Transcriptomic analyses of *Pinus contorta* responses to *Grosmannia clavigera* under contrasting levels of nitrogen availability

By

Louisa Charlotte Normington

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

Master of Science

in

Plant Biology

Department of Biological Sciences
University of Alberta

Abstract

In western Canada alone, the current mountain pine beetle (MPB) outbreak has devastated nearly 19 million hectares of pine forest, negatively impacting communities, industries and ecosystems. Reforestation efforts can benefit from a deeper understanding of the impact of nitrogen (N)-based fertilization on pine trees under attack by bark beetles and their microbial cohorts. To this end, we employed transcriptomic tools to explore how N availability affects the molecular response of lodgepole pine to the pathogenic MPB fungal associate *Grosmannia clavigera*.

The first objective was to find the optimal approach to the *de novo* assembly of a lodgepole pine reference transcriptome in preparation for RNA-Seq. We hypothesized that the CLC Genomics Workbench would produce a faster assembly, but that Trans-ABySS, given its multiple k-mer de Bruijn graph algorithm, would produce a higher quality assembly. Following our findings, we suggest using Trans-ABySS for constructing reference transcriptomes using Illumina next-generation sequence data from non-model species.

The second objective was to conduct a controlled environment experiment to evaluate the effects of N availability on the responses of lodgepole pine seedlings to *G. clavigera* infection. We hypothesized that higher N availability would result in an increase in foliar N concentration in the lodgepole pine seedlings. In addition, we hypothesized that fertilization with 1 mM or 10 mM NH_4NO_3 would impact

lodgepole pine defense against *G. clavigera*, as measured by differences in lesion development. We demonstrated that increased fertilization significantly impacted foliar N content, resulting in an increase in foliar N concentration. Furthermore, the higher concentration fertilizer seemed to elicit a stronger defense response through the creation of a longer lesion, possibly in response to N-stimulated fungal growth.

The third objective was to use RNA-Seq to identify the defense response patterns of lodgepole pine seedlings fertilized with either low (0.3 mM NH_4NO_3) or high (10 mM NH_4NO_3) levels of N followed by inoculation with *G. clavigera* in growth chamber conditions. This portion of the thesis project produced a comprehensive lodgepole pine master transcriptome with accompanying annotations. Differential expression analysis was accompanied by data mining of significantly differentially expressed transcripts to uncover patterns of gene expression influenced by N availability, in accordance with the fourth objective, followed by a network analysis approach to identify genes that were co-expressed with key transcription factors. We hypothesized that N availability would: (1) affect known components of lodgepole pine defense against *G. clavigera,* such as monoterpene synthesis, (2) modulate expression of genes thought to be important in mediating *G. clavigera*-elicited responses in lodgepole pine, and (3) alter the ratio of N-based to carbon (C)-based defense-related genes that were up-regulated in response to *G. clavigera* inoculation, with a greater proportion of N-based defense genes up-regulated in response to higher N availability.

Quantification of monoterpene levels in volatile organic compounds (VOCs), foliage and phloem determined that levels of a relatively small number of monoterpenes were significantly impacted by N fertilization and *G. clavigera* inoculation, including the VOC α-pinene. Significantly differentially expressed defense-associated genes, such as terpene synthase and chitinase genes, displayed a greater fold change induction in the 0.3 mM NH_4NO_3-treated seedlings compared with the 10 mM NH_4NO_3-treated seedlings. It is possible that more resources were allocated to growth in 10 mM NH_4NO_3 treated seedlings resulting in a lower intensity defense response, though without the appropriate growth measurements, we can only speculate. A greater proportion of N-based defenses were up-regulated compared with the up-regulated C-based defenses in the high N-treated tissues compared with the low N-treated tissues. Finally, network analysis identified JAZ and WRKY transcription factors as hub genes that represented potential regulators of lodgepole pine's molecular response to *G. clavigera*.

The analyses carried out in this study suggest an intersection between N use and defense in lodgepole pine seedlings challenged by *G. clavigera* and grown under low and high N applications in growth chamber conditions.

Preface

Some of the research conducted for this thesis forms part of a collaborative effort within the lab of Janice Cooke. The tissue used for Chapters 2 and 4 came from an experiment performed by Adriana Arango-Velez with the help of Kim Lam, Miranda Meents, Leonardo Galindo-González and other Cooke lab members. Tissue processing was performed by me with the help of Colleen Fortier and Anh Dao. RNA-extractions and cDNA library preparations were performed by Chandra McAllister. Libraries were sequenced by Sofie Dang at the Molecular Biology Service Unit of the University of Alberta's Department of Biological Sciences (Edmonton, Alberta, Canada). The master reference transcriptome used for Chapter 4 was constructed by Rhiannon Peery. The monoterpene data were collected and analyzed by Mehvash Malik and Inka Lusebrink, respectively. For Chapters 3 and 4, the tissue preparation and nitrogen concentration testing were performed by Ekaterina Stolnikova and the Biogeochemical Analytical Service Laboratory at the University of Alberta (Edmonton, Alberta, Canada), respectively. All foliar nitrogen concentration and lesion length statistical analyses were performed by me. The tissue used for Chapter 3 came from an experiment performed by me with the help of Eden Kate St. Onge, McPeak, Bianca Sacchi, Colleen Fortier, Deborah Adesegun and other Cooke lab members. The RNA-Seq analysis in Chapter 4 and all of the Bash and R scripts implemented for this thesis project are my original work. The data analysis in Chapter 2 along with Chapters 1 and 5 are my original work.

Dedication

This thesis is dedicated to my grandmothers,
Joan Normington and Lynnie Weinlos.

May their memories be for infinite blessings.

Acknowledgements

Firstly, I would like to thank my graduate supervisor, Dr. Janice Cooke, for providing me with the opportunity to carry out this research in her lab. She has been a significant support, and I have learned and grown so much through my time working with her. I would like to recognize the assistance provided by Dr. Cooke and the other members of my supervisory committee, Dr. Neil Harris and Dr. David Wishart. Thank you to the past members of the Cooke lab for generating the tissue used for my research and to the current lab members for their assistance, guidance and friendship. Thank you to Oksana Cheypesh, Brenda Reitsman and Maria Kotovych for helping me to succeed. I would like to acknowledge my friends, Toby Bernstein and Winston Lim, for supporting me through the writing process. The University of Alberta and the Department of Biological Sciences provided financial support during my time as a teaching assistant. The NSERC TRIA-Net grant provided funding for my research assistantship. TRIA-Net funding also covered my international student fee differential. I would like to acknowledge WestGrid and Compute Canada that provided computational resources for my thesis work. And finally, many thanks are owed to my partner, Jeremy Woods, and my family back home for their unwavering support.

Table of Contents

List of tables

List of figures

List of abbreviations

2°	Secondary
2P	Secondary phloem
2X	Secondary xylem
ABA	Abscisic acid
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
BC	British Columbia
BHLH	Basic helix-loop-helix transcription factor
Bp	Base pair
BZIP	Basic leucine zipper domain transcription factor
C	Carbon
CD	Defense-related enzymes synthesizing carbon-based products
cDNA	Complementary deoxyribonucleic acid
CDS	Coding sequence
CHO	Carbohydrate metabolism
CNB	Carbon-nutrient balance
Contig	Contiguous sequence
Cpm	Count(s) per million
Csv	Comma separated values
CW	Cell wall-related genes

DE	Differentially expressed
DEG	Differentially expressed gene
Df	Degrees of freedom
DNA	Deoxyribonucleic acid
Dpi	Day(s) post inoculation
Ds	Double-stranded
dTTP	Deoxynucleotide thymidine triphosphate
dUTP	Deoxynucleotide uridine triphosphate
EIN	Ethylene-insensitive protein
ER	Endoplasmic reticulum
ERF	Ethylene response factor
ERT	Ethylene receptor
EST	Expressed sequence tag
ETI	Effector-triggered immunity
FC	Fold change
GC	Gas chromatography
GDB	Growth-differentiation balance
GGC	Gluconeogenesis/glyoxylate cycle
GI	GenInfo Identifier
GO	Gene ontology
Gtf	General transfer format
GUI	Graphical user interface

H	Hormone-related genes
HR	Hypersensitive response
ID	Identifier
IUPAC	International Union of Pure and Applied Chemistry
JA	Jasmonic acid
JA-Ile	Jasmonic acid-isoleucine
JAZ	Jasmonate (zinc-finger expressed in inflorescence meristem)-domain transcription factor
KEGG	Kyoto Encyclopedia of Genes and Genomes
log_2FC	Log base two fold change
LRR	Leucine-rich repeat
LS	Low sample
MANOVA	Multivariate analysis of variance
MAPK	Mitogen-activated protein kinase
MeJA	Methyl-jasmonate
MLO	Mildew resistance locus O
MPB	Mountain pine beetle
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MYB	Myeloblastosis oncogenes transcription factor
MYC	Myelocytomatosis transcription factor
N	Nitrogen

NAC	No apical meristem, *Arabidopsis* transcription activation factor and cup-shaped cotyledon transcription factor
NB-ARC	Nucleotide binding adaptor shared by nucleotide-binding oligomerization domain with leucine-rich repeat proteins, apoptotic protease activating factor 1, resistance proteins and cell-death protein 4
NBS-LRR	Nucleotide-binding site - leucine-rich repeat gene
NCBI	National Center for Biotechnology Information
ND	Nitrogen-based defense proteins, including enzymes
NGS	Next-generation sequencing
NH_4^+	Ammonium
NH_4NO_3	Ammonium nitrate
NM	Nitrogen metabolism-related genes
NO_3^-	Nitrate
NT	Nitrogen transport-related genes
OD	Optimal defense
OPP	Oxidative pentose pathway
ORF	Open reading frame
PAMP	Pathogen-associated molecular pattern
PAR	Photosynthetically active radiation
Pbs	Portable batch system
PC1	The first principal component
PC2	The second principal component

PCA	Principal component analysis
PCR	Polymerase chain reaction
PP	Polyphenolic parenchyma
PR	Pathogenesis-related
PTI	Pattern-triggered immunity
Q	Quality score
QC	Quality control
Q-Q	Quartile-quartile
qPCR	Quantitative polymerase chain reaction
R	Other regulators of genetic activity
RDA	Redundancy analysis
R-gene	Resistance gene
RNA	Ribonucleic acid
RNA-Seq	Ribonucleic acid sequencing
ROS	Reactive oxygen species
Rpm	Revolutions per minute
rRNA	Ribosomal ribonucleic acid
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
RuBisCO	Ribulose 1,5-bisphosphate carboxylase/oxygenase
S	Signaling genes
Sam	Sequence alignment map

SCF^{COI1}	S-phase kinase-associated protein 1 - cell division control protein 53 - F-box protein - coronatine-insensitive protein 1 complex
$SCF^{EBF1/2}$	S-phase kinase-associated protein 1 - cell division control protein 53 - F-box protein - ethylene-insensitive protein 3 binding F-box 1 and 2 complex
Ss	Single-stranded
Sum Sq	Sum squared values
TCA	Tricarboxylic acid cycle/organic acid transformation
TF	Transcription factors
TIR	Toll/interleukin-1 receptor-like
tRNA	Transfer ribonucleic acid
USD	United States dollar
VOC	Volatile organic compound
Vs.	Versus
WGCNA	Weighted Gene Correlation Network Analysis
WRKY	Transcription factor with a WRKY amino acid sequence deoxyribonucleic acid binding domain at the N-terminus

1.0 Chapter 1: Introduction and background

1.1 *Pinus contorta*

Representing nearly 9% of the world's overall forest cover, Canadian forests cover 347 million hectares of land (Natural Resources Canada 2017b; Natural Resources Canada 2020). Canada hosts eight forest regions, which include many coniferous genera belonging to the *Pinaceae* family, such as *Pinus* (pines), *Picea* (spruce) and *Abies* (firs) (Natural Resources Canada 2017b). Canadian pines are predominantly found in the montane and subalpine forest regions (Natural Resources Canada 2017b) and are characterized as long-lived conifers that throughout the year retain pronounced needle like foliage, typically found in bundles of two, three or five needles (Richardson 2000). They are some of the most economically and ecologically influential trees in all of Canada (Richardson 2000).

Two subspecies of *Pinus contorta* occur in Canada, namely shore pine (*P. contorta* Douglas ex Loudon var. *contorta*) and lodgepole pine (*P. contorta* Douglas ex Loudon var. *latifolia*) (Figure 1.1). Both varieties are two needled pines (Lotan and Critchfield 1990). While shore pine is confined to the coast and islands of British Columbia (BC), lodgepole pine ranges throughout BC, extending north to the Yukon and Northwest territories, east across the Rocky Mountains into Alberta and south into the United States (Figure 1.2; Lotan and Critchfield 1990). Since the last glaciation period, about 18,000 years before the present time, the lodgepole pine range has expanded into north central Alberta and hybridized with jack pine (*Pinus banksiana* Lambert) (Cullingham *et al.* 2012). Lodgepole pines'

extensive range can be attributed to their tolerance of ecological disturbances (Richardson 2000).

Figure 1.1. Lodgepole pine trees growing in the Norris Geyser Basin at Yellowstone National Park in Wyoming, United States. Acknowledgement goes to Christopher Earle, Gymnosperm Database, July 22, 2007. This image was utilized with permission from Christopher Earle.

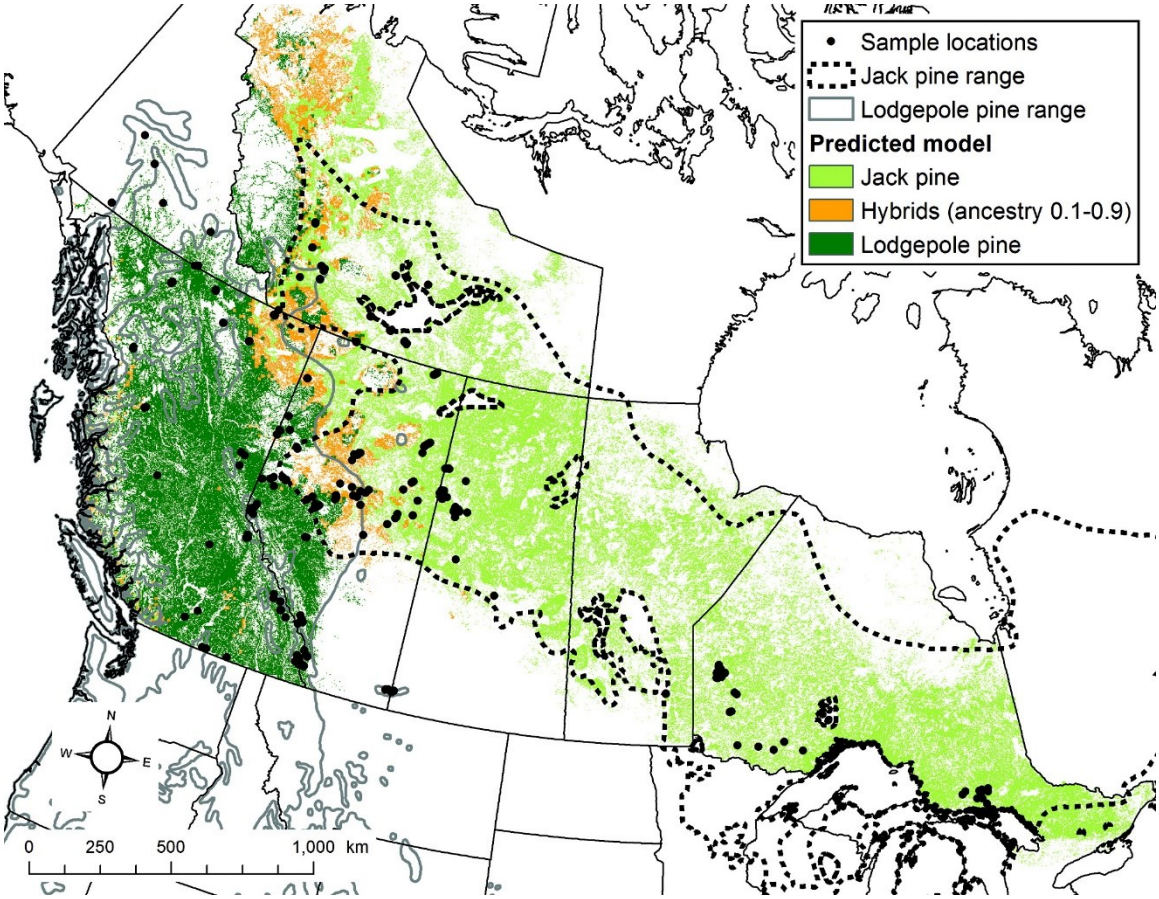


Figure 1.2. Predicted distribution of lodgepole pine, jack pine and lodgepole-jack pine hybrids in Canada. Map is based on a spatial model, which included geographic location. Dark green represents lodgepole pine, light green represents jack pine, and orange indicates hybrid pine. Black dots indicate sample sites utilized in the study conducted by Burns *et al.* (2019). Grey lines and dotted black lines indicate the proposed distributions of lodgepole pine and jack pine, respectively, in Canada taken from Little (1971). Acknowledgement goes to Burns, I., James, P. M. A., Coltman, D. W., Cullingham, C. I. (2019). Spatial and genetic structure of the lodgepole × jack pine hybrid zone. *Canadian Journal of Forest Research*, 49, pg 851. This figure was reprinted with permission from © 2020 Canadian Science Publishing or its licensors.

Lodgepole pine grow under a variety of topological and climatic conditions (Lotan and Critchfield 1990) and are a mesophytic species, growing successfully in soils with high or low water content (Carlson *et al.* 1999). Lodgepole pine are also

tolerant of both nutrient rich and limited soils (Brockley 2001). Soil quality has implications for the timing of lodgepole pine phenological events, such as dormancy that is characterized by a cessation of growth (Dougherty *et al.* 1994). Perennial trees like lodgepole pine transition into dormancy in late summer, early fall and recommence active growth the following spring (Rohde *et al.* 2007). Lodgepole pine bloom early to mid-summer and pollen is wind born (Lotan and Critchfield 1990). Seeds are held in the cones on the tree until the scales open in response to high temperatures, like those caused by wildfires (Bancroft 2008). Because lodgepole pine retain many dead branches, they are particularly susceptible to wildfire (Richardson 2000). In addition, mature trees are more susceptible to attack by pests, such as bark beetles (Natural Resource Canada 2004).

1.2 The mountain pine beetle (MPB) outbreak

1.2.1 MPB ecology and life cycle

The MPB (*Dendroctonus ponderosae* Hopkins) is a bark beetle indigenous to western North America, ranging from northern Mexico to southwestern Canada (Figure 1.3; Carroll *et al.* 2004; Fettig *et al.* 2007). In recent outbreaks, lodgepole pine has been the primary host of MPB, though other vulnerable species include jack pine, western white pine (*Pinus monticola* Douglas ex D. Don), whitebark pine (*Pinus albicaulis* Engelmann) and ponderosa pine (*Pinus ponderosa* Douglas ex

Loudon), as well as lodgepole-jack pine hybrids (Safranyik and Carroll 2006; Rice *et al.* 2007a; Rice and Langor 2009; Cullingham *et al.* 2011). Eruptive forest pests like MPB can exert widespread disturbance when biological and ecological thresholds are surpassed and constraints, such as forest heterogeneity, no longer influence outbreak dynamics (Raffa *et al.* 2008). While low populations correspond with the endemic phase, large scale population density corresponds with the epidemic phase (Raffa *et al.* 2008). Within an epidemic outbreak, such as what is currently occurring in regions of North America with MPB, populations can exist at endemic, incipient (rising), epidemic and hyperepidemic (historically unprecedented) levels (Safranyik and Carroll 2006; Sambaraju *et al.* 2019). In western Canada, adult beetles disperse mid to late summer (Natural Resources Canada 2017a). Long distance dispersal often occurs above the forest canopy, where the beetles do not fly but are carried by the wind (Jones *et al.* 2019). Short distance dispersal usually occurs by flight under the forest canopy and is semiochemical-mediated (Jones *et al.* 2019). Female beetles use host defense chemicals, particularly α-pinene enantiomers, as precursors for the aggregation pheromone trans-verbenol (Figure 1.4; Chiu *et al.* 2019). Trans-verbenol facilitates the mass attack strategy employed by MPB to overcome host defenses (Bentz *et al.* 2005).

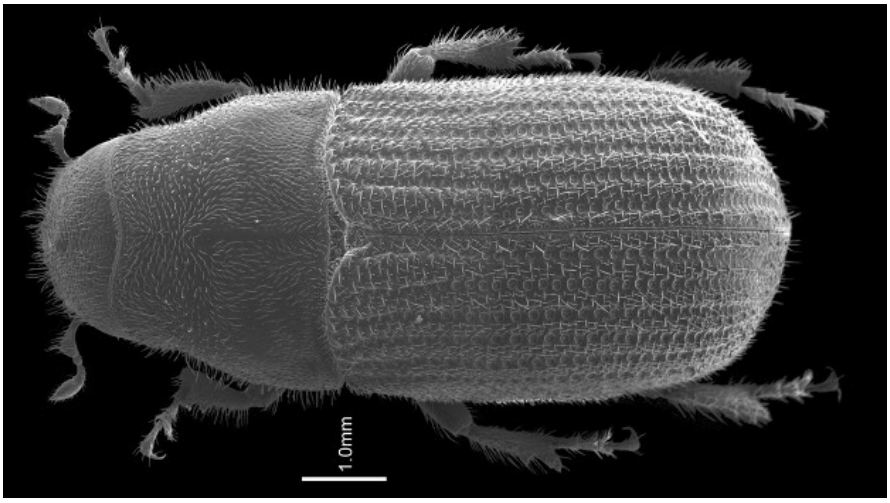


Figure 1.3. Electron micrograph of an adult MPB. Acknowledgement goes to Jack Scott at the University of Alberta. This image was utilized with permission from Janice Cooke, University of Alberta.

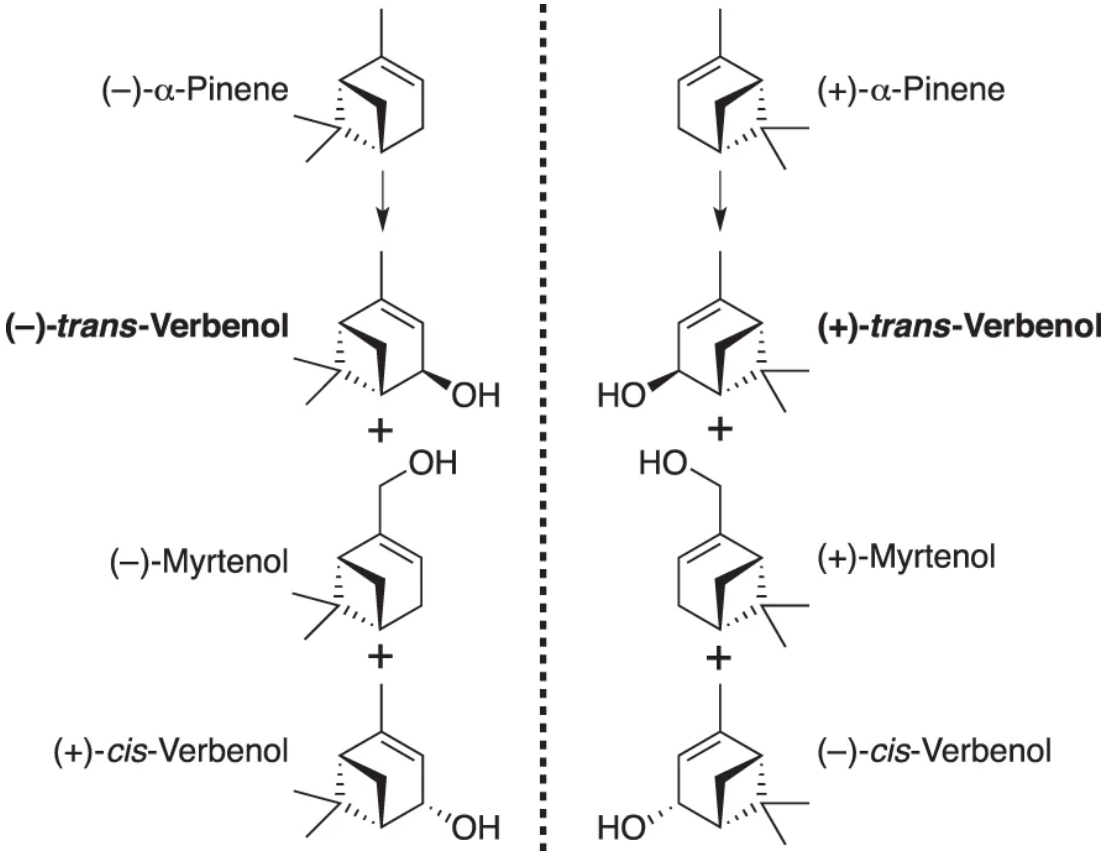


Figure 1.4. The two enantiomers of α-pinene, trans-verbenol, myrtenol and cis-verbenol. (–)-α-Pinene is the precursor of (–)-trans-verbenol, (-)-myrtenol and (+)-cis-verbenol. (+)-α-Pinene is the precursor of (+)-trans-verbenol, (+)-myrtenol and (-)-cis-verbenol. Acknowledgement goes to Chiu, C. C., Keeling, C. I., Bohlmann, J. (2019). The cytochrome P450 CYP6DE1 catalyzes the conversion of α-pinene into the mountain pine beetle aggregation pheromone trans-verbenol. *Scientific Reports*, 9, pg 2. This figure was reprinted with permission from the publisher, Nature Research. This figure is licensed under the Creative Commons Attribution 4.0 International License.

Attacking beetles bore into the tree phloem tissue and excavate nuptial chambers for copulation (Gibson *et al.* 2009). The female then burrows vertically up the tree, creating galleries underneath the bark in which to lay her eggs (Gibson *et al.* 2009). The eggs hatch into larvae that excavate tunnels that terminate in pupal chambers (Natural Resources Canada 2017a). The larvae spend the winter in these chambers and continue to feed on phloem tissues into the spring (Natural Resources Canada 2017a). If temperatures fall below approximately -40 °C for several days, the eggs and larvae do not survive the winter (Government of Alberta 2010). Eggs and larvae are also at risk early spring and late fall if temperatures drop below -20 °C for several days (Government of Alberta 2010). Larvae complete their development after the fourth instar before pupating (Gibson *et al.* 2009). Pupae transform into adult beetles the following summer (Natural Resources Canada 2017a). From mid-summer to early fall fully formed adults begin to emerge from the bark and fly to new host trees (Natural Resources Canada 2017a).

1.2.2 MPB fungal associate *Grosmannia clavigera*

Dispersing MPB disseminate a microbial complex of bacteria, yeast and fungi that contribute to beetle success (Mercado *et al.* 2014). Some of the most pathogenic MPB-vectored fungi belong to the *Ophiostomataceae* family (Roe *et al.* 2010). This includes *Grosmannia clavigera* (Robinson-Jeffrey and Davidson), *Leptographium longiclavatum* (Lee, Kim, and Breuil) and *Ophiostoma montium* (Rumbold) (Rice *et al.* 2007b; McAllister *et al.* 2018). Fungal spores are localized to the MPB mycangia, which are specialized invaginations in the exoskeleton of the beetle's head (Six 2003). The timing of fungal dispersion to a new host, fungal growth into the host wood, the development of fungal spores and the packing of those spores into the MPB mycangia is timed in a manner that contributes to the success of the fungus (Figure 1.5).

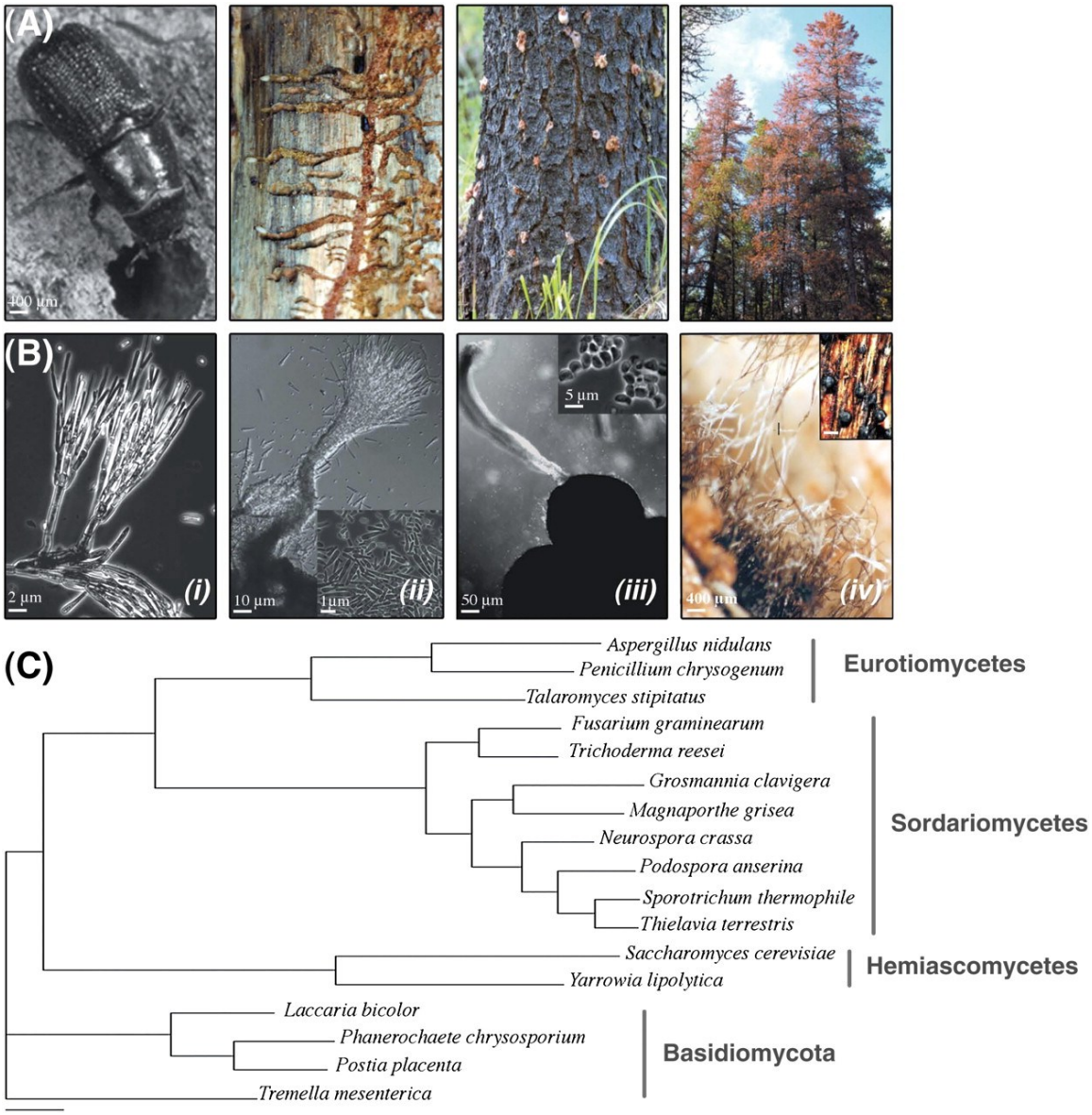


Figure 1.5. The life cycle of MPB and associated *G. clavigera*. (**A**) MPB transport *G. clavigera* to new host trees mid-summer to early fall. While penetrating the tree, MPB deposit the fungus which begins to colonize the tree phloem tissue. Shortly thereafter, the wood-staining *G. clavigera* penetrates the xylem tissue. The MPB larvae feed on phloem in excavated galleries. During feeding, the larvae and beetles accumulate *G. clavigera* spores on their exoskeletons and in their mycangia. This ensures that the fungus will be transported to the next host tree. When MPB entered the tree bark, the host responded with the production of resin that exuded from the wounds. These pitch tubes remain on the tree after the beetles fly to new hosts. MPB attack in conjunction with fungal growth can cause tree mortality, resulting in characteristic

red foliage. (**B**) (**i**) Electron micrograph of *G. clavigera* in its asexual stage characterized by mononematous and synnematous conidiophores which form from single hyphae and bundles of hyphae, respectively. (**ii**) Conidiophores reproducing fungal spores known as conidia (inset). (**iii**) Light micrograph of *G. clavigera* in its sexual phase characterized by a spherical ascocarp oozing fungal spores known as ascospores (inset). (**iv**) Stereomicrograph of conidiophores that grow inside the MPB gallery and ascocarps (inset) on the inner bark of lodgepole pine. (**C**) Phylogenetic tree showcasing the position of *G. clavigera* relative to other pezizomycotina fungi. Acknowledgement goes to DiGuistini, S., Wang, Y., Liao, N. Y., Taylor, G., Tanguay, P., Feau, N., Henrissat, B., Chan, S. K., Hesse-Orce, U., Alamouti, S. M., Tsui, C. K. M., Docking, R. T., Levasseu, A., Haridas, S., Robertson, G., Birol, I., Holt, R. A., Marra, M. A., Hamelin, R. C., Hirst, M., Jones, S. J. M., Bohlman, J., Breuil, C. (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*, 108, pg 2505. This figure was reprinted with permission from the publisher, Proceedings of the National Academy of Sciences.

G. clavigera exists in an obligate symbiotic relationship with MPB (Addison *et al.* 2015). The fungus relies upon MPB to transport spores to susceptible coniferous hosts (Addison *et al.* 2015). *G. clavigera* aids the beetle in overcoming tree defenses by detoxifying hazardous chemicals (DiGuistini *et al.* 2011; Wang *et al.* 2012). This blue-stain pathogen may lower the critical threshold of attack density required for successful beetle colonization (Guérard *et al.* 2000). Furthermore, *G. clavigera* mycelium concentrates tree nutrients in pupal chambers (Paine *et al.* 1997; Goodsman *et al.* 2012) and serves as food for larvae and emerging beetles (Paine *et al.* 1997; Bleiker and Six 2007).

When in the epidemic phase, MPB are preferentially drawn to healthy hosts (Nelson *et al.* 2018). Though MPB have the most reproductive success in healthy trees (Nelson *et al.* 2018), their necrotrophic fungal associates actively destroy tree

cells (Ballard *et al.* 1984). The severe tissue damage inflicted by necrotrophic pathogens commonly results in the activation of jasmonic acid (JA) signaling pathways (Glazebrook 2005; Zhang *et al.* 2017). JA increased around *G. clavigera* inoculation points in both lodgepole and jack pine (Arango-Velez *et al.* 2016), distinguishing the fungus as a necrotroph (Glazebrook 2005; Zhang *et al.* 2017).

Lesions develop around sites of inoculation as a symptom of infection (Rice *et al.* 2007a; Arango-Velez *et al.* 2016). Evidence suggests that lesions are an indicator of the lodgepole pine defense response rather than an indicator of fungal spread (McAllister *et al.* 2018). In grand fir (*Abies grandis* Douglas ex D. Don), lesions are the result of degenerative metabolism in infected tissues and is meant to rob the fungus of nutrition (Wong and Berryman 1977). In grand fir and red pine (*Pinus resinosa* Aiton), jack pine and lodgepole pine, these lesions serve to contain fungal growth (Wong and Berryman 1977; Raffa and Smalley 1988; McAllister *et al.* 2018) and can become necrotic in mature trees, demonstrating the severity of infection (Lusebrink *et al.* 2013).

The severity of plant diseases caused by living organisms indicates the relative resistance or susceptibility of the host and the relative avirulence or virulence of the infecting agents (Pagán and García-Arenal 2018). *G. clavigera* is the most virulent of all MPB fungal associates (Lee *et al.* 2006; Rice *et al.* 2007b). Via hyphal growth, the fungus can rapidly colonize the phloem of susceptible trees (Ballard *et al.* 1984; DiGuistini *et al.* 2011). Overcoming the tree defenses allows mycelium to spread into host xylem tissue (DiGuistini *et al.* 2011). In response to

the intrusion, trees produce ray and axial parenchyma cell ingrowths called tyloses (Clérivet and El Modafar 1994; Clérivet *et al.* 2000). Pectin based gels are also produced by parenchyma cells and deposited in xylem vessels (Clérivet and El Modafar 1994; Clérivet *et al.* 2000). Meant to prevent the axial spread of the fungus via compartmentalization, these structures disrupt the transport of water (Ballard *et al.* 1982; Morris *et al.* 2016; Arango-Velez *et al.* 2016). It has been suggested that infected hosts ultimately die from water deprivation (Hubbard *et al.* 2013). Even in the absence of MPB, *G. clavigera* infection can result in tree mortality (Yamaoka *et al.* 1995). However, mass attacked trees have weakened defenses, augmenting the lethality of *G. clavigera* (Lieutier *et al.* 2009). Therefore, the ability of a host to resist MPB attack is vital for the conifer's survival.

1.2.3 The current outbreak

In western North America, lodgepole pine and MPB ranges overlap extensively (Safranyik *et al.* 2010; Cullingham *et al.* 2011). The plant-pest system has co-evolved such that both species adapt and counter adapt via detection and evasion mechanisms, establishing a state of equilibrium between host and pest that assures the survival of both (Flor 1971; Raffa and Berryman 1987). MPB-invaded stands serve as fuel for wildfire, which results in favorable conditions for lodgepole pine regeneration and, consequently, renewed habitats for MPB (Lotan and Critchfield 1990). The MPB's current prolonged epidemic phase and range expansion are thought to be the result of climate change combined with harvest

and fire suppression programs (Stahl *et al.* 2006; Kurz *et al.* 2008; Raffa *et al.* 2008). Warmer winters resulting from climate change have encouraged the survival of more individuals (Carroll *et al.* 2004). Regional wildfire suppression and harvesting programs in BC have mostly focused on other conifer species, such as ponderosa pine and Douglas fir (*Pseudotsuga menziesii* (Mirbel) Franco) (Klenner *et al.* 2008). Large expanses of lodgepole pine have been allowed to mature and become more susceptible to MPB attack (Taylor and Carroll 2004).

The current MPB epidemic started in the 1990s. As of 2020, the epidemic has spread through pine forests on both sides of the Great Divide from the southern Yukon and Northwest Territories to Northern Mexico, and from the Pacific Ocean to eastern Alberta - including the Cypress Hills that straddle the Alberta-Saskatchewan border – the Dakotas, Nebraska, Colorado and New Mexico (Taylor and Carroll 2004; Cullingham *et al.* 2011; Negrón and Cain 2019; *https://www.fs.fed.us/foresthealth/docs/Range_Maps/FDAR-Mountain-Pine-Beetle-Summary.png*). In western Canada, over 18 million hectares of pine forests have been decimated, including close to 1.5 million hectares in Alberta alone (Hodge *et al.* 2017). Mortality rate of MPB infected trees per year is estimated to be >50% and has been recorded to be as high as 79% in regions of North America (Reed *et al.* 2014). It has been suggested that an attack density of >40 beetles is the threshold for causing mortality in jack pine and lodgepole-jack pine hybrids (Everden and Musso, unpublished data), and lodgepole pine trees that possess

larger traumatic resin ducts have been shown to have greater success at fighting off the beetle (Zhao and Erbilgin 2019).

