTATAATTTGGTTTCAATTTTTTCACTGATAATTGATTA   1022

Chia4-1cds       424 --------------------ATGTGCTACATCAATGAGATAAACCCTCAA    453
                                         ||||||||||||||||||.||||||||.|
Chia4-1refere   1023 ATTTTTGTGTCATGTCAGGAATGTGCTACATCAATGAGAGAAACCCTCCA   1072

Chia4-1cds       454 AGTAACTACTGTAATAGTTCAGCCACGTGGCCATGTGCGAGTGGTAAGAG    503
                     |..||||||||||..||||||||||||||||.||||||||||||||||||
Chia4-1refere   1073 ATGAACTACTGTATGAGTTCAGCCACGTGGCCGTGTGCGAGTGGTAAGAG   1122

Chia4-1cds       504 CTACCATGGACGGGGCCCTCTCCAGTTGAGCT------------------    535
                     |||||||||.||.||||||||||||||||||
Chia4-1refere   1123 CTACCATGGTCGTGGCCCTCTCCAGTTGAGCTGGTATGTATGCGAAAACT   1172

Chia4-1cds       536 -------------------------------------------------    535

Chia4-1refere   1173 CTACACCCACTAATCACATCTCATCTCCTCTGTAATGATAAAAAAGATTG   1222

Chia4-1cds       536 ------------------------GGAATTACAATTATGGAGCGGCTGGGC    562
                                             ||||||||||||||||||||||||||||
```

```
Chia4-1refere   1223 AGACTAACAATTAGATTGTTACAGGAATTACAATTATGGAGCGGCTGGGC   1272

Chia4-1cds       563 AGACCATTGGGTTCGATGGGGTGAACAACCCAGAGAAGGTGGGGCAAGAC    612
                     |||.|||||||||||||.||||||||||||||||||||||||||||||||
Chia4-1refere   1273 AGAGCATTGGGTTCGACGGGGTGAACAACCCAGAGAAGGTGGGGCAAGAC   1322

Chia4-1cds       613 CCCACCATTTCGTTCAAGACGGCTGTCTGGTTCTGGATGAAGAACAGCAA    662
                     |||||||||||||||||||||||.|||||||||||||||||||||||||||
Chia4-1refere   1323 CCCACCATTTCGTTCAAGACGGCGGTCTGGTTCTGGATGAAGAACAGCAA   1372

Chia4-1cds       663 CTGTCACTCCGCCATAACGTCAGGGCAGGGCTTCGGTGGCACAATCAAAG    712
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Chia4-1refere   1373 CTGTCACTCCGCCATAACGTCAGGGCAGGGCTTCGGTGGCACAATCAAAG   1422

Chia4-1cds       713 CCATCAACAGCCAGGAATGCAACGGTGGAAAGACTGGGGAGGTGAATAAC    762
                     |||||||||||||||||||||||||||||||.|.|||||||||||||.|
Chia4-1refere   1423 CCATCAACAGCCAGGAATGCAACGGTGGAAACAGTGGGGAGGTGAATAGC   1472