Strategies for combatting the spread of MPB include restructuring forest composition to diversify species and age classes and the detection and removal of infested trees exhibiting symptoms of attack (Hodge *et al.* 2017). MPB-infested trees usually retain green foliage up to one year following attack, after which needles gradually fade to yellow then red (Wulder *et al.* 2006; Page *et al.* 2012). One year following attack, over 90% of infested trees have red needles (Wulder *et al.* 2006). Two to four years following the initial infestation, red needles fall off completely (Natural Resources Canada 2017a). The current epidemic is negatively impacting timber and pulp industries, water quality and quantity, biodiversity, wildlife populations, recreation, real estate values, and cultural resources (Shore *et al.* 2006; Kurz *et al.* 2008; Corbett *et al.* 2016). Since 2004, Alberta and Saskatchewan provincial governments have invested over $500 million dollars to mitigate the impact of the destructive pest and its harmful fungal associates (Hodge *et al.* 2017).

1.3 Lodgepole pine defense responses to *G. clavigera*

1.3.1 Constitutive and inducible defenses

To constrain *G. clavigera* colonization, conifers exhibit both constitutive and inducible defense mechanisms (Franceschi *et al.* 2005; Ott *et al.* 2012; Keeling

and Bohlmann 2006; Lieutier *et al.* 2002). Constitutive defenses are always present and serve as the first responders to invaders (Agrios 2005). Most defense mechanisms are in place constitutively and then are induced and become strengthened upon pest or pathogen perception (Nagy *et al.* 2000; Franceschi *et al.* 2005; Van Loon *et al.* 2006). MPB begins its attack by boring through the tree bark, which is composed of all tissues outside of the vascular cambium, including phloem (Taiz *et al.* 2015). The beetle causes damage to both phloem and xylem tissues when excavating nuptial chambers and egg galleries underneath the bark. Upon herbivore damage, radial resin ducts found in the xylem and phloem immediately release resin (Krokene and Nagy 2012). Rapid resin flow can prevent attacking MPB from entering the tree by gluing together their mouthparts, flushing out the beetles or engulfing them entirely (Taiz *et al.* 2015). Bark beetles have been shown to preferentially attack ponderosa pines that possess fewer resin ducts (Kane and Kolb 2010). Furthermore, smaller resin duct area and density were significant factors for predicting lodgepole and limber pine (*Pinus flexilis* E. James) mortality in the field (Ferrenberg *et al.* 2014; Zhao and Erbilgin 2019). MPB fungal associate-induced defenses include the production of traumatic resin ducts that form a continuous network with constitutive radial resin ducts (Nagy *et al.* 2000; Hudgins and Franceschi 2004). In Norway spruce, there is evidence that traumatic resin ducts impart acquired resistance to subsequent pathogen challenge (Christiansen *et al.* 1999).

Though resin flow from both constitutive and traumatic resin ducts is a component of physical defenses, the resin itself contains several toxic carbon (C)-based secondary metabolites (Franceschi *et al.* 2005; Keeling and Bohlman 2006; Kovalchuk *et al.* 2013) that are highly plant species-specific (Heldt and Piechulla 2010). Examples include products of the isoprenoid (terpenoids) and phenylpropanoid (phenolics) pathways (Franceschi *et al.* 2005; Keeling and Bohlman 2006; Kovalchuk *et al.* 2013). Conifer resin is composed of a mixture of diverse terpenoids including roughly equal parts mono- and diterpenes, along with smaller concentrations of sesquiterpenes (Keeling *et al.* 2008). Terpenoids possess antimicrobial, -fungal and -feedant properties (Bohlmann and Keeling 2008). Their modes of action are diverse (Guimarães *et al.* 2019; Sadasivam and Thayumanayan 2019). Miron *et al.* (2014) found that the human fungal pathogen *Trichophyton rubrum* (Castellani) Sabouraud experienced cell wall destabilization in the presence of monoterpenes *in vitro*. Robert *et al.* (2010) suggested that volatile monoterpenes may be fumigant toxins that infiltrate the respiratory system of pests, such as the white pine weevil (*Pissodes strobe* Peck).

Phenolics also possess anti-microbial, -fungal and -feedant properties (Franceschi *et al.* 2005). For example, Hunter (1974) found that the fungal pathogen *Rhizoctonia solani* J.G. Kühn exuded polygalacturonase, a pectin hydrolase that can degrade plant cell walls. Polygalacturonase was deactivated by induced polyphenolics in cotton (*Gossypium hirsutum* Linnaeus) (Hunter 1974). *G. clavigera* hyphal growth caused phloem polyphenolic parenchyma (PP) cells to

swell in the sapwood of lodgepole and jack pine (Arango-Velez *et al.* 2014). In Norway spruce, swollen PP cells collapsed adjacent sieve cells, creating a physical barrier to fungal spread in the phloem tissue (Franceschi *et al.* 2000). Other physical defense measures include investment in lignin, which, like phenolics, are produced through phenylpropanoid metabolism (Zabel and Morell 1992; Franceschi *et al.* 2005; Kovalchuk *et al.* 2013). Conifer colonizing blue stain fungi like *G. clavigera* have not evolved the full capacity to degrade lignin and other structural elements of wood, such as cellulose (Zabel and Morell 1992; Kovalchuk *et al.* 2013).

To compliment the anti-feedant and -fungal action of C-based defense measures, nitrogen (N)-based defense-associated proteins enhance plant resistance to invading pathogens and pests. In response to pathogen attack, pathogenesis-resistance (PR) proteins in conifers are highly up-regulated (Fossdal *et al.* 2001, 2006; Adomas *et al.* 2007; Islam *et al.* 2010). Conifer PR proteins include chitinolytic chitinases (Neuhaus 1999; Liu *et al.* 2005; Kolosova *et al.* 2014), glucancases (Boller 1985; Asiegbu *et al.* 1995) and thaumatin-like proteins (Osmond *et al.* 2001; Selitrennikoff 2001), all of which inhibit hyphal growth by degrading fungal cell wall components like chitin and glucan. PR protein modes of action have been explored in both conifer and non-conifer species (Fossdal *et al.* 2001; Monteiro *et al.* 2003; Ralph *et al.* 2006a; Chowdhury *et al.* 2015). Chitinases (Fossdal *et al.* 2006) and glucanases (Asiegbu *et al.* 1995) were both produced in Norway spruce (*Picea abies* Miller) tissues in response to *Heterobasidion*

annosum (Fries) Brefeld infection. Galindo *et al.* 2012 found that in white spruce (*Picea glauca* (Moench) Voss) stems, chitinases and thaumatin-like proteins were up-regulated during the transition from active growth to dormancy, possibly as constitutive defense elements. Osmotin-like proteins, which belong to the same class of PR proteins as thaumatins, showed antifungal activity in *Solanum nigrum* Linnaeus var. *indica* against necrotrophic fungi by inhibiting hyphal growth and spore germination, as well as reducing the viability of fungal spores (Chowdhury *et al.* 2015).

Other N-based defense-associated proteins include peroxidases, which have been implicated in the enhancement of cell wall toughness (Zhong and Ye 2014). Warinowski *et al.* (2016) isolated extracellular lignin-bound peroxidases from Norway spruce tissue cultures that could catalyze the polymerization of lignin monomers (monolignols). Peroxidases were induced in the roots of Norway spruce in response to *Pythium dimorphum* (F. F. Hendrix and W. A. Campbell) infection (Fossdal *et al.* 2001), implicating these enzymes as a defense response to parasitic oomycetes. Dirigent proteins are also involved in lignin production and are thought to direct the stereospecific coupling of phenolic precursors (Davin and Lewis 2000; Ralph *et al.* 2006a). Dirigent-like proteins were highly up-regulated in response to pathogen challenge in conifers (Ralph *et al.* 2006a). Defensins, which are small cysteine-rich defense-associated peptides, inhibited the growth of filamentous fungi *R. solani* and *Phaeoisariopsis personata* (Berk and M. A. Curtis) van Arx *in vitro* (Olli and Kirti 2006). Defensins can bind and interact with the

negatively charged cell membranes of pathogens, causing an increase in membrane permeability that results in the leakage of cell contents (Lacerda *et al.* 2014).

Perception of pest or pathogen invasion activates inducible defenses local to the inflicted tissue and systemically, throughout the plant (Agrios 2005; Franceschi *et al.* 2005). The intense localized induced response of plants to wounding is often termed the hypersensitive response (HR), which has been associated with rapid cell death through the production of reactive oxygen species (ROS; Lieutier 2002; Balint-Kurti 2019). By sacrificing a moderate number of cells to save the rest of the plant, the HR may limit a pest's or pathogen's access to water and nutrients (Balint-Kurti 2019). However, programmed cell death as an inducible defense measure is generally less effective against, and may sometimes be beneficial to, necrotrophic pathogens, which require dead host tissue to complete their life cycle (Laluk and Mengiste 2010; Balint-Kurti 2019). Nonetheless, the lesion that forms following *G. clavigera* infection may, in part, indicate an HR around the point of inoculation (Franceschi *et al.* 2005).

1.3.2 Pathogen recognition and signaling

Identification of pathogen detection and signaling mechanisms is necessary to understand the regulatory networks associated with induced defense responses in lodgepole pine to the pathogen *G. clavigera*. Plant pattern recognition receptor

proteins, such as receptor kinases with leucine-rich repeat (LRR) domains or lysine motifs, recognize highly conserved avirulent features know as pathogen-associated molecular patterns (PAMPs; Jwa and Hwang 2017; Liang and Zhou 2018). The fungal cell wall component chitin is a PAMP (Jwa and Hwang 2017). Identification of PAMPs by plasma membrane receptor kinases can trigger mitogen-activated protein kinase (MAPK) cascades (Jwa and Hwang 2017). This occurs via the phosphorylation of proteins throughout the cytosol and into the nucleus to modulate the expression of transcription factors and regulators (Meng and Zhang 2013). This pattern triggered immunity (PTI) includes the up-regulation of genes involved in innate immune responses, such as transcription factors, enzymes, hormones, peptides and antimicrobial chemicals (Tena *et al.* 2011).

Pathogens that have adapted to overcome host defenses produce and release effector proteins that interfere with PTI (Jwa and Hwang 2017). Necrotrophs produce necrotrophic effectors, a diverse group of molecules that induce cell death in host tissues (Tan *et al.* 2010; Tan *et al.* 2015). Virulence effectors modify the plant's structure, metabolism or hormonal regulation to the advantage of the pathogen (Taiz *et al.* 2015). For example, effectors can inhibit plant cysteine proteases that promote programmed cell death in response to pathogen attack (Stergiopoulos and de Wit 2009; Zamyatnin 2015). Recognition of virulence effectors by resistance proteins can elicit effector-triggered immunity (ETI; Jwa and Hwang 2017), which provides a more robust and amplified defense

response compared with PTI (Cui *et al.* 2015; Jwa and Hwang 2017). Resistance genes are very specific to the plant-pathogen system (Agrios 2005). The most prevalent resistance proteins contain a nucleotide binding site leucine-rich repeat (NBS-LRR) domain, with accompanying domains, such as Toll/interleukin-1 receptor-like (TIR; McHale *et al.* 2006) The NBS domain is also called the NB-ARC domain (which stands for nucleotide binding adaptor shared by nucleotide-binding oligomerization domain with leucine-rich repeat proteins, apoptotic protease activating factor 1, resistance proteins and cell-death protein 4) (McHale *et al.* 2006). MAPK cascades, oxidative (respiratory) burst and the activation of various defense response genes are all aspects of ETI (Jwa and Hwang 2017). Respiratory burst results in the generation of ROS that facilitate apoptosis as part of the HR (Jwa and Hwang 2017). ROS in conjunction with peroxidase enzymes can strengthen the plant cell wall by catalyzing cross-linkages between glycoproteins and polymers (Tenhaken 2015; Jwa and Hwang 2017).

1.3.3 Hormone signaling modulates defense-related gene expression

Another component of ETI is the induction of hormone signaling pathways (Taiz *et al.* 2015). Both hormone production and the availability of active vs. inactive forms of hormone compounds can be up-regulated by herbivore, parasite or pathogen challenge (Pieterse *et al.* 2012; Vos *et al.* 2013; Taiz *et al.* 2015). JA and ethylene are plant hormones commonly associated with the induced defense response (Pieterse *et al.* 2009; Ruan *et al.* 2019; Broekgaarden *et al.* 2015). JA and

ethylene signaling mechanisms often work in tandem, yielding a stronger front to pathogen challenge (Taiz *et al.* 2015). In conifers, exogenous application of methyl-jasmonate (MeJA), a volatile JA derivative, increased expression of the ethylene biosynthesis gene 1-aminocyclopropane-1-carboxylate oxidase (Hudgins and Franceschi 2004; Ralph *et al.* 2006b; Hudgins *et al.* 2006). When applied exogenously to conifer stems, both ethylene and MeJA induced the swelling of PP cells, accumulation of phenolics and formation of traumatic resin ducts (Franceschi *et al.* 2002; Hudgins *et al.* 2004). MeJA application resulted in massive up-regulation of terpene synthase genes and the accumulation of terpenoids in the stems of Norway and Sitka spruce (*Picea sitchensis* Bongard) (Martin *et al.* 2002; Miller *et al.* 2005). In slash and loblolly pine (*Pinus taeda* Linnaeus) inoculated with pathogenic fungus *Ophiostoma minus* (Hedgc.) H.P. Sydow, production of ethylene was associated with the biosynthesis of toxic monoterpenes in the lesion tissue (Popp *et al.*1995).

In model plants, the JA signaling pathway comprises two branches termed the myelocytomatosis transcription factor (MYC) branch and the ethylene response factor (ERF) branch (Figure 1.6 A; Pieterse *et al.* 2012; Broekgaarden *et al.* 2015). Upon insect herbivory, JA signaling activates the MYC branch and simultaneously suppresses of the ERF branch (Broekgaarden *et al.* 2015). The basic helix-loop-helix transcription factor, MYC2, has been well categorized as a positive regulator of flavonoid biosynthesis in both *Arabidopsis thaliana* (Linnaeus) and *Oryza sativa* (Linnaeus) (Dombrecht *et al.* 2007; Ogawa *et al.*

2017). Upon attack by a necrotrophic pathogen, JA signaling frees ERF1, which has been shown to initiate the transcription of chitinases and defensins in non-conifer species (Solano *et al.* 1998; Pré *et al.* 2008). Activation of the JA and ethylene pathways by necrotrophic pathogens suppresses the MYC branch (Broekgaarden *et al.* 2015). Ethylene stimulates a cascade of transcriptional regulation that includes nuclear proteins ethylene-insensitive 3 (EIN3) and ERF1 (Figure 1.6 B; Qiao *et al.* 2012).

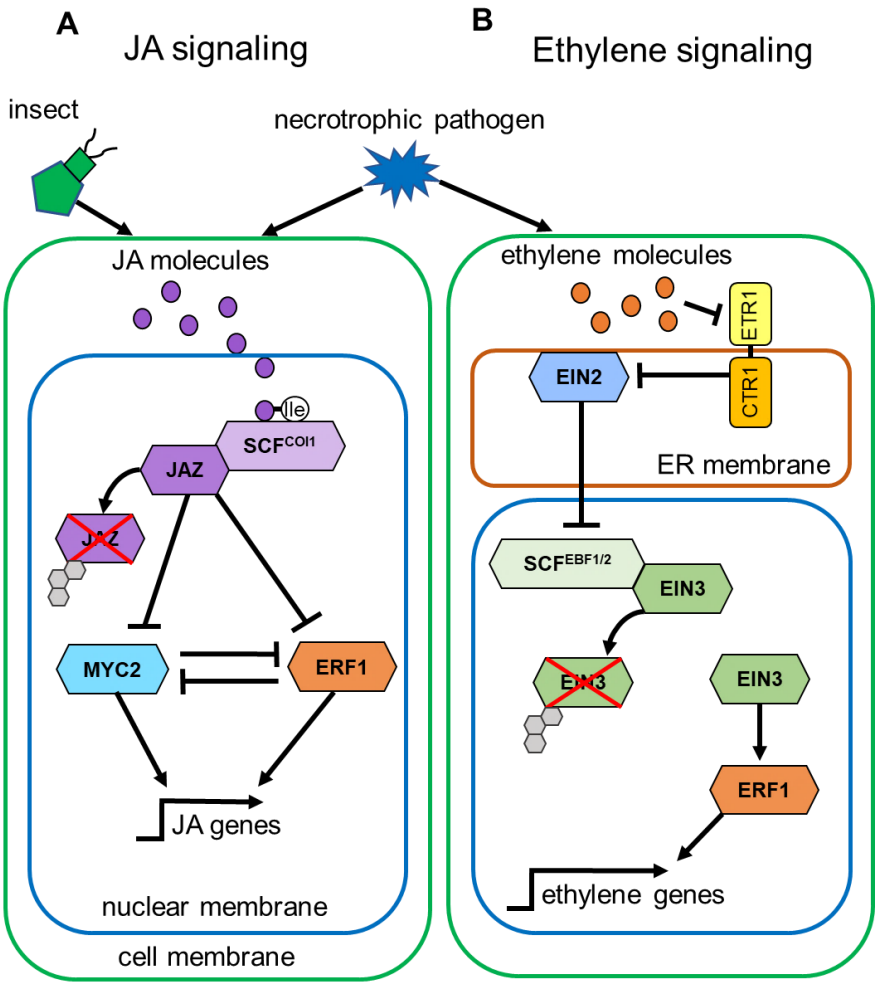


Figure 1.6. Hormone signaling modulates defense-related gene expression. (**A**) JA is produced in response to pest or necrotrophic pathogen attack. JA is conjugated with the amino acid isoleucine to produce JA-isoleucine (JA-Ile), which binds to the SCF^{COI1} protein complex that targets members of the JAZ protein family (Thines *et al.* 2007). This interaction results in the ubiquitination of JAZs, which frees MYC2 to activate JA-dependent wounding-responsive genes (Thines *et al.* 2007). The degradation of JAZs also frees ERF1 to activate expression of JA-dependent pathogen-responsive genes (Pieterse *et al.* 2012; Broekgaarden *et al.* 2015). (**B**) Volatile ethylene molecules are produced in response to necrotrophic pathogen attack. Ethylene binds to ERT1 located in the endoplasmic reticulum (ER) membrane, deactivating CRT1 (Qiao *et al.* 2012). Once the CRT1 repressor is disabled, signaling through the positive regulator EIN2 is activated (Qiao *et al.* 2012). Carboxyl-terminal EIN2 fragments enter the nucleus to suppress the ubiquitination of EIN3 transcription factors. This causes EIN3 to build-up and induces the expression of ERF transcription factors, which initiate transcription of ethylene-dependent defense-associated genes (Qiao *et al.* 2012). The arrows and bars indicate activating and inhibiting regulatory relationships, respectively. Abbreviations include jasmonic acid (JA), S-phase kinase-associated protein 1 - cell division control protein 53 - F-box protein - coronatine-insensitive protein 1 complex (SCF^{COI1}), jasmonate (zinc-finger expressed in inflorescence meristem)-domain transcription factor (JAZ), myelocytomatosis transcription factor 2 (MYC2), ethylene response factor 1 (ERF1), ethylene receptor 1 (ERT1), calreticulin 1 (CRT1), ethylene-insensitive protein 2 (EIN2), S-phase kinase-associated protein 1 - cell division control protein 53 - F-box protein - ethylene-insensitive protein 3 binding F-box 1 and 2 complex ($SCF^{EBF1/2}$), and ethylene-insensitive protein 3 (EIN3).

1.3.4 Other regulators of the defense response

Aside from JAZs, ERFs and MYCs, plant transcription factors with known roles in the defense response include basic leucine zipper domains (bZIPs), myeloblastosis oncogenes (MYBs), transcription factors with a WRKY amino acid sequence deoxyribonucleic acid (DNA) binding domain at the N-terminus (WRKYs), and no apical meristem, *Arabidopsis* transcription activation factor and

cup-shaped cotyledon transcription factors (NACs). Transcription factors recognize specific DNA-binding motifs and activate or repress a particular gene (Alves *et al.* 2014). The bZIP domain family is the largest of the eukaryotic transcription factors and is known to regulate genes involved in abiotic stress mitigation, flower development and pathogen defense (Jakoby *et al.* 2002). Recognition of pathogen induced-elicitors may enable bZIPs to up-regulate HR- and innate immunity-related genes (Alves *et al.* 2014). MYB transcription factors in plants are involved in regulation of gene expression related to primary and secondary metabolism, development, and responses to abiotic and biotic stress (Ambawat *et al.* 2013). MYBs have been shown to induce lignification in loblolly pine (Patzlaff *et al.* 2003) and contribute to the accumulation of sesquiterpenes in white spruce and loblolly pine (Bedon *et al.* 2010).

Transcription factors unique to plants include WRKYs and NACs. These transcription factors are key modulators of the plant response to abiotic and biotic stresses (Alves *et al.* 2014; Schluttenhofer and Yuan 2015). Initiation of MAPK cascades in response to pathogen invasion has been shown to induce expression of WRKYs in *Arabidopsis thaliana* (Linnaeus) (Ishihama and Yoshioka 2012). In response to pathogen invasion, WRKYs can interact and form protein complexes to bolster defense phenotypes (Alves *et al.* 2014). In maritime pine (*Pinus pinaster* Aiton), Pascual *et al.* (2015) found that the NAC transcription factor family was involved in various aspects of growth, development and stress responses. A Norway spruce NAC was found to be responsive to *Heterobasidion annosum* (Fr.)

Bref. inoculation and MeJA application and appeared to participate in the control of flavonoid production (Dalman *et al.* 2017).

1.4 N availability impacts tree physiology

1.4.1 N acquisition and assimilation

Plants require N in larger quantities than any other essential mineral (Taiz *et al.* 2015). Slow growing perennial species like conifers may rely heavily on internal cycling, especially at the beginning of the growing season (Lupi *et al.* 2013). Trees in the family *Pinaceae* rely upon fungal symbionts called ectomycorrhizae to efficiently acquire and reduce inorganic N in the soil into organic compounds (Lupi *et al.* 2013). Fungal hyphae transfer organic N compounds like amino acids to the tree roots in exchange for carbohydrates that sustain the fungus (Chalot and Brun 1998). There is evidence that boreal plants, such as Scots pine (*Pinus sylvestris* Loureiro) and Norway spruce, also take up amino acids from the soil directly (Gruffman *et al.* 2012). Assimilated N is transported through xylem tissue to living plant cells, where it is used to produce cellular structures required for plant growth, development and defense (Tegeder and Masclaux-Daubresse 2018). Plants require N to produce photosynthetic machinery, such as ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) and chlorophyll (Taiz *et al.* 2015). The production of N-based RuBisCO can be increased by fertilization (Cheng and Fuchigami 2000). The RuBisCO active state

is decreased by increases in both fertilization concentrations and duration of application, suggesting that RuBisCO can serve as a storage protein (Cheng and Fuchigami 2000). Additionally, an excess of RuBisCO may result in higher steady-state carbon dioxide assimilation and water use efficiency (Cheng and Fuchigami 2000). Therefore, fertilization has implications for primary (growth- and development-related) and secondary metabolism.

1.4.2 Growth is impacted by N availability

N has been shown to be an essential growth limiting nutrient for northern temperate tree species (Brockley 2001; Vadeboncoeur 2010; Högberg *et al.* 2013). Without sufficient N reserves, proteins, nucleic acids and hormones required for growth processes cannot be produced (Taiz *et al.* 2015). In *Eucalyptus* and *Melaleuca* species, deficiency in N has been shown to result in the reduction of stem elongation and above-ground biomass (Nguyen *et al.* 2003). N deficiency results in a remobilization of assimilated N that can negatively impact photosynthetic capabilities and growth rate (Tang *et al.* 2019). For deciduous trees, N deficiency symptoms appear first on the oldest leaves as a uniform lightening in color (Broschat 2017). As the deficiency progresses, the entire plant quickly becomes light green in color and growth rate declines sharply (Broschat 2017). N-limited *Eucalyptus* species, Norway spruce and silver birch (*Betula pendula* Roth) translocated additional photosynthates to the roots to allow for an increase in the

size of the root system (Ericsson 1995; Nguyen *et al.* 2003; Miller and Cramer 2004).

Biomass is accumulated through the replication and differentiation of meristematic cells (Taiz *et al.* 2015). Meristem tissue can be found in the roots, leaves and stems of vascular plants (Taiz *et al.* 2015). In stem tissue, meristematic cells make up the vascular cambium, which lies between phloem and xylem tissue and provides partially undifferentiated cells for secondary structures (Taiz *et al.* 2015). Stockfors and Linder (1998) found that fertilization of young Norway spruce increased tree stem diameter possibly by stimulating vascular cambial activity. Hacke *et al.* (2010) suggested that increased concentrations of ammonium nitrate (NH_4NO_3) fertilizer enhanced secondary xylem growth in hybrid poplar saplings (*Populus trichocarpa* (Torrey and Gray) × *deltoides* Bartram ex Marshall) by stimulating cambial activity and increasing cell size. N impacts on cellular proliferation in the cambial zone can be observed by microscopy (Bossinger and Spokevicius 2018).

A surplus of N can lead to an overabundance of growth, resulting in weakened stems (Fuller and Jellings 2003). Plant N saturation is reached when the availability of inorganic N exceeds the plant's nutritional demand (Wilson and Skeffington 1994). N input that is in excess of vegetation requirements can result in alterations in soil pH that can be harmful for forest ecosystems (Wilson and Skeffington 1994). Increased N availability also impacts annual growth cycles of perennial plants and can delay phenological events like flowering and growth

cessation (Wang and Tang 2019). Knowledge of the maximum and minimum N fertilization thresholds of trees is imperative for enhanced forest management (Government of Alberta 2006).

1.4.3 Defense is impacted by N availability

The literature regarding pathogen attacked trees has highlighted the influence of N availability on the production of defense-related compound precursors and proteins (Fagard *et al.* 2014; Sun *et al.* 2020). In response to increased NH_4NO_3 availability, Tomova *et al.* (2005) discovered that there was a decrease of fungistatic phenolic compounds in the roots of beech trees (*Fagus sylvatica* Linnaeus) and Norway spruce. In lower compared to higher N treatments, Sitka spruce inoculated with the fungal pathogen *Phacidium coniferarum* (G.G. Hahn) produced higher concentrations of resin and polyphenols in infected stem tissue (Wainhouse *et al.* 1998). Though research on gymnosperms is lacking, alterations in N availability impacts defensive protein production in angiosperms (Sun *et al.* 2020). Verly *et al.* (2020) treated *A. thaliana* with one of three nitrate (NO_3^-) concentrations, 2 mM, 10 mM, or 26 mM NO_3^-, followed by the application of BION© (Syngenta AG, Basel, Switzerland) to simulate pathogen attack. They found that induced levels of PR protein expression in BION©-treated *A. thaliana* were significantly higher under 10 mM NO_3^- conditions when compared with the 2 mM and 26 mM fertilization regimes (Verly *et al.* 2020). Furthermore, when NO_3^--treated *A. thaliana* were infected with the

necrotrophic pathogen *Dickeya dadantii* (Samson, Legendre, Christen, Fischer-Le Saux, Achouak and Gardan), exogenous application of MeJA resulted in a decrease of maceration symptoms for the 10 mM NO_3^- plants only (Verly *et al.* 2020).

Higher levels of C-based herbivore-deterring chemicals in N-deficient trees may correlate with the retention of N for other processes, such as growth (Mihaliak and Lincoln 1985). Waring and Pitman (1985) found that improved N nutrition resulted in an increase in MPB attack per square meter of lodgepole pine bark surface. Interestingly, the increased N availability also hastened lodgepole pine recovery, as indicated by significantly increased growth following the invasion (Waring and Pitman 1985). Cook *et al.* (2015) found that fertilization of mature lodgepole pine with low levels of urea, a N-based compound, resulted in an increase in constitutive resin flow in response to wounding compared with the application of high levels of urea. The authors suggested that the reduction in resin flow accompanied increased growth in lodgepole pine resulting in a negative correlation between the two parameters (Cook *et al.* 2015).

1.4.4 Nutrient-defense balance hypotheses

Observed trade-offs between growth and defense have been summarized into a variety of hypotheses that contextualize the allocation of nutrients to different plant processes (Stamp 2003). The carbon-nutrient balance (CNB) hypothesis states that resource allocation towards growth, defense and other

processes is influenced by genetic and environmental factors (Tuomi *et al.* 1988; Tuomi *et al.* 1991). The phenotypic plasticity of a given plant species determines the extent to which resource allocation is relatively fixed or flexible (Tuomi *et al.* 1988; Tuomi *et al.* 1991). The flexible component of resource allocation to growth vs. defense is influenced by environmental factors, such as nutrient availability, that impact the internal (assimilated) C:N ratios (Matyssek *et al.* 2002; Matyssek *et al.* 2005). A species' phenotypic plasticity determines the degree to which the plant's internal C:N ratio influences the allocation of C and N resources towards defense (Tuomi *et al.* 1988; Tuomi *et al.* 1991).

The growth-differentiation balance (GDB) hypothesis views internal resource reserves as affecting the balance between the investment of C in the production of biomass (growth) and the chemical and structural modification of biomass (differentiation) (Herms and Mattson 1992). While growth comprises any process that requires substantial cell division and elongation, differentiation includes the production and maintenance of defense mechanisms and storage structures (Stamp 2004). The GDB hypothesis states that growth processes and defense-related metabolism respond in terms of a trade-off (Herms and Mattson 1992). Nutrient deprivation restricts growth more than photosynthetic capacity (Herms and Mattson 1992). Therefore, nutrient-limited conditions increase allocation of photosynthates towards defense processes (Herms and Mattson 1992).

The optimal defense (OD) hypothesis also states that there is a tradeoff between defense and other plant functions, such as the construction of reproductive tissues and the accumulation of biomass (Stamp 2003). The OD hypothesis predicts that plants experiencing abiotic stress, like nutrient deficiency, are less able to defend themselves against pathogens and pests (Stamp 2003). Because they divert resources away from growth, constitutive defenses against herbivores and pathogens are costly in terms of construction and maintenance (Stamp 2003). Phenotypic plasticity allows plants to maximize fitness by optimizing the expression of defensive traits (Bakhtiari *et al.* 2019). A limited supply of constitutive defense elements is concentrated in the most vulnerable tissues, such as young leaves (Herms and Mattson 1992; Meldau *et al.* 2012). Studies have shown that this applies to phenolic compounds (Moreira *et al.* 2012; Massad *et al.* 2014), volatile organic compounds (Radhika *et al.* 2008) and defensive proteins (Shudo and Iwasa 2002). This runs counter to the GDB hypothesis, which states that defense processes are amplified under nutrient-limited conditions (Herms and Mattson 1992).

1.5 Transcriptomics to analyze the lodgepole pine defense response to *G. clavigera*

1.5.1 A transcriptomics approach to study the current MPB outbreak

Scientific research is a critical component of the response effort to minimize the impact of the current MPB epidemic (Hodge *et al.* 2017). Comprehensive assessments of the underlying mechanisms that contribute to host susceptibility are important for developing innovative management strategies (Cullingham *et al.* 2019). A deeper understanding of the impacts of N-based fertilization on pine trees under attack by MPB and their fungal associates may increase the efficacy of reforestation efforts (Government of British Columbia 2006). The application of powerful transcriptomic tools allows for the exploration of transcriptional regulatory mechanisms impacting the lodgepole pine response to the MPB-vectored *G. clavigera* under varying levels of N availability (Wang *et al.* 2009).

1.5.2 Next-generation sequencing (NGS)

The first step towards implementing these transcriptomic tools is the extraction and sequencing of nucleic acids from lodgepole pine tissue (Kukurba and Montgomery 2015). Early strategies for sequencing complex genomes relied on low-throughput, high quality sequencing (Sanger *et al.* 1977). Chain-termination or Sanger sequencing was first described in the 1970s (Sanger *et al.* 1977) and was the primary methodology until approximately 2005 (Margulies *et al.* 2005). The method includes many rounds of chain-termination polymerase chain reaction (PCR) on specific genes of interest, size-based sequence separation using gel electrophoresis, and gel analysis for sequence determination (Sanger *et al.* 1977). Sanger sequencing is still considered the gold standard because of its low

error rate and long read length (> 700 base pairs) (Thomas *et al.* 2012). Some drawbacks of Sanger sequencing are the labor-intensive cloning process and the overall cost per gigabase (approximately $400,000 USD) (Thomas *et al.* 2012). Practical NGS began in 2005 with the development of automated pyrosequencing (Margulies *et al.* 2005). Numerous methods have been and continue to be developed. Thus far, most processes produce over a million short reads (< 300 base pairs) from a single cell, tissue, or environmental sample at low per-base costs (Shendure and Ji 2008). This places the focus on greater depth of coverage and creates significant challenges for down-stream processes (Shendure and Ji 2008; Roy *et al.* 2016). Commercially available NGS platforms include Illumina sequencers, such as the MiSeq, HiSeq and NextSeq systems (Illumina, Inc., San Diego, California, United States). Each sequencing approach has its own unique biases (Huse *et al.* 2007; Wang *et al.* 2010). NGS can generate FASTQ files that contain the sequence data along with the quality of each base pair called (Cock *et al.* 2010). NGS has become a method of choice for many different applications (Wang *et al.* 2010).

1.5.3 RNA sequencing (RNA-Seq) and bioinformatic analysis

Extracted ribonucleic acid (RNA) can serve as a template to produce complementary DNA (cDNA) libraries, which can then be sequenced using NGS. Sequencing all cDNA fragments in each library yields a snapshot of the library's transcriptome (Wang *et al.* 2009). The transcriptome comprises the whole set of

transcripts of a cell or collection of cells expressed at a given time and under certain conditions, and the genome is the entire set of DNA, which includes all genes (Wang *et al.* 2009). NGS yields high amounts of messenger RNA (mRNA) fragments (reads) that can be counted by alignment to a reference transcriptome or genome (Vijay *et al.* 2013). The relative abundance of reads can be indicative of the level of proteins produced in response to genetic differences or varied environmental factors (Vogel and Marcotte 2012). If a reference is unavailable for read enumeration, it can be assembled *de novo* using a variety of techniques, including the de Bruijn graph method (Conesa *et al.* 2016; Zhang *et al.* 2011). De Bruijn graph transcriptome assembly uses small pieces of sequences called k-mers to arrange reads into contiguous sequences with the goal of reconstructing full-length transcripts (Pevzner *et al.* 2001). Lodgepole pine lacks both a comprehensive genome and transcriptome reference, and *de novo* assembly must be performed prior to read abundance enumeration. If counts vary between samples grouped by treatment type, computational and statistical tools can determine if that variation is significant. When this is done for all reads in the samples, it is called differential expression analysis.

A variety of approaches exist for performing differential expression analysis, including variations in count normalization techniques and dispersion approximation protocols (Kvam *et al.* 2012; Soneson and Delorenzi 2013). Given that sequencing depth and library sizes will differ between samples, the removal of low abundance reads, the application of normalization both within and between

different treatment types, and the calculation of the common dispersion are all required before significance testing. Some common differential expression tools used today are edgeR (Robinson *et al.* 2010) and DESeq2 (Love *et al.* 2014), which both employ a negative binomial probability distribution for detecting significantly differentially expressed genes. To support differential expression results, reverse transcription quantitative polymerase chain reaction (RT-qPCR) data can be used to yield a snapshot of transcript information at specific time points (Nolan *et al.* 2006). RT-qPCR of genes implicated in the lodgepole pine defense response to *G. clavigera* can be used to validate the RNA-Seq data generated for our study.

1.6 Current study

The goal of this project was to identify the gene expression response patterns of young lodgepole pine trees to *G. clavigera* inoculation and contrasting levels of N availability in growth chamber conditions using an RNA-Seq approach. To approach that goal, the following specific objectives were addressed:

(1) We optimized and compared two state of the art transcriptome assembly platforms. We focused on two de Bruijn graph assemblers, the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) and Trans-ABySS (Robertson *et al.* 2010). The outcomes of assembling with distinct k-mer values were compared using the following criteria: assembly time, N50 length, maximum contig length,

the number of contigs and percentage of reads mapped back to the respective assemblies.

We hypothesized that the CLC Genomics Workbench v9.5.2 (*www.qiagenbioinformatics.com*) would produce a faster assembly, but that Trans-ABySS v1.5.5 (Robertson *et al.* 2010), given its multiple k-mer de Bruijn graph algorithm, would produce a higher quality assembly.

(2) We conducted a controlled environment experiment to evaluate the effects of varying levels of N availability on the responses to *G. clavigera* infection in three-year-old lodgepole pine. The experiment was partially carried out to provide experimental materials for future analyses that will complement the work presented in this thesis.

We hypothesized that higher N availability would result in an increase in foliar N concentration in the lodgepole pine seedlings. In addition, we hypothesized that different levels of N fertilization, such as low (1 mM) or high (10 mM) concentrations of NH_4NO_3, would impact the lodgepole pine defense against *G. clavigera*, as measured by differences in lesion development.

(3) We identified the defense response patterns of young lodgepole pine trees given low or high soil N supply and inoculated with *G. clavigera* in growth chamber conditions. An RNA-Seq experiment was carried out with samples obtained from an earlier Cooke lab experiment, which tested the responses of lodgepole and jack pine to *G. clavigera* inoculation under control (0 mM), low (0.3

mM) or high (10 mM) NH_4NO_3 conditions. cDNA sequencing from 32 libraries (1 organism (lodgepole) × 2 inoculation treatments (mock- or fungal-inoculated) × 2 N treatments (low or high) × 2 tissue types (phloem or xylem) × 4 collection points = 32) was performed using Illumina NGS. This project produced a comprehensive lodgepole pine transcriptome with accompanying annotations, along with a robust analysis of significantly differentially expressed defense-related transcripts.

(4) We data mined differentially expressed transcripts identified in objective (3) to uncover patterns of gene expression. These analyses included a network analysis approach to identify genes that ware co-expressed with key transcription factors.

We hypothesized that N availability: (1) affects well-characterized components of lodgepole pine defense against *G. clavigera,* such as monoterpene synthesis, (2) modulates expression of genes thought to be important in mediating *G. clavigera*-elicited responses in lodgepole pine, and (3) alters the ratio of N-based to C-based defense-related genes that are up-regulated in response to *G. clavigera* inoculation, with a greater proportion of N-based defense genes up-regulated in response to higher N availability.