Chia4-1cds       763 AGAGTGAACTACTACAAGAATATCTGTAGCCA    794
                     |||||||||||||||||||||||.|||||||
Chia4-1refere   1473 AGAGTGAACTACTACAAGAATATCGGTAGCCA   1504
```

Appendix 6.6

Table 6.6.1: Allelic richness based on four variant loci in jack and lodgepole pine class I chitinase, *chia1-1,* sequenced from gDNA of seedlings from 12 provenances collected across Canada.

	Minto YK	Baldy Hughes BC	Edgewood BC	Hudson Bay SK	London ON	Weagamow Lake ON	Nose Mountain AB	Blue Ridge AB	Stoney Mountain AB	Cypress Hills AB	Despres Lake NB	Overall
SNV170	2.00	2.00	2.00	1.95	2.00	1.93	2.00	2.00	2.00	2.00	2.00	2.00
Indel1	2.00	3.00	2.00	2.95	2.00	2.00	2.00	2.00	2.00	1.93	2.97	2.35
SNV324	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.93	1.97	2.00
SNV357	2.00	2.00	2.00	2.00	2.00	1.95	2.00	2.00	2.00	2.00	2.00	2.14

Table 6.6.2: Matrix of pairwise Fst values estimated on the basis of four variant loci found in *chia1-1* among gDNA of seedlings from 12 provenances collected across Canada.

	Minto YK	Baldy Hughes BC	Edgewood BC	Hudson Bay SK	London ON	Weagamow Lake ON	Nose Mountain AB	Blue Ridge AB	Stoney Mountain AB	Cypress Hills AB	Despres Lake NB
Minto, YK	0.000	0.101	0.134	0.211	0.185	0.370	0.048	0.143	0.258	0.082	0.343
Baldy Hughes, BC	0.101	0.000	0.000	0.088	0.025	0.146	0.000	0.000	0.020	0.379	0.125
Edgewood, BC	0.134	0.000	0.000	0.054	0.000	0.140	0.013	0.000	0.004	0.422	0.079
Hudson Bay, SK	0.211	0.088	0.054	0.000	0.000	0.070	0.151	0.024	0.006	0.526	0.000
London, ON	0.185	0.025	0.000	0.000	0.000	0.082	0.099	0.000	0.000	0.494	0.000
Weagamow Lake, ON	0.370	0.146	0.140	0.070	0.082	0.000	0.207	0.039	0.000	0.668	0.048
Nose Mountain, AB	0.048	0.000	0.013	0.151	0.099	0.207	0.000	0.000	0.097	0.279	0.226
Blue Ridge, AB	0.143	0.000	0.000	0.024	0.000	0.039	0.000	0.000	0.000	0.441	0.057
Stoney Mountain, AB	0.258	0.020	0.004	0.006	0.000	0.000	0.097	0.000	0.000	0.563	0.000
Cypress Hills, AB	0.082	0.379	0.422	0.526	0.494	0.668	0.279	0.441	0.563	0.000	0.650
Despres Lake, NB	0.343	0.125	0.079	0.000	0.000	0.048	0.226	0.057	0.000	0.650	0.000

Table 6.6.3: Allelic richness based on four variant loci in jack and lodgepole pine class II chitinase, *chia2-1,* sequenced from gDNA of seedlings from 12 provenances collected across Canada.

	Minto YK	Baldy Hughes BC	Edgewood BC	Hudson Bay SK	London ON	Weagamow Lake ON	Nose Mountain AB	Blue Ridge AB	Stoney Mountain AB	Cypress Hills AB	Despres Lake NB	Overall
SNV -284	2.00	2.00	2.00	1.96	1.99	1.76	2.00	2.00	1.79	2.00	1.98	2.00
SNV -141	1.00	1.00	2.00	2.00	2.00	2.00	1.88	2.00	1.79	1.00	2.00	2.00
InDel 6	1.98	2.00	2.00	2.00	2.00	2.00	1.88	1.97	1.95	1.00	2.00	1.99
SNV -413	1.99	2.00	2.00	2.00	2.00	2.00	1.82	2.00	1.99	2.00	2.00	2.00

Table 6.6.4: Matrix of pairwise Fst values estimated on the basis of four variant loci found in *chia2-1* among gDNA of seedlings from 12 provenances collected across Canada.

	Minto YK	Baldy Hughes BC	Edgewood BC	Hudson Bay SK	London ON	Weagamow Lake ON	Nose Mountain AB	Blue Ridge AB	Stoney Mountain AB	Cypress Hills AB	Despres Lake NB
Minto, YK	0.00	0.00	0.00	0.39	0.42	0.47	0.00	0.00	0.60	0.00	0.37
Baldy Hughes, BC	0.00	0.00	0.00	0.33	0.33	0.39	0.05	0.00	0.57	0.00	0.30
Edgewood, BC	0.00	0.00	0.00	0.23	0.26	0.31	0.01	0.00	0.45	0.01	0.22
Hudson Bay, SK	0.39	0.33	0.23	0.00	0.02	0.00	0.43	0.31	0.04	0.46	0.00
London, ON	0.42	0.33	0.26	0.02	0.00	0.00	0.47	0.34	0.21	0.48	0.00
Weagamow Lake, ON	0.47	0.39	0.31	0.00	0.00	0.00	0.51	0.39	0.13	0.53	0.00
Nose Mountain, AB	0.00	0.05	0.01	0.43	0.47	0.51	0.00	0.00	0.62	0.00	0.42
Blue Ridge, AB	0.00	0.00	0.00	0.31	0.34	0.39	0.00	0.00	0.50	0.00	0.30
Stoney Mountain, AB	0.60	0.57	0.45	0.04	0.21	0.13	0.62	0.50	0.00	0.65	0.19
Cypress Hills, AB	0.00	0.00	0.01	0.46	0.48	0.53	0.00	0.00	0.65	0.00	0.46
Despres Lake, NB	0.37	0.30	0.22	0.00	0.00	0.00	0.42	0.30	0.19	0.46	0.00

Table 6.6.5: Allelic richness based on eight variant loci in jack and lodgepole pine class IV chitinase, *chia4-1,* sequenced from gDNA of seedlings from 12 provenances collected across Canada.

	Minto YK	Baldy Hughes BC	Edgewood BC	Hudson Bay SK	London ON	Weagamow Lake ON	Nose Mountain AB	Blue Ridge AB	Stoney Mountain AB	Cypress Hills AB	Despres Lake NB	Overall
SNV1 -167	2.00	2.00	2.82	1.79	1.00	1.00	2.00	2.00	2.00	2.00	1.82	2.10
SNV2 -119	1.82	2.00	1.82	1.00	1.00	1.95	2.00	1.98	1.95	1.79	1.82	1.81
SNV -45	2.00	2.00	2.70	1.96	1.00	1.76	2.00	1.98	2.00	2.00	1.82	2.07
Indel 36	2.00	2.00	1.99	1.00	1.00	1.00	1.85	1.00	1.73	1.76	1.00	1.77
SNV 56	2.00	2.00	2.00	1.79	1.00	1.76	2.00	2.00	2.00	2.00	1.79	2.00
SNV 737	2.82	2.00	2.00	1.79	1.76	1.00	2.88	2.00	2.00	2.96	1.97	2.41
SNV 793	1.97	2.00	2.00	1.00	1.00	1.00	2.00	2.00	2.76	2.00	2.00	2.10
SNV 867	2.82	2.00	1.97	1.00	1.00	1.00	1.99	1.82	1.96	1.99	1.00	2.01

Table 6.6.6: Matrix of pairwise Fst values estimated on the basis of four variant loci found in *chia4-1* among gDNA of seedlings from 12 provenances collected across Canada.

	Minto YK	Baldy Hughes BC	Edgewood BC	Hudson Bay SK	London ON	Weagamow Lake ON	Nose Mountain AB	Blue Ridge AB	Stoney Mountain AB	Cypress Hills AB	Despres Lake NB
Minto, YK	0.00	0.03	0.09	0.53	0.60	0.57	0.00	0.14	0.26	0.08	0.44
Baldy Hughes, BC	0.03	0.00	0.01	0.46	0.55	0.50	0.01	0.06	0.18	0.07	0.37
Edgewood, BC	0.09	0.01	0.00	0.30	0.37	0.34	0.02	0.00	0.06	0.01	0.19
Hudson Bay, SK	0.53	0.46	0.30	0.00	0.01	0.00	0.48	0.28	0.07	0.32	0.01
London, ON	0.60	0.55	0.37	0.01	0.00	0.02	0.55	0.37	0.14	0.39	0.04
Weagamow Lake, ON	0.57	0.50	0.34	0.00	0.02	0.00	0.51	0.32	0.10	0.37	0.02
Nose Mountain, AB	0.00	0.01	0.02	0.48	0.55	0.51	0.00	0.07	0.19	0.00	0.37
Blue Ridge, AB	0.14	0.06	0.00	0.28	0.37	0.32	0.07	0.00	0.04	0.06	0.16
Stoney Mountain, AB	0.26	0.18	0.06	0.07	0.14	0.10	0.19	0.04	0.00	0.08	0.01
Cypress Hills, AB	0.08	0.07	0.01	0.32	0.39	0.37	0.00	0.06	0.08	0.00	0.23
Despres Lake, NB	0.44	0.37	0.19	0.01	0.04	0.02	0.37	0.16	0.01	0.23	0.00

6.7 Appendix

Script.sh: script used to quality trim Sanger generated sequences using fastq trimmer, convert trimed sequences to fasta files, align sequences using MAFFT (gap opening penalty of -3.0), and format alingments onto a single line to be assembled by hand into contigs. Sequences genereated for each individual were organized into sub directories. Main.go: Go script which runs script.sh in all sub-directories.

Script.sh