1.7 References

Addison, A., Powell, J. A., Bentz, B. J., Six, D. L. (2015). Integrating models to investigate critical phenological overlaps in complex ecological interactions: The

mountain pine beetle-fungus symbiosis. *Journal of Theoretical Biology*, 368, 55–66.

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.

Agrios, G. N. (2005). *Plant pathology*. (5th ed.). Elsevier Academic Press, New York, New York, United States.

Alves, M., Dadalto, S., Gonçalves, A., Souza, G. D., Barros, V., Fietto, L. (2014). Transcription factor functional protein-protein interactions in plant defense responses. *Proteomes*, 2, 85-106.

Ambawat, S., Sharma, P., Yadav, N. R., Yadav, R. C. (2013). MYB transcription factor genes as regulators for plant responses: An overview. *Physiology and Molecular Biology of Plants*, 19, 307-321.

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

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

Asiegbu, F.O., M. Denekamp, G. Daniel and M. Johansson. 1995. Immunocytochemical localization of pathogenesis-related proteins in roots of Norway spruce infected with *Heterobasidion annosum*. *European Journal of Forest Pathology*, 25, 169–178.

Bakhtiari, M., Formenti, L., Caggía, V., Glauser, G., Rasmann, S. (2019). Variable effects on growth and defence traits for plant ecotypic differentiation and phenotypic plasticity along elevation gradients. *Ecology and Evolution*, 9, 3740-3755.

Balint-Kurti, P. (2019). The plant hypersensitive response: Concepts, control and consequences. *Molecular Plant Pathology*, 20, 1163-1178.

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.

Ballard, R. G., Walsh, M. A., Cole, W. E. (1984). The penetration and growth of blue-stain fungi in the sapwood of lodgepole pine attacked by mountain pine beetle. *Canadian Journal of Botany*, 62, 1724-1729.

Bancroft, B. (2008). *Fundamentals of natural lodgepole pine regeneration and drag scarification*. B.C. Ministry of Forests, Forest Renewal Section Silviculture Practices Branch, Victoria, British Columbia, Canada. pp. 1-30.

Bedon, F., Bomal, C., Caron, S., Levasseur, C., Boyle, B., Mansfield, S. D., Schmidt, A., Gershenzon, J., Grima-Pettenati, J., Séguin, A., 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*, 61, 3847-3864.

Bentz, B. J., Kegley, S., Gibson, K., Thier, R. (2005). A test of high-dose verbenone for stand-level protection of lodgepole and whitebark pine from mountain pine beetle (*Coleoptera: Curculionidae: Scolytinae*) attacks. *Journal of Economic Entomology*, 98, 1614-1621.

Bleiker, K. P., Six, D. L. (2007). Dietary benefits of fungal associates to an eruptive herbivore: Potential implications of multiple associates on host population dynamics. *Environmental Entomology*, 36, 1384-1396.

Bohlmann, J., Keeling, C. I. (2008). Terpenoid biomaterials. *The Plant Journal,* 54, 656–669.

Boller, T. (1985). Induction of hydrolases as a defense reaction against pathogens. *UCLA Symposium on Molecular and Cellular Biology,* 22, 247-262.

Bossinger, G., Spokevicius, A. V. (2018). Sector analysis reveals patterns of cambium differentiation in poplar stems. *Journal of Experimental Botany*, 69, 4339-4348.

Brockley, R. P. (2001). Fertilization of lodgepole pine in western Canada. In Bamsey, C. (Ed.), *Enhanced forest management: Fertilization and economics conference*. March 1-2, 2001, Edmonton, Alberta, Canada. pp. 44-55.

Broekgaarden, C., Caarls, L., Vos, I. A., Pieterse, C. M., Wees, S. C. (2015). Ethylene: Traffic controller on hormonal crossroads to defense. *Plant Physiology*, 169, 2371-2379.

Broschat, T. K. (2017). *Nutrient deficiency symptoms of woody ornamental plants in south Florida*. U. S. Department of Agriculture, UF/IFAS Extension, Gainesville, Florida, United States, ENH1098.

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.

Carroll, A., Taylor, S., Régnière, J., Safranyik, L. (2004). Effects of climate change on range expansion by the mountain pine beetle in British Columbia. In Shore, T. L., Brooks, J. E., Stone, J. E. (Eds.), *Mountain pine beetle symposium: Challenges and solutions*. October 30-31, 2003, Kelowna, British Columbia, Canada. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Information Report BC-X-399, Victoria, British Columbia, Canada. pp. 223-232.

Chalot, M., Brun, A. (1998). Physiology of organic nitrogen acquisition by ectomycorrhizal fungi and ectomycorrhizas. *FEMS Microbiology Reviews*, 22, 21-44.

Cheng, L., Fuchigami, L. H. (2000). Rubisco activation state decreases with increasing nitrogen content in apple leaves. *Journal of Experimental Botany*, 51, 1687-1694.

Chiu, C. C., Keeling, C. I., Bohlmann, J. (2019). The cytochrome P450 CYP6DE1 catalyzes the conversion of α-pinene into the mountain pine beetle aggregation pheromone trans-verbenol. *Scientific Reports*, 9, 1-10.

Chowdhury, S., Basu, A., Kundu, S. (2015). Cloning, characterization, and bacterial over-expression of an osmotin-like protein gene from *Solanum nigrum* L. with antifungal activity against three necrotrophic fungi. *Molecular Biotechnology*, 57, 371-381.

Christiansen, E., Krokene, P., Berryman, A. A., Franceschi, V. R., Krekling, T., Lieutier, F., Lönneborg, A., Solheim, H. (1999). Mechanical injury and fungal infection induce acquired resistance in Norway spruce. *Tree Physiology*, 19, 399-403.

Clérivet, A., Déon, V., Alami, I., Lopez, F., Geiger, J., Nicole, M. (2000). Tyloses and gels associated with cellulose accumulation in vessels are responses of plane tree seedlings (*Platanus* × *acerifolia*) to the vascular fungus *Ceratocystis fimbriata* f. sp *platani*. *Trees*, 15, 25-31.

Clérivet, A., El Modafar, C. (1994). Vascular modifications in *Platanus acerifolia* seedlings inoculated with *Ceratocystis fimbriata* f. sp. *platani*. *European Journal of Forest Pathology*, 24, 1–10.

Cock, P. J., Fields, C. J., Goto, N., Heuer, M. L., Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, 38, 1767-1771.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. J., Gaffney, D., Elo, L. L., Zhang, X., Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology,* 17, 1-19.

Cook, S., Carroll, A., Kimsey, M., Shaw, T. (2015). Changes in a primary resistance parameter of lodgepole pine to bark beetle attack one year following fertilization and thinning. *Forests*, 6, 280-292.

Corbett, L. J., Withey, P., Lantz, V. A., Ochuodho, T. O. (2016). The economic impact of the mountain pine beetle infestation in British Columbia: Provincial estimates from a CGE analysis. *Forestry*, 89, 100-105.

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., 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.

Cullingham, C. I., James, P. M. A., Cooke, J. E. K., 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., Janes, J. K., Hamelin, R. C., James, P. M., Murray, B. W., Sperling, F. A. (2019). The contribution of genetics and genomics to understanding the ecology of the mountain pine beetle system. *Canadian Journal of Forest Research*, 49, 721-730.

Dalman, K., Wind, J. J., Nemesio-Gorriz, M., Hammerbacher, A., Lundén, K., Ezcurra, I., Elfstrand, M. (2017). Overexpression of PaNAC03, a stress induced NAC gene family transcription factor in Norway spruce leads to reduced flavonol biosynthesis and aberrant embryo development. *BMC Plant Biology*, 17, 1-17.

Davin, L.B., Lewis, N.G. (2000). Dirigent proteins and dirigent sites explain the mystery of specificity of radical precursor coupling in lignan and lignin biosynthesis. *Plant Physiology*, 123, 453-462.

DiGuistini, S., Wang, Y., Liao, N. Y., Taylor, G., Tanguay, P., Feau, N., Henrissat, B., Chan, S. K., Hesse-Orce, U., Alamouti, S. M., Tsui, C. K. M., Docking, R. T., Levasseu, A., Haridas, S., Robertson, G., Birol, I., Holt, R. A., Marra, M. A., Hamelin, R. C., Hirst, M., Jones, S. J. M., Bohlman, J., Breuil, C. (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*, 108, 2504–2509.

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

Dougherty, P. M., Whitehead, D., Vose, J. M. (1994). Environmental influences on the phenology of pine. *Ecological Bulletins (Copenhagen)*, 43, 64-75.

Ericsson, T. (1995). Growth and shoot-root ratio of seedlings in relation to nutrient availability. *Plant Soil*, 168, 205-214.

Fagard, M., Launay, A., Clement, G., Courtial, J., Dellagi, A., Farjad, M., Krapp, A., Soulié, M. C., Masclaux-Daubresse, C. (2014). Nitrogen metabolism meets phytopathology. *Journal of Experimental Botany*, 65, 5643-5656.

Ferrenberg, S., Kane, J. M., Mitton, J. B. (2014). Resin duct characteristics associated with tree resistance to bark beetles across lodgepole and limber pine. *Oecologia*, 174, 1283–1292.

Fettig, J. C., Klepzig, K. D., Billings, R. F., Munson, A. S., Nebeker, T. E., Negrón, J. F., Nowak, J. T. (2007). The effectiveness of vegetation management practices for prevention and control of bark beetle infestations in coniferous forests of the western and southern United States. *Forest Ecology and Management*, 238, 24–53.

Flor, H. H. (1971). Current status of the gene-for-gene concept. *Annual Review of Phytopathology*, 9, 275-296.

Fossdal, C. G., Hietala, A. M., Kvaalen, H., Solheim, H. (2006). Changes in host chitinase isoforms in relation to wounding and colonization by *Heterobasidion annosum*: Early and strong defense response in 33-year-old resistant Norway spruce clone. *Tree Physiology*, 26, 169-177.

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., 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.

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, 353-375.

Franceschi, V. R., Krokene, P., Krekling, T., Christiansen, E. (2000). Phloem parenchyma cells are involved in local and distant defense responses to fungal inoculation or bark-beetle attack in Norway spruce (*Pinaceae*). *American Journal of Botany*, 87, 314-326.

Fuller M. P., Jellings A. J. (2003). Crop physiology. In Soffe, R. (Ed.), *Agriculture notebook*. Butterworths & Co Ltd, London, England.

Galindo-González, L. M., Kayal, W. E., Ju, C. J., Allen, C. C., King-Jones, S., Cooke, J. E. K. (2012). Integrated transcriptomic and proteomic profiling of white spruce stems during the transition from active growth to dormancy. *Plant, Cell and Environment,* 35, 682-701.

Gibson, K., Kegley, S., Bentz, B. (2009). *Forest insect & disease leaflet 2*. U. S. Department of Agriculture, Forest Service, Pacific Northwest Region (R6), Portland, Oregon, United States.

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

Goodsman, D. W., Erbilgin, N., Lieffers, V. J. (2012). The impact of phloem nutrients on overwintering mountain pine beetles and their fungal symbionts. *Environmental Entomology*, 41, 478-486.

Government of Alberta (2010). *Mountain pine beetle & cold temperatures: The facts*. Ministry of Environment and Park, Environment and Sustainable Resource Development, Edmonton, Alberta, Canada.

Government of British Columbia (2006). *Forest fertilization in British Columbia*. British Columbia Forest Service, Ministry of Forests and Range, Victoria, British Columbia, Canada.

Gruffman, L., Ishida, T., Nordin, A., Näsholm, T. (2012). Cultivation of Norway spruce and Scots pine on organic nitrogen improves seedling morphology and field performance. *Forest Ecology and Management*, 276, 118-124.

Guérard, N., Dreyer, E., Lieutier, F. (2000). Interactions between Scots pine*, Ips acuminatus* (Gyll.) and *Ophiostoma brunneociliatum* (Math.): Estimation of the critical thresholds of attack and inoculation densities and effects on hydraulic properties in the stem. *Annals of Forest Science, 57*, 681-690.

Guimarães, A. C., Meireles, L. M., Lemos, M. F., Guimarães, M. C., Endringer, D. C., Fronza, M., Scherer, R. (2019). Antibacterial activity of terpenes and terpenoids present in essential oils. *Molecules*, 24, 1-12.

Hacke, U. G., Plavcova, L., Almeida-Rodriguez, A., King-Jones, S., Zhou, W., Cooke, J. E. K. (2010). Influence of nitrogen fertilization on xylem traits and aquaporin expression in stems of hybrid poplar. *Tree Physiology, 30*, 1016-1025.

Heldt, H.-W., Piechulla, B. (2010). *Plant biochemistry* (4th ed.). Elsevier Academic Press, Cambridge, Massachusetts, United States.

Herms, D. A., Mattson, W. J. (1992). The dilemma of plants: To grow or defend. *The Quarterly Review of Biology*, 67, 283–335.

Hodge, J., Cooke, B., McIntosh, R. (2017). *A strategic approach to slow the spread of mountain pine beetle across Canada*. Canadian Council of Forest Ministers, Forest Pest Working Group.

Högberg, P., Näsholm, T., Franklin, O., Högberg, M. N. (2017). Tamm review: On the nature of the nitrogen limitation to plant growth in Fennoscandian boreal forests. *Forest Ecology and Management*, 403, 161-185.

Hubbard, R. M., Rhoades, C. C., Elder, K., Negron, J. (2013). Changes in transpiration and foliage growth in lodgepole pine trees following mountain pine beetle attack and mechanical girdling. *Forest Ecology and Management,* 289, 312-317.

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., 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., 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.

Hunter, R. (1974). Inactivation of pectic enzymes by polyphenols in cotton seedlings of different ages infected with *Rhizoctonia solani*. *Physiological Plant Pathology,* 4, 151-159.

Huse, S. M., Huber, J. A., Morrison, H. G., Sogin, M. L., Welch, D. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*, 8, R143.1-R143.9.

Ishihama, N., Yoshioka, H. (2012). Post-translational regulation of WRKY transcription factors in plant immunity. *Current Opinion in Plant Biology*, 15, 431-437.

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

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., Parcy, F. (2002). BZIP transcription factors in *Arabidopsis*. *Trends in Plant Science*, 7, 106-111.

Jones, K. L., Shegelski, V. A., Marculis, N. G., Wijerathna, A. N., Evenden, M. L. (2019). Factors influencing dispersal by flight in bark beetles (Coleoptera: Curculionidae: Scolytinae): From genes to landscapes. *Canadian Journal of Forest Research*, 49, 1024-1041.

Jwa, N., Hwang, B. K. (2017). Convergent evolution of pathogen effectors toward reactive oxygen species signaling networks in plants. *Frontiers in Plant Science*, 8, 1-12.

Kane, J. M., Kolb, T. E. (2010). Importance of resin ducts in reducing Ponderosa pine mortality from bark beetle attack. *Oecologia,* 164, 601-609.

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

Keeling, C., Weisshaar, S., Lin, R. P. C., Bohlmann, J. (2008). Functional plasticity of paralogous diterpene synthases involved in conifer defense. *PNAS*, 105, 1085–1090.

Klenner, W., Walton, R., Arsenault, A., Kremsater, L. (2008). Dry forests in the southern interior of British Columbia: Historic disturbances and implications for restoration and management. *Forest Ecology and Management,* 256, 1711-1722.

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

Kovalchuk, A., Keriö, S., Oghenekaro, A. O., Jaber, E., Raffaello, T., Asiegbu, F. O. (2013). Antimicrobial defenses and resistance in forest trees: Challenges and perspectives in a genomic era. *Annual Review of Phytopathology*, 51, 221–244.

Krokene, P., Nagy, N. E. (2012). 5. Anatomical aspects of resin-based defences in pine. In Fett-Neto, A. G., Rodrigues-Corrêa, K. C. S. (Eds.), *Pine Resin: Biology, Chemistry and Applications*. Research Signpost, Thiruvananthapuram, India. pp. 67-86.

Kukurba, K. R., Montgomery, S. B. (2015). RNA sequencing and analysis. *Cold Spring Harbor Protocols*, 2015, 951–969.

Kurz, W. A., Dymond, C. C., Stinson, G., Rampley, G. J., Neilson, E. T., Carroll, A. L., Ebata, T., Safranyik, L. (2008). Mountain pine beetle and forest carbon feedback to climate change. *Nature*, 452, 987-990.

Kvam, V. M., Liu, P., Si, Y. (2012). A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. *American Journal of Botany, 99*, 248-256.

Lacerda, A. F., Vasconcelos, Ã. A., Pelegrini, P. B., Sa, M. F. (2014). Antifungal defensins and their role in plant defense. *Frontiers in Microbiology, 5*, 1-10.

Laluk, K., Mengiste, T. (2010). Necrotroph attacks on plants: Wanton destruction or covert extortion? *The Arabidopsis Book*, 8, 1-24.

Lee, S., Kim, J., Breuil, C. (2006). Pathogenicity of *Leptographium longiclavatum* associated with *Dendroctonus ponderosae* to *Pinus contorta*. *Canadian Journal of Forest Research,* 36, 2864-2872.

Liang, X., Zhou, J. (2018). Receptor-like cytoplasmic kinases: Central players in plant receptor kinase–mediated signaling. *Annual Review of Plant Biology*, 69, 267-299.

Lieutier, F. (2002). Mechanisms of resistance in conifers and bark beetle attack strategies. In Wagner, M. R., Clancy, K. M., Lieutier, F., Paine, T. D. (Eds.), *21st International congress of entomology*. Springer, Iguassu Falls, Brazil. pp. 105–130.

Lieutier, F., Yart, A., Salle, A. (2009). Stimulation of tree defenses by *Ophiostomatoid* fungi can explain attack success of bark beetles on conifers. *Annals of Forest Science,* 66, 801-801.

Little, E. L., Jr. (1971). *Atlas of United States trees. Volume 1: Conifers and important hardwoods*. US Department of Agriculture Forest Service, Washington, D.C., United States. Miscellaneous Publication Number 1146.

Liu, J.-J., Ekramoddoullah, A. K. M., 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.

Lotan, J. E., Critchfield, W. B. (1990). Lodgepole pine. In Burns, R. M., Barbara H. H. (Tech. Coords.), *Silvics of North America volume 1. Conifers. Agriculture handbook 654*. U.S. Department of Agriculture, Forest Service, Washington, DC, United States.

Love, M. I., Huber, W., Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*, 1-21.

Lupi, C. (2013). Role of soil nitrogen for the conifers of the boreal forest: A critical review. *International Journal of Plant Soil Science*, 2, 155-189.

Lusebrink, I., Erbilgin, N., Evenden, M. L. (2013). The lodgepole × Jack pine hybrid zone in Alberta, Canada: A stepping stone for the mountain pine bBeetle on its journey east across the boreal forest? *Journal of Chemical Ecology*, 39, 1209-1220.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J.-B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F., Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437, 376-380.

Martin, D., Tholl, D., Gershenzon, J., Bohlmann, 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.

Massad, T. J., Trumbore, S. E., Ganbat, G., Reichelt, M., Unsicker, S., Boeckler, A., Gleixner, G., Gershenzon, J., Ruehlow, S. (2014). An optimal defense strategy for phenolic glycoside production in *Populus trichocarpa*- Isotope labeling demonstrates secondary metabolite production in growing leaves. *New Phytologist*, 203, 607-619.

Matyssek, R., Agerer, R., Ernst, D., Munch, J.-C., Osswald, W., Pretzsch, H., Priesack, E., Schnyder, H., Treutter, D. (2005). The plant's capacity in regulating resource demand. *Plant Biology*, 7, 560–580.

Matyssek, R., Schnyder, H., Elstner, E.-F., Munch, J.-C., Pretzsch, H., Sandermann, H. (2002). Growth and parasite defence in plants: The balance between resource sequestration and retention. *Plant Biology*, 4, 133–136.

McAllister, C. H., Fortier, C. E., Onge, K. R. S., Sacchi, B. M., Nawrot, M. J., Locke, T., Cooke, J. E. K. (2018). A novel application of RNase H2-dependent quantitative PCR for detection and quantification of *Grosmannia clavigera*, a mountain pine beetle fungal symbiont, in environmental samples. *Tree Physiology*, 38, 485–501.

McHale, L., Tan, X., Koehl, P., Michelmore, R. W. (2006). Plant NBS-LRR proteins: Adaptable guards. *Genome Biology*, 7, 212.1-212.11.

Meldau, S., Erb, M., Baldwin, I. T. (2012). Defence on demand: Mechanisms behind optimal defence patterns. *Annals of Botany*, 110, 1503-1514.

Meng, X., Zhang, S. (2013). MAPK cascades in plant disease resistance signaling. *Annual Review of Phytopathology*, 51, 245–266.

Mercado, J. E., Hofstetter, R. W., Reboletti, D. M., Negrón, J. F. (2014). Phoretic symbionts of the mountain pine beetle (*Dendroctonus ponderosae* Hopkins). *Forest Science*, 60, 512-526.

Mihaliak, A. C., Lincoln, D. E. (1985). Growth pattern and carbon allocation to volatile leaf terpenes under nitrogen-limiting conditions. *Journal of Chemical Ecology*, 13, 2059–2067.

Miller, A. J., Cramer, M. D. (2004). Root nitrogen acquisition and assimilation. *Plant Soil*, 274, 1-36.

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.

Miron, D., Battisti, F., Silva, F. K., Lana, A. D., Pippi, B., Casanova, B., Gnoattoc, S., Fuentefriab, A., Mayorgad, P., Schapoval, E. E. (2014). Antifungal activity and mechanism of action of monoterpenes against dermatophytes and yeasts. *Revista Brasileira de Farmacognosia*, 24, 660-667.

Monteiro, S., Barakat, M., Piçarra-Pereira, M. A., Teixeira, A. R., Ferreira, R. B. (2003). Osmotin and thaumatin from grape: A putative general defense mechanism against pathogenic fungi. *Phytopathology, 93*, 1505-1512.

Moreira, X., Zas, R., Sampedro, L. (2012). Differential allocation of constitutive and induced chemical defenses in pine tree juveniles: A test of the optimal defense theory. *PLoS ONE*, 7, 1-8.

Morris, H., Brodersen, C., Schwarze, F. W. M. R., Jansen, S. (2016). The parenchyma of secondary xylem and its critical role in tree defense against fungal decay in relation to the CODIT model. *Frontiers in Plant Science*, 7, 1-18.

Nagy, N. E., Franceschi, V. R., Solheim, H., Krekling, T., Christiansen, E. (2000). Wound-induced traumatic resin duct development in stems of Norway spruce (*Pinaceae*): Anatomy and cytochemical traits. *American Journal of Botany*, 87, 302-313.

Negrón, J. F., Cain, B. (2019). Mountain pine beetle in Colorado: A story of changing forests. *Journal of Forestry, 117*, 144-151.

Näsholm, T., Ekblad, A., Nordin, A., Giesler, R., Högberg, M., Högberg, P. (1998). Boreal forest plants take up organic nitrogen. *Nature*, 392, 914-916.

Natural Resources Canada (2004). *Mountain pine beetle management, A guide for small woodland operations*. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia, Canada.

Natural Resources Canada (2017a, February 21). Mountain pine beetle (factsheet). *https://www.nrcan.gc.ca/forests/fire-insects-disturbances/top-insects/13397.* (Last accessed April 13, 2020).

Natural Resources Canada (2017b, June 14). Forest classification. *https://www.nrcan.gc.ca/our-natural-resources/forests-forestry/sustainable-forest-management/measuring-reporting/forest-classification/13179*. (Last accessed April 13, 2020).

Natural Resources Canada (2020, May 20). How much forest does Canada have? *https://www.nrcan.gc.ca/our-natural-resources/forests-forestry/how-much-forest-does-canada-have/17601*. (Last accessed July 1, 2020).

Nelson, M. F., Murphy, J. T., Bone, C., Altaweel, M. (2018). Cyclic epidemics, population crashes, and irregular eruptions in simulated populations of the mountain pine beetle, *Dendroctonus ponderosae*. *Ecological Complexity*, 36, 218-229.

Neuhaus, J. M. (1999). Plant chitinases (PR-3, PR-4, PR-8, PR-11). In Datta, S. K., Muthukrishnan, S. (Eds.), *Pathogenesis-related proteins in plants*. CRC Press, Boca Ranton, Florida, United States. pp. 77–105.

Nguyen, N. T., Nakabayashi, K., Mohapatra, P. K., Thompson, J., Fujita, K. (2003). Effect of nitrogen deficiency on biomass production, photosynthesis, carbon partitioning, and nitrogen nutrition status of *Melaleuca* and *Eucalyptus* species. *Soil Science and Plant Nutrition*, 49, 99-109.

Nolan, T., Hands, R. E., Bustin, S. A. (2006). Quantification of mRNA using real-time RT-PCR. *Nature Protocols*, 1, 1559-1582.

Ogawa, S., Miyamoto, K., Nemoto, K., Sawasaki, T., Yamane, H., Nojiri, H., Okada, K. (2017). OsMYC2, an essential factor for JA-inductive sakuranetin production in rice, interacts with MYC2-like proteins that enhance its transactivation ability. *Scientific Reports*, 7, 1-11.

Olli, S., Kirti, P. (2006). Cloning, characterization and antifungal activity of defensin Tfgd1 from *Trigonella foenum-graecum* L. *Journal of Biochemistry and Molecular Biology*, 39, 278-283.

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.

Ott, D. S., Yanchuk, A. D., Huber, D. P. W., Wallin, K. F. (2012). Genetic variation of lodgepole pine (*Pinus contorta* var. *latifolia*) chemical and physical defenses that affect mountain pine beetle, *Dendroctonus ponderosae,* attack and tree mortality. *Journal of Chemical Ecology*, 37, 1002–1012.

Pagán, I., García-Arenal, F. (2018). Tolerance to plant pathogens: Theory and experimental evidence. *International Journal of Molecular Sciences*, 19, 1-17.

Page, W. G., Jenkins, M. J., Runyon, J. B. (2012). Mountain pine beetle attack alters the chemistry and flammability of lodgepole pine foliage. *Canadian Journal of Forest Research*, 42, 1631-1647.

Paine, T. D., Raffa, K. F., Harrington, T. C. (1997). Interactions among scolytid bark beetles, their associated fungi, and live host conifers. *Annual Review of Entomology*, 42, 179–206.

Pascual, M. B., Cánovas, F. M., Ávila, C. (2015). The NAC transcription factor family in maritime pine (*Pinus Pinaster*): Molecular regulation of two genes involved in stress responses. *BMC Plant Biology*, 15, 2-15.

Patzlaff, A., Mcinnis, S., Courtenay, A., Surman, C., Newman, L. J., Smith, C., Bevan, M. W., Mansfield, S., Whetten, R. W., Sederoff, R. R., Campbell, M. M. (2003). Characterisation of a pine MYB that regulates lignification. *The Plant Journal*, 36, 743-754.

Pevzner, P. A., Tang, H. X., Waterman, M. S. (2001). An Eulerian path approach to DNA fragment assembly. *Proceedings of the National Academy of Sciences of the United States of America,* 98, 9748-9753.

Pieterse, C. M., Does, D. V., Zamioudis, C., Leon-Reyes, A., Wees, S. C. (2012). Hormonal modulation of plant immunity. *Annual Review of Cell and Developmental Biology,* 28, 489-521.

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

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.

Pré, M., Atallah, M., Champion, A., Vos, M. D., Pieterse, C. M., Memelink, J. (2008). The AP2/ERF domain transcription factor ORA59 integrates jasmonic acid and ethylene signals in plant defense. *Plant Physiology*, 147, 1347-1357.

Qiagen, CLCBio. *www.qiagenbioinformatics.com*. (Last accessed March 16, 2020).

Qiao, H., Shen, Z., Huang, S. C., Schmitz, R. J., Urich, M. A., Briggs, S. P., Ecker, J. R. (2012). Processing and subcellular trafficking of ER-tethered EIN2 control response to ethylene gas. *Science*, 338, 390-393.

Radhika, V., Kost, C., Bartram, S., Heil, M., Boland, W. (2008). Testing the optimal defense hypothesis for two indirect defenses: Extrafloral nectar and volatile organic compounds. *Planta*, 228, 449-457.

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., Berryman, A. A. (1987). Interacting selective pressures in conifer-bark beetle systems: A basis for reciprocal adaptations? *The American Naturalist*, 129, 234-262.

Raffa, K. F., Smalley, E. B. (1988). Seasonal and long-term responses of host trees to microbial associates of the pine engraver, *Ips pini*. *Canadian Journal of Forest Research*, 18, 1624–1634.

Ralph, S. G., Park, J., Bohlmann, J., Mansfield, S. D. (2006a). 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.

Ralph, S. G., Yueh, H., Friedmann, M., Aeschliman, D., Zeznik, J. A., Nelson, C. C., Butterfield, Y. S. N., Kirkpatrick, R., Liu, J., Jones, S. J. M., Marra, M. A., Douglas, C. J., Ritland, K., Bohlmann, J. (2006b). 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 and Environment*, 29, 1545-1570.

Reed, D. E., Ewers, B. E., Pendall, E. (2014). Impact of mountain pine beetle induced mortality on forest carbon and water fluxes. *Environmental Research Letters*, 9, 1-12.

Rice, A. V., Langor, D. W. (2009). Mountain pine beetle-associated blue-stain fungi in lodgepole × jack pine hybrids near Grande Prairie, Alberta (Canada). *Forest Pathology*, 39, 323-334.

Rice A. V., Thormann M. N., Langor D. W. (2007a). Mountain pine beetle associated blue-stain fungi cause lesions on jack pine, lodgepole pine, and lodgepole x jack pine hybrids in Alberta. *Canadian Journal of Botany,* 85, 307-315.

Rice A. V., Thormann M. N., Langor D. W. (2007b). Virulence of, and interactions among, mountain pine beetle associated blue-stain fungi on two pine species and their hybrids in Alberta. *Canadian Journal of Botany*, 85, 316-323.

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

Robert, J. A., Madilao, L. L., White, R., Yanchuk, A., King, J., Bohlmann, J. (2010). Terpenoid metabolite profiling in Sitka spruce identifies association of dehydroabietic acid, (+)-3-carene, and terpinolene with resistance against white pine weevil. *Botany*, 88, 810-820.

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., Mungall, K., Lee, S., Okada, H., Qian, J., Griffith, M., Raymond, A., Thiessen, N., Cezard, T., Butterfield, Y., Newsome, R., Chan, S., She, R., Varhol, R., Birol, I., Birol, I. (2010). *De novo* assembly and analysis of RNA-seq data. *Nature Methods*, 7, U909-U962.

Robinson, M. D., McCarthy, D. J., Smyth, G. K. (2010). EdgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139-140.

Roe A. D., Rice A. V., Bromilow S. E., Cooke J. E. K., Sperling F. A. H. (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.

Rohde, A., Bhalerao, R. P. (2007). Plant dormancy in the perennial context. *Trends in Plant Science,* 12, 217-223.

Roy, S., Laframboise, W. A., Nikiforov, Y. E., Nikiforova, M. N., Routbort, M. J., Pfeifer, J., Nagarajan, R., Carter, A. B., Pantanowitz, L. (2016). Next-generation sequencing informatics: Challenges and strategies for implementation in a clinical environment. *Archives of Pathology Laboratory Medicine,* 140, 958-975.

Ruan, J., Zhou, Y., Zhou, M., Yan, J., Khurshid, M., Weng, W., Cheng, Z., Zhang, K. (2019). Jasmonic acid signaling pathway in plants. *International Journal of Molecular Sciences*, 20, 2479.

Sadasivam, S., Thayumanayan, B. (Eds.) (2019). *Molecular host plant resistance to pests*. CRC Press, Boca Raton, Florida, United States.

Safranyik, L., Carroll, A. L. (2006). Chapter 1 The biology and epidemiology of the mountain pine beetle in lodgepole pine forests. In Safranyik, L., Wilson, B. (Eds.), *The mountain pine beetle, A synthesis of biology, management, and impacts on lodgepole pine*. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia, Canada. pp. 3-66.

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.

Sambaraju, K. R., Carroll, A. L., Aukema, B. H. (2019). Multiyear weather anomalies associated with range shifts by the mountain pine beetle preceding large epidemics. *Forest Ecology and Management, 438*, 86-95.

Sanger, F., Nicklen, S., Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74, 5463-5467.

Schluttenhofer, C., Yuan, L. (2015). Regulation of specialized metabolism by WRKY transcription factors. *Plant Physiology*, 167, 295–306.

Selitrennikoff, C. P. (2001). Antifungal proteins. *Applied and Environmental Microbiology*, 67, 2883-2894.

Shendure, J., Ji, H. L. (2008). Next-generation DNA sequencing. *Nature Biotechnology,* 26, 1135-1145.

Shore, T. L., Safranyik, L., Hawkes, B. C., Taylor, S. W. (2006). Chapter 3 Effects of the mountain pine beetle on lodgepole pine stand structure and dynamics. In Safranyik, L., Wilson, W.R. (Eds.). *The mountain pine beetle: A synthesis of biology, management, and impacts on lodgepole pine*. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia. pp. 95-114.

Shudo, E., Iwasa, Y. (2002). Optimal defense strategy: Storage vs. new production. *Journal of Theoretical Biology*, 219, 309-323.

Six, D. L. (2003). A comparison of mycangial and phoretic fungi of individual mountain pine beetles. *Canadian Journal of Forest Research*, 33, 1331–1334.

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 and Development*, 12, 3703-3714.

Soneson, C., Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, 14, 1-18.

Stahl, K., Moore, R., Mckendry, I. (2006). Climatology of winter cold spells in relation to mountain pine beetle mortality in British Columbia, Canada. *Climate Research*, 32, 13-23.

Stamp, N. (2003). Out of the quagmire of plant defense hypotheses. *The Quarterly Review of Biology*, 78, 23–55.

Stamp, N. (2004). Can the growth-differentiation balance hypothesis be tested rigorously? *Oikos*, 107, 439-448.

Stergiopoulos, I., Wit, P. J. (2009). Fungal effector proteins. *Annual Review of Phytopathology*, 47, 233-263.

Stockfors, J., Linder, S. (1998). Effect of nitrogen on the seasonal course of growth and maintenance respiration in stems of Norway spruce trees. *Tree Physiology*, 18, 155-166.

Sun, Y., Wang, M., Mur, L. A. J., Shen, Q., Guo, S. (2020). Unravelling the roles of nitrogen nutrition in plant disease defences. *International Journal of Molecular Sciences*, 21, 1–20.

Taiz, L., Zeiger, E., Møller, I. M., Murphy, A. (Eds.). (2015). *Plant physiology and development* (6th ed.). Sinauer Associates, Inc., Publishers, Sunderland, Massachusetts, United States.

Tan, K., Oliver, R. P., Solomon, P. S., Moffat, C. S. (2010). Proteinaceous necrotrophic effectors in fungal virulence. *Functional Plant Biology*, 37, 907.

Tan, K., Phan, H. T., Rybak, K., John, E., Chooi, Y. H., Solomon, P. S., Oliver, R. P. (2015). Functional redundancy of necrotrophic effectors – consequences for exploitation for breeding. *Frontiers in Plant Science, 6*, 1-9.

Tang, J., Sun, B., Cheng, R., Shi, Z., Luo, D., Liu, S., Centritto, M. (2019). Effects of soil nitrogen (N) deficiency on photosynthetic N-use efficiency in N-fixing and non-N-fixing tree seedlings in subtropical China. *Scientific Reports*, 9, 1-14.

Taylor, S. W., Carroll, A. L. (2004). Disturbance, forest age, and mountain pine beetle outbreak dynamics in BC: A historical perspective. In Shore, T. L., Brooks, J. E., Stone, J. E. (Eds.), *Mountain pine beetle symposium: Challenges and solutions*. October 30-31, 2003, Kelowna, British Columbia, Canada. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Information Report BC-X-399, Victoria, British Columbia, Canada. pp. 41-51.

Tegeder, M., Masclaux-Daubresse, C. (2018). Source and sink mechanisms of nitrogen transport and use. *New Phytologist*, 217, 35-53.

Tena, G., Boudsocq, M., Sheen, J. (2011). Protein kinase signaling networks in plant innate immunity. *Current Opinion in Plant Biology*, 14, 519-529.

Tenhaken, R. (2015). Cell wall remodeling under abiotic stress. *Frontiers in Plant Science*, 5, 1-9.

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

Thomas, T., Gilbert, J., Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, 2, 1-12.

Tomova, L., Braun, S., Fluckiger, W. (2005). The effect of nitrogen fertilization on fungistatic phenolic compounds in roots of beech (*Fagus sylvatica*) and Norway spruce (*Picea abies*). *Forest Pathology*, 35, 262–276.

Tuomi, J., Fagerstrom, T., Niemela, P. (1991). Carbon allocation, phenotypic plasticity, and induced defenses. In Tallamy, D. W., Raupp, M. J. (Eds.), *Phytochemical Induction by Herbivores*. Wiley, New York, New York, United States. pp. 85–104.

Tuomi, J., Niemelä, P., Chapin, III, F. S., Bryant, J. P., Sirén, S. (1988). Defensive responses of trees in relation to their carbon/nutrient balance. In Mattson, W. J.

et al. (Eds.), *Mechanisms of Woody Plant Defenses Against Insects: Search for Pattern*. Springer, New York, New York, United States. pp. 57–72.

Vadeboncoeur, M. A. (2010). Meta-analysis of fertilization experiments indicates multiple limiting nutrients in northeastern deciduous forests. *Canadian Journal of Forest Research,* 40, 1766-1780.

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

Verly, C., Djoman, A. C., Rigault, M., Giraud, F., Rajjou, L., Saint-Macary, M., Dellagi, A. (2020). Plant defense stimulator mediated defense activation is affected by nitrate fertilization and developmental stage in *Arabidopsis thaliana*. *Frontiers in Plant Science,* 11, 1-15.

Vijay, N., Poelstra, J. W., Künstner, A., Wolf, J. B. (2012). Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive *in silico* assessment of RNA-seq experiments. *Molecular Ecology*, 22, 620-634.

Vogel, C., Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, 13, 227-232.

Vos, I. A., Pieterse, C. M., Wees, S. C. (2013). Costs and benefits of hormone-regulated plant defences. *Plant Pathology,* 62, 43-55.

Wainhouse, D., Ashburner, R., Ward, E., Rose, J. (1998). The effect of variation in light and nitrogen on growth and defense in young Sitka spruce. *Functional Ecology*, 12, 561-572.

Wang, C., Tang, Y. (2019). Responses of plant phenology to nitrogen addition: A meta-analysis. *Oikos*, 128, 1243-1253.

Wang, L., Li, P., Brutnel. T. P. (2010). Exploring plant transcriptomes using ultra high-throughput sequencing. *Briefings in Functional Genomics*, 9, 118–128.

Wang, Y., Lim, L., Diguistini, S., Robertson, G., Bohlmann, J., Breuil, C. (2012). A specialized ABC efflux transporter GcABC-G1 confers monoterpene resistance to *Grosmannia clavigera*, a bark beetle-associated fungal pathogen of pine trees. *New Phytologist*, 197, 886–898.