```
#!/bin/sh


# Merge the files.
cat M038* > merged.fastq


# Trim the merged file.
fastq_quality_trimmer -t 28 -Q33  -i merged.fastq -o
merged.trimmed.fastq


# Convert merged/trimmed file to fasta.
fastq_to_fasta -n -i merged.trimmed.fastq -o trimmed.fasta -Q33


# Align the fasta file.
mafft --reorder --op 5.0 --auto trimmed.fasta >
trimmed.fasta.aligned
```

```
# Format alignment to one line.
fasta_formatter -i trimmed.fasta.aligned -o final.fasta -w 0

# Clean up.
rm -rf merged.fastq merged.trimmed.fastq trimmed.fasta
trimmed.fasta.aligned
```

Main.go

```
package main

import (
	"io/ioutil"
	"log"
	"os"
	"os/exec"
)

func main() {
	dirs := getDirectories()
	for i, dir := range dirs {
		log.Printf("Running %v of %v", i, len(dirs))
		run(dir.Name())
	}
}

func run(dir string) {
	log.Println("running script.sh in", dir)
```

```
	cmd := exec.Command("../script.sh")
	cmd.Dir = dir
	out, err := cmd.CombinedOutput()
	if err != nil {
		log.Fatal(err, out)
	}
	log.Println("finished script.sh in", dir)
}

func getDirectories() []os.FileInfo {
	files, err := ioutil.ReadDir(".")
	if err != nil {
		log.Fatal(err)
	}

	var directories []os.FileInfo

	for _, file := range files {
		if file.IsDir() {
			directories = append(directories, file)
		}
	}
	return directories
}
```

6.8 Appendix

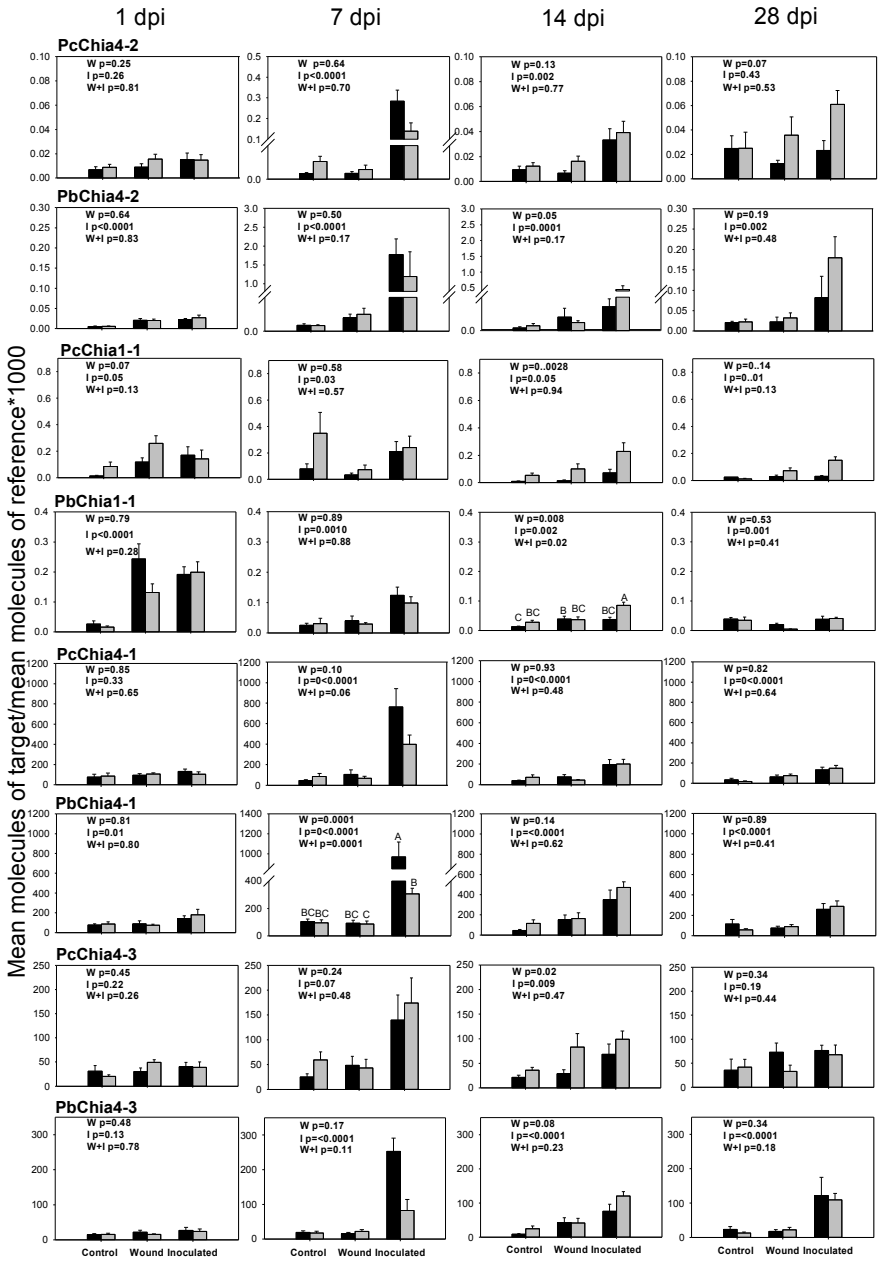


Figure 6.8.1: Relative transcript abundance of highly induced chitinase genes in lodgepole (Pc) and jack (Pb) pine seedlings in response to inoculation by MPB fungal associate *G. clavigera* during 1, 7, 14 and 28 dpi under either well watered (black) or water deficit (grey) conditions.

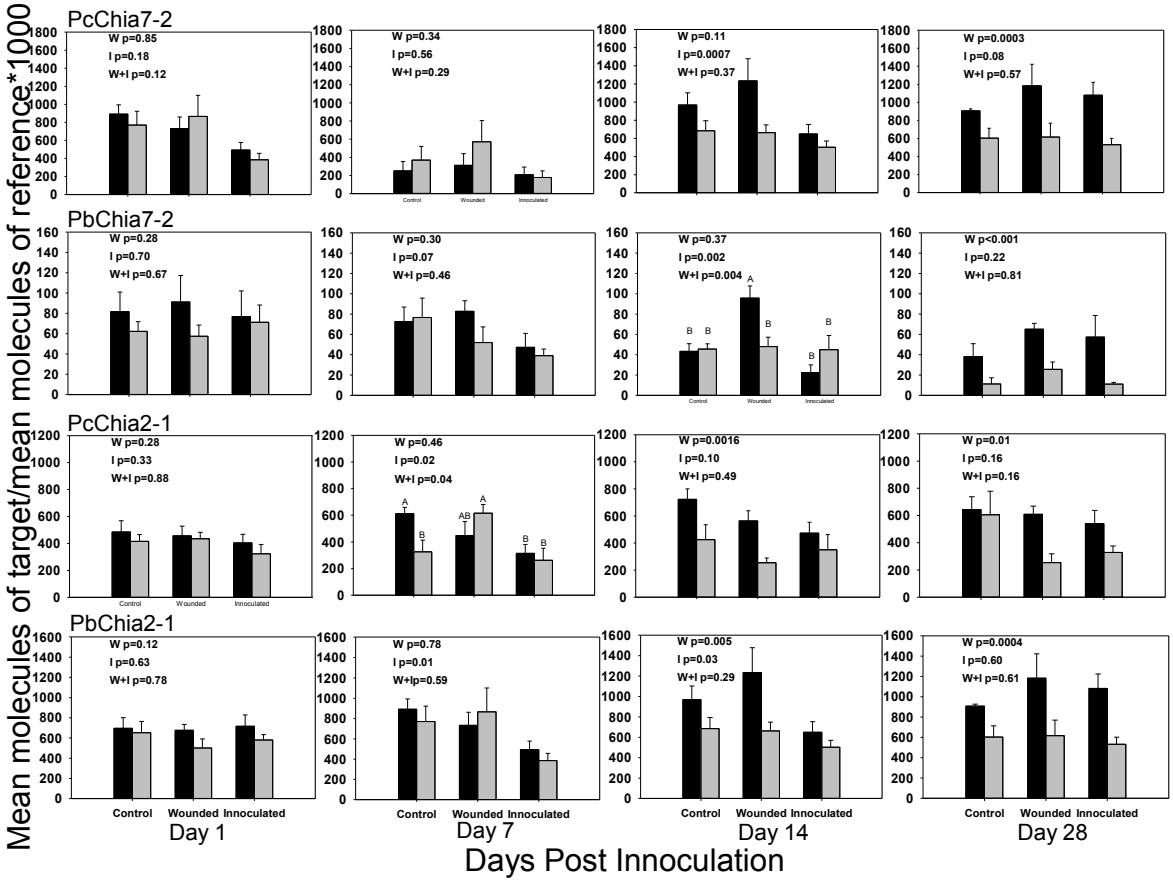


Figure 6.8.2: Relative transcript abundance of constitutively expressed chitinase genes in lodgepole (Pc) and jack (Pb) pine seedlings in response to inoculation by MPB fungal associate *G. clavigera* during 1, 7, 14 and 28 dpi under either well watered (black) or water deficit (grey) conditions.

Appendix 6.9

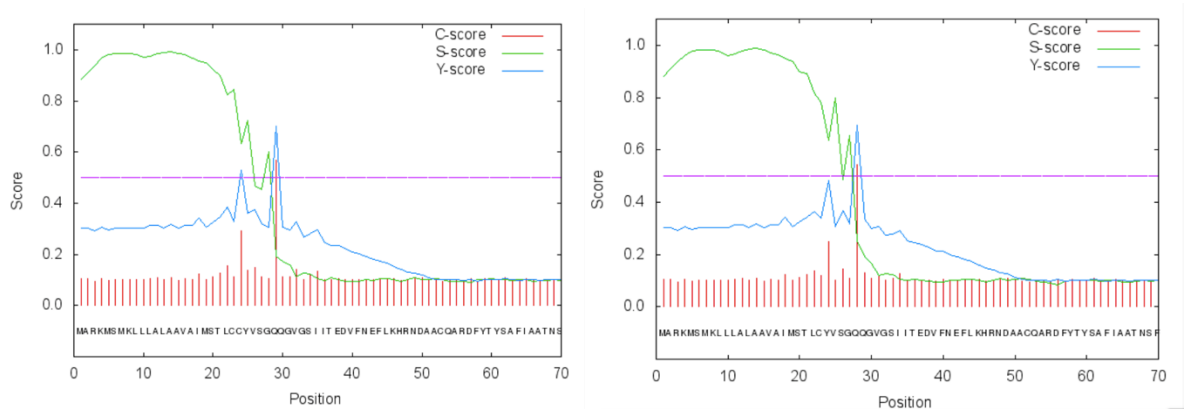