Wang, Z., Gerstein, M., Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics,* 10, 57-63.

Waring, H. R., Pitman, G. B. (1985). Modifying lodgepole pine stands to change susceptibility to mountain pine beetle attack. *Ecology*, 66, 889–897.

Warinowski, T., Koutaniemi, S., Kärkönen, A., Sundberg, I., Toikka, M., Simola, L. K., Kilpeläinen, I., Teeri, T. H. (2016). Peroxidases bound to the growing lignin polymer produce natural like extracellular lignin in a cell culture of Norway spruce. *Frontiers in Plant Science,* 7, 1-12.

Wilson, E., Skeffington, R. (1994). The effects of excess nitrogen deposition on young Norway spruce trees. Part I the soil. *Environmental Pollution*, 86, 141-151.

Wong, B. L., Berryman, A. A. (1977). Host resistance to the fir engraver beetle. 3. Lesion development and containment of infection by resistant *Abies grandis* inoculated with *Trichosporium symbioticum*. *Canadian Journal of Botany*, 55, 2358–2365.

Wulder, M. A., White, J. C., Bentz, B. J., Ebata, T. (2006). Augmenting the existing survey hierarchy for mountain pine beetle red-attack damage with satellite remotely sensed data. *The Forestry Chronicle*, 82, 187-202.

Yamaoka, Y., Hiratsuka, Y., Maruyama, P. J. (1995). The ability of *Ophiostoma clavigerum* to kill mature lodgepole pine trees. *Forest Pathology*, 25, 401-404.

Zabel, R. A., Morell, J. J. (1992). Wood stains and discolorations. In Zabel R. A., Morell J. J. (Eds.), *Wood microbiology: Decay and its prevention*. Academic Press, San Diego, California, United States. pp. 326–343.

Zamyatnin, A. A. (2015). Plant proteases involved in regulated cell death. *Biochemistry*, 80, 1701-1715.

Zhang, W. Y., Chen, J. J., Yang, Y., Tang, Y. F., Shang, J., Shen, B. R. (2011). A practical comparison of *de novo* genome assembly software tools for next-generation sequencing technologies. *PLoS ONE,* 6, 1-12.

Zhang, L., Zhang, F., Melotto, M., Yao, J., He, S. Y. (2017). Jasmonate signaling and manipulation by pathogens and insects. *Journal of Experimental Botany,* 68, 1371-1385.

Zhao, S., Erbilgin, N. (2019). Larger resin ducts are linked to the survival of lodgepole pine trees during mountain pine beetle outbreak. *Frontiers in Plant Science*, 10, 1-14.

Zhong, R., Ye, Z. (2014). Secondary cell walls: Biosynthesis, patterned deposition and transcriptional regulation. *Plant and Cell Physiology,* 56, 195-214.

2.0 Ch 2: A comparison between the CLC Genomics Workbench and Trans-ABySS

2.1 Introduction

The term RNA-Seq refers to the sequencing and bioinformatic analysis of complementary deoxyribonucleic acid (cDNA) libraries that are synthesized from extracted ribonucleic acid (RNA) transcripts (Wang *et al.* 2009). The relative abundance of transcripts reveals the functional elements of the genome in a very specific temporal and physiological context (Brady *et al.* 2006). Sequencing technologies have improved at an exceptional rate, with next-generation sequencing (NGS) taking the lead in cost and time efficiency (Shendure and Ji 2008). Using high-throughput collection methods, NGS technologies generate massive amounts of data, though their sequence lengths are shorter than those produced by earlier technologies like Sanger sequencing (Sanger *et al.* 1977; Shendure and Ji 2008). Computational advances in working with short sequences enable an NGS approach for efficient identification of molecular markers and transcripts involved in important biological processes (Wang *et al.* 2009). An initial step in obtaining a large and reliable transcriptomic data set to be used in downstream analyses is deep sequencing (Wang *et al.* 2009). The acquisition of large quantities of sequenced cDNA fragments allows for the analysis of lodgepole pine (*Pinus contorta* Douglas ex Loudon *var. latifolia*) transcriptomes from secondary phloem (2P) and secondary xylem (2X) tissue libraries.

To obtain RNA-Seq data from cDNA, isolated messenger ribonucleic acid (mRNA) goes through reverse-transcription using the enzyme reverse transcriptase to produce single-stranded (ss) cDNA molecules. These ss cDNA molecules are in the same orientation as the coding sequences (CDS) and are without introns. Single-stranded cDNA is made double-stranded (ds) by employment of the polymerase-chain reaction (PCR; Nolan *et al.* 2006). Deoxynucleotide uridine triphosphate (dUTP) can be used instead of deoxynucleotide thymidine triphosphate (dTTP) during this process to ensure that the second strand of cDNA contains uracil and not thymine. This creates a sequence that cannot be later amplified by deoxyribonucleic acid (DNA)-dependent DNA polymerase, which is unable to extend beyond an uracil nucleotide (Nolan *et al.* 2006). This library preparation approach produces "stranded" sequence data with the strand orientation of the CDS maintained (Parkhomchuk *et al.* 2009). Stranded sequences are essential for the analysis of expression data since genes can be located on either strand and may even be complementary with one another (Parkhomchuk *et al.* 2009; Zhao *et al.* 2015). When stranded cDNA is used for sequencing, there is also a reduction in computation time for downstream analyses (Martin and Wang 2011; Agarwal *et al.* 2015).

Prior to sequencing, additional sequences called indexed adapters are ligated to the 3' ends of the ds cDNA sequences (Bentley *et al.* 2008). Indexed adapters contain three main types of sequences (Figure 2.1). The first portion is called the adapter sequence, and it is the same for each library. The adapter

sequence and its complement are complementary to short ss DNA sequences, called oligonucleotides, that adhere to the sequencer's acrylamide-coated glass flow cell. The adapter sequence and its complement hybridize with the oligonucleotides, effectively tethering the cDNA fragments to the flow cell. The second portion is the index sequence, which differs between samples. Indices are used to tag samples so that their libraries can be sequenced together and split up later. Lastly, there are primer regions, which are the same for each library. Primer sequences from the sequencer hybridize with these regions, allowing the DNA polymerase to add fluorescently tagged nucleotides in a process called sequencing by synthesis (Bentley *et al.* 2008). When a nucleotide hybridizes to the sequence, it fluoresces a base-specific wavelength, and the sequencer detects the signal to determine not only which nucleotide hybridized, but also the quality of the hybridization (Bentley *et al.* 2008).

5' ...cDNA... GATCGGAAGAGCACACGTCTGAACTCCAGTCAC (INDEX) ATCT CGTATGCCGTCTTCTGCTTG 3'

3' ...sequenced cDNA... GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC 5'

Figure 2.1. TruSeq Index Adapter structure post-sequencing with the Illumina Next-Seq 500. TruSeq Index Adapters were ligated to the 3' ends of ds cDNA. In preparation for sequencing by synthesis, the adapter region (underlined text) was hybridized with the flow cell. The index sequence (in parentheses) was a six base pairs sequence unique to each sample on the flow cell. The primer sequence used by the sequencer is in blue text. The complement of the primer region (red text) remained attached to the 5' end of the sequenced cDNA. This sequencing artifact can be fully sequenced, partially sequenced or absent.

NGS comes with several shortcomings, such as limitations on the length of the cDNA fragment that can be sequenced and variations in the accuracy of bases called by the sequencer (Huse *et al.* 2007). The NextSeq 500 platform (Illumina, Inc., San Diego, California, United States), which we used for this thesis project, can sequence up to 300 base pairs (bp) by sequencing 150 bp from each 3' end of the ds cDNA fragment rather than across the entire length of the fragment. The 150 bp sequences produced by the Illumina NextSeq 500 system with strand information preserved are called stranded, paired-end reads. Following sequencing, the digital expression data are in the form of read identifiers, which are strings of International Union of Pure and Applied Chemistry (IUPAC) nucleotide codes (Cornish-Bowden 1985) and quality scores that accompany each base called by the sequencer. NGS often results in sequencing errors, such as low-quality or ambiguous base calls (Huse *et al.* 2007). The Illumina platform registers quality using a Phred-score like algorithm, which specifies the probability that a given base is called incorrectly (Ewing and Green 1998; Ewing *et al.* 1998). Ambiguous bases, represented with the IUPAC symbol N, are sequencing errors that occur when the detection of one base-specific wavelength is unsuccessful (Huse *et al.* 2007).

In an attempt to reconstruct full-length transcripts, the process of transcriptome assembly orients reads into contiguous sequences (contigs; Pevzner *et al.* 2001). By clarifying the relative position of reads, the production of stranded, paired-end reads assists in the assembly of full-length contigs (Corley *et al.* 2017).

Sequenced reads may include all or portions of the reverse complement of the indexed adapter sequences' primer regions and reads must be pre-processed *in silico* before being assembled into contigs (Figure 2.1). This pre-processing is often called trimming because the read length is reduced by the removal of not only adapter sequences, but also portions with low quality or multiple ambiguous bases (Martin 2011; Bolger *et al.* 2014). Pre-processing also includes the disposal of sequences below or above a certain threshold length that has been specified by the researcher. Following trimming, a sequence that is below a certain length is indicative of a poor quality read, while a sequence that is longer than the read length is indicative of multiple sequencing errors (Huse *et al.* 2007). The choice of read length thresholds and the optimization of other pre-processing parameters is paramount in obtaining a reliable transcriptomic data set.

Once reads have been pre-processed, they are ready for assembly into contigs. Assembly of sequence data can be carried out using *de novo* approaches or alignment to a reference transcriptome or genome (Conesa *et al.* 2016). *De novo* refers to the creation of an assembly that does not rely on a reference to orient reads, but rather relies on overlaps of sequences to orient them relative to one another (Schatz *et al.* 2010). A *de novo* assembly method using an adaptation of de Bruijn graphs became popular in 2007 and continues to be the algorithm of choice for the assembly of short reads produced by NGS technology (Zhang *et al.* 2011). Prior to assembly, trimmed reads are cut computationally into smaller components of length k, designated by the researcher. These k-mers are

overlapped by k-1 bp to assemble the reads into contigs. Contigs can be turned into scaffolds by using paired-end read information to resolve long repeats that are spanned not by individual reads, but by read pairs (Earl *et al.* 2011). Sequencing errors not resolved by pre-processing complicate *de novo* assembly (Pevzner *et al.* 2001). Errors in reads will create extra k-mers in the de Bruijn graph, which increases computer memory usage (Schatz *et al.* 2010). However, *de novo* assembly of stranded sequences reduces processing time since reverse compliment k-mers are not generated when assembling paired-end read data (Martin and Wang 2011).

Errors in reads will cause forks or bubbles in the de Bruijn graph that, if not handled properly, will prevent longer contigs from being produced (Schatz *et al.* 2010). Biological phenomena, such as heterozygosity and the assembly of non-genetically uniform individuals, will also generate erroneous graph structures (Chaisson *et al.* 2015). Forks with sequence lengths below a certain threshold must be removed and discarded, while forks with sequence lengths above a certain threshold must be separated from the graph and used to create distinct contigs (Schatz *et al.* 2010). If the bubbled sequences are below a certain length threshold, often set by the researcher, bubbles that form in the de Bruijn graph are collapsed on the basis of k-mer coverage (Schatz *et al.* 2010). Bubble sequences above a certain length threshold are removed from the graph and used to create a distinct contig (Schatz *et al.* 2010). The assembler must resolve these errors as part of the transcriptome assembly process.

Bioinformatic analyses based on two state-of-the-art assemblers, the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) and Trans-ABySS (Robertson *et al.* 2010), can be used to assemble expression data. The CLC Genomics Workbench (*www.qiagenbioinformatics.com*) is a commercial product, while Trans-ABySS (Robertson *et al.* 2010) is freely distributed. The CLC Genomics Workbench (*www.qiagenbioinformatics.com*) utilizes proprietary bioinformatic techniques that are not completely known to the user, whereas the Trans-ABySS (Robertson *et al.* 2010) algorithm is made public via the GitHub webpage (*https://github.com/bcgsc/transabyss*). Though both employ a similar de Bruijn graph approach, each tool treats sequencing errors, resolves ambiguities, and utilizes read pair information differently, which results in different assembly outcomes (Jung *et al.* 2016). Prior to the comparison of assembly outcomes, parameters used by each assembler, such as k-mer size and minimum contig length, must be optimized. The CLC Genomics Workbench (*www.qiagenbioinformatics.com*) and Trans-ABySS (Robertson *et al.* 2010) use single k-mer and multiple k-mer methods, respectively. The multiple k-mer approach used by Trans-ABySS (Robertson *et al.* 2010) first creates separate assemblies using designated k-mer sizes, and then merges those assemblies by discarding a contig if it is contained within a longer one (Robertson *et al.* 2010). The influence of single and multiple k-mers on assembly performance can be assessed. This helps the user gain insight into transcriptome construction from short reads, which can inform decisions made when setting the parameters of an

RNA-Seq experiment (Robertson *et al.* 2010). In addition, ease of use must be considered when choosing which software to employ. The CLC Genomics Workbench (*www.qiagenbioinformatics.com*) has a comprehensive read trimming tool, whereas the choice of Trans-ABySS (Robertson *et al.* 2010) requires the selection of a separate trimming algorithm when constructing transcriptomes. The CLC Genomics Workbench (*www.qiagenbioinformatics.com*) interface is user friendly, whereas Trans-ABySS (Robertson *et al.* 2010) is implemented through command line.

Given the size and complexity of the lodgepole pine transcriptome (De La Torre *et al.* 2014; Suren *et al.* 2016), it was unclear which platform would create the more accurate and comprehensive assembly. Therefore, the objective of the study outlined in this chapter was to optimize and compare lodgepole pine transcriptome assemblies made using the CLC Genomics Workbench v9.5.2 (*www.qiagenbioinformatics.com*) and Trans-ABySS v1.5.5 (Robertson *et al.* 2010). To achieve this end, we used 32 tissue samples obtained from an experiment that tested the responses of lodgepole pine to *Grosmannia clavigera* (Robinson-Jeffrey and Davidson) inoculation under varying levels of nitrogen (N) fertilization. After grinding the tissue samples, RNA was isolated and used to construct 32 cDNA libraries. cDNA sequencing was carried out using the Illumina NextSeq 500 platform (Illumina, Inc., San Diego, California, United States). Sequence data were trimmed using the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) and a comprehensive comparison of one

library's assembly made with the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) and Trans-ABySS (Robertson *et al.* 2010) was performed using the following metrics: assembly time, N50 length, maximum contig length and the number of contigs. Assembly quality was also assessed based on the percentage of reads mapped back to the respective assemblies. Excluding total contig number, it is generally acknowledged that larger values of these criteria imply better assembly performance (Jung *et al.* 2016). We hypothesized that the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) would produce a faster assembly using proprietary algorithms, but that Trans-ABySS (Robertson *et al.* 2010), given its multiple k-mer de Bruijn graph algorithm, would produce a higher quality and more comprehensive assembly.

2.2 Materials and methods

2.2.1 Lodgepole pine experiment

Tissue used for the RNA-Seq experiment was obtained from a Cooke lab experiment carried out in 2010 by Dr. Adriana Arango-Velez, which tested the responses of lodgepole and jack pine (*Pinus banksiana* Lambert) to *G. clavigera* inoculation under varying concentrations of ammonium nitrate (NH_4NO_3) fertilizer. One-year-old dormant lodgepole pine seedlings provided by PRT Armstrong Nursery (Armstrong, British Columbia, Canada) were removed from cold storage and repotted in 3.78 L pots with Sunshine Mix #4 (SunGro Horticulture, Agawam, Massachusetts, United States). Seedlings were grown in a

completely randomized block design under controlled environment growth in rooms at 19 °C constant temperature, 35% relative humidity, 15 h day / 9 h night photoperiod, and approximately 200-250 μmol photosynthetically active radiation (PAR) light intensity. Beginning the second week following repotting, liquid fertilizer solution was applied each week to seedlings until soil saturation. For the first two weeks, all seedlings received 0.5 g/L 15:30:15 (N:P:K) fertilizer (Plant Products Ltd, Brampton, Ontario, Canada), followed by two weeks of 0.5 g/L 20:20:20 (N:P:K) fertilizer (Plant Products Ltd, Brampton, Ontario, Canada). At six weeks after repotting, seedlings were fertilized weekly with Hocking's complete nutrient solution (Hocking 1971) containing either 0 mM (no N), 0.3 mM (low N) or 10 mM (high N) NH_4NO_3 until the conclusion of the experiment. Five weeks after beginning the differential N treatments (i.e. 11 weeks after repotting), mock inoculation and *G. clavigera* inoculation treatments were applied as previously described (Arango-Velez *et al.* 2016) with minor modifications. A spore suspension (~140 spores μL^{-1}) for *G. clavigera* isolate M001-03-03-07-UC04DL09, initially described by Roe *et al.* (2010, 2011), was prepared according to Arango-Velez *et al.* (2016). Fungal inoculations were performed by adding 10 μL of *G. clavigera* spores suspended in Milli-Q water (Merck Millipore, Burlington, Massachusetts, United States) into three small wounds per tree generated at 4-5 cm intervals along both the first and second seasons' growth using a syringe needle (23G1 PrecisionGlide, Becton, Dickinson and Company, Mississauga, Ontario, Canada). The mock-inoculated group was mechanically wounded with the syringe needle in an

identical fashion, but Milli-Q water (Merck Millipore, Burlington, Massachusetts, United States) without the fungal inoculum was applied. Phloem and xylem tissues were collected at 0, 7, 14 and 28 days post inoculation (dpi). There were six biological replicates for each combination of N, inoculation and day of collection.

2.2.2 Tissue processing

Since lodgepole pine indicators for *G. clavigera* resistance or susceptibility have been shown to occur shortly after inoculation (Arango-Velez *et al.* 2016), bioinformatic analysis was carried out for 7 dpi samples only. This resulted in 48 lodgepole pine tissue samples (8 treatments = 1 day × phloem or xylem × high or low N × mock- or fungal-inoculated, and 6 biological replicates per treatment). The following table explains each sample in more detail (Table 2.1).

Table 2.1. Lodgepole pine stem tissue samples for RNA-Seq analysis. Forty-eight trees were assigned treatments as part of the experiment's multifactorial design. Samples were taken from the trees at 7 dpi. Lodgepole pine seedlings grown under 0.3 mM (Low) or 10 mM (High) NH_4NO_3 conditions were either mechanically wounded and inoculated with *G. clavigera* (Wound+Fungi) or mechanically wounded and mock-inoculated (Wound). Two tissue types were collected from each tree, secondary phloem (2° phloem) and secondary xylem (2° xylem). Note that 3-4 biological replicates (trees that received the same treatments) for each treatment combination were required for RNA-Seq. Samples used for RNA-Seq were assigned names that highlighted treatment type.

Library code	N treatment	Inoculation treatment	Tissue type	Biological replicate	Used for RNA-Seq?	Sample name
14-01-2P	Low	Wound	2° phloem	1	yes	LowN.Wound.2P.1
14-02-2P	Low	Wound	2° phloem	2	yes	LowN.Wound.2P.2
14-03-2P	Low	Wound	2° phloem	3	yes	LowN.Wound.2P.3
14-04-2P	Low	Wound	2° phloem	4	yes	LowN.Wound.2P.4

Table 2.1. Lodgepole pine stem tissue samples for RNA-Seq analysis. Continued.

Library code	N treatment	Inoculation treatment	Tissue type	Biological replicate	Used for RNA-Seq?	Sample name
14-05-2P	Low	Wound	2° phloem	5		
14-06-2P	Low	Wound	2° phloem	6		
14-01-2X	Low	Wound	2° xylem	1	yes	LowN.Wound.2X.1
14-02-2X	Low	Wound	2° xylem	2	yes	LowN.Wound.2X.2
14-03-2X	Low	Wound	2° xylem	3		
14-04-2X	Low	Wound	2° xylem	4	yes	LowN.Wound.2X.3
14-05-2X	Low	Wound	2° xylem	5	yes	LowN.Wound.2X.4
14-06-2X	Low	Wound	2° xylem	6		
15-01-2P	Low	Wound+Fungi	2° phloem	1	yes	LowN.Fungus.2P.1
15-02-2P	Low	Wound+Fungi	2° phloem	2	yes	LowN.Fungus.2P.2
15-03-2P	Low	Wound+Fungi	2° phloem	3	yes	LowN.Fungus.2P.3
15-04-2P	Low	Wound+Fungi	2° phloem	4	yes	LowN.Fungus.2P.4
15-05-2P	Low	Wound+Fungi	2° phloem	5		
15-06-2P	Low	Wound+Fungi	2° phloem	6		
15-01-2X	Low	Wound+Fungi	2° xylem	1		
15-02-2X	Low	Wound+Fungi	2° xylem	2	yes	LowN.Fungus.2X.1
15-03-2X	Low	Wound+Fungi	2° xylem	3	yes	LowN.Fungus.2X.2
15-04-2X	Low	Wound+Fungi	2° xylem	4		
15-05-2X	Low	Wound+Fungi	2° xylem	5	yes	LowN.Fungus.2X.3
15-06-2X	Low	Wound+Fungi	2° xylem	6	yes	LowN.Fungus.2X.4
17-01-2P	High	Wound	2° phloem	1	yes	HighN.Wound.2P.1
17-02-2P	High	Wound	2° phloem	2	yes	HighN.Wound.2P.2
17-03-2P	High	Wound	2° phloem	3	yes	HighN.Wound.2P.3
17-04-2P	High	Wound	2° phloem	4	yes	HighN.Wound.2P.4
17-05-2P	High	Wound	2° phloem	5		
17-06-2P	High	Wound	2° phloem	6		
17-01-2X	High	Wound	2° xylem	1	yes	HighN.Wound.2X.1
17-02-2X	High	Wound	2° xylem	2	yes	HighN.Wound.2X.2
17-03-2X	High	Wound	2° xylem	3		
17-04-2X	High	Wound	2° xylem	4		
17-05-2X	High	Wound	2° xylem	5	yes	HighN.Wound.2X.3
17-06-2X	High	Wound	2° xylem	6	yes	HighN.Wound.2X.4
18-01-2P	High	Wound+Fungi	2° phloem	1	yes	HighN.Fungus.2P.1
18-02-2P	High	Wound+Fungi	2° phloem	2	yes	HighN.Fungus.2P.2
18-03-2P	High	Wound+Fungi	2° phloem	3	yes	HighN.Fungus.2P.3
18-04-2P	High	Wound+Fungi	2° phloem	4	yes	HighN.Fungus.2P.4
18-05-2P	High	Wound+Fungi	2° phloem	5		
18-06-2P	High	Wound+Fungi	2° phloem	6		
18-01-2X	High	Wound+Fungi	2° xylem	1		
18-02-2X	High	Wound+Fungi	2° xylem	2	yes	HighN.Fungus.2X.1
18-03-2X	High	Wound+Fungi	2° xylem	3		
18-04-2X	High	Wound+Fungi	2° xylem	4	yes	HighN.Fungus.2X.2
18-05-2X	High	Wound+Fungi	2° xylem	5	yes	HighN.Fungus.2X.3
18-06-2X	High	Wound+Fungi	2° xylem	6	yes	HighN.Fungus.2X.4

To extract RNA from our lodgepole pine stem tissue, the tissue must be ground into a fine powder either manually or with a grinding machine, such as a

Geno/Grinder 2010 (SPEX SamplePrep, Metuchen, New Jersey, United States). The 24 secondary phloem and 24 secondary xylem samples, each individually wrapped in tin foil, were retrieved from the -80 °C freezer where they had been stored since their collection in 2010. The frozen tissue was ground to a fine powder using metal bead agitation. First, the Cryostation 2600 (SPEX SamplePrep, Metuchen, New Jersey, United States) was turned on and left to cool. Then, the following materials were collected: 4" by 4" weighing paper for transferring tissue, cotton gloves for wear under nitrile laboratory gloves, the frozen samples wrapped in tin foil placed into a Dewar containing liquid N, metal tongs for retrieving the frozen samples from the liquid N, 15 mL Falcon tubes (Becton Dickinson Labware, Franklin Lakes, New Jersey, United States) to hold the tissue, a 24 well Cryo-Block (SPEX SamplePrep, Metuchen, New Jersey, United States) and 11 mm stainless-steel grinding balls. To prevent the thawing of the tissue and subsequent deterioration of the mRNA transcripts (Opitz *et al.* 2010), samples wrapped in tin foil were kept frozen in a liquid N Dewar, while Falcon tubes, weighing paper, grinding balls and the Cryo-Block were kept frozen by the Cryostation 2600. Samples wrapped in tin foil were first hammered to break up the phloem or xylem tissue. Care was taken not to expose or contaminate any tissue. Between 2 mL and 3.5 mL of hammered, frozen tissue was then transferred to a clean, frozen Falcon tube using weighing paper and placed in the Cryostation 2600. Two frozen, 11 mm grinding balls were added to each tube, and eight tubes at a time were arranged in a symmetrical manner in the Geno/Grinder Cryo-Block. The Cryo-Block was

clamped into the Geno/Grinder 2010 in a manner consistent with the operating manual (SPEX SamplePrep 2013). Every sample was ground three to four times, 30 seconds per round, at 1100 revolutions per minute (rpm). The Falcon tubes holding the tissue and metal balls, along with the Cryo-Block were cooled in the Cryostation 2600 for 10 minutes between each round to keep all materials frozen. When the tissue became a fine, homogenous powder, the Falcon tubes containing tissue and metal balls were returned to the Dewar of liquid N and then transferred into storage boxes and stowed at -80 °C to await RNA extraction.

Most likely due to a higher lignin and cellulose content, some samples treated with this procedure did not grind into a fine powder (Miedes *et al.* 2014; Lourenço *et al.* 2016). This occurred in two fungal-inoculated phloem and ten xylem samples that were treated with additional hand grinding described below. First, the following materials were collected: cotton gloves for wear under nitrile laboratory gloves, a mortar, a pestle, a mortar cozy, two foam coolers containing liquid N, one with the Falcon tubes holding the tissue samples to be ground and one which was used to hold the Falcon tubes once the tissue was ground, a clean Dewar filled with liquid N, a metal scoop for transferring liquid N, metal tongs for handling frozen Falcon tubes and clean plastic scoopulas for transferring ground tissue. The mortar with pestle resting inside was nestled into a mortar cozy, and the metal ladle scoop was used to transfer liquid N into the mortar to cool the mortar and the pestle. Once cooled, tongs were used to retrieve a sample in its Falcon tube, and the tissue with metal beads was poured into the mortar with

additional liquid N added to the mortar as needed to keep the tissue frozen. Optimal grinding occurred when a small amount of liquid N was present in the mortar during the pulverizing process. Once the tissue was ground into a fine, homogenous powder, a plastic scoopula was used to transfer the ground tissue back into the Falcon tubes, which were returned to the Dewar of liquid N and then transferred into storage boxes and stowed at -80 °C to await RNA extraction.

2.2.3 RNA extraction

Biological replicates enhance the power and reduce the false-discovery rate of RNA-Seq analyses (Lamarre *et al.* 2018). Liu *et al.* (2014) found that four biological replicates are sufficient for RNA-Seq. Therefore, RNA extractions were performed on 32 of the 48 samples (32 samples = 4 biological replicates x 8 treatment combinations). RNA extractions were performed by Dr. Chandra McAllister using the procedure developed by Chang *et al.* (1993), modified according to Pavy *et al.* (2008). The extraction buffer, warmed in the water bath to 65 °C, contained only 2% hexadecyltrimethylammonium bromide and 2% β-mercaptoethanol. Ground tissue was obtained from the -80 °C freezer and was transferred into 1.5 mL microfuge tubes up to the 500-600 µL mark. Precisely 900 µL of buffer was added to ground tissue, and tubes were then incubated in a 65 °C water bath for 10 minutes, with occasional vortexing. Two chloroform extractions were performed. Nitrile gloves were doubled to prevent skin exposure before adding 600 µL of the chloroform:isoamyl alcohol (24:1) to each microfuge tube.

Tubes were inverted for 1-2 minutes before being centrifuged at 14,000 rpm at room temperature for 15 minutes. The supernatant was transferred to a clean 1.5 mL microfuge tube. This extraction process was performed again on the ground tissue, and the supernatant was added to the supernatant from the previous extraction. After discarding the original tubes that contained the ground tissue, the tubes of supernatant were measured on a scale to garner the approximate volume. A 1 mL of liquid weighs 1 g approximation was used. One-third the volume of the supernatant of 10 M lithium chloride was added and the tubes were inverted several times. The RNA precipitated for 1 hour in the -20 °C freezer and was then harvested by use of the Benchmark Z446-K Refrigerated Centrifuge (Benchmark Scientific, Sayreville, New Jersey, United States), which spun the microfuge tubes at 14,000 rpm for 15 minutes while maintaining the temperature of 4 °C. After removal of the supernatant, 600 µL of 4 °C 80% ethanol was added to each tube to wash the RNA. The tubes were flicked to dislodge the RNA pellet, and the tubes were spun quickly in the table-top centrifuge at 14,000 rpm for 5 minutes to collect the supernatant at the bottom of the tubes. The supernatant was then removed, and the pellet of RNA was left to dry for at most 5 minutes. Finally, the pellet was resuspended in Milli-Q water (Merck Millipore, Burlington, Massachusetts, United States) on ice, between 10 µL and 40 µL depending on the size of the pellet. The pellet was resuspended by the take-up and release of the solution.

RNA quantity was checked with the NanoQuant Spectrophotometer (Tecan, Männedorf, Switzerland) following the manufacturer's instructions, and quality

was checked with the Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, California, United States) following manufacturer's instructions. High quantity was specified as an RNA concentration between 700 and 800 ng/μL, whereas high quality was specified as an A260/A280 absorbance ratio between 2 and 2.2 and an RNA integrity number above 7.5. Four biological replicates for each of the eight treatments with the highest quality and quantity RNA extractions were chosen to build cDNA libraries.

2.2.4 Library preparation

The 32 cDNA libraries were constructed using the Illumina TruSeq Stranded mRNA Low Sample (LS) Protocol (Illumina, Inc. 2013) by Dr. Chandra McAllister. No alterations to the protocol were made. Up to 12 libraries were prepared at a time. In summary, extracted ss RNA was first diluted in Milli-Q water (Merck Millipore, Burlington, Massachusetts, United States) to a final volume of 50 μl. Highly abundant RNA molecules, such as ribosomal RNA (rRNA) and transfer RNA (tRNA), were extracted along with the mRNA. The rRNA and tRNA were undesirable and removed because they diluted the functionally relevant protein-coding transcripts (Herbert *et al.* 2018). This was done by taking advantage of the poly(A) tails at the 3' end of the mRNA sequences, which consisted of a sequence of adenosine monophosphates. The undesirable RNA molecules did not have a poly(A) tail. Magnetic beads with covalently attached poly(T) oligonucleotides consisting of a sequence of thymine monophosphates

were mixed with the extracted RNA, enabling the mRNA poly(A) tails to hybridize to the poly(T) sequences. The suspension of mRNA hybridized to the magnetic beads was centrifuged to create a pellet, allowing all other elements of the extracted RNA, such as rRNA and tRNA, to be discarded in the supernatant. During the second enrichment process, mRNA strands were fragmented to lengths of 300-500 bp using chemical fragmentation, followed by cDNA synthesis. The fragmented mRNA was primed for cDNA synthesis with random hexamers, and the mRNA fragments were used as the template for the synthesis of the first strand of cDNA. Next, the first strand of cDNA and mRNA were separated, and the first strand was used as the template for the second strand of cDNA. So that only the antisense strand (i.e. corresponding to the gene) would be amplified later, dUTP was used rather than dTTP when generating the second strand *in vivo*. This method produced "stranded" cDNA where the orientation of the expressed gene was maintained. The ds cDNA was adenylated at the 5' ends with a single adenosine monophosphate. The 63 bp TruSeq Indexed Adapters (Illumina, Inc., San Diego, California, United States) were then ligated to the 5' ends of the ds cDNA. PCR followed so that the cDNA strands were amplified in anticipation of sequencing. Libraries were validated for quality using the Agilent DNA 1000 kit (Agilent Technologies, Inc., Santa Clara, California, United States).

2.2.5 Illumina sequencing

Once indexed, the 32 cDNA libraries were combined into four separate pools in a random fashion, with each pool containing one biological replicate for each of the eight treatments. Each of the four pools were separately prepped using the NextSeq System Denature and Dilute Libraries Protocol (Illumina, Inc. 2015). No alterations to the protocol were made. Once diluted to 1.8 pM, each pool was sequenced by Sophie Dang on the Illumina NextSeq 500 platform (Illumina, Inc., San Diego, California, United States) located at the Molecular Biology Service Unit of the University of Alberta's Department of Biological Sciences (Edmonton, Alberta, Canada) according to the NextSeq 500 System Guide (Illumina, Inc. 2018). In summary, ss cDNA was washed over the four-lane acrylamide-coated glass flow cell. The cDNA strands hybridized to a lawn of two types of oligonucleotide sequences that were randomly distributed in equal concentrations across the flow cell lanes. These oligonucleotides were complementary to the TruSeq adapter sequence and the reverse-complement of the TruSeq adapter sequence, respectively. DNA-dependent DNA polymerase along with primers complementary to the primer regions of the cDNA fragments were used to create ds DNA. The double strands were denatured, and the original sequences washed away, leaving behind tethered strands that were complementary to the original cDNA fragments. Since both ends of these tethered strands had adapter regions, both ends of the sequences hybridized to the flow cell forming bridges. DNA polymerase and primer sequences were used to create ds DNA bridges. The double strands were then denatured, resulting in two complementary strands tethered to

the flow cell. This cloning process was run on millions of cDNA sequences simultaneously and resulted in millions of clusters less than 1 micrometer in diameter composed of cloned sequences. The reverse strands were then cleaved and washed away, leaving the forward strand tethered to the flow cell. The forward strand had the same orientation as the extracted mRNA. The exposed 3' ends were capped to prevent bridging, and DNA-polymerase and primer sequences were used to sequence the forward reads from the forward strand. These reads were sequenced by a proprietary technique known as sequencing by synthesis. Fluorescently tagged nucleotides competed for hybridization to the tethered sequences. When a nucleotide hybridized, a certain wavelength particular to the fluorescent tag of that base was emitted. The quality of the emission was interpreted by the sequencer as the quality of the hybridization of that specific base. Signals from clusters of cloned sequences were strong enough to be read by the sequencer, which effectively transformed a chemical signal into a digital signal. This process was repeated until the desired read length of 150 bp was obtained. The forward reads were then washed off. Then, primers specific to the indices were used to sequence the index sequences. This was how the sequencer knew which clusters went with which libraries. The 3' ends were uncapped, and the sequences formed bridges. Bridges were made ds with primers and DNA polymerase. The double strands were denatured, and the forward sequences were cleaved from the flow cell lanes and washed off. Primers were used to sequence the reverse reads, followed by the sequencing of the index sequences. Paired-end, stranded, 150 bp

reads were generated using all four lanes of the NextSeq flow cell, and data were parsed and clustered by index. Each lane produced its own sequence data file, which included read identifiers, the sequence and quality scores associated with each base in the sequence. This generated eight separate sequence files per library in binary base call format, with four forward sequence files, file names tagged R1, and four reverse sequence files, file names tagged R2. The forward sequences had the same orientation as the gene sequence without intron regions, while the reverse sequences had the same orientation as the mRNA sequence. Sequence data were then converted to paired-end read FASTQ files (Cock *et al.* 2010) using the Illumina BaseSpace Sequence Hub bcl2FASTQ Conversion Software (Illumina, Inc., San Diego, California, United States).

2.2.6 Quality control

Raw, paired-end read FASTQ files generated from Illumina NextSeq 500 (Illumina, Inc., San Diego, California, United States) technology consisting of four forward and four reverse sequences per library were assessed for quality using the CLC Genomics Workbench v9.5.2 (*www.qiagenbioinformatics.com*) on a personal server. First, the FASTQ files were imported into the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) using the *Import* tool for Illumina reads along with the following parameters: paired-end (forward-reverse) reads, minimum paired-end distance=1, maximum paired-end distance=1000, Remove failed reads=TRUE. For each library, all eight sequence data files were imported

together, and the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) compiled a single sequence data element for that library automatically. The quality control reports were produced as follows: The quality control (QC) tool was accessed by going to Toolbox | NGS Core Tools | Create Sequencing QC Report, and the library was selected. Both of the boxes for a quality analysis and an over-representation analysis were checked, and the following outputs were selected: "Create a graphical report", "Create supplementary report", and "Create duplicated sequence list". This was done for all 32 libraries. Quality was measured by the following statistics: sequence length distribution, base coverage distribution, base-wise nucleotide contributions, sequence-wise and base-wise GC distribution, sequence-wise and base-wise ambiguous nucleotide distribution, and per-sequence and per-base quality distribution. The quality distribution metric was given by the quality scores that were generated by the Illumina BaseSpace Sequence Hub (Illumina, Inc., San Diego, California, United States) when sequence data were converted into FASTQ files.

2.2.7 Trimming optimization

The FASTQ files were pre-processed using the CLC Genomics Workbench v.9.5.2 (*www.qiagenbioinformatics.com*) to remove nonsense sequences in the following order: low-quality sequence determined by the quality score cutoff, reads with more than two ambiguous bases, adapter sequence artifacts and lastly reads below 75 bp. The adapter sequence was the 34 bp reverse complement of the

primer region of the TruSeq Index Adapter (Illumina, Inc., San Diego, California, United States). The CLC Genomics Workbench (*www.qiagenbioinformatics.com*) looked for adapter matches on the minus strand at two different positions in the read sequence called internal matches or end matches. The definition of an end match was that the alignment of the adapter started at the read's 5' end, and the definition of an internal match was that the alignment of the adapter did not start at the 5' end. Both the internal match and end match values were chosen for optimization. The third parameter chosen for optimization was the quality score cutoff for trimming low-quality bases.

The first step to optimizing these parameters was to create a the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) analysis workflow, which became the pipeline used to test different combinations of trimming parameter values (Figure 2.3). The workflow took a single library as input and a particular set of trim parameters and trimmed and assembled the library *de novo* using default assembly parameters. Both a trim report and a *de novo* assembly report were produced as output. These two reports were used to make comparisons between each set of trim parameters and to empirically choose the optimal set. The workflow was created by clicking File | New | Workflow and adding the following elements: an input for the library that was to be optimized, an input for the list of adapter detection parameters, NGS Core Tools | Sample reads, NGS Core Tools | Trim Sequences, *De Novo* Sequencing | *De Novo* Assembly, and outputs for the trimmed sequences, the trim report, the assembled contigs and the *de novo*

assembly report. Elements other than “Workflow Input” and “Workflow Output” had three components (Figure 2.2). Elements were connected by either clicking on the “Workflow Input” element and dragging the mouse to the correct input option for the next tool, by clicking on the output option of an element box and dragging the mouse to the correct input option for the next tool, or by clicking on the correct output option of an element box and dragging the mouse to the proper “Workflow Output” element. Each element was then configured by double clicking on the element box and following the dialog box to set the desired parameters.

Figure 2.2. A CLC Genomics Workbench workflow element. The CLC Genomics Workbench user interface allows the researcher to develop a workflow pipeline to run bioinformatic analyses. Workflows are composed of elements that represent computational processes with an input and an output. At the top of the element box are the input options, the middle of the box makes explicit the bioinformatic process and the output options are located at the bottom of the box. Each element box has associated parameters set by the researcher.

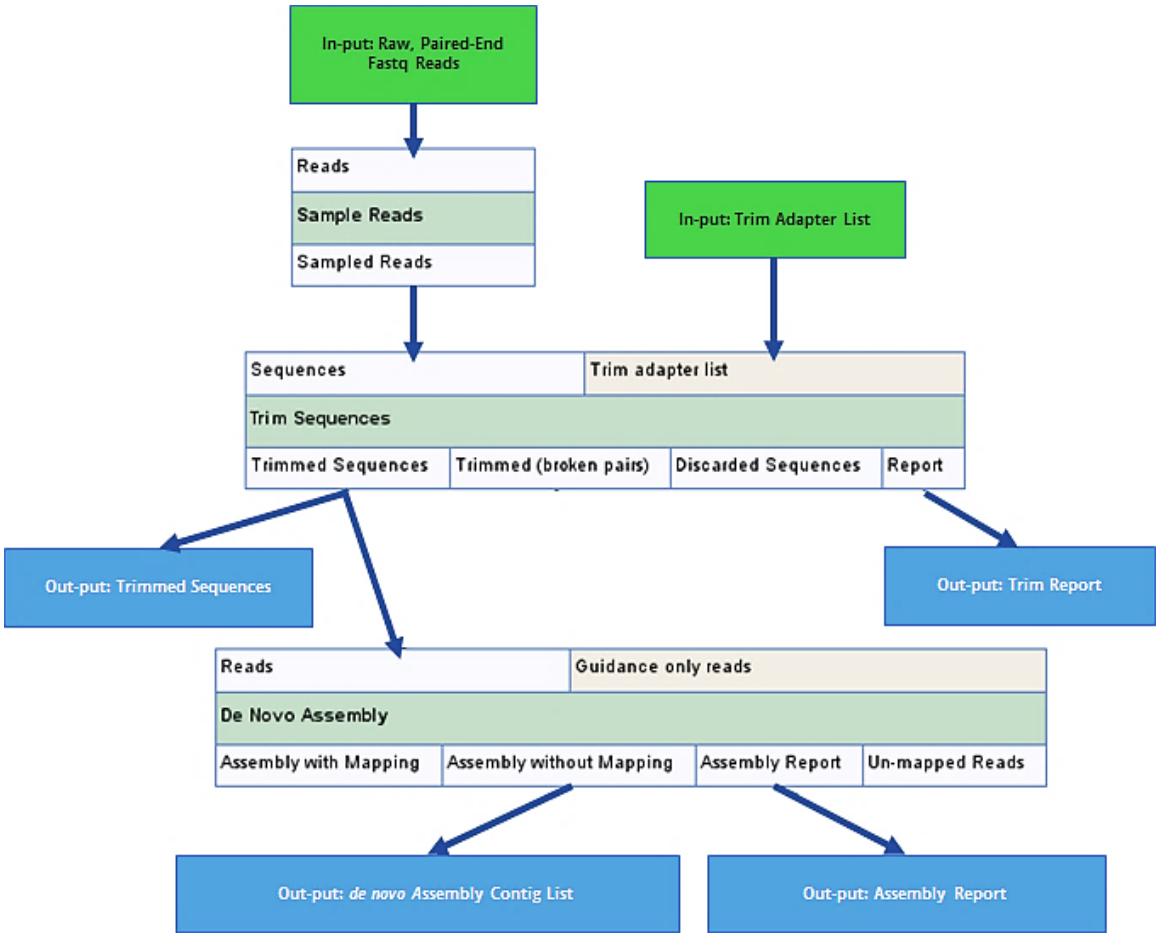


Figure 2.3. The CLC Genomics Workbench trimming optimization workflow. The CLC Genomics Workbench can be used to create workflows that are then implemented to carry out various processes that occur in a sequential manner. The element boxes in this workflow diagram show the processes used for read trimming optimization, namely sampling reads, trimming reads and *de novo* assembly. The workflow input was 150 bp paired-end, stranded reads generated from the LowN.Wound.2P.1 library. The outputs were trimmed sequences, a trim report, contigs and a *de novo* assembly report. This workflow diagram was executed as a pipeline to optimize trim parameters after setting parameters for each element.

One million paired-end reads (two million reads in total) were sampled from the LowN.Wound.2P.1 library to decrease processing time during the optimization procedure. The TruSeq Index Adapter (Illumina, Inc., San Diego,

California, United States) sequences, ligated during library preparation, were garnered from the Illumina Adapter Sequences document (Illumina, Inc. 2016) and used to craft the Trim Adapter List. The Trim Adapter List was produced by going to File | New | Trim Adapter List and clicking the *Add row* button. Each row was given the following information: a unique name, the sequence to be removed (i.e. the reverse-complement of the primer region of the TruSeq Index Adapter), the strand the sequence was on, the action to be performed upon sequence detection, the alignment score costs for a mismatch and a gap, and the alignment score thresholds for matches at the ends of reads (the end match) and near the ends of the reads (the internal match; Table 2.2). Illumina TruSeq library preparation resulted in an additional adenine before the adapter sequence on the 5' end (Illumina, Inc. 2013). This produced an extra thymine at the 3' end of the reverse-complement of the primer region of the TruSeq Index Adapter that also needed to be removed. The trim and *de novo* assembly parameters for the workflow were configured by double clicking on the trim or assembly element, following the dialog box and choosing the appropriate parameters (Table 2.3). The trim report results of each parameter combination were judged by the number of adapters trimmed and the number of reads after trimming. Maximizing the number of adapters trimmed and the number of reads after trimming was optimal. *De novo* assembly report results were judged by the N50 length and the number of contigs post assembly. Maximizing the N50 length and minimizing the number of contigs was optimal (Surget-Groba and Montoya-Burgos 2010).

Table 2.2. The CLC Genomics Workbench Trim Adapter List parameters chosen for trimming optimization. Parameters used for the Trimming Optimization Workflow (Figure 2.3) are detailed. The 34 bp adapter sequence was the reverse-complement of the primer region of the TruSeq Indexed Adapter with an additional thymine at the 3' end. The sequence orientation was on the minus strand. The error costs were set to default values and the match cutoff values were optimized before choosing the best parameter combination. Optimization included the comparison of trim and assembly outcomes.

Parameter	Value
Sequence	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Strand	minus
Action	remove adapter
Mismatch cost	2
Gap cost	3
Allow internal matches	TRUE
Minimum score	To be optimized*
Allow end matches	TRUE
Minimum score at end	To be optimized*

*Value optimized using the methods described in this section.

Table 2.3. The CLC Genomics Workbench trim and *de novo* assembly parameters chosen for trimming optimization. Parameters listed were used when creating the Trimming Optimization Workflow (Figure 2.3). The quality score cutoff was optimized along with the internal match score and end match score for adapter detection. The trim parameters, aside from those being optimized, were left as default as were all assembly parameters. Trim and assembly outcomes were compared following the execution of the workflow.

Parameter	Value
Trim Sequences	
Trim adapter list	Trim Adapter List
Also search on reversed sequence	FALSE
Ambiguous trim	TRUE
Ambiguous limit	2
Quality trim	TRUE
Quality limit	To be optimized*

Table 2.3. The CLC Genomics Workbench trim and *de novo* assembly parameters chosen for trimming optimization. Continued.

Parameter	Value
Trim Sequences	
Remove 5' terminal nucleotides	FALSE
Remove 3' terminal nucleotides	FALSE
Maximum number of nucleotides in reads	151
Minimum number of nucleotides in reads	75
Discard short reads	TRUE
Discard long reads	TRUE
De Novo Assembly	
Mapping mode	Create simple contigs
Automatic bubble size	TRUE
Automatic word size	TRUE
Minimum contig length	200
Perform scaffolding	FALSE
Auto-detect paired distances	TRUE
Create report	TRUE

*Value optimized using the methods described in this section.

2.2.8 Assembly optimization

With the goal to produce the best the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) *de novo* assembly, two parameters were optimized: k-mer length and bubble size. Assemblies were run by clicking on Toolbox | *De Novo* Sequencing | *De Novo* Assembly and selecting the appropriate parameters (Table 2.4). A comprehensive comparison of generated *de novo* assembly reports was made using the following metrics: N50 length, average contig length and the number of contigs, while continuing to maximize both the N50 and

average contig lengths and to minimize the number of contigs (Surget-Groba and Montoya-Burgos 2010).

Table 2.4. The CLC Genomics Workbench *de novo* assembly parameters chosen for assembly optimization. All parameters, aside from those being optimized, were default values. Mapping mode indicated if the assembly was by mapping to a reference genome or transcriptome or was being done *de novo*. The mapping mode value for this analysis indicates a *de novo* approach. Bubble size and k-mer selection (word size) were manually altered to create different parameter combinations for assembly optimization. Scaffolding was not performed, and distances between the pair-end sequences were registered as part of the *de novo* assembly report. The assembly report for each parameter combination was used to compare outputs in order to select the best possible combination.

Assembly parameter	Value
Mapping mode	Create simple contig sequences
Automatic bubble size	FALSE
Automatic word size	FALSE
Bubble size	To be optimized*
Word size	To be optimized*
Minimum contig length	200
Perform scaffolding	FALSE
Auto-detect paired distances	TRUE
Create report	TRUE

*Value optimized using the methods described in this section.

2.2.9 Master phloem transcriptome construction

Following trim and assembly optimization, the CLC Genomics Workbench v9.5.2 (*www.qiagenbioinformatics.com*) was used to attempt to assemble a master phloem transcriptome from the reads of all phloem libraries derived from

the low N, mock-inoculated tree RNA samples. The assembly method described above was employed together with the optimized trimming and assembly parameters determined from the comparisons described in Sections 2.2.7 and 2.2.8, respectfully.

2.2.10 Assembler comparison

Reads from sample LowN.Wound.2P.1 were trimmed by the CLC Genomics Workbench v9.5.2 (*www.qiagenbioinformatics.com*) using methods described in Section 2.2.7 with optimized parameters and a minimum read length of 51 bp. Retaining reads ≥51 bp in length was determined by Dr. Rhiannon Peery after a literature review of trim parameters. The commercial software the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) and the open-source assembler Trans-ABySS v1.5.5 (Robertson *et al.* 2010), were used to assemble the data. Four assemblies were compared, two made with Trans-ABySS (Robertson *et al.* 2010) and two made with the CLC Genomics Workbench (*www.qiagenbioinformatics.com*). The minimum contig length and Trans-ABySS (Robertson *et al.* 2010) k values were optimized by Dr. Rhiannon Peery. The CLC Genomics Workbench (*www.qiagenbioinformatics.com*) *de novo* assembly was performed as described in Section 2.2.8 using a minimum contig length of 500 bp and optimized assembly parameters. While the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) was run as a graphical user interface on a private computer, Trans-ABySS (Robertson *et al.* 2010) was run on a Bash Unix

shell on the Compute Canada WestGrid server Hungabee (*www.westgrid.ca; www.computecanada.ca*). The portable batch system (pbs) script had the following header:

```
#!/bin/bash
#PBS -S /bin/bash
#PBS -l procs=32                    number of threads needed
#PBS -l pmem=8190mb                 amount of memory needed
#PBS -l walltime=60:00:00           time needed to run job
#PBS -N transabyss                  name of the job
#PBS -m bea                         send email notifications when the run
                                    begins, ends and is aborted
#PBS -M normingt@ualberta.ca        email address
```

The script then identified and installed programs required for the proper implementation of “transabyss” and “transabyss-merge” using the command “module load”:

```
module load application/Trans-ABySS/1.5.5
module load application/ABySS/1.5.2
module load application/gmap/2014-12-02
module load application/samtools/0.1.19
module load application/blat/3.5
module load application/python/2.7.3
module load library/igraph/0.7.1
```

The following code was then implemented for each k value as part of the pbs script:

```
transabyss
        --pe ${FASTQ reads}             paired-end reads
        --SS                            strand-specific library preparation
        --outdir ${assembly directory}  directory for output file
        --name ${output file name}      name of output file
        -k ${kmer size}                 size of k-mer
        --threads 16                    number of threads required
        --island 0                      minimum island size
```

-s 50	seed length
-c 2	minimum mean k-mer coverage

An island was a high quality contig with unique sequence data. Setting island equal to 0 told Trans-ABySS (Robertson *et al.* 2010) to not remove any assembled sequences that were completely unique. The seed length (argument -s) was the minimum portion of a sequence required for building complete contigs. Setting the seed length to 50 told Trans-ABySS (Robertson *et al.* 2010) that FASTQ reads were a minimum of 50 bp long. The argument -c indicated the minimum mean k-mer coverage required of a contig. Setting this parameter to 2 told Trans-ABySS (Robertson *et al.* 2010) that the minimum number of k-mers for each k bp section of sequence must be at least two. Then, to merge the assemblies, the following code was implemented:

```
transabyss-merge kmer1-final.fa kmer2-final.fa kmer3-final.fa kmer4-final.fa kmer5-final.fa --mink kmer1 --maxk kmer5 --threads 16 --SS --out ./kmer1-kmer5.merged.fa
```

Finally, assembly quality was assessed with the command “abyss-fac” followed by the assembly FASTA file (Pearson and Lipman 1988). These Trans-ABySS (Robertson *et al.* 2010) scripts are available in Appendix B.

A comprehensive comparison was made using the following metrics: N50 length, maximum contig length, the number of contigs and run time (Figure 2.4). An added element of the comparison was the percentage of reads that aligned with their respective assemblies. Alignments were performed using the CLC Genomics

Workbench v9.5.2 (*www.qiagenbioinformatics.com*). First the two Trans-ABySS v1.5.5 (Robertson *et al.* 2010) assemblies in FASTA file format were imported by clicking Import | Standard Import, opening the FASTA file, and clicking *Automatic import*. Then, Toolbox | NGS Core Tools | Map Reads to Reference was clicked and an assembly was chosen as the reference. The "No masking" option was selected, and all default mapping parameters were employed as part of the workflow (Table 2.5).

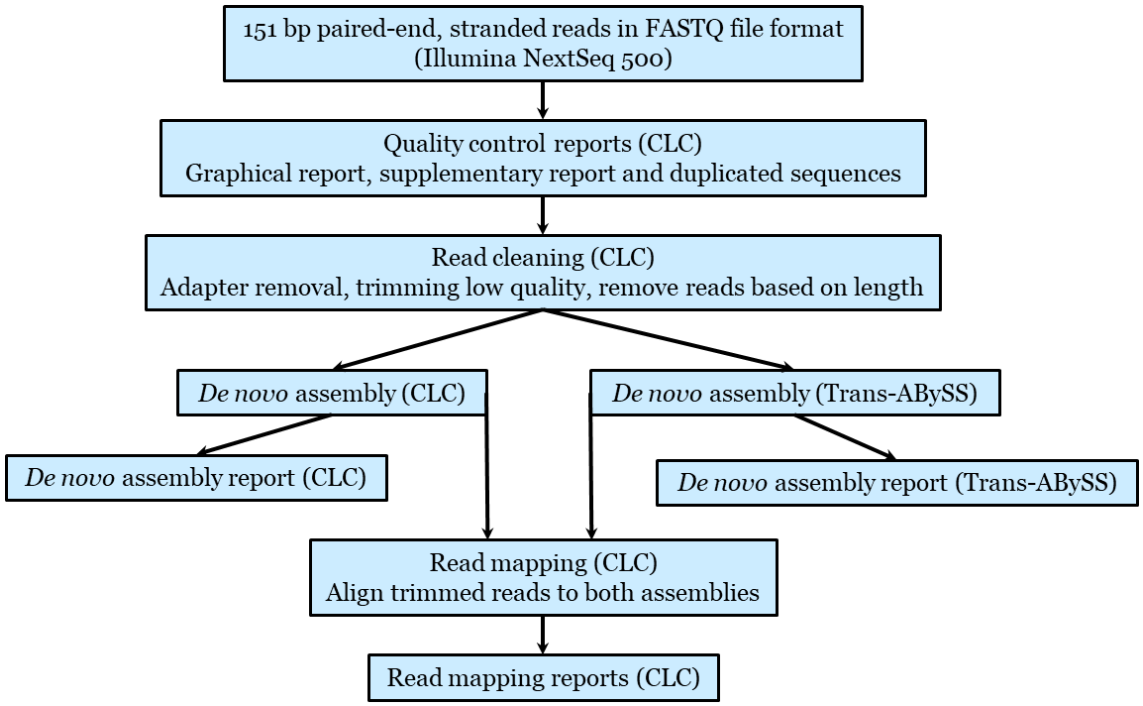


Figure 2.4. Bioinformatic pipeline for the comparison of *de novo* assemblies by the CLC Genomics Workbench and Trans-ABySS. Each step indicates which technology or software was used for its execution. Metrics for comparison included N50 length, maximum contig length, the number of contigs, the percentage of reads mapped back to the assembly and run time. The CLC Genomics Workbench was used on a personal computer, while Trans-ABySS was available for use on the WestGrid server Hungabee.

Table 2.5. The CLC Genomics Workbench mapping parameters for assembly comparison with Trans-ABySS. Reads generated from the LowN.Wound.2P.1 library using Illumina NextSeq 500 technology were mapped back to their assemblies using the CLC Genomics Workbench as a component of assembler comparison. All parameters chosen were default values.

Mapping parameter	Value
No masking	TRUE
Match score	1
Mismatch cost	2
Linear gap cost	TRUE
Insertion cost	3
Deletion cost	3
Length fraction	0.5
Similarity fraction	0.8
Global alignment	FALSE
Auto-detect paired distances	TRUE
Map randomly	TRUE
Ignore	FALSE
Create reads track	FALSE
Create stand-alone read mappings	TRUE
Create report	TRUE
Collect un-mapped reads	TRUE

2.3 Results

2.3.1 Sequencing

The sequencing of 32 lodgepole pine cDNA libraries produced an average of 47.28 GB of high quality paired-end read data per library, yielding a combined data set of 1.51 TB. On average, 73,872,029 clusters were produced for each library during the sequencing process, and the NextSeq 500 platform (Illumina, Inc., San

Diego, California, United States) generated on average 147,744,058 reads per library. Sequencing results are described in detail in Appendix A (Table 6.1). The CLC Genomics Workbench (*www.qiagenbioinformatics.com*) quality control tool assessed each data set using metrics, such as average quality score and variations in position-specific ambiguity. Per sequence and per base quality QC results are detailed in Tables 6.2 and 6.3, respectively, in Appendix A. All libraries passed quality control. The maximum read length for all libraries was 151 bp. The extra base pair was an artifact of sequencing that was necessary to achieve a read length of 150 bp. The minimum read length for all libraries was 35 bp. The average read length among all sequenced libraries was 128.9 bp, and an average of 99.87% of all sequences per library had no ambiguous bases. The mean percentage of sequences per library with an average quality score greater than or equal to 30 was 81.73%. The NextSeq 500 platform generated an average of 18,813,133,982 bases per library. The mean percentage of non-ambiguous bases per library was 97.92% and the mean percentage of bases with a median quality score greater than or equal to 30 was 92.41% per library. These statistics, garnered from the CLC Quality Control Supplementary Report, highlighted the high quality of the lodgepole pine sequence data.

2.3.2 Trimming optimization

Trimming raw reads using the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) required a thorough optimization process in

which different parameters were compared based on the quality of both the reads post trim and their subsequent *de novo* assembly. Three different match values were selected for adapter detection optimization for both the internal match (11, 14 or 17) and the end match (11, 14 or 17). The CLC Genomics Workbench Manual (*http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/754/index.php?manual=Introduction_CLC_Genomics_Workbench.html*) highlights the benefits of searching for both internal and end matches (*http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/755/index.php?manual=Adapter_trimming.html*). Specific values were chosen by the careful examination of the duplicated sequence list. The duplicated sequence list was produced as a component of the CLC quality report for LowN.Wound.2P.1. A portion of the adapter sequence, GTGACTGGAGT, was found to be overrepresented amongst the sampled unprocessed reads. This sequence was the first 11 nucleotides at the 3' end of the adapter sequence, thus informing the choice of 11 as an end match value for optimization. To observe the impact of increased match value stringency on the loss of valuable sequence data, larger end match values of 14 and 17 were also chosen for comparison. The same three values were selected for the internal match.

Phred-based quality scores were assigned to each base called by the NextSeq 500 system (Illumina, Inc., San Diego, California, United States). Quality scores were optimized as p-values using the CLC Genomics Workbench (*www.qiagenbioinformatics.com*). P-values, associated with quality scores (Q) by

the equation $p = 10^{(-Q/10)}$ (Ewing and Green 1998; Ewing *et al.* 1998), indicated the probability that a base was incorrectly called by the sequencer (*http://resources.qiagenbioinformatics.com/manuals /clcgenomicsworkbench/755/index.php?manual=Quality_trimming.html*). Quality score values for trimming optimization were inspired by MacManes (2014) who compared the transcriptome assemblies produced by a variety of quality score cutoffs. Therefore, p=0.05 (Q=13) and p=0.01 (Q=20) were selected for optimization of the LowN.Wound.2P.1 sample.

The quality of the trimmed reads and their *de novo* assemblies were compared for the 18 trim parameter combinations (3 internal match values x 3 end match values x 2 p-values = 18). Quality was determined by the examination of the following statistics: the number of reads after trim, the number of reads adapter trimmed, the total number of contigs and N50 length. N50 is one of the most basic measurements of assembly quality. To calculate the N50 length, the contig lengths are sorted from largest to smallest. The contig lengths are summed starting with the smallest length until the sum is larger than 50% of the total length of all contigs combined. The N50 length is the last contig length added that pushes the sum over the 50% threshold.

Maximizing the number of reads post trim, the number of reads adapter trimmed and the N50 length, while minimizing the number of contigs, was desirable when optimizing trim parameters (Del Fabbro *et al.* 2013; Didion *et al.* 2017; Surget-Groba and Montoya-Burgos 2010). Using the CLC Genomics

Workbench (*www.qiagenbioinformatics.com*) trim sequences tool, an internal match value of 17 with an end match of 11 was optimal based on the trim results (Figure 2.5). The adapter detection parameter combination maximized the number of reads post trim with the third highest number of reads adapter trimmed. The internal match value of 17 with an end match of 11 was also optimal based on *de novo* assembly results (Figure 2.6). The adapter detection parameter combination minimized the number of contigs and yielded the second highest N50 value. A p-value of 0.01 was chosen to reflect the high quality of the sequence data.

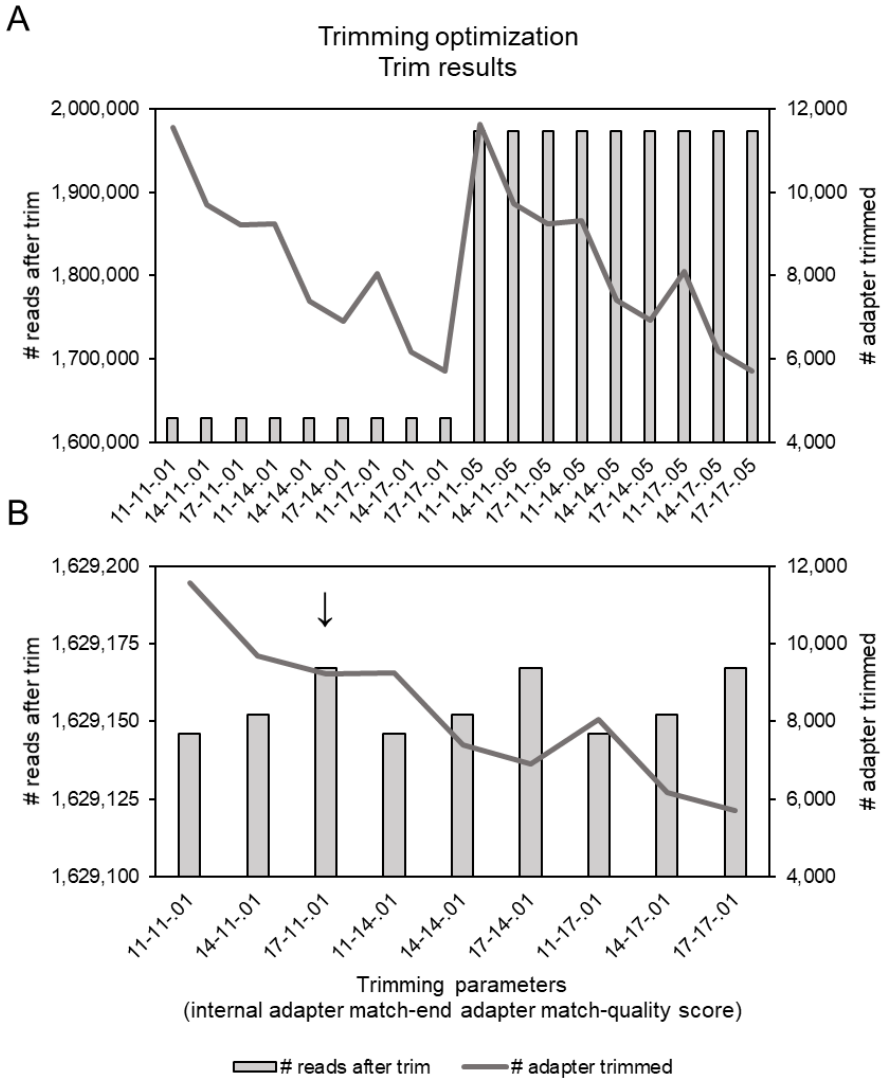


Figure 2.5. Comparison of trim statistics for the optimization of read trimming. The pre-processing of one million paired-end, stranded reads from sample LowN.Wound.2P.1 using the CLC Genomics Workbench was optimized for adapter match and quality score cutoff. Eighteen different parameter combinations produced different trim outcomes for comparison. The number of reads post trim and the number of reads adapter trimmed were used for quality assessment, where maximizing both values was optimal. (**A**) Comparison of trim statistics using the CLC Genomics Workbench trim report with all parameter combinations, (**B**) comparison of trim results with quality score cutoff p=0.01. The optimized statistics are marked with an arrow (↓), showing the internal match of 17, end match of 11 and p-value cutoff of p=0.01 to be optimal.

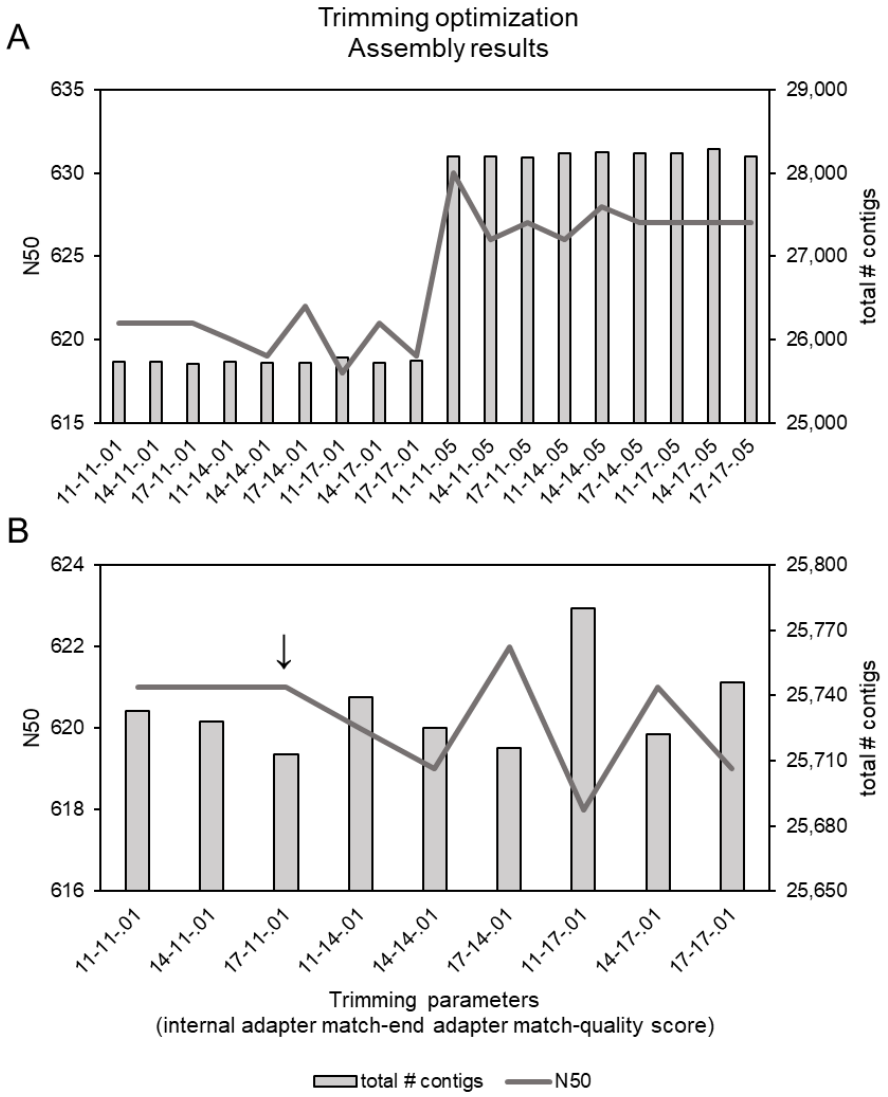


Figure 2.6. Comparison of assembly statistics for the optimization of read trimming. The pre-processing of one million paired-end reads from sample LowN.Wound.2P.1 using the CLC Genomics Workbench was optimized for adapter match and quality score cutoff. Eighteen different parameter combinations produced different assembly outcomes that were then compared. Default assembly parameters (k=20, bubble=50 and minimum contig length=200 bp) were utilized for all assemblies. The N50 length and the number of contigs were used for quality assessment, where maximizing both values was optimal. (**A**) Comparison of assembly statistics using the CLC Genomics Workbench *de novo* assembly report with all parameter combinations, (**B**) comparison of assembly results with p=0.01. The optimized statistics are marked with an arrow (↓), showing the internal match of 17, end match of 11 and p-value cutoff of p=0.01 to be optimal.

2.3.3 *De novo* assembly optimization

For de Bruijn graph-based assemblers, such as the commercial software the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) and the open source assembler Trans-ABySS (Robertson *et al.* 2010), the k-mer size can have a significant impact on the quality of the assembly (Miller *et al.* 2010). To optimize the k-mer length for the CLC Genomics Workbench (*www.qiagenbioinformatics.com*), 1 million paired-end reads generated from the LowN.Wound.2P.1 library were assembled with multiple k-mer lengths. K-mer lengths ranged from the minimum (k=20) to the maximum (k=60) settings, with a 10-mer difference in each setting using the default minimum contig length of 200 bp and the default bubble size of 50. The following statistics were compared using the *de novo* assembly reports produced by the CLC Genomics Workbench (*www.qiagenbioinformatics.com*): N50 length, average contig length, and the number of contigs. The 20-mer yielded both the highest N50 value and the largest average contig length while keeping the number of contigs relatively low (Figure 2.7). Bubble length was also optimized by choice of the default value (50 bp) to the average read length post trimming (around 130 bp), with a 20 bp difference between each selection (Figure 2.8). For bubble resolution optimization, differences between number of contigs was minimal, and N50 and average contig lengths varied only slightly. The default 50 bp bubble length, optimized for high quality short reads, was selected as the optimized parameter.

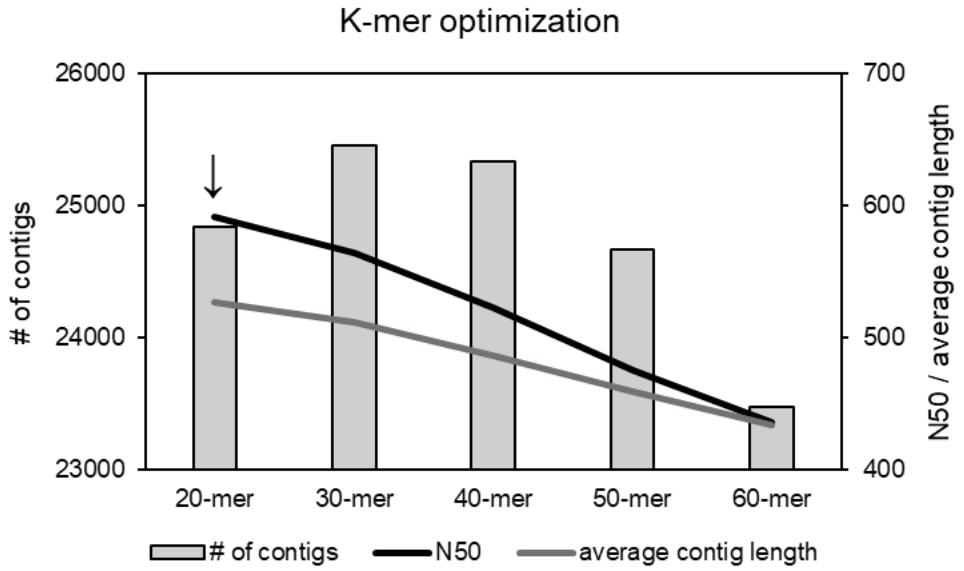


Figure 2.7. The CLC Genomics Workbench *de novo* assembly k-mer length comparison. One million paired-end, stranded reads were sampled from the LowN.Wound.2P.1 sequence data and assembled using the CLC Genomics Workbench. A default minimum contig length of 200 bp was chosen for all assemblies. Results presented in the CLC Genomics Workbench *de novo* assembly reports were compared for the number of contigs, the average contig length and the N50 length. Five different k-mer lengths were tested for assembly optimization, starting with the default value and increasing to the maximum k-mer size: 20, 30, 40, 50 and 60. The default bubble size of 50 was chosen for k-mer optimization. The optimized statistics are marked with an arrow (↓), showing k=20 was optimal for *de novo* assembly.

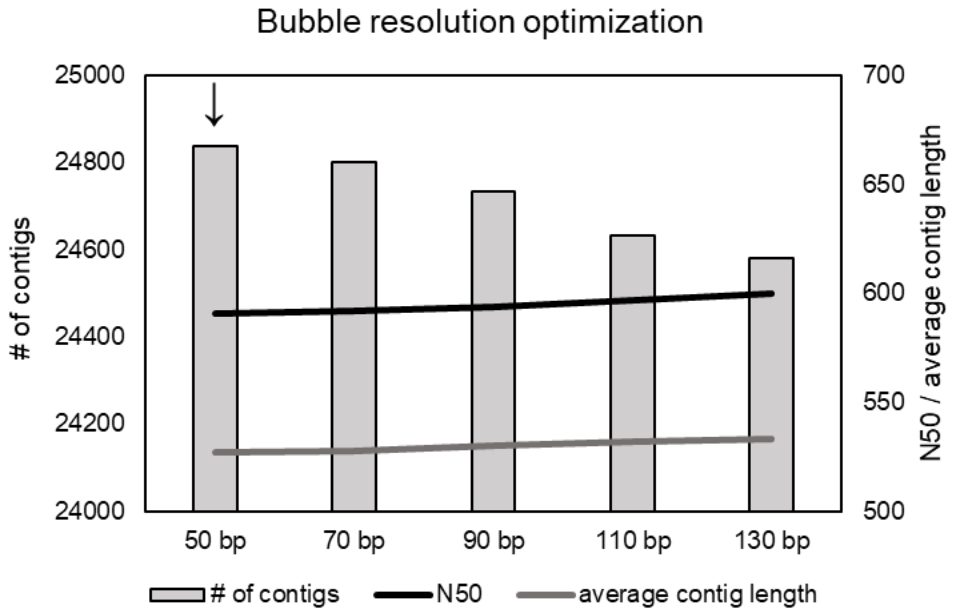


Figure 2.8. The CLC Genomics Workbench *de novo* assembly bubble size comparison. One million paired-end reads were sampled from the LowN.Wound.2P.1 cDNA library and assembled using the CLC Genomics Workbench. A default minimum contig length of 200 bp was chosen for all assemblies. Results presented in the CLC Genomics Workbench *de novo* assembly reports were compared for the number of contigs, the average contig length and the N50 length. Five different bubble lengths were tested for assembly optimization starting with the default and increasing to the average contig length: 50, 70, 90, 110 and 130. The default k-mer size of 20 was chosen for bubble optimization. The optimized statistics are marked with an arrow (↓), showing a bubble size of 50 was optimal for *de novo* assembly.

2.3.4 Master assembly

In anticipation of differential expression analysis (Chapter 4) and given that lodgepole pine lacks an available well-annotated genome reference, we chose to construct a reference transcriptome *de novo* from all sequenced cDNA libraries, termed the "master transcriptome". Differential expression analysis relies on

sequence alignment maps, produced by the mapping of reads to a reference genome or transcriptome, to properly count the number of contigs expressed by each biological sample. We reasoned that a more complete and accurate master transcriptome would result from assembling reads from a greater number and diversity of libraries. Therefore, an assembly consisting of more than one secondary phloem library was attempted using the optimized trim and *de novo* assembly parameters (Table 2.6). The CLC Genomics Workbench (*www.qiagenbioinformatics.com*) could not assemble more than three secondary phloem samples combined even when ample computational resources were provided (Table 2.7). Therefore, the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) displayed a lack of computing prowess when the number of reads exceeded a certain threshold.

Table 2.6. Optimized parameters used for master transcriptome assembly with the CLC Genomics Workbench. Trim and *de novo* assembly parameters were optimized by procedures detailed in Sections 2.2.7 and 2.2.8, respectively, and the results of optimization are consolidated below. All values were default aside from those that were optimized, namely the internal match score and the end match score for adapter detection, the quality score cutoff for read trimming, and the k-mer and bubble size for *de novo* assembly.

Parameter	Value
Trim Adapter List	
Sequence	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Strand	minus
Action	remove adapter
Mismatch cost	2
Gap cost	3
Allow internal matches	TRUE
Minimum score	17
Allow end matches	TRUE
Minimum score at end	11

Table 2.6. Optimized parameters used for master transcriptome assembly with the CLC Genomics Workbench. Continued.