Figure 6.9.1: SignalP predicition of signal peptide in deduced amino acid sequences of *chia2-1* allelic varients. Left: variant carries a deletion of a cysteine residue in the signal peptide. Right: variant contains an insertion of a cysteine residue in the signal peptide. C-score: raw cleavage site score. S-score: raw signal peptide score. Y-score: combined C-score and S-score which better predicts cleavage site.

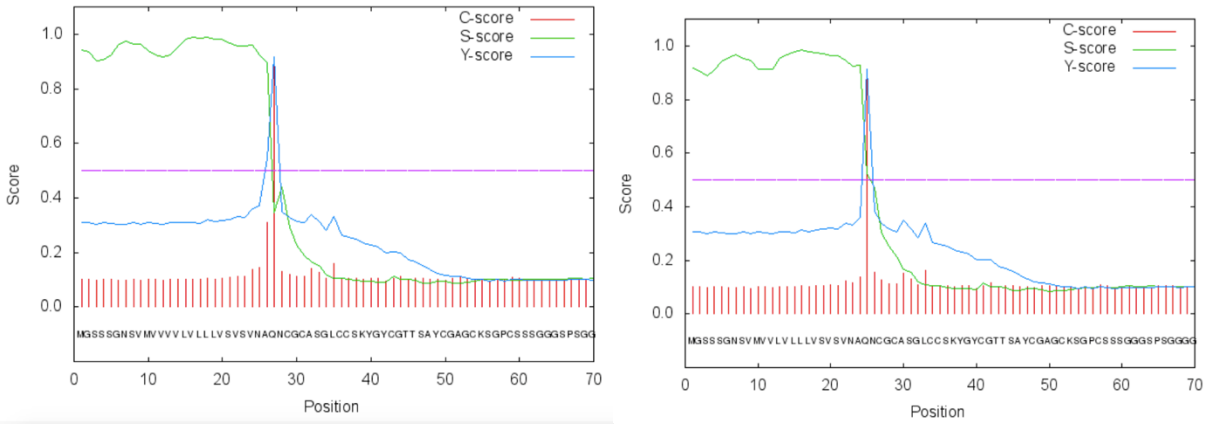


Figure 6.9.2: SignalP predicition of signal peptide in deduced amino acid sequences of *chia2-1* allelic varients. Left: variant carries a deletion of two valine residues in the signal peptide. Right: variant contains an insertion of two valine residues in the signal peptide. C-score: raw cleavage site score. S-score: raw signal peptide score. Y-score: combined C-score and S-score which better predicts cleavage site.