Parameter	Value
Trim Sequences	
Trim adapter list	FALSE
Also search on reversed sequence	TRUE
Ambiguous trim	2
Ambiguous limit	TRUE
Quality trim	0.01
Quality limit	FALSE
Remove 5' terminal nucleotides	FALSE
Remove 3' terminal nucleotides	FALSE
Maximum number of nucleotides	151
Minimum number of nucleotides	75
Discard short reads	TRUE
Discard long reads	TRUE
De Novo Assembly	
Mapping mode	Create simple contig sequences
Automatic bubble size	FALSE
Bubble size	50
Automatic word size	FALSE
Word size	20
Minimum contig length	500
Perform scaffolding	FALSE
Auto-detect paired distances	TRUE
Create report	TRUE

Table 2.7. *De novo* assembly of a master transcriptome using the CLC Genomics Workbench. Optimized assembly parameters (k=20, bubble size=50) were used to construct an assembly of more than one sequenced library. First, reads from four biological replicates of low N, mock-inoculated, secondary phloem samples were pooled, and an assembly was attempted. When the assembly failed, three biological replicates from the same treatment group were pooled and successfully assembled.

Assembly	Number of reads	N50	Max contig length	Average contig length	Number of contigs	Run time
LowN.Wound.2P.1	130,210,622	1,082	13,972	664	75,734	0:49:18
Master transcriptome (3 biological replicates)	1,465,724,192	640	16,669	541	208,704	34:21:06
Master transcriptome (4 biological replicates)	1,943,000,172	—	—	—	—	28:15:44

2.3.5 Assembler comparison

The same library used for the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) trimming and assembly optimization, LowN.Wound.2P.1, was assembled using Trans-ABySS (Robertson *et al.* 2010) on the WestGrid server Hungabee (*www.westgrid.ca; www.computecanada.ca*). Optimized k values (25, 29, 33, 36, 41 merged and 33 stand-alone) and the minimum contig length (500 bp) were provided by Dr. Rhiannon Peery. The same minimum contig length was used for the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) so that a proper comparison could be made. The merged assembly of the five k-mer values chosen by Dr. Peery was optimal and compared with the optimized k=20 the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) assembly (Table 2.8). The k=33 parameter was the optimal stand-alone k value for Trans-ABySS (Robertson *et al.* 2010) assembly and was therefore compared with a the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) assembly utilizing the same k value (Table 2.8). The following statistics were used for the comprehensive comparison of the two assemblers: N50 length, maximum contig length, the number of contigs, the percentage of reads mapped back to the respective assemblies and the run time. Statistics for the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) were produced by the *de novo* assembly report, whereas statistics for Trans-ABySS (Robertson *et al.* 2010) were produced using the command abyss-fac. The percentage of reads mapped back was a supplemental metric for assembly quality

that had not been applied for the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) *de novo* assembly optimization. Run time was also included as an important metric for the comparison of the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) and Trans-ABySS (Robertson *et al.* 2010). The CLC Genomics Workbench (*www.qiagenbioinformatics.com*) outperformed Trans-ABySS (Robertson *et al.* 2010) with respect to minimizing the number of contigs and the assembly time (Table 2.8). The k=20 the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) assembly displayed a slight decrease in assembly quality with respect to the percentage of reads mapped back when compared with the k=33 CLC Genomics Workbench (*www.qiagenbioinformatics.com*) assembly. Trans-ABySS (Robertson *et al.* 2010) showcased remarkable accuracy based on the percentage of reads mapped back to the two assemblies (Table 2.8).

Table 2.8. The CLC Genomics Workbench vs. Trans-ABySS *de novo* assembly of a low N, mock-inoculated, secondary phloem sample. Reads from the LowN.Wound.2P.1 sample were trimmed using the CLC Genomics Workbench optimized parameters (internal match=17, end match=11, p=0.01) with a minimum read length of 51 bp. Trimmed reads were assembled by the CLC Genomics Workbench using a bubble size of 50 and either k=20 or k=33. Trimmed reads were also assembled by Trans-ABySS either a merged assembly with k=25, 29, 33, 36 and 41 or k=33. A minimum contig length of 500 bp was used by both assemblers, and assembly statistics were compared. The minimum read length, the Trans-ABySS k values and the minimum contig length were optimized by Dr. Rhiannon Peery.

Statistic	Optimized assembly parameters		Equal k values	
	Trans-ABySS-merge	CLC	Trans-ABySS	CLC
min contig length	500	500	500	500
k	25-29-33-36-41	20	33	33
N50	1,875	1,766	1,685	1,757
max length	10,566	9,612	10,431	9,460
# contigs	52,926	29,469	46,067	30,079
% mapped back	99.90%	87.47%	99.89%	88.72%
run time	21:45:31	0:46:04	4:21:06	0:52:53

2.4 Discussion

When working with non-model systems, the selection and optimization of a *de novo* assembler is essential for obtaining a reliable transcriptome (Cahais *et al.* 2012). This study sought to use optimized parameters to effectively compare two state-of-the-art bioinformatic tools for the *de novo* assembly of lodgepole pine NGS data. The CLC Genomics Workbench (*www. qiagenbioinformatics.com*) and Trans-ABySS (Robertson *et al.* 2010) were evaluated using the following metrics: assembly time, N50 length, maximum contig length, the number of contigs and the number of reads mapped back to respective assemblies. We also compared ease of use and the computational requirements of both tools. Using the CLC Genomics Workbench (*www.qiagenbioinformatics.com*), our intention was to map reads to a master transcriptome that was constructed from multiple libraries. However, the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) was unable to construct a master reference, and we were unable to make this comparison. Our

study highlights and reinforces the importance of the optimization and assessment of bioinformatic tools to achieve specific research goals.

2.4.1 Trimming optimization

The goal of read trimming is to reduce sequence length by removing all bases from the error prone 3′ end of the read while retaining the longest, highest quality sequence (Williams *et al.* 2016). Sequence data errors include low quality bases, ambiguous base calls and adapter sequence artifacts (Huse *et al.* 2007). Low quality base calls have a higher probability of being erroneous and can be caused by several factors, such as spot-specific signal noise or equipment malfunctions (Del Fabbro *et al.* 2013). Illumina quality scores use the Phred-based metric that ranges from 0 to 41, which is equivalent to a p-value ranging from 1 to 7.94e-5 (Ewing and Green 1998; Ewing *et al.* 1998). Though the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) offers p=0.01 as the default cutoff, this value can be optimized by comparing trim and assembly outcomes following the application of different levels of quality stringency.

Ambiguous base calls must also be removed. They are the result of no base being called by the sequencer at that position throughout the entire flow cycle, and their removal has been shown to greatly improve the overall quality of the data set (Huse *et al.* 2007). The CLC Genomics Workbench (*www. qiagenbioinformatics.com*) default number of ambiguous bases allowed in a read

post trimming is two. In a manner consistent with techniques used to optimize the adapter match values and the p-value cutoff, this value can be optimized depending on the quality of the raw sequence data and the research objectives.

Adapter artifacts must be detected prior to removal. Adapter contamination occurs when a fragment is shorter than the read length designated by the sequencer (Sturm *et al.* 2016). The CLC Genomics Workbench (*www.qiagenbioinformatics.com*) detects adapters using a semi-global match algorithm that begins at the 3' or 5' end, as designated by the researcher (*http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/755/index.php?manual=Adapter_trimming.html*). The algorithm tries to align the adapter sequence at the very end of the read. If no end alignment is possible, the assessment will continue down the sequence looking for an internal match. If no internal match is possible, the read passes inspection. If a match is made, either an end match or an internal match, gaps and mismatches in the alignment are given costs and each type of match is given a score cutoff, which must be optimized for the specific data set. Optimal is defined as maximizing both the number of high-quality reads adapter trimmed and the number of reads post adapter trimming (Didion *et al.* 2017).

A read with multiple errors is more likely to be of questionable overall quality and should be removed from the data set by the trimming software (Huse *et al.* 2007). It is also more likely to be longer or shorter than expected (Huse *et al.* 2007). Trimming based on read length is an essential component of pre-

processing, and it has direct implications for *de novo* assembly (Schatz *et al.* 2010; Surget-Groba and Montoya-Burgos 2010). Longer reads generally produce better assemblies because they can span repeat regions and thus prevent contig breakage (Schatz *et al.* 2010). This breakage increases the number of contigs, an indicator of a poor-quality assembly (Surget-Groba and Montoya-Burgos 2010). Therefore, the removal of sequences that fall below a certain threshold is critical. The read must be long enough to be unique, but the minimum read length cutoff should be short enough so as to not remove valuable sequence data. The choice of 51 bp was informed by Dr. Rhiannon Peery who used the 51 bp minimum read cutoff when preparing the master reference transcriptome (Chapter 4).

Trimming read data inappropriately can increase processing time and memory requirements downstream (Didion *et al.* 2017). Sequence quality drops towards the 3' end of the read, so it is assumed that once an error is encountered the remainder of the sequence is of poor quality (Bolger *et al.* 2014). When a read is completely removed, its pair may be retained. Maintaining both pairs assists in *de novo* assembly and read mapping (Lehri *et al.* 2017). When attempting *de novo* assembly, inappropriately trimmed data can cause errors, such as erroneous k-mers (Zerbino and Birney 2008), misalignments when mapping to a reference sequence and spurious variant calls when genotyping (Didion *et al.* 2017).

2.4.2 *De novo* assembly optimization

The process of transcriptome assembly attempts to reconstruct full-length transcripts (Pevzner *et al.* 2001). One technique is the application of the de Bruijn graph method (Zhang *et al.* 2011). Sequenced reads are cut into fragments of length k. When two fragments match by k-1 bases, they are oriented next to one another and the contig is extended by adjoining the last nucleotide of the second k-mer to the end of the first k-mer. The quality and length of a contig is altered by the choice of length k (Miller *et al.* 2010). Smaller values of k are often used for under-represented reads that have minimal k-mer coverage to minimize contig fragmentation (Miller *et al.* 2010). These ultimately reduce the number of contigs in the final assembly (Miller *et al.* 2010). Since requiring a larger k-1 overlap ensures a more accurate match, larger values of k in higher coverage regions reduce the number of sequencing errors (Miller *et al.* 2010).

the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) suggests a default k-mer size based on the number of nucleotides in the data set (*http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/755/index.php?manual=How_it_works.html*). In addition, it allows for even k values when other assemblers, such as the open source assembler Velvet (Zerbino and Birney 2008), do not. Odd k values are suggested because such k-mers cannot be their own reverse complement. K-mers that are their own reverse complement are called palindromic (Orenstein and Shamir 2013), and they have the potential to cause the de Bruijn graph to fold back on itself (Miller *et al.* 2010). The CLC Genomics Workbench (*www.qiagenbioinformatics.com*) avoids the folding back

of the de Bruijn graph by halting k-mer extension when either a palindromic even length k-mer or a palindromic superstring, created by two odd value k-mers, is detected (*http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/755/index.php?manual=How_it_works.html*).

The second parameter to be optimized for *de novo* assembly is bubble size. Bubbles, defined as bifurcations in the de Bruijn graph that merge back again, appear when reads have heterozygous single nucleotide polymorphisms (SNPs) or sequencing errors, such as insertions or deletions (Fasulo *et al.* 2002). When multiple SNPs or sequencing errors are in close proximity to one another such that the distance between them is smaller than the assigned k-mer size, a larger bubble occurs (*http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/755/index.php?manual=Bubble_resolution.html*). If the length of a bubble is below an assigned threshold, the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) will resolve the bubble by choosing the sequence that has the higher k-mer coverage. If the bubble length is greater than the assigned threshold, the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) will break up the contig into separate fragments. Larger bubble sizes increase the chance of misassembly by collapsing contigs that should be kept distinct (*http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/755/index.php?manual=Bubble_resolution.html*).

When creating a high-quality *de novo* assembly, the choice of minimum contig length is dictated by the research objectives and computing resources. If the intention is to construct full-length transcripts, then contigs that are deemed too short based on knowledge of transcriptome structure should not be included in the final assembly (Simão *et al.* 2015). Furthermore, minimum contig length should also be based on the availability of time and resources for downstream analyses. For example, if the number of contigs is very large, the mapping of reads back to an assembly can be extremely time and resource consuming (*http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/755/index.php?manual=De_novo_assembly_parameters.html*). In this type of situation, raising the minimum contigs length is advantageous.

Following the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) parameter optimization for trimming and *de novo* assembly, described in Section 2.2.7 and Section 2.2.8, respectively, this pipeline consolidates the results that are detailed in Section 2.3.2 and Section 2.3.3 (Figure 2.9).

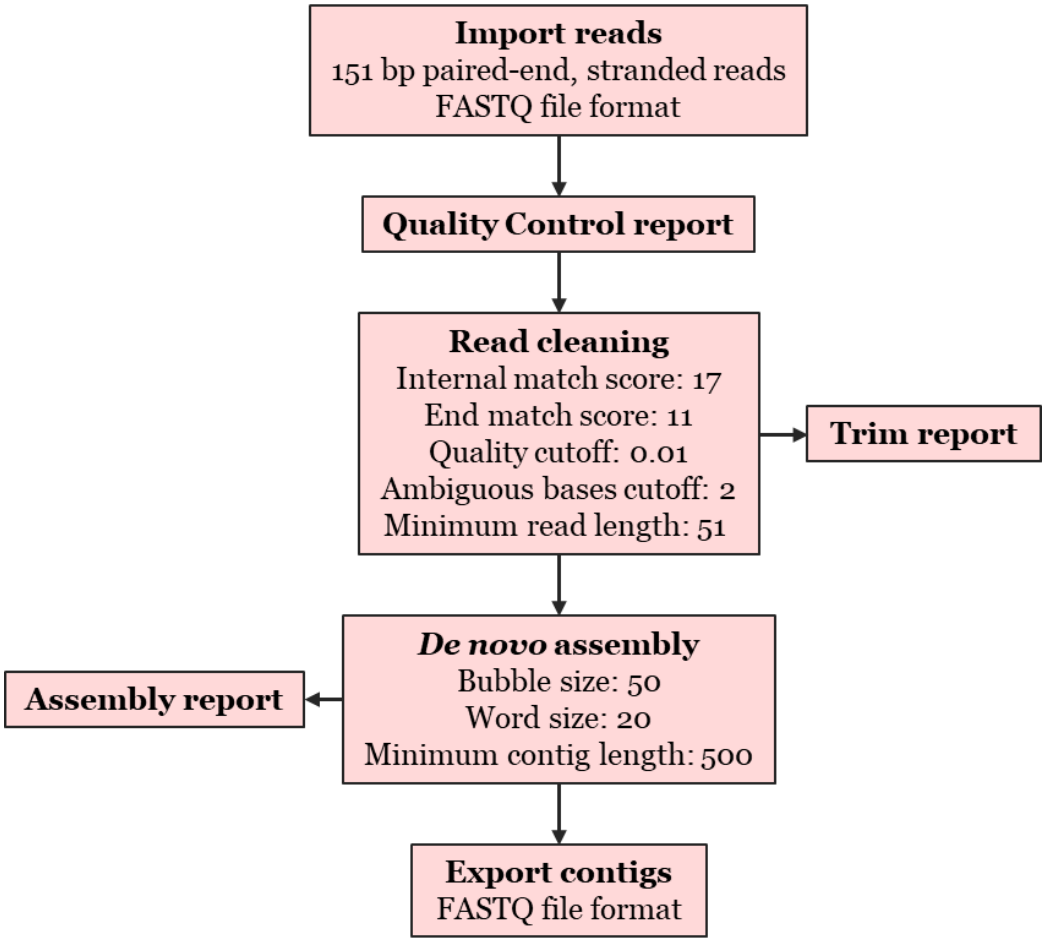


Figure 2.9. Optimized trim and assembly pipeline for use with the CLC Genomics Workbench, based on the results of this study. Optimized parameters embellish this bioinformatic pipeline, offering specifications that can be applied to the trimming and *de novo* assembly of 151 bp paired-end, stranded reads that were prepared using the Illumina TruSeq Indexed Adapters and generated using the Illumina NextSeq 500 platform.

2.4.3 Comparison of two assemblers

Choosing a suitable assembler and applying optimum parameters is critical to achieving the best possible assembly performance. This chapter exemplifies the

importance of assembly optimization, culminating in the comparison of the commercial the CLC Genomics Workbench software (*www.qiagenbioinformatics.com*) and the open-source Trans-ABySS (Robertson *et al.* 2010). Various assembly statistics and the quality of downstream processes can inform the choice of assembler. Assembly size can be described by the maximum contig length, average contig length, total number of contigs and the N50 length (Miller *et al.* 2010). When using either multiple k values or a single k value, Trans-ABySS (Robertson *et al.* 2010) had a larger maximum contig length than the CLC Genomics Workbench (*www.qiagenbioinformatics.com*). This may indicate that Trans-ABySS (Robertson *et al.* 2010), when compared with the CLC Genomics Workbench (*www.qiagenbioinformatics.com*), more accurately constructed contiguous sequences. The abyss-fac tool does not produce an average contig length value, so this statistic was not included in the assembler comparison. The number of contigs produced by Trans-ABySS (Robertson *et al.* 2010) was much higher than the amount produced by the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) regardless of the number of k values or the values themselves. Though it is often advantageous to reduce the number of contigs in an assembly, a larger number of contigs when building a transcriptome *de novo* can indicate that less expressed transcripts were effectively assembled (Surget-Groba and Montoya-Burgos 2010).

Because it indicates how much of the transcriptome is covered by relatively large contigs, the N50 length is often used as a measure of assembly quality (Schatz

et al. 2010). Taking into account the length and number of contigs for assembly evaluation, it is a measure of de Bruijn graph contiguity (Molina-Mora *et al.* 2020). So long as assemblers use similar graph strategies, N50 length can be used to not only compare assemblies performed by one *de novo* assembler, but also between assemblers (Earl *et al.* 2011). Even though the assembly techniques used by the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) are proprietary and not fully known to the user, both the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) and Trans-ABySS (Robertson *et al.* 2010) use an application of de Bruijn graphs, which facilitated their comparison (Earl *et al.* 2011). Trans-ABySS (Robertson *et al.* 2010), when optimized, produced a higher N50 value than the optimized CLC Genomics Workbench (*www.qiagenbioinformatics.com*) assembly. When the same single k value was applied to both assemblers, the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) produced a higher N50 value. Partially supporting our hypothesis, this indicates that when a single k value was employed, the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) outperformed Trans-ABySS (Robertson *et al.* 2010), but that it fell short when the Trans-ABySS (Robertson *et al.* 2010) multiple k-mer procedure was applied.

The mapped back percentage of trimmed reads can also be used when comparing assemblers. It is indicative not only of the quality of trimmed reads (Del Fabbro *et al.* 2013), but also of post *de novo* assembly contig or scaffold quality (Earl *et al.* 2011). If appropriately stringent mapping parameters are employed,

high quality reads will not align with misassembled contigs, and inaccurate reads will not align with correctly constructed contigs (*http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/755/index.php?manual=Mapping_parameters.html*). As indicated by the greater percentage of reads aligned, Trans-ABySS (Robertson *et al.* 2010) produced higher quality contigs when compared with the CLC Genomics Workbench (*www.qiagenbioinformatics.com*). Furthermore, the k=33 CLC Genomics Workbench (*www.qiagenbioinformatics.com*) assembly produced higher quality contigs compared with the k=20 assembly. The 1.25% difference indicates that the percentage of reads mapped back is an important statistic that should be used when optimizing assembly parameters. It is possible that a different optimized k value would have been chosen had the mapped back percentage been included in the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) optimization analyses.

Run time, ease of use and memory requirements are of great concern and should be considered when comparing assemblers. Trans-ABySS (Robertson *et al.* 2010) utilizes a combination of graph algorithms that add to run time, and the transabyss-merge tool must produce individual assemblies for each k-mer designated, store the assembly files and then merge the assemblies, which further increases the amount of run time required (Robertson *et al.* 2010). CLC Genomic Workbench (*www.qiagenbioinformatics.com*), on the other hand, utilizes a proprietary algorithm that prides itself on its assembly speed (CLC Bio 2013). It

has an extremely well curated interactive online manual (*http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/755/index.php?manual=Introduction_CLC_Genomics_Workbench.html*), and contains a multitude of pre-processing tools, such as QC and read trimming, that can be included in the assembly workflow. On the other hand, Trans-ABySS (Robertson *et al.* 2010) does not have pre-processing functionalities, and the user must employ other tools to QC and trim reads prior to assembly. Furthermore, Trans-ABySS (Robertson *et al.* 2010) is entirely command line implemented, and the user must be familiar with basic commands to assemble trimmed reads. By comparison, the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) has a well-developed graphical user interface that, when used in conjunction with the manual, is user friendly. Finally, both the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) and Trans-ABySS (Robertson *et al.* 2010) can be run on remote servers that have greater access to computing power and memory allocation. However, Trans-ABySS (Robertson *et al.* 2010) requires more computational resources to construct a *de novo* assembly, especially when implementing the merge capability. Therefore, there are many considerations when producing a comprehensive *de novo* assembly. To achieve research goals, comparing different tools after parameter optimization is paramount.

2.5 Conclusion

De novo assembly is the method of choice for constructing transcriptomes for non-model species like conifers (Cahais *et al.* 2012). Lodgepole pine possesses a relatively large and complex transcriptome (De La Torre *et al.* 2014; Suren *et al.* 2016), and it was unclear which platform would create the more accurate and comprehensive assembly. Therefore, the objective of this study was to optimize and compare lodgepole pine transcriptome assemblies made using two state-of-the-art bioinformatic tools, the CLC Genomics Workbench v9.5.2 (*www.qiagenbioinformatics.com*) and Trans-ABySS v1.5.5 (Robertson *et al.* 2010). This study also showcased how to optimize read trimming and assembly parameters. We hypothesized that the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) would produce a faster assembly, but that Trans-ABySS (Robertson *et al.* 2010) would produce a more contiguous and accurate assembly. Our results supported our hypothesis and emphasized that there is a tradeoff between accuracy and time allocation that researchers must consider when choosing between *de novo* assemblers for the analysis of short reads. Even though the CLC Genomics Workbench (*www.qiagenbioinformatics.com*) was more user friendly, Trans-ABySS (Robertson *et al.* 2010) produced a higher quality assembly. Following these findings, Trans-ABySS v1.5.5 (Robertson *et al.* 2010) was chosen for the creation of a master reference transcriptome to be used for differential expression analysis (Chapter 4).

2.6 References

Agarwal, S., Macfarlan, T. S., Sartor, M. A., Iwase, S. (2015). Sequencing of first-strand cDNA library reveals full-length transcriptomes. *Nature Communications*, 6, 1-12.

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

Bentley, D. R., Balasubramanian, S., Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456, 53-59.

Bolger, A. M., Lohse, M., Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114-2120.

Brady, S. M., Long, T. A., Benfey, P. N. (2006). Unraveling the dynamic transcriptome. *The Plant Cell Online*, 18, 2101-2111.

Cahais, V., Gayral, P., Tsagkogeorga, G., Melo-Ferreira, J., Ballenghien, M., Weinert, L., Chiari, Y., Belkhir, K., Ranwez, V., Galtier, N. (2012). Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Molecular Ecology Resources*, 12, 834-845.

Chaisson, M. J., Wilson, R. K., Eichler, E. E. (2015). Genetic variation and the *de novo* assembly of human genomes. *Nature Reviews Genetics*, 16, 627-640.

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.

CLC Bio (2013). *De novo assembly with CLC Assembly Cell 4.0: White paper executive summary*. Author, Cambridge, Massachusetts, United States.

Cock, P. J., Fields, C. J., Goto, N., Heuer, M. L., Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, 38, 1767-1771.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. J., Gaffney, D., Elo, L. L., Zhang, X., Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology,* 17, 1-19.

Corley, S. M., Mackenzie, K. L., Beverdam, A., Roddam, L. F., Wilkins, M. R. (2017). Differentially expressed genes from RNA-Seq and functional enrichment results are affected by the choice of single-end versus paired-end reads and stranded versus non-stranded protocols. *BMC Genomics*, 18, 1-13.
Cornish-Bowden, A. (1985). Nomenclature for incompletely specified bases in nucleic-acid sequences - Recommendations 1984. *Nucleic Acids Research,* 13, 3021-3030.

De La Torre, A. R., Birol, I., Bousquet, J., Ingvarsson, P. K., Jansson, S., Jones, S. J., Keeling, C. I., MacKay, J., Nilsson, O., Ritland, K., Street, N., Yanchuk, A., Zerbe, P., Bohlmann, J. (2014). Insights into conifer giga-genomes. *Plant Physiology,* 166, 1724-1732.

Del Fabbro, C., Scalabrin, S., Morgante, M., Giorgi, F. M. (2013). An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PLoS ONE*, 8, 1-13.

Didion, J. P., Martin, M., Collins, F. S. (2017). Atropos: Specific, sensitive, and speedy trimming of sequencing reads. *PeerJ*, 5, 2-19.

Earl, D., Bradnam, K., John, J. S., Darling, A., Lin, D., Fass, J., Paten, B. (2011). Assemblathon 1: A competitive assessment of *de novo* short read assembly methods. *Genome Research*, 21, 2224-2241.

Ewing, B., Green, P. (1998). Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Research*, 8, 186-194.

Ewing, B., Hillier, L., Wendl, M. C., Green, P. (1998). Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Research*, 8, 175-185.

Fasulo, D., Halpern, A., Dew, I., Mobarry, C. (2002). Efficiently detecting polymorphisms during the fragment assembly process. *Bioinformatics*, 18, S294–S302.

Herbert, Z. T., Kershner, J. P., Butty, V. L., Thimmapuram, J., Choudhari, S., Alekseyev, Y. O., Fan, J., Podnar, J. W., Wilcox, E., Gipson, J., Gillaspy, A., Jepsen, K., Splinter, S., Durant, B., Morris, K., Berkeley, M, LeClerc, A., Simpson, S. D., Sommerville, G., Grimmett, L., Adams, M., Levine, S. S. (2017). Cross-site comparison of ribosomal depletion kits for Illumina RNAseq library construction. *BMC Genomics*, 19, 2-10.

Hocking, D. (1971). Preparation and use of a nutrient solution for culturing seedlings of lodgepole pine and white spruce, with selected bibliography. *Northern Forest Research Centre Information Report Nor-X-1*. Canadian Forest Service, Department of the Environment, Edmonton, Alberta, Canada.

Huse, S. M., Huber, J. A., Morrison, H. G., Sogin, M. L., Welch, D. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*, 8, R143.1-R143.9.

Illumina, Inc. (2013). *Illumina TruSeq stranded mRNA low sample (LS) protocol*. Author, San Diego, California, United States.

Illumina, Inc. (2015). *NextSeq system denature and dilute libraries protocol*. Author, San Diego, California, United States.

Illumina, Inc. (2016). *Illumina TruSeq indexed adapters*. Author, San Diego, California, United States.

Illumina, Inc. (2018). *NextSeq 500 system guide*. Author, San Diego, California, United States.

Jung, H., Yoon, B. H., Kim, W. J., Kim, D. W., Hurwood, D. A., Lyons, R. E., Salin, K. R., Kim, H.-S., Baek, I., Chand, V., Mather, P. B. (2016). Optimizing hybrid *de novo* transcriptome assembly and extending genomic resources for giant

freshwater prawns (*Macrobrachium rosenbergii*): The identification of genes and markers associated with reproduction. *International Journal of Molecular Sciences,* 17, 1-21.

Lamarre, S., Frasse, P., Zouine, M., Labourdette, D., Sainderichin, E., Hu, G., Berre-Anton, V. L., Bouzayen, M., Maza, E. (2018). Optimization of an RNA-Seq differential gene expression analysis depending on biological replicate number and library size. *Frontiers in Plant Science,* 9*,* 1-18.

Lehri, B., Seddon, A. M., Karlyshev, A. V. (2017). The hidden perils of read mapping as a quality assessment tool in genome sequencing. *Scientific Reports*, 7, 1-8.

Liu, Y. W., Zhou, J., White, K. P. (2014). RNA-seq differential expression studies: More sequence or more replication? *Bioinformatics,* 30, 301-304.

Lourenço, A., Rencoret, J., Chemetova, C., Gominho, J., Gutiérrez, A., Río, J. C., Pereira, H. (2016). Lignin composition and structure differs between xylem, phloem and phellem in *Quercus suber* L. *Frontiers in Plant Science*, 7, 1-14.

MacManes, M. D. (2014). On the optimal trimming of high-throughput mRNA sequence data. *Frontiers in Genetics*, 5, 1-7.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17, 1-10.

Martin, J. A., Wang, Z. (2011). Next-generation transcriptome assembly. *Nature Reviews Genetics*, 12, 671-682.

Miedes, E., Vanholme, R., Boerjan, W., Molina, A. (2014). The role of the secondary cell wall in plant resistance to pathogens. *Frontiers in Plant Science*, 5, 1-13.

Miller, J. R., Koren, S., Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics*, 95, 315-327.

Molina-Mora, J. A., Campos-Sánchez, R., Rodríguez, C., Shi, L., García, F. (2020). High quality 3C *de novo* assembly and annotation of a multidrug resistant ST-111 Pseudomonas aeruginosa genome: Benchmark of hybrid and non-hybrid assemblers. *Scientific Reports*, 10, 1-16.

Nolan, T., Hands, R. E., Bustin, S. A. (2006). Quantification of mRNA using real-time RT-PCR. *Nature Protocols*, 1, 1559-1582.

Opitz, L., Salinas-Riester, G., Grade, M., Jung, K., Jo, P., Emons, G., Ghadimi, B. M., Beißbarth, T., Gaedcke, J. (2010). Impact of RNA degradation on gene expression profiling. *BMC Medical Genomics, 3*, 1-14.

Orenstein, Y., Shamir, R. (2013). Design of shortest double-stranded DNA sequences covering all k-mers with applications to protein-binding microarrays and synthetic enhancers. *Bioinformatics*, 29, 71-79.

Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H., Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Research,* 37, 1-7.

Pavy, N., Boyle, B., Nelson, C., Paule, C., Giguère, I., Caron, S., Parsons, L. S., Dallaire, N., Bedon, F., Bérubé, H., Cooke, J., Mackay, J. (2008). Identification of conserved core xylem gene stets: Conifer cDNA microarray development, transcript profiling and computational analyses. *New Phytology*, 180, 766–786.

Pearson, W. R., Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences*, 85, 2444-2448.

Pevzner, P. A., Tang, H. X., Waterman, M. S. (2001). An Eulerian path approach to DNA fragment assembly. *Proceedings of the National Academy of Sciences of the United States of America,* 98, 9748-9753.

Qiagen, CLCBio. *www.qiagenbioinformatics.com*. (Last accessed March 16, 2020).

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., Mungall, K., Lee, S., Okada, H., Qian, J., Griffith, M., Raymond, A., Thiessen, N., Cezard, T., Butterfield, Y., Newsome, R., Chan, S., She, R., Varhol, R., Birol, I., Birol, I. (2010). *De novo* assembly and analysis of RNA-seq data. *Nature Methods*, 7, U909-U962.

Roe A. D., Rice A. V., Bromilow S. E., Cooke J. E. K., Sperling F. A. H. (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.

Roe A. D., Rice A. V., Coltman D. W., Cooke J. E. K., Sperling F. A. H. (2011). Comparative phylogeography, genetic differentiation, and contrasting reproductive modes in three fungal symbionts of a multipartite bark beetle symbiosis. *Molecular Ecology*, 20, 584–600.

Sanger, F., Nicklen, S., Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74, 5463-5467.

Schatz, M. C., Delcher, A. L., Salzberg, S. L. (2010). Assembly of large genomes using second-generation sequencing. *Genome Research*, 20, 1165-1173.

Shendure, J., Ji, H. L. (2008). Next-generation DNA sequencing. *Nature Biotechnology,* 26, 1135-1145.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 3, 3210–3212.

SPEX SamplePrep. (2013). Geno/Grinder 2010 Operating Manual. Author, Metuchen, New Jersey, USA.

Sturm, M., Schroeder, C., Bauer, P. (2016). SeqPurge: Highly-sensitive adapter trimming for paired-end NGS data. *BMC Bioinformatics*, 17, 1-7.

Suren, H., Hodgins, K. A., Yeaman, S., Nurkowski, K. A., Smets, P., Rieseberg, L. H., Aitken, S. N., Holliday, J. A. (2016). Exome capture from the spruce and pine giga-genomes. *Molecular Ecology Resources*, 16, 1136-1146.

Surget-Groba, Y., Montoya-Burgos, J. I. (2010). Optimization of *de novo* transcriptome assembly from next-generation sequencing data. *Genome Research*, 20, 1432-1440.

Wang, Z., Gerstein, M., Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics,* 10, 57-63.

Williams, C. R., Baccarella, A., Parrish, J. Z., Kim, C. C. (2016). Trimming of sequence reads alters RNASeq gene expression estimates. *BMC Bioinformatics*, 17, 1-13.

Zerbino, D. R., Birney, E. (2008). Velvet: Algorithms for d*e novo* short read assembly using de Bruijn graphs. *Genome Research*, 18, 821-829.

Zhang, W. Y., Chen, J. J., Yang, Y., Tang, Y. F., Shang, J., Shen, B. R. (2011). A practical comparison of *de novo* genome assembly software tools for next-generation sequencing technologies. *PLoS ONE,* 6, 1-12.

Zhao, S. R., Zhang, Y., Gordon, W., Quan, J., Xi, H. L., Du, S., von Schack, D., Zhang, B. (2015). Comparison of stranded and non-stranded RNA-seq transcriptome profiling and investigation of gene overlap. *BMC Genomics,* 16, 1-14.

3.0 Chapter 3: Lesion development in *Grosmannia clavigera*-inoculated lodgepole pine responds to varying levels of nitrogen availability

3.1 Introduction

The mountain pine beetle (MPB; *Dendroctonus ponderosae* Hopkins) is a bark beetle indigenous to western North America. Lodgepole pine (*Pinus contorta* Douglas ex Loudon *var. latifolia*) is one of the main hosts of MPB, sharing an evolutionary history with the pest (Raffa and Berryman 1987). In western Canada, adult beetles disperse in mid- to late summer, with the female beetles being the first to penetrate the lodgepole pine bark (Natural Resources Canada 2017a). The pioneer females use aggregation pheromones to call more beetles to the attacked trees (Bentz *et al.* 2005; Chiu *et al.* 2019). While chewing vertical egg galleries in the inner bark of the tree, the MPB inoculates the tree with a fungal complex that includes pathogenic *Ophiostomatoid* fungi, such as *Grosmannia clavigera* (Robinson-Jeffrey and Davidson) (Roe *et al.* 2010). The microbial community hosted by MPB facilitates the beetle in overcoming tree defenses and successfully colonizing the tree (Guérard *et al.* 2000; DiGuistini *et al.* 2011). *G. clavigera* detoxifies tree resin meant to prevent the MPB from borrowing further into the bark (DiGuistini *et al.* 2011; Wang *et al.* 2012). In lodgepole pines, research has shown that *G. clavigera* concentrate phloem nitrogen (N) in their hyphae and conidia, increasing the availability of N for colonizing MPB adults and their larvae

(Ayres *et al.* 2000; Cook *et al.* 2010; Goodsman *et al.* 2012). *G. clavigera* hyphae grow into host sapwood causing phloem polyphenolic parenchyma cells to swell and accumulate toxic phenolics meant to inhibit fungal growth (Arango-Velez *et al.* 2014). *G. clavigera* mycelium also spread into host xylem tissue causing the reactive production of ray and axial parenchyma cell ingrowths called tyloses, meant to prevent the axial spread of the fungus (Clérivet and El Modafar 1994; Clérivet *et al.* 2000). These ultimately disrupt water transport resulting in host mortality (Ballard *et al.* 1982; Hubbard *et al.* 2013; Arango-Velez *et al.* 2016; Morris *et al.* 2016).

As *G. clavigera* grows into the infected lodgepole pine sapwood, it causes the host tree to produce lesions around points of inoculation (Rice *et al.* 2007a; Arango-Velez *et al.* 2016). Relative to control treatments, Arango-Velez *et al.* (2016) found that levels of the defense-signaling hormone jasmonic acid (JA) increased inside and around lesions of *G. clavigera*-inoculated lodgepole and jack pine indicating that *G. clavigera* is a necrotrophic rather than a biotrophic pathogen (Glazebrook 2005). Particularly in mature trees, these lesions can become necrotic (Lusebrink *et al.* 2013). Lesions grow over time in length, and their size has been correlated with the extent of fungal invasion into the lodgepole pine stem tissues (Rice and Langor 2008). Lesion development is a function of pine defense against the fungus, an effort by the tree to contain and wall off the fungal infection, rather than a function of colonization (Wong and Berryman 1977; Raffa and Smalley 1988; McAllister *et al.* 2018). In this context, lesion size

indicates the strength of the defense response to the fungus as the pathogen spreads both vertically and axially through the lodgepole pine vasculature. *G. clavigera*-induced lesions contain cells full of carbon (C)-based chemical defense compounds, such as flavonoids and terpenoids (Franceschi *et al.* 2005; Keeling and Bohlman 2006; Kovalchuk *et al.* 2013). Acting as a barrier to fungal spread, phenolic compounds, such as lignin, can serve as a physical defense (Zabel and Morell 1992; Franceschi *et al.* 2005; Kovalchuk *et al.* 2013). In addition, cells within lesions contain N-based defense-related proteins, such as chitinases that hydrolyze fungal cell walls and peroxidases that enhance host cell wall toughness (Fossdal *et al.* 2001; Kovalchuk *et al.* 2013; Kolosova *et al.* 2014). Using reverse transcription quantitative polymerase chain reaction (RT-qPCR; Nolan *et al.* 2006) for the gene expression profiling of defense-related genes, Arango-Velez *et al.* (2014) provided evidence that both C- and N-based defenses in mature lodgepole pine × jack pine hybrids are induced by *G. clavigera* inoculation, and that the magnitude of induction for these defense genes is lessened by water deficit.

In lodgepole pine and other annual and perennial plant species, N availability is an additional factor that is known to affect defense responses (Hoffland *et al.* 1999; Hoffland *et al.* 2000; Leser and Treutter 2005; Cook *et al.* 2015). Plants appear to balance nutrient allocation to growth and reproduction processes with allocation to defense and stress acclimation responses (Matyssek *et al.* 2002; Matyssek *et al.* 2005). Hoffland *et al.* (1999) found that increasing N availability for tomato plants (*Solanum lycopersicum* Linnaeus) grown in nutrient

solutions had decreased susceptibility to the fungal pathogen *Botrytis cinerea* (Persoon). Follow-up experiments showed that increasing N availability for tomato plants grown in nutrient solutions significantly increased their susceptibility to *Pseudomonas syringae* (van Hall) and *Oidium lycopersicum* (Cooke and Massee) (Hoffland *et al.* 2000). When compared with those grown with lower N supply, Lesser and Treutter (2005) found that 'Golden Delicious' apple trees (*Malus domestica* Barkhausen) that were grown in greenhouse conditions responded to higher soil N supply with a reduced accumulation of leaf scab caused by the pathogen *Venturia inaequalis* (Cooke). In addition, they found that trees that received higher compared to lower concentrations of N fertilizer had increased shoot growth (Leser and Treutter 2005). In poplar trees (*Populus cathayana* Rehder), alterations in soil N has been shown to affect plant growth, increasing root and decreasing stem and leaf biomass in response to 0.25 mM compared with 5 mM ammonium nitrate (NH_4NO_3; Luo *et al.* 2019). When compared with plants grown in higher nitrate conditions (10 mM and 35 mM NO_3^-), Marín *et al.* (2011) demonstrated that *Arabidopsis thaliana* (Linnaeus) grown in nitrate-limiting conditions (1 mM NO_3^-) had accelerated flowering.

Lodgepole pine typically grow in the Pacific Maritime and Montane Cordillera ecoregions of western Canada (Natural Resources Canada 2017b), where soil has been shown to be N deficient (Brockley 2001). Furthermore, N has been shown to be the nutrient most limiting to the growth of northern temperate tree species (Brockley 2001; Vadeboncoeur 2010; Högberg *et al.* 2013). For 6 years

after application, lodgepole pine in the forests of interior British Columbia (BC) showed increased height and diameter in response to N fertilization when compared with non-fertilized trees (Prescott *et al.* 2019). Mature lodgepole pine growing in the BC interior that were treated with yearly N soil enrichment had increased mean basal area (Amponsah *et al.* 2004). When compared with non-fertilized trees, mature lodgepole pine fertilized with urea, an N-based compound, experienced an increase in constitutive resin flow (Cook *et al.* 2015). Soil N augmentation may have enhanced the ability of lodgepole pine to successfully resist MPB infestation (Cook *et al.* 2015). Induced resin following MPB attack contains a higher concentration of secondary defense compounds and defense-related proteins compared with constitutive resin (Franceschi *et al.* 2005; Ott *et al.* 2011; Kovalchuk *et al.* 2013). Following the application of a comprehensive fertilizer, Klepzig *et al.* (2005) showed that mature loblolly pine (*Pinus taeda* Linnaeus) inoculated with *Ophiostoma minus* (Hedgecock) had a significant increase in induced resin flow compared with non-fertilized trees. However, they did not find a significant difference in lesion size (Klepzig *et al.* 2005). A greater understanding of the influence of varying levels of N application on the defense response of lodgepole pine seedlings to *G. clavigera* is warranted, providing motivation for this master's thesis project.

This portion of the thesis project had three objectives: (1) to determine whether soil application of different levels of NH_4NO_3 nutrient solution applied to lodgepole pine seedlings grown in controlled growth environment conditions

translated into differential accumulation of foliar N, (2) to assess the impact of varying N availability on lesion development in lodgepole pine seedlings inoculated with *G. clavigera*, and (3) to generate tissue that will be used in the future for (a) transcript profiling by RT-qPCR, (b) quantification of *G. clavigera* colonization by quantitative polymerase chain reaction (qPCR) of fungal deoxyribonucleic acid (DNA), and (c) histochemical analyses of host defense responses. We hypothesized that higher N availability would result in an increase in foliar N concentration in the lodgepole pine seedlings. In addition, we hypothesized that different levels of N fertilization, such as low (1 mM) or high (10 mM) concentrations of NH_4NO_3, would impact the lodgepole pine defense against *G. clavigera*, as measured by differences in lesion development.

3.2 Materials and methods

3.2.1 Plant material

To complement the experiment performed in 2010 (Chapter 2), an experiment was conducted during the summer of 2016 to test the responses of *P. contorta* to *G. clavigera* inoculation under low vs. high levels of N fertilization in growth chamber conditions. In June 2016, Mr. David Swindlehurst (Weyerhaeuser Co., Drayton Valley, Alberta, Canada) provided us with lodgepole pine trees in their third growth cycle. These lodgepole pine seedlings were from a provenance in the eastern slopes of Alberta, south of Hinton and northwest of Drayton Valley.

Therefore, we could be confident that they are genetically pure lodgepole pine without introgression from jack pine (*Pinus banksiana* Lambert) (Cullingham *et al.* 2012). Upon delivery of trees in 3.78L pots, they were initially fertilized to field capacity with 0.5 g/L 20-20-20 (N:P:K) fertilizer (Plant Products Ltd, Brampton, Ontario, Canada). The plants were grown in a complete randomized block design in controlled environment growth rooms, 15 h day / 9 h night photoperiod, and under fluorescent lights at approximately 200 µmol photosynthetically active radiation (PAR) light intensity. This light level is lower than what is typical in greenhouses and outdoors indicating that our seedlings were likely light-limited. All seedlings were watered with approximately 400 mL of deionized water every two to three days leading up to the fertilization treatments described in Section 3.2.2.1.

3.2.2 Experimental design

The multifactorial experimental design comprised of two fertilization treatments, three inoculation treatments and four time points. The 1 mM and 10 mM NH_4NO_3 treatments are referred to as low N and high N, respectively. Plants were randomly assigned to one of three inoculation treatments: control (no wound or inoculation), wound plus inoculation with water (mock-inoculated), and wound plus *G. clavigera* inoculation (fungal-inoculated). Seedlings were grown in one of two growth chambers, each arranged in a randomized complete block design. The first growth chamber contained trees used for (a) lesion development

measurements, and (b) generation of stem tissues for RT-qPCR transcript abundance analyses. Eight biological replicates were used for each N, inoculation and day treatment combination in the first growth chamber. The second growth chamber contained trees used to generate (a) foliar tissue for N concentration analyses and (b) stem tissues for microscopy and qPCR quantification of *G. clavigera* colonization. In the second growth chamber, four biological replicates were used for each N, inoculation and day treatment combination. Trees were destructively harvested at 1, 7, 15 and 28 days post inoculation (dpi). The experiment yielded a total of 288 trees [(1 organism × 3 inoculation treatments × 2 N treatments × 4 time points × 8 biological replicates) + (1 organism × 3 inoculation treatments × 2 N treatments × 4 time points × 4 biological replicates) = 288].

3.2.2.1 N fertilization treatment

Three weeks following tree delivery, we replaced the twice-a-week application of deionized water with a modified fertilization regime of approximately 400 mL of Hocking's modified complete nutrient solution (Hocking 1971) containing either 1 mM or 10 mM NH_4NO_3. For the two weeks prior to inoculation treatments, the fertilization treatment was applied to the soil twice a week. Fertilization continued twice a week, every week until the conclusion of the experiment.

3.2.2.2 Wounding and *G. clavigera* inoculation

G. clavigera isolate M002-12-03-03-UC10G11 (Roe *et al.* 2010; Roe *et al.* 2011) was cultured on malt extract agar plates composed of 1.5% malt extract (Difco Laboratories, Detroit, Michigan, United States) and 1.5% agar (Bioshop, Inc., Burlington, Ontario, Canada) as described in Rice *et al.* (2007a). To suspend the spores, plates with actively growing mycelia were flooded with sterile Milli-Q water (Merck Millipore, Burlington, Massachusetts, United States). The spore suspensions (~150 spores μL^{-1}) were used as the inoculum. For the mock and *G. clavigera* inoculation treatments, four punctures of the bark along the main stem of the tree were made using an 18-gauge blunt end syringe needle. Punctures penetrated the bark until the cambial zone but did not enter the xylem (wood). Three of the punctures were made in the main stem of the previous seasons' growth, while a single puncture was made in the current season's growth. These punctures are referred to as wounds. For the wound plus fungal-inoculated treatments, inoculation was completed by adding 1 µL of *G. clavigera* spore suspension (~150 spores μL^{-1}) into the tiny cavity created by the puncture. For the wound plus mock-inoculated treatment, 1 µL of deionized water was added to the puncture cavity.

3.2.3 Foliar N concentration analysis

Foliar samples collected 1, 7, 15 and 28 dpi were cut up and collected in 1.5 ml microfuge tubes that were then placed with lids open inside a dehydration oven. Needles were left to dry for 24 hours at 60 °C after which the lids of the microfuge tubes were closed, and the needles were stored at room temperature. Between four and seven biological replicates from each treatment combination were selected for N concentration analysis. Needles were ground with the Mixer Mill 301 (Retsch GmbH, Haan, Germany) at a frequency of 1/s to obtain a fine powder by Ekaterina Stolnikova. Three mg of each sample were used for dry combustion at the Biogeochemical Analytical Service Laboratory at the University of Alberta (Edmonton, Alberta, Canada), where total C and N content were detected by thermal conductivity using the CE-440 Elemental Analyzer (Exeter Analytical, Inc., North Chelmsford, Massachusetts, United States) and the US EPA Test Method 440.0 protocol (Zimmerman *et al.* 1997). The positive control, which also served as the calibration standard, was 99.9% acetanilide. The negative control was no sample or standard. Foliar N concentration was calculated by dividing foliar N content by dry weight of the needles.

3.2.4 Lesion measurement

Ten cm segments of whole stems were harvested from the current year growth (i.e. the leader), such that the wound was at the midpoint of this 10 cm section (Figure 3.1). An equivalent piece of stem was harvested from control trees. After needle removal, a scalpel was used to produce a longitudinal slice through

the bark. Bark (with living tissues that are predominantly secondary phloem) was peeled and separated from the wood (secondary xylem) at the cambial zone. Lesions that formed along the inside of the secondary phloem and outside of the secondary xylem of the fungal-inoculated trees were measured along the longitudinal axis of the stem using a digital caliper. Measurements were taken at 1, 7, 15 and 28 dpi. Eight biological replicates were measured for each N and day treatment combination.

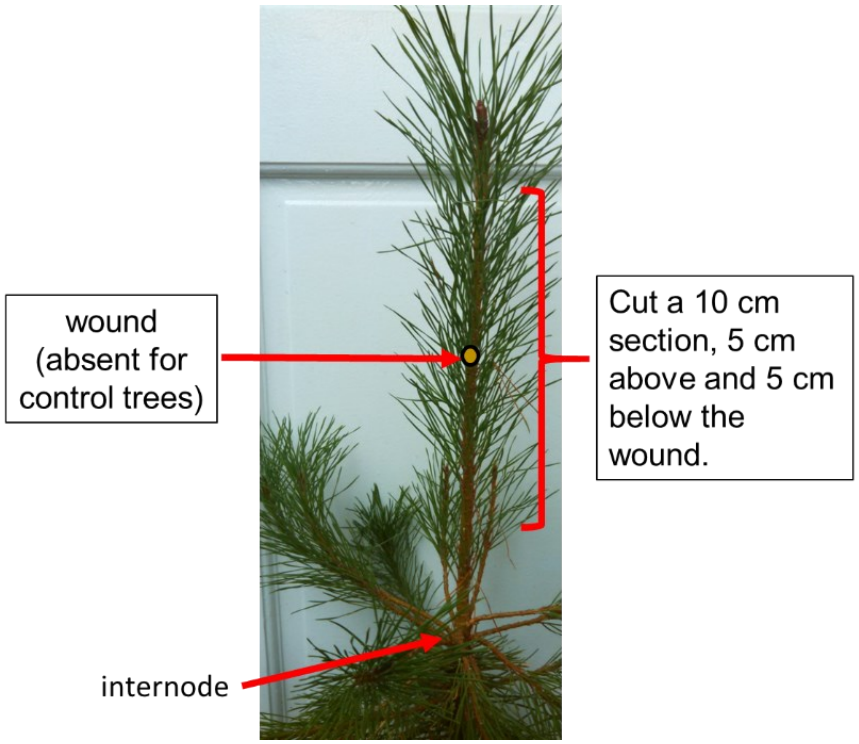


Figure 3.1. Stem harvest diagram. A 10 cm section of lodgepole pine stem from the current season growth was harvested 3 cm above the internode. Trees were assigned one of three inoculation treatments, no wound, wounded plus mock-inoculated and wounded plus inoculated with *G. clavigera*. Wounds were made 8 cm above the internode.

3.2.5 Tissue collection

Tissue samples for different analyses were collected at 1, 7, 15 and 28 dpi. These analyses are not included in this thesis, but a description of the collection of these tissues has been included for completeness.

Transcript profiling by RT-qPCR

Secondary phloem and secondary xylem tissue, whose collection was detailed in Section 3.2.4, were flash frozen separately in liquid N prior to storage at -80 °C.

G. clavigera quantification by qPCR

Five cm segments of stems were harvested from the current year growth with the wound at the midpoint of the segment. An equivalent segment of stem was harvested from control trees. This 5 cm segment of stem was further cut into 1 cm sections, such that one of the 1 cm sections flanked the wound, and two 1 cm sections were obtained on either side of the wound (Figure 3.2). In total, five 1 cm sections were obtained for each of the stems, with one 1 cm section containing the wound, two 1 cm sections immediately flanking this central section, and two additional 1 cm sections flanking these. Thus, these more distal sections were the furthest from the wound. Each of these 1 cm sections was cut transversely into two

0.5 cm pieces. One of the two 0.5 cm pieces from each of the five 1 cm sections was placed individually into 1.5 mL microfuge tubes and flash frozen in liquid N prior to storage at -80 °C.

Microscopy

From the stems harvested for qPCR fungal quantification, the second 0.5 cm piece from each of the five 1 cm sections was placed into individual vials of fresh fixative [2% (v/v) glutaraldehyde, 1% (w/v) caffeine buffered in 0.1 M sodium phosphate buffer, pH 7.2] (Figure 3.2). Samples, without vial lids, were placed in a vacuum at -8 mmHg for two days. Old fixative was discarded and fresh fixative [2% (v/v) glutaraldehyde, 1% (w/v) caffeine buffered in 0.1 M sodium phosphate buffer, pH 7.2] was added to the vials before returning the samples to the vacuum, without vial lids, at -8 mmHg for 16 hours. Fixative was discarded, and samples were washed with 0.1 M sodium phosphate buffer followed by immersion in 70% ethanol. Samples were shaken at room temperature for 16 hours, after which the 70% ethanol was removed from each vial and replaced with fresh 70% ethanol. Samples were then shaken at 4 °C for 64 hours. Samples were taken out of the vials and transferred to tissue cassettes and embedded in paraffin at the University of Alberta Microscopy Service Unit (Edmonton, Alberta, Canada) in anticipation of light microscopy.

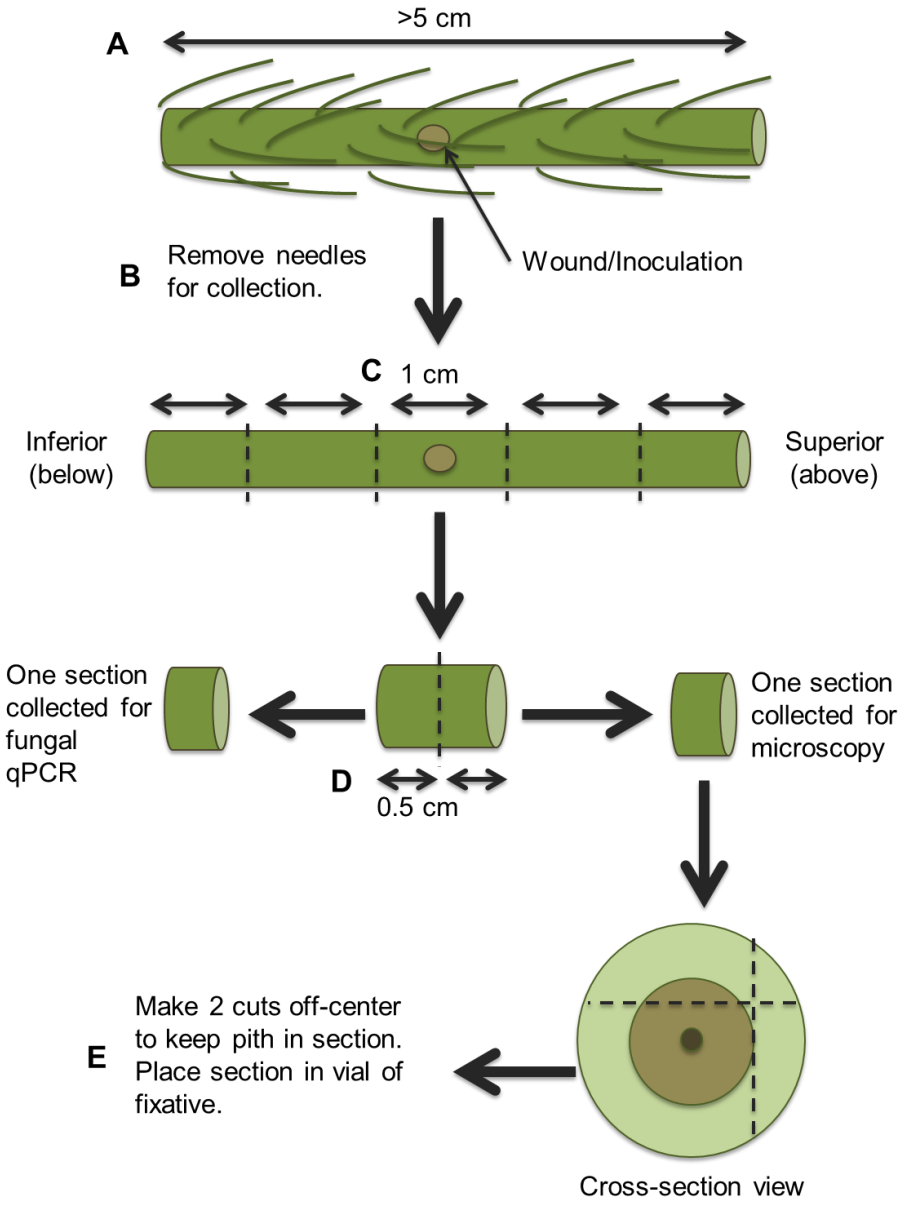


Figure 3.2. Harvest plan for microscopy and fungal qPCR. (**A**) A 5 cm section of lodgepole pine stem from the current season growth was harvested so that the wound was located in the middle. (**B**) Needles were collected using a single-edged blade for foliar N concentration analysis, described in Section 3.3.2. Care was taken to not rip the needles from the phloem. (**C**) Five 1 cm cross sections were cut using a double-edged blade. A fresh blade was used for each incision. (**D**) Each 1 cm cross section was cut into two 0.5 cm cross sections using a double-edged blade. A fresh blade was used for each incision. One section was flash frozen for qPCR. (**E**) The other 0.5 cm section was prepared for microscopy.

3.2.6 Statistical analysis

Foliar N concentration and lesion length data were analyzed with the R packages outliers v0.14 (Komsta 2011), car v3.0-7 (Fox and Weisberg 2019), multcompView v0.1-8 (Grave *et al.* 2019) and emmeans v1.4.5 (Lenth 2020) in R v3.6.1 (RStudio team 2015; R Core Team 2017) using similar methods. An outlier was removed from the foliar N data after verification using the Grubb's test for one outlier (Grubbs 1950). Both data sets were tested for normality using Shapiro-Wilk's test (Shapiro and Wilk 1965) and homogeneity of variance using Levene's test (Levene 1960). To meet the assumptions of normality and homogeneity of variance, Box-Cox transformation (Box and Cox 1964) was applied to the lesion length values. Statistical significance of the N concentration data was tested using a three-way analysis of variance (ANOVA; Fisher 1934) with $p<0.05$. Statistical significance of the lesion length data following the Box-Cox transformation (Box and Cox 1964) was tested using a two-way ANOVA (Fisher 1934) with $p<0.05$. Tukey's multiple comparison test (Tukey 1949) was leveraged to detect significant differences in means at an α value of 0.05 for both data sets.

3.3 Results

3.3.1 Tissue collection for future analyses

The experiment successfully produced 147 secondary phloem samples and 147 secondary xylem samples for RT-qPCR. Also collected were 280 stem samples

for DNA qPCR quantification of *G. clavigera* colonization of stem tissues paired with 280 stem cross-sections for microscopy. The stem samples were taken at defined distances from the point of wounding so that future analyses can examine (a) how N availability affects the extent of fungal colonization relative to lesion development, as analyzed by qPCR, and (b) how N availability alters the defense response along the length of the lesion, including effects on cambial activity and tissue development.

3.3.2 Foliar N concentration

Since proteins represent a major reservoir of N in plants, and foliage-localized ribulose 1,5-bisphosphate carboxylase/oxygenase accounts for a substantial fraction of the total plant protein content (Evans and Seemann 1989; Buchanan and Wolosiuk 2015), the concentration of N in leaves is often measured to determine the effect of N fertilization on a plant's N economy (Muñoz-Huerta *et al.* 2013). Accordingly, total N was measured in foliage as a means to determine whether low vs. high NH_4NO_3 nutrient solution applications were sufficient to significantly alter the N concentration of foliar tissue (Figure 3.3, Table 3.1). N availability levels significantly impacted foliar N concentration ($p<0.05$), and the dpi x N interaction term was also significant ($p<0.05$; Table 3.1). Foliar N concentration was significantly different at all time points measured in this experiment ($p<0.05$; Table 3.1). The N concentration in foliage did not

significantly differ between control, mock-inoculated or *G. clavigera*-inoculated samples at any of these time points (p>0.05; Table 3.1).

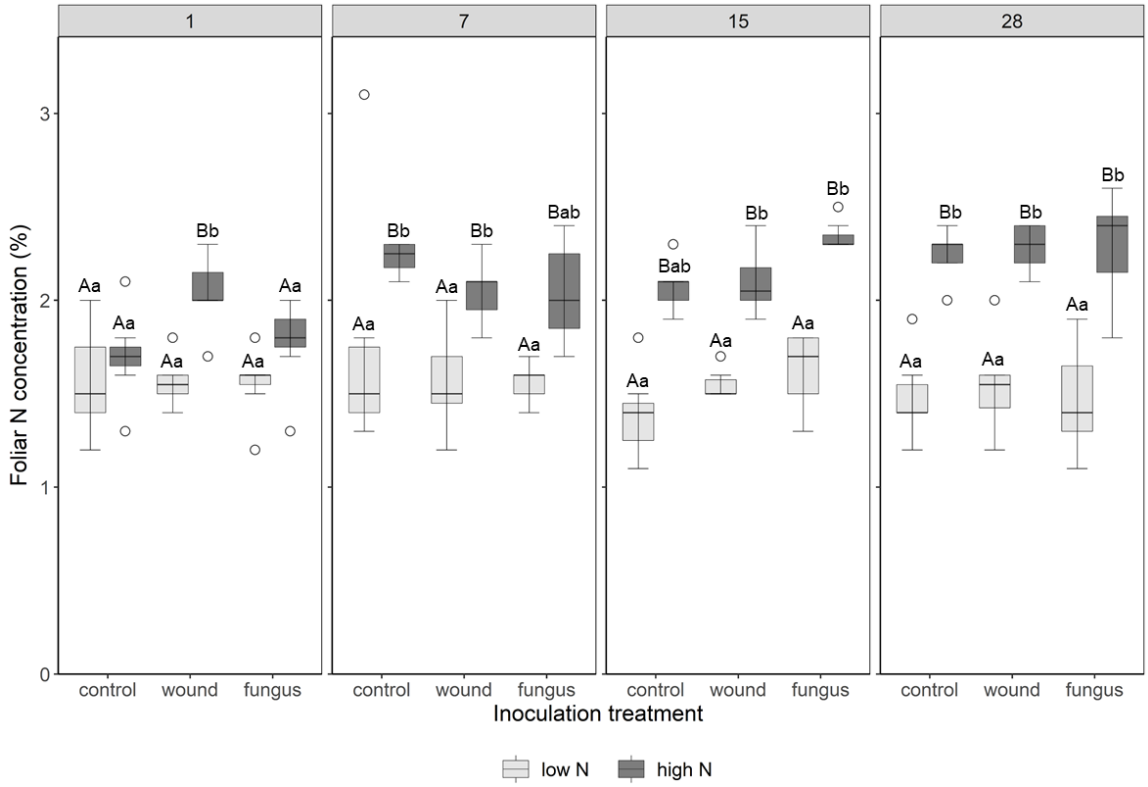


Figure 3.3. Lodgepole pine foliar N concentration. Lodgepole pine seedlings were treated with one of two levels of NH_4NO_3 fertilization, 1 mM (low N) or 10 mM (high N). Seedlings were also given one of three inoculation treatments, no wound (control), mechanically wounded and mock-inoculated (wound), and mechanically wounded plus inoculation with *G. clavigera* (fungus). The concentration of N (total N content/tissue dry weight) was determined for foliar tissue collected at 1, 7, 15 or 28 dpi, n=4-7. An outlier value from a low N control tree at 7 dpi was removed from the data set for statistical analyses. Within each inoculation type, pair-wise comparisons were performed between N treatments for each time point (upper case letters) and between time points for each N treatment (lower case letters), α=0.05. Different letters represent significantly different means.

Table 3.1. Three-way analysis of variance examining effects of time, N fertilization levels and inoculation treatment on foliar N concentration. Dpi refers to the length of time following inoculation that foliar samples were collected, while N refers to the level of NH_4NO_3 (low or high) used to fertilize the plants. Inoculation refers to one of three treatments, no wound, mechanically wounded and mock-inoculated, and mechanically wounded plus inoculation with *G. clavigera*. Degrees of freedom (df) for each factor are shown. Sum squared values (Sum Sq) express the total variation that can be attributed to each factor. The F value determines the ratio of explained variance to unexplained variance. The probability that the F value is significant is given as the P value, and the P value significance threshold is 0.05.

Factor	df	Sum Sq	F value	P value
dpi	3	7.840	5.926	8.290e-04
N	1	1.111	251.900	1.870e-31
inoculation	2	0.157	1.785	0.172
dpi x N	2	1.336	10.099	5.422e-06
dpi x inoculation	6	0.480	1.814	0.101
N x inoculation	2	3.523e-04	3.994e-03	0.996
dpi x N x inoculation	6	0.347	1.310	0.258
residuals	122	5.381		

3.3.3 Lesion length

Lesions in response to inoculation with *G. clavigera* developed in both xylem and phloem tissue of the lodgepole pine seedlings (Rice *et al.* 2007a), and lengths were measured at 1, 7, 15 and 28 dpi. To apply the power of parametric statistical testing, the lesion length data distribution was assessed for normality. First, the Shapiro-Wilk test (Shapiro and Wilk 1965) was applied. Based on the Shapiro-Wilk test results, raw lesion length data were not normally distributed for secondary phloem ($p<0.05$; Table 3.2). In contrast, raw lesion length data were normally distributed for secondary xylem ($p>0.05$; Table 3.2). A Box-Cox

transformation (α=0.05; Box and Cox 1964) was applied to both phloem and xylem data sets, and the resulting transformed data had a normal distribution for both tissue types (p>0.05; Table 3.2).

Table 3.2 Shapiro-Wilk normality test for lesion length data. Data were either untransformed or transformed using the Box-Cox method (α=0.05). The W value is the test statistic, and the probability that the W value is significant is given as the P value. The P value significance threshold is 0.05.

Lesion data	W value	P value
Phloem	0.90845	9.243e-4
Phloem (transformed)	0.98109	0.5989
Xylem	0.97737	0.6400
Xylem (transformed)	0.9615	0.2254

Along with the Shapiro-Wilk test (Shapiro and Wilk 1965), normality was also assessed visually with density plots (Figures 3.4 and 3.5) and Quartile-Quartile (Q-Q) plots (Figures 3.6 and 3.7). Density plots were used to assess for a classic bell-shaped curve or Gaussian distribution. Skewness and kurtosis measurements of asymmetry and "tailedness" yielded quantitative data to accompany the density plots. Values closer to zero indicated normally distributed data. Gaussian distribution Q-Q plots used linear regression analysis to assess normality. Plots for the untransformed data show right skew since the plotted points curve up away from the normal line, which indicates a long tail to the right (Figures 3.6 A and 3.7 A). This coincides with the right-skewness of the density plots (Figures 3.4 A and 3.5 A). Plotted points for the transformed data fall within

the 95% confidence zone, indicating that the transformed data are normal (Figures 3.6 B and 3.7 B). R-squared values gave quantitative data to accompany Q-Q plots, confirming normality of the Box-Cox transformed data (Figures 3.6 B and 3.7 B).

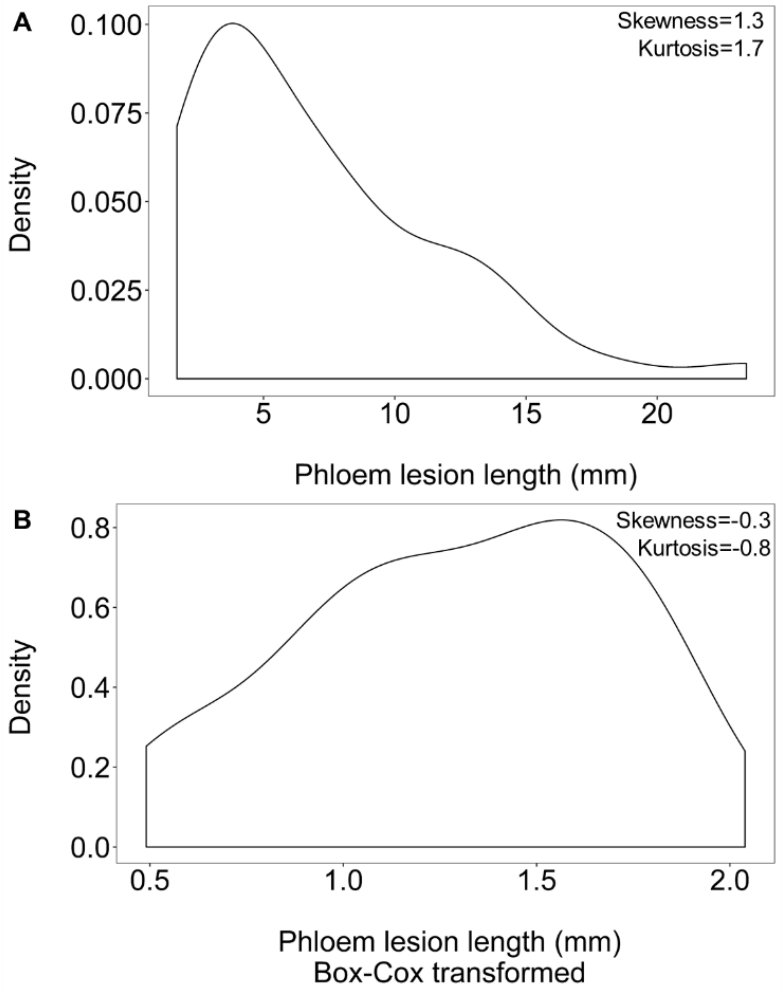


Figure 3.4. Lodgepole pine seedling secondary phloem lesion length density plots. Lesions developed in three-year-old lodgepole pine trees following inoculation with *G. clavigera*. The density of phloem lesion length measurements is plotted, and kurtosis and skewness calculations are listed in the top right corner. (**A**) Lodgepole pine phloem and (**B**) lodgepole pine phloem after data transformation using the Box-Cox method (α=0.05), *n*=6-7.

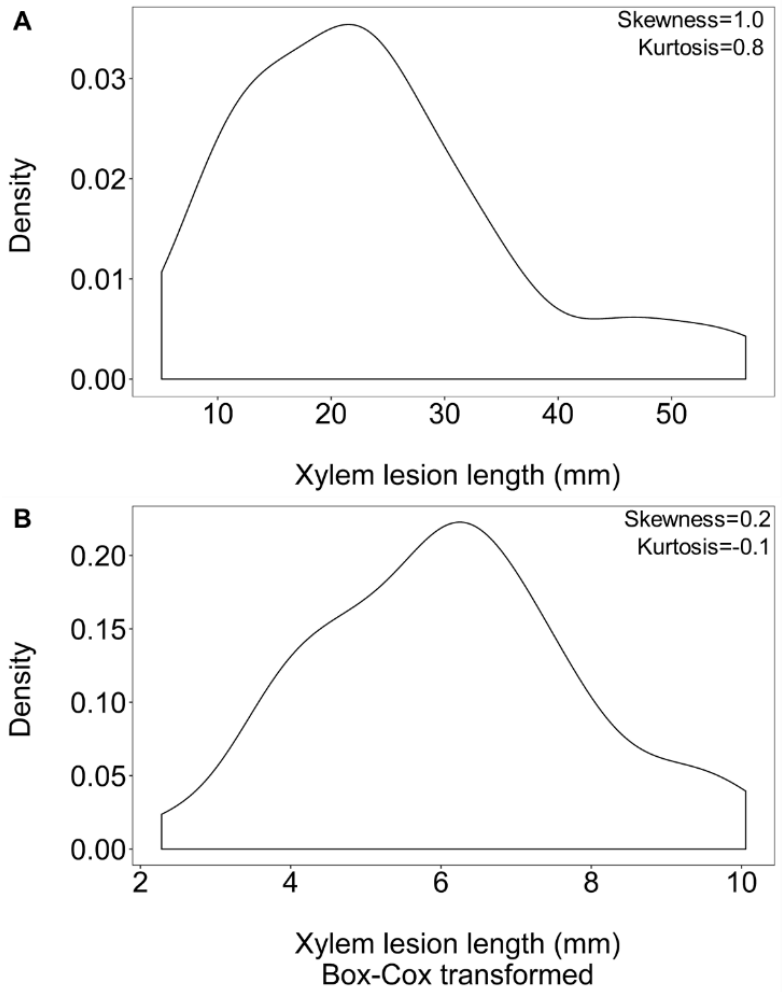


Figure 3.5. Lodgepole pine seedling secondary xylem lesion length density plots. Lesions developed in three-year-old lodgepole pine trees following inoculation with *G. clavigera*. The density of xylem lesion length measurements is plotted, and kurtosis and skewness calculations are listed in the top right corner. (**A**) Lodgepole pine xylem and (**B**) lodgepole pine xylem after data transformation using the Box-Cox method (α=0.05), *n*=6-7.

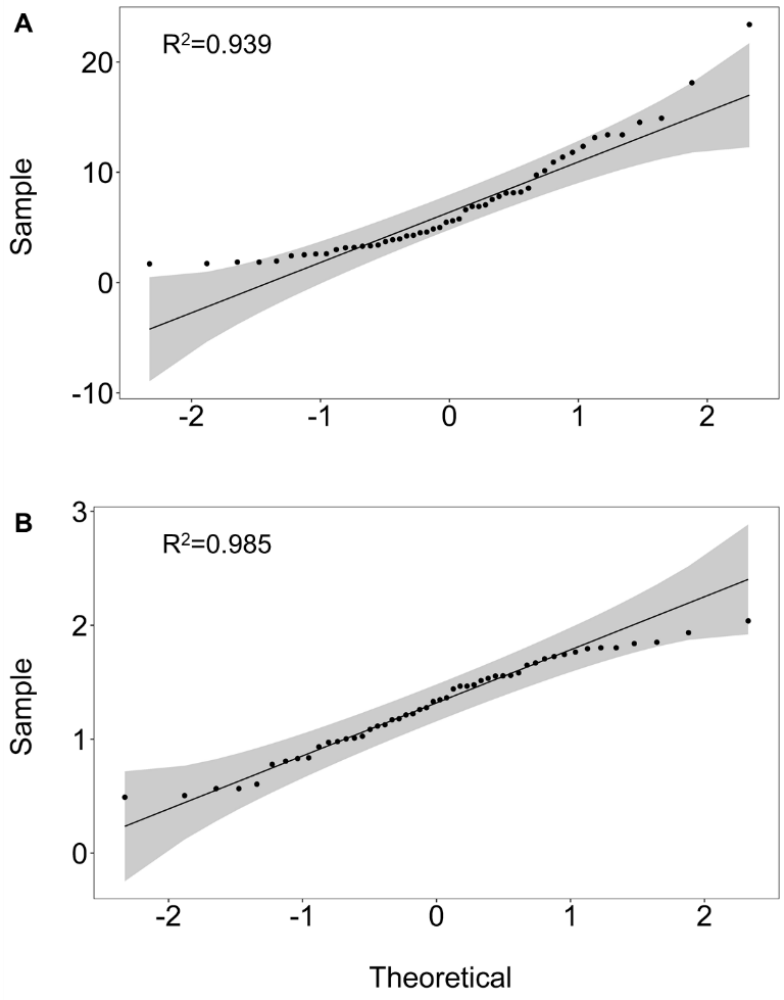


Figure 3.6. Lodgepole pine seedling secondary phloem lesion length Q-Q plots. Lesions developed in three-year-old lodgepole pine trees following inoculation with *G. clavigera*. A linear regression between theoretically normalized data and the actual phloem lesion length data are accompanied by the R-square correlation value in the top left corner. The data points in the grey region fall between the 95% upper and lower confidence bounds. (**A**) Lodgepole pine phloem and (**B**) lodgepole pine phloem after data transformation using the Box-Cox method (α=0.05), *n*=6-7.

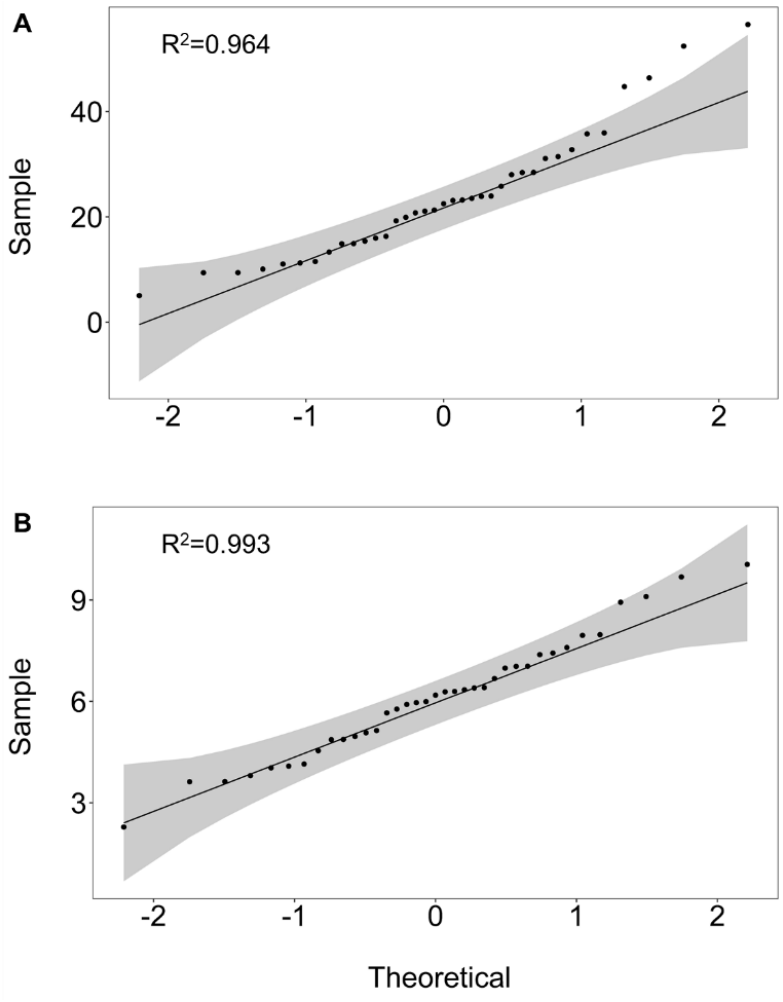


Figure 3.7. Lodgepole pine seedling secondary xylem lesion length Q-Q plots. Lesions developed in three-year-old lodgepole pine trees following inoculation with *G. clavigera*. A linear regression between theoretically normalized data and the actual xylem lesion length data are accompanied by the R-square correlation value in the top left corner. The data points in the grey region fall between the 95% upper and lower confidence bounds. (**A**) Lodgepole pine xylem and (**B**) lodgepole pine xylem after data transformation using the Box-Cox method (α=0.05), *n*=6-7.

The parametric two-way ANOVA (Fisher 1934) test requires data to be homogeneous as well as normally distributed. Levene's test (Levene 1960) was applied to test for homogeneity of variance (Table 3.3). In both tissue types, based on Levene's test results, the raw lesion data were not homogeneous ($p<0.05$; Table 3.3). Following the Box-Cox transformation ($\alpha=0.05$; Box and Cox 1964), the resulting transformed data were homogenous for variance for both phloem and xylem ($p>0.05$; Table 3.3). The normal, homogeneous, Box-Cox transformed lesion length data were then analyzed for variance consisting of a two-way comparison of the main effects: dpi with *G. clavigera* and N fertilization level (Table 3.4). The two-way ANOVA found that lesion length varied significantly with dpi for phloem ($p<0.05$), but not for xylem ($p>0.05$, Table 3.4). N availability levels did not significantly impact lesion length for either tissue type ($p>0.05$), and the dpi x N interaction term was also not significant ($p>0.05$, Table 3.4). A post hoc Tukey test (Tukey 1949) was used to assess main effects of phloem treatment factors (Table 3.5). The resulting groups were used when labelling a boxplot of the raw lesion length data (Figure 3.8). Differences between low and high N treatments were reflected by increases in lesion length across time points.

Table 3.3. Levene's test for homogeneity of variance of lesion length data. Data were either raw or transformed using the Box-Cox method ($\alpha=0.05$). The F value determines the ratio of explained variance to unexplained variance. The probability that the F value is significant is given as the P value, and the P value significance threshold is 0.05.

Lesion data	df	F value	P value
Phloem (raw)	7	5.855	8.8e-05
Phloem (transformed)	7	1146	0.3537
Xylem (raw)	5	2.7845	0.0344
Xylem (transformed)	5	1.443	0.2368

Table 3.4. Two-way analysis of variance examining effects of time and N fertilization levels on phloem and xylem lesion length. Dpi refers to the length of time following inoculation that lesion lengths were recorded, while N refers to the level of NH_4NO_3 (low or high) used to fertilize the plants. Box-Cox-transformed data were used for the analyses. Degrees of freedom (df) for each factor are shown. Sum squared values (Sum Sq) express the total variation that can be attributed to each factor. The F value determines the ratio of explained variance to unexplained variance. The probability that the F value is significant is given as the P value, and the P value significance threshold is 0.05.

Tissue	Factor	df	Sum Sq	F value	P value
Phloem	dpi	3	6.363	45.652	2.797e-13
	N	1	2.196e-8	4.727e-7	0.999
	dpi x N	3	3.181e-3	0.228	0.876
	residuals	42	1.951		
Xylem	dpi	2	16.280	2.898	0.0701
	N	1	8.857	3.154	0.0856
	dpi x N	2	1.680	0.299	0.744
	residuals	33	87.060		

Table 3.5. Tukey's honestly significant difference test for lesion length data. Secondary phloem data were first transformed using the Box-Cox method (α=0.05). Treatment combinations were assigned letters for factor groups. Mean is given by the absolute value of the difference between pairs of means, divided by the standard error as determined by a one-way ANOVA test, α=0.05.

Tissue	Time point factor	Fertilization factor	Mean	Group
Phloem	1 dpi	Low N	0.7910961	A
		High N	0.7409426	A
	7 dpi	Low N	1.2008691	B
		High N	1.2729697	B
	15 dpi	Low N	1.5699642	BC
		High N	1.5897861	B

Table 3.5. Tukey's honestly significant difference test for lesion length data. Continued.

Tissue	Time point factor	Fertilization factor	Mean	Group
Phloem	28 dpi	Low N	1.6975315	C
		High N	1.6539000	B

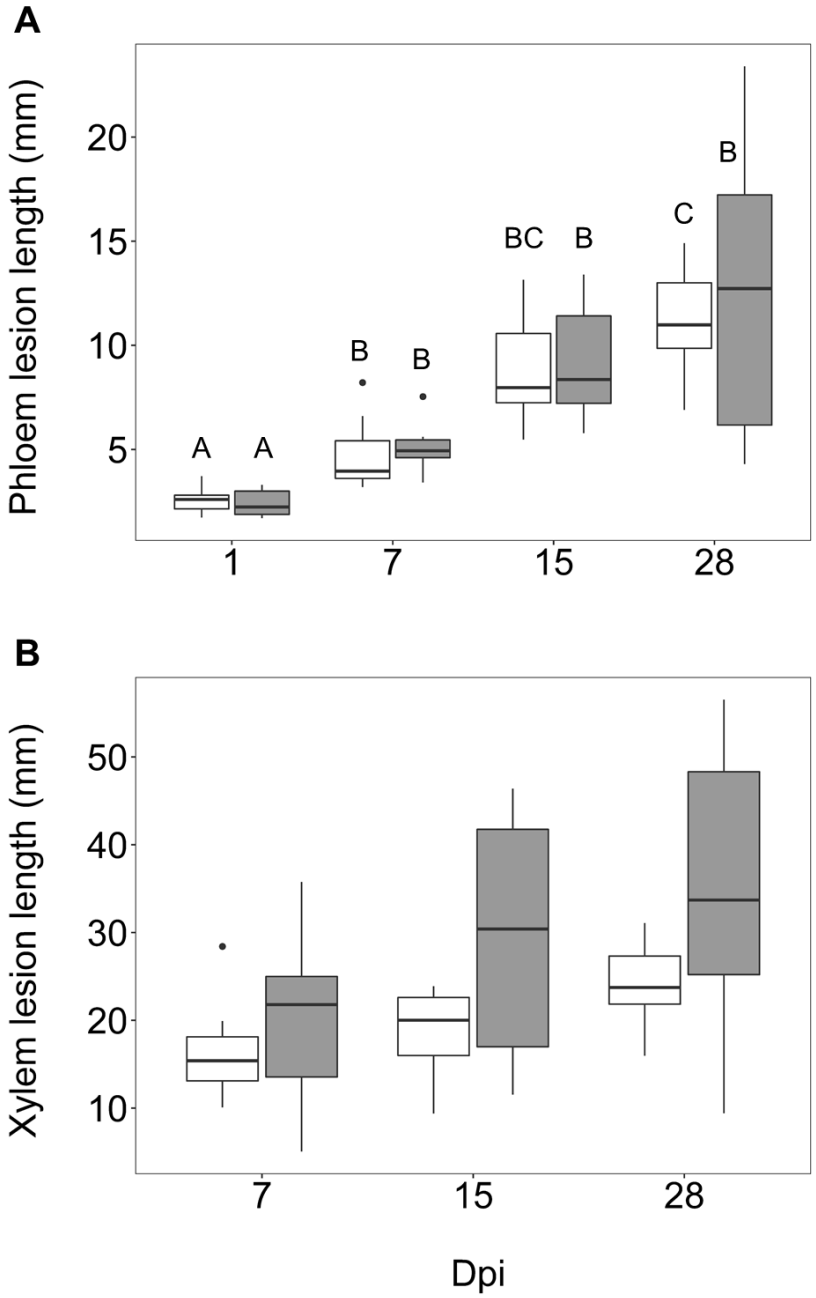


Figure 3.8. Dpi has significant effect on lodgepole pine seedling secondary phloem, but not secondary xylem. Lodgepole pine seedlings formed lesions following inoculation with *G. clavigera*. In both graphs, white bars indicate low N treatments (1 mM NH_4NO_3) while grey bars indicate high N treatments (10 mM NH_4NO_3). Medians are shown by the crossbars, upper and lower quartiles are shown by the top and bottom of the boxes, respectively, upper and lower extremes are indicated by the whiskers, and outliers are visible. Pair-wise comparisons between time points were performed for the phloem lesion data, α=0.05. Different letters represent differences between time points within each fertilization treatment. (**A**) Lodgepole pine phloem lesion length, *n*=6-7 and (**B**) lodgepole pine xylem lesion length, *n*=6-7.

Foliar N concentration and lesion length measurements were taken from the same experiment, but not the same trees. Foliar N concentration values for low N and high N plants, respectively, averaged 1.53 ± 0.28% and 2.09 ± 0.27% for all time points and inoculation treatments. The positive correlation between foliar N concentration for fungal-inoculated trees and lesion length is showcased for both secondary phloem (Figure 3.9 A) and secondary xylem (Figure 3.9 B).

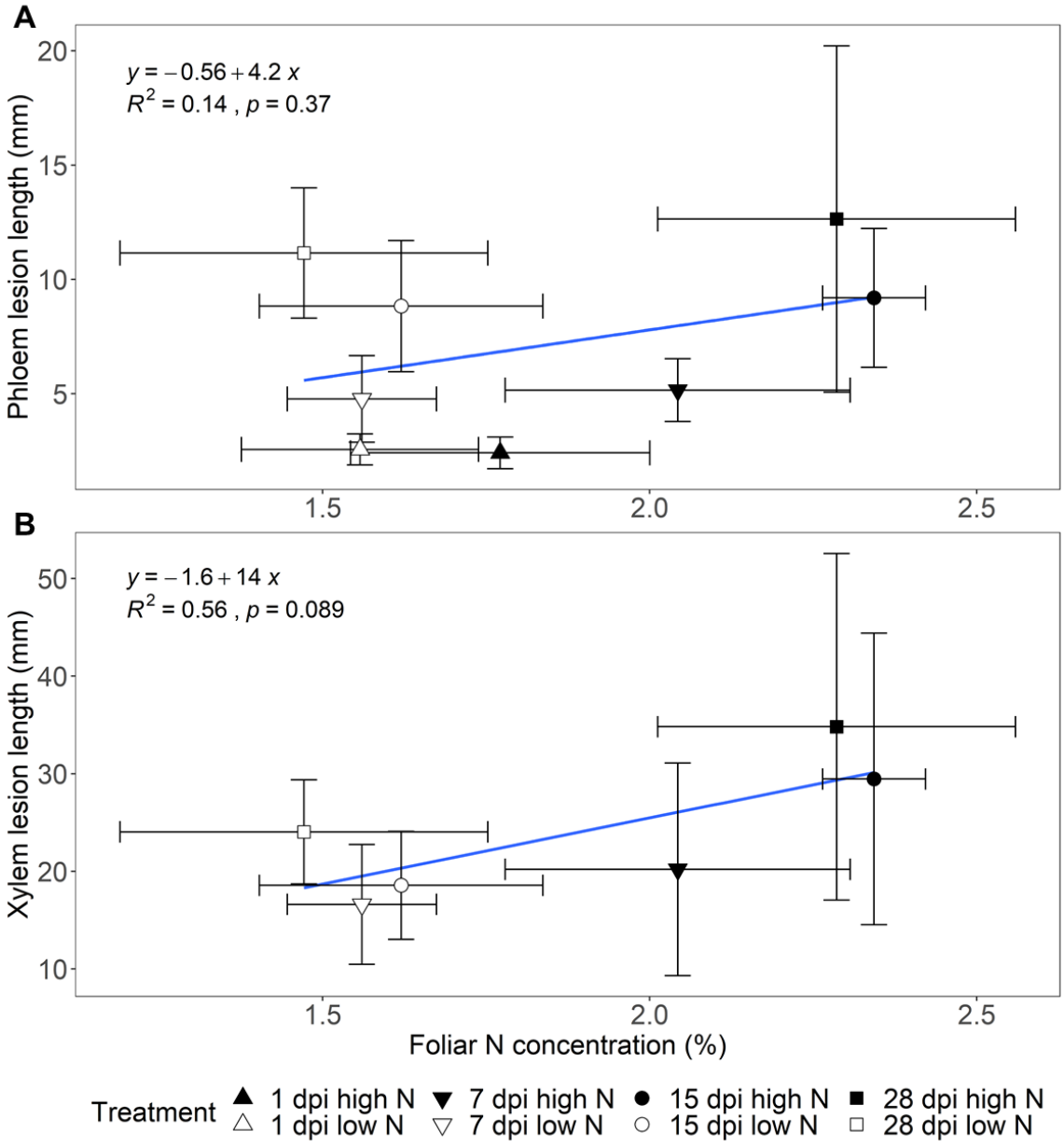


Figure 3.9. Effect of foliar N concentration on lesion length for lodgepole pine seedlings. Total foliar N and lesion length were measured for G. *clavigera*-inoculated lodgepole pine grown with 10 mM (high; filled symbols) or 1 mM (low; open symbols) NH_4NO_3 fertilization. Mean values (±SE) are shown, n=4-7. Linear regression equations are displayed, as are correlation R-squared values with accompanying significance values. (**A**) Phloem tissue was measured at 1 (triangles), 7 (inverted triangles), 15 (circles) and 28 (squares) dpi. (**B**) Xylem tissue was measured at 7 (inverted triangles), 15 (circles) and 28 (squares) dpi.

3.4 Discussion

This study had two main objectives: (1) to determine whether application of 1 mM vs. 10 mM NH_4NO_3 nutrient solution would result in differential N assimilation in fertilized lodgepole pine seedlings grown in controlled growth chamber conditions as reflected by foliar N concentration, and (2) to assess the impact of N availability on lodgepole pine defenses by comparing lesion development in *G. clavigera*-inoculated lodgepole pine fertilized with 1 mM vs. 10 mM NH_4NO_3 nutrient solution. Analysis of lesion length in lodgepole pine seedlings in response to N fertilization is currently lacking, especially in the context of lesion development as a result of the defense response of the lodgepole pine to *G. clavigera*. This provided motivation for this study.

Lodgepole pine has been shown to take up both ammonium (NH_4^+) and nitrate (NO_3^-; Hawkins *et al.* 2008). Foliar N levels in trees, including mature lodgepole pine, have been utilized to predict the influence of fertilization on plant N economy (Brockley 2000; Muñoz-Huerta *et al.* 2013). N remobilization is closely linked to phenological events in annual growth cycles (Millard and Gretlet 2010). Though N-rich storage proteins in stem tissues should be considered when calculating N concentrations later in the season (Wetzel *et al.* 1989), foliar N measurements were most appropriate in our study given the phenological stage of our trees. N concentration analysis of needles collected from lodgepole pine seedlings at four different time points following three different inoculation treatments were tested for alterations in N concentration. We demonstrated that

trees treated with low N (1 mM NH_4NO_3) nutrient solution accumulated significantly less foliar N per needle dry weight than trees treated with high N (10 mM NH_4NO_3) nutrient solution. This provided important confirmation that these levels of NH_4NO_3 are appropriate for studying the effect of N fertilization on the defense response of lodgepole pine under the controlled growth environment conditions. The 10-fold difference between the high N treatment and the low N treatment was selected to mimic the variation of soil N concentrations that might be encountered in the field (Prescott and Preston 1994; Köchy and Wilson 2005). Under the short time frames used for this study, these analyses revealed that wounding or *G. clavigera* inoculation of lodgepole pine seedling stems had no significant effect on foliar N concentration. This may in part be attributed to the fact that our trees were light-limited throughout the duration of this study. Diminished light levels reduce a plant's ability to incorporate inorganic N into amino groups (Champigny 1995; Kaiser *et al.* 1999). Extending the duration of the N treatment application under increased light levels may have resulted in inoculation type having a significant effect on foliar N concentration.

The production of primary and secondary metabolites in woody plants has been shown to respond to varying levels of N fertilization (Haukioja *et al.* 1998). Lodgepole pine lesions are the result of chemical and physical defense response to the growth of *G. clavigera* into the phloem and xylem tissue. They are laden with defense-related secondary metabolites (Franceschi *et al.* 2005; Rice and Langor 2008; Arango-Velez *et al.* 2016; McAllister 2018). We found that the progression

of lesion development in both 1 mM and 10 mM NH_4NO_3-treated trees was similar to that reported by Arango-Velez *et al.* (2016). Lesion length increased over time in secondary phloem and secondary xylem. This difference in lesion length over time was significant for the lesions that developed in phloem ($p<0.01$) but was only nearly significant for lesions measured in xylem ($p=0.07$). The lack of significant difference in lesion growth over time in xylem may be due to its vascular cambium producing new cells, which accumulated over the lesion. This accumulation was principally at the edges where cambial cell divisions are less influenced by defense responses (Fortier, St. Onge and Cooke, unpublished data). When lesion length is measured at the surface of the woody layer that is revealed by peeling off the bark, this new cell production, which is primarily manifested at later time points, masks the true extent of lesion growth. The relationship between stem secondary growth and lesion development in lodgepole pine needs further examination.

Lesion length in secondary phloem and secondary xylem increased as a function of N availability with the effect of N availability on lesion length being greatest at later time points. While the effect of N availability on lesion length in xylem was nearly significant ($p=0.086$), there was no significant effect of N availability on lesion length in phloem ($p=0.999$). In both xylem and phloem, increased variance in lesion length under the high N treatment likely contributed to the lack of significant differences in lesion length between 1 mM and 10 mM NH4NO3-treated seedlings. Increasing the number of seedlings sampled and access to more genetically uniform seedlings might lead to a significant statistical

difference between the two N levels. This is particularly true for the xylem, which is the tissue that is primarily colonized by *G. clavigera* (Ballard *et al.* 1982).

While we have found a positive but not statistically significant relationship between N availability and *G. clavigera*-induced lesion development, other studies that have looked at the effect of N availability on pathogen-induced lesion development in conifers do not report a consistent pattern. Following inoculation with *Sphaeropsis sapinea* (Fries), N fertilization significantly increased lesion size in mature red pine (*Pinus resinosa* Aiton) (Blodgett *et al.* 2005). Sitka spruce (*Picea sitchensis* Bongard) inoculated with the fungal pathogen *Phacidium coniferarum* (G.G. Hahn) produced higher concentrations of resin and polyphenols in lower N treatments when compared with higher N treatments (Wainhouse *et al.* 1998). Interestingly, N was not a significant factor for lesion size in response to *P. coniferarum* infection (Wainhouse *et al.* 1998). In response to *Ceratocystis polonica* (Siemaszko) inoculation in Norway spruce (*Picea abies* Miller), resistance parameters, such as lesion length and resin phenolic concentrations, were unaffected by N fertilization (Kytö *et al.* 1996). However, Kytö *et al.* (1996) found a significant positive correlation between lesion length and stem diameter increment. In conjunction with these analyses, our results, suggest that a finer scale analyses of defense responses, rather than the coarse-scale measurement of lesion length, is required to resolve the relationship between N availability and the tree's ability to effectively defend itself against pathogen attack.

There are many contradictions in the literature concerning the impact of plant N status on defense capacity against pathogen attack. Low N environments can increase a plant's susceptibility to disease. During growth under nutrient-limited conditions, several bacterial and fungal genes, which are thought to be involved in plant pathogenicity, are induced *(*Snoeijers *et al.* 2000). For example, the fungal pathogen *Cladosporium fulvum* (Cooke) avirulence gene *Avr9* is both induced in tomato plants and during N starvation *in vitro* (Van den Ackerveken *et al.* 1994). Low N conditions can reduce production of defensive proteins. When Dietrich *et al.* (2004) applied BION© (Syngenta AG, Basel, Switzerland), a chemical resistance elicitor, to *A. thaliana* to simulate pathogen attack, they found that the constitutive and induced levels of peroxidase and chitinase activity were significantly lower under limited N supply (Dietrich *et al.* 2004). Conversely, high concentrations of N in host tissues can increase susceptibility of plants to diseases (Snoeijers *et al.* 2000; Huber *et al.* 2012; Mur *et al.* 2016). On barley (*Hordeum vulgare* Linnaeus), Jensen and Munk (1997) found that powdery mildew (*Erysiphe graminis* f. sp. *hordei* Jaczewski) colony density was significantly increased with increasing concentrations of NH_4NO_3 fertilization. Similarly, *G. clavigera*-inoculated lodgepole pine seedlings may have been more susceptible to fungal attack in nutrient-rich conditions. Larger lesions in high compared to low N-treated trees may have resulted from an elevated defense response.

The growth-differentiation balance (GDB) hypothesis views the availability of resources, such as N, as affecting the balance between the investment of C in the

production of biomass (growth) and in the chemical and structural modification of biomass (differentiation) (Herms and Mattson 1992). Differentiation includes the maintenance of defense chemicals and mechanisms (Lorio 1986). GDB predicts that defense is favored at low resource availability at the expense of growth, and growth is favored at high resource availability when defense is low (Herms and Mattson 1992). In this context, high N fungal-inoculated lodgepole pine would allocate more resources into growth rather than defense when compared with low N fungal-inoculated lodgepole pine that would allocate more resources into defense rather than growth. However, given that lodgepole pine lesions are a measure of the strength of the defense response, our results run counter to this theory. We found that lesion length was increased with a higher level of soil N fertilization.

The carbon-nutrient balance (CNB) hypothesis explains that the C to nutrient ratio in the environment influences the variation of phenotypic expression in defense (Stamp 2003). The CNB hypothesis assumes that resources are first allocated to growth, then any remaining goes towards defense or storage. In this context, both low and high N treated trees may have first allocated N towards growth. Increased lesion length in high N seedlings may have been the result of a surplus of N available to support the defense response.

The optimal defense (OD) hypothesis states that constitutive defenses are costly, necessitating mechanisms for a timely induced defense response (Herms and Mattson 1992; Stamp 2003). This is showcased in this study by lesion

development occurring shortly after inoculation with *G. clavigera*. High N treated trees had more "nutrient currency" to contribute to induced responses compared with low N treated trees. This may have resulted in the formation of longer lesions in the high N seedlings.

It is also possible that the high N- compared to the low N-treated seedlings experienced delayed growth cessation (Toca *et al.* 2017). This would allow for the allocation of defense compounds to the lesion zone of the high N trees rather than to other cellular processes associated with growth cessation, such as secondary cell wall thickening and lignification (Dougherty *et al.* 1994; Arango-Velez *et al.* 2014).

Our results reveal that the positive relationship between lesion length and N fertilization is nearly significant for xylem, the main tissue inhabited by *G. clavigera* (Ballard *et al.* 1982). Goodsman *et al.* (2012) found that fungal growth in the phloem of lodgepole pine was not impacted by increased fertilization. However, in response to increased N availability, the extent of hyphal expansion in lodgepole pine xylem was not measured (Goodsman *et al.* 2012). The high N-treated tissue in our study may have offered a more nutritious environment for fungal colonization compared with the low N-treated tissue. It is therefore possible that we observed increased lesion length in the xylem in response to increased N-stimulated fungal growth. As has been observed in other plant-pathogen systems, the fungus may have benefitted from increased fertilization more than the host (Lahr and Krokene 2013; Mur *et al.* 2016). Fungal samples for qPCR analysis collected from this experiment will allow for further investigation of the correlation

between lesion progression and fungal growth within the developing lesion (McAllister *et al.* 2018).

3.5 Conclusion

At the onset of this study, we hypothesized that levels of N fertilization would be positively correlated with the concentration of N found in lodgepole pine foliage and that trees supplied with 1 mM compared to 10mM NH_4NO_3 nutrient solution would have lower foliar N concentrations. Our results support this hypothesis. We also hypothesized that the lodgepole pine defense against *G. clavigera*, as measured by differences in lesion development, would vary between these two NH_4NO_3 levels. We found that although there was an overall positive relationship between N availability and lesion length, this relationship was not significant in either tissue type. The N concentration of foliage was positively correlated with phloem and xylem lesion development, with the relationship being nearly significant in xylem. Running counter to the GDB hypothesis, our results indicated that increased N availability induced a stronger defense response in lodgepole pine to *G. clavigera* inoculation. As explained by the OD hypothesis, the fungal-inoculated lodgepole pine seedlings grown with high N fertilization had more nutrient currency to allocate to defense. Increased lesion length may have resulted from a surplus of N being allocated to growth as predicted by the CNB hypothesis. The analyses carried out in this study lay the foundation for a

comprehensive investigation into the relationship between soil N availability and the lodgepole pine defense response to MPB-vectored *G. clavigera*.

3.6 References

Amponsah, I. G., Lieffers, V. J., Comeau, P. G., Brockley, R. P. (2004). Growth response and sapwood hydraulic properties of young lodgepole pine following repeated fertilization. *Tree Physiology*, 24, 1099-1108.

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

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

Ayres, M. P., Wilkens, R. T., Ruel, J. J., Lombardero, M. J., Vallery, E. (2000). Nitrogen budgets of phloem-feeding bark beetles with and without symbiotic fungi. *Ecology*, 81, 2198–2210.

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.

Bentz, B. J., Kegley, S., Gibson, K., Thier, R. (2005). A test of high-dose verbenone for stand-level protection of lodgepole and whitebark pine from mountain pine beetle (*Coleoptera: Curculionidae: Scolytinae*) attacks. *Journal of Economic Entomology*, 98, 1614-1621.

Blodgett, J. T., Herms, D. A., Bonello, P. (2005). Effects of fertilization on red pine defense chemistry and resistance to *Sphaeropsis sapinea*. *Forest Ecology and Management*, 208, 373-382.

Box, G. E. P., Cox, D. R. (1964). An analysis of transformation. *Journal of the Royal Statistical Society, Series B (Methodological)*, 26, 211-252.

Brockley, R. P. (2000). Using foliar variables to predict the response of lodgepole pine to nitrogen and sulphur fertilization. *Canadian Journal of Forest Research*, 30, 1389-1399.

Brockley, R. P. (2001). Fertilization of lodgepole pine in western Canada. In Bamsey, C. (Ed.), *Enhanced forest management: Fertilization and economics conference*. March 1-2, 2001, Edmonton, Alberta, Canada. pp. 44-55.

Buchanan, B. B., Wolosiuk, R. A. (2015). Photosynthesis: The carbon reactions. In Taiz, L., Zeiger, E., Møller, I. M., Murphy, A. (Eds.), *Plant physiology and development* (6th ed.). Sinauer Associates, Inc., Publishers, Sunderland, Massachusetts, United States. pp. 203-241.

Champigny, M. L. (1995). Integration of photosynthetic carbon and nitrogen metabolism in higher plants. *Photosynthesis Research*, 46, 117-127.

Chiu, C. C., Keeling, C. I., Bohlmann, J. (2019). The cytochrome P450 CYP6DE1 catalyzes the conversion of α-pinene into the mountain pine beetle aggregation pheromone trans-verbenol. *Scientific Reports*, 9, 1-10.

Clérivet, A., Déon, V., Alami, I., Lopez, F., Geiger, J., Nicole, M. (2000). Tyloses and gels associated with cellulose accumulation in vessels are responses of plane tree seedlings (*Platanus* × *acerifolia*) to the vascular fungus *Ceratocystis fimbriata* f. sp *platani*. *Trees*, 15, 25-31.

Clérivet, A., El Modafar, C. (1994). Vascular modifications in *Platanus acerifolia* seedlings inoculated with *Ceratocystis fimbriata* f. sp. *platani*. *European Journal of Forest Pathology*, 24, 1–10.

Cook, S. P., Shirley, B. M., Zambino, P. J. (2010). Nitrogen concentration in mountain pine beetle larvae reflects nitrogen status of the tree host and two fungal associates. *Environmental Entomology*, 39, 821-826.

Cook, S., Carroll, A., Kimsey, M., Shaw, T. (2015). Changes in a primary resistance parameter of lodgepole pine to bark beetle attack one year following fertilization and thinning. *Forests*, 6, 280-292.

Cullingham, C. I., James, P. M. A., Cooke, J. E. K., 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.

Dietrich, R., Ploss, K., Heil, M. (2004). Constitutive and induced resistance to pathogens in *Arabidopsis thaliana* depends on nitrogen supply. *Plant, Cell and Environment*, 27, 896–906.

DiGuistini, S., Wang, Y., Liao, N. Y., Taylor, G., Tanguay, P., Feau, N., Henrissat, B., Chan, S. K., Hesse-Orce, U., Alamouti, S. M., Tsui, C. K. M., Docking, R. T., Levasseu, A., Haridas, S., Robertson, G., Birol, I., Holt, R. A., Marra, M. A., Hamelin, R. C., Hirst, M., Jones, S. J. M., Bohlman, J., Breuil, C. (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*, 108, 2504–2509.

Dougherty, P. M., Whitehead, D., Vose, J. M. (1994). Environmental influences on the phenology of pine. *Ecological Bulletins (Copenhagen)*, 43, 64-75.

Evans J. R., Seemann J. R. (1989). The allocation of protein nitrogen in the photosynthetic apparatus: Cost, consequences and control. In Briggs, W. R. (Ed.), *Photosynthesis*. Alan R. Liss, Inc., New York, New York, United States. pp. 183–205.

Fisher, R. A. (1934). *Statistical methods for research workers* (5th ed.). Oliver and Boyd, Edinburgh, Scotland.

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.

Fox, J., Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage Publications, Inc., Thousand Oaks, California, United States.

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, 353-375.

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

Goodsman, D. W., Erbilgin, N., Lieffers, V. J. (2012). The impact of phloem nutrients on overwintering mountain pine beetles and their fungal symbionts. *Environmental Entomology*, 41, 478-486.

Guérard, N., Dreyer, E., Lieutier, F. (2000). Interactions between Scots pine*, Ips acuminatus* (Gyll.) and *Ophiostoma brunneociliatum* (Math.): Estimation of the critical thresholds of attack and inoculation densities and effects on hydraulic properties in the stem. *Annals of Forest Science, 57*, 681-690.

Haukioja, E., Ossipov, V., Koricheva, J., Honkanen, T., Larsson, S., Lempa, K. F. (1998). Biosynthetic origin of carbon-based secondary compounds: Cause of variable responses of woody plants to fertilization? *Chemoecology*, 8, 133–139.

Hawkins, B. J., Boukcim, H., Plassard, C. (2008). A comparison of ammonium, nitrate and proton net fluxes along seedling roots of Douglas-fir and lodgepole pine grown and measured with different inorganic nitrogen sources. *Plant, Cell and Environment*, 31, 278–287.

Herms, D. A., Mattson, W. J. (1992). The dilemma of plants: To grow or defend. *The Quarterly Review of Biology*, 67, 283–335.

Hocking, D. (1971). Preparation and use of a nutrient solution for culturing seedlings of lodgepole pine and white spruce, with selected bibliography. *Northern*

Forest Research Centre Information Report Nor-X-1. Canadian Forest Service, Department of the Environment, Edmonton, Alberta, Canada.

Hoffland, E., van Beusichem, M. L., Jeger, M. J. (1999). Nitrogen availability and susceptibility of tomato leaves to *Botrytis cinerea*. *Plant and Soil*, 210, 263–272.

Hoffland, E., van Beusichem, M. L., Jeger, M. J. (2000). Effect of nitrogen supply rate on disease resistance in tomato depends on the pathogen. *Plant and Soil*, 218, 239–247.

Högberg, P., Näsholm, T., Franklin, O., Högberg, M. N. (2017). Tamm review: On the nature of the nitrogen limitation to plant growth in Fennoscandian boreal forests. *Forest Ecology and Management*, 403, 161-185.

Hubbard, R. M., Rhoades, C. C., Elder, K., Negron, J. (2013). Changes in transpiration and foliage growth in lodgepole pine trees following mountain pine beetle attack and mechanical girdling. *Forest Ecology and Management,* 289, 312-317.

Huber, D., Römheld, V., Weinmann, M. (2012). Relationship between nutrition, plant diseases and pests. In Marschner, H. (Ed.), *Marschner's mineral nutrition of higher plants*. Academic Press, Cambridge, Massachusetts, USA. pp. 283–298.

Jensen, B., Munk, L. (1997). Nitrogen-induced changes in colony density and spore production of *Erysiphe graminis* f. sp. *hordei* on seedlings of six spring barley cultivars. *Plant Pathology*, 46, 191–202.

Kaiser, W. M., Weiner, H., Huber, S. C. (1999). Nitrate reductase in higher plants: A case study for transduction of environmental stimuli into control of catalytic activity. *Physiologia Plantarum*, 105, 384-389.

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

Klepzig, K. D., Robison, D. J., Fowler, G., Minchin, P. R., Hain, F. P., Allen, H. L. (2005). Effects of mass inoculation on induced oleoresin response in intensively managed loblolly pine. *Tree Physiology*, 25, 681–688.

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

Komsta, L. (2011). Outliers: Tests for outliers. R package version 0.14. *https://CRAN.R-project.org/package=outliers*.

Kovalchuk, A., Keriö, S., Oghenekaro, A. O., Jaber, E., Raffaello, T., Asiegbu, F. O. (2013). Antimicrobial defenses and resistance in forest trees: Challenges and perspectives in a genomic era. *Annual Review of Phytopathology*, 51, 221–244.

Kytö, M., Niemelä, P., Annila, E. (1996). Vitality and bark beetle resistance of fertilized Norway spruce. *Forest Ecology and Management*, 84, 149–157.

Lahr, E. C., Krokene, P. (2013). Conifer stored resources and resistance to a fungus associated with the spruce bark beetle *Ips typographus*. *PLoS ONE*, 8, 1-8.

Leser, C., Treutter, D. (2005). Effects of nitrogen supply on growth, contents of phenolic compounds and pathogen (scab) resistance of apple trees. *Physiologia Plantarum*, 123, 49–56.

Lenth, R. (2020). emmeans: Estimated marginal means, aka least-squares means. R package version 1.4.5. *https://CRAN.R-project.org/package=emmeans*.

Levene, H. (1960). *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (Olkin, I., Ghurye, S. G., Hoeffding, W., Madow, W. G., Mann, H. B, Eds.). Stanford University Press, Palo Alto, California, United States. pp. 278-292.

Lorio, P. L. (1986). Growth-differentiation balance: A basis for understanding southern pine beetle-tree interactions. *Forest Ecology and Management*, 14, 259-273.

Luo, J., Zhou, J.-J., Masclaux-Daubresse, C., Wang, N., Wang, H., Zheng, B. (2019). Morphological and physiological responses to contrasting nitrogen regimes in *Populus cathayana* is linked to resources allocation and carbon/nitrogen partition. *Environmental and Experimental Botany*, 162, 247–255.

Lusebrink, I., Erbilgin, N., Evenden, M. L. (2013). The lodgepole × Jack pine hybrid zone in Alberta, Canada: A stepping stone for the mountain pine beetle on its journey east across the boreal forest? *Journal of Chemical Ecology*, 39, 1209-1220.

Marín, I. C., Loef, I., Bartetzko, L., Searle, I., Coupland, G., Stitt, M., Osuna, D. (2011). Nitrate regulates floral induction in *Arabidopsis*, acting independently of light, gibberellin and autonomous pathways. *Planta*, 233, 539–552.

Matyssek, R., Agerer, R., Ernst, D., Munch, J.-C., Osswald, W., Pretzsch, H., Priesack, E., Schnyder, H., Treutter, D. (2005). The plant's capacity in regulating resource demand. *Plant Biology*, 7, 560–580.

Matyssek, R., Schnyder, H., Elstner, E.-F., Munch, J.-C., Pretzsch, H., Sandermann, H. (2002). Growth and parasite defence in plants: The balance between resource sequestration and retention. *Plant Biology*, 4, 133–136.

McAllister, C. H., Fortier, C. E., Onge, K. R. S., Sacchi, B. M., Nawrot, M. J., Locke, T., Cooke, J. E. K. (2018). A novel application of RNase H2-dependent quantitative PCR for detection and quantification of *Grosmannia clavigera*, a mountain pine beetle fungal symbiont, in environmental samples. *Tree Physiology*, 38, 485–501.

Morris, H., Brodersen, C., Schwarze, F. W. M. R., Jansen, S. (2016). The parenchyma of secondary xylem and its critical role in tree defense against fungal decay in relation to the CODIT model. *Frontiers in Plant Science*, 7, 1-18.

Muñoz-Huerta, R., Guevara-Gonzalez, R., Contreras-Medina, L., Torres-Pacheco, I., Prado-Olivarez, J., Ocampo-Velazquez, R. (2013). A review of methods for sensing the nitrogen status in plants: Advantages, disadvantages and recent advances. *Sensors*, 13, 10823–10843.

Mur, L. A. J., Simpson, C., Kumari, A., Gupta, A. K., Gupta, K. J. (2016). Moving nitrogen to the centre of plant defense against pathogens. *Annals of Botany*, 119, 703–709.

Natural Resources Canada (2017, February 21). Mountain pine beetle (factsheet). *https://www.nrcan.gc.ca/forests/fire-insects-disturbances/top-insects/13397*. (Last accessed April 13, 2020).

Natural Resources Canada (2017, June 14). Forest classification. *https://www.nrcan.gc.ca/our-natural-resources/forests-forestry/sustainable-forest-management/measuring-reporting/forest-classification/13179*. (Last accessed April 13, 2020).

Nolan, T., Hands, R. E., Bustin, S. A. (2006). Quantification of mRNA using real-time RT-PCR. *Nature Protocols*, 1, 1559-1582.

Ott, D. S., Yanchuk, A. D., Huber, D. P. W., Wallin, K. F. (2012). Genetic variation of lodgepole pine (*Pinus contorta* var. *latifolia*) chemical and physical defenses that affect mountain pine beetle, *Dendroctonus ponderosae,* attack and tree mortality. *Journal of Chemical Ecology*, 37, 1002–1012.

Prescott, C., de Montigny, L., Harper, G. (2019). Eighteen-year growth responses to thinning and fertilization of a height-repressed lodgepole pine stand in interior British Columbia. *The Forestry Chronicle*, 95, 207-221.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *https://www.R-project.org*.

Raffa, K. F., Berryman, A. A. (1987). Interacting selective pressures in conifer-bark beetle systems: A basis for reciprocal adaptations? *The American Naturalist*, 129, 234-262.

Raffa, K. F., Smalley, E. B. (1988). Seasonal and long-term responses of host trees to microbial associates of the pine engraver, *Ips pini*. *Canadian Journal of Forest Research*, 18, 1624–1634.

Rice A. V., Thormann M. N., Langor D. W. (2007a). Mountain pine beetle associated blue-stain fungi cause lesions on jack pine, lodgepole pine, and lodgepole x jack pine hybrids in Alberta. *Canadian Journal of Botany,* 85, 307-315.

Rice, A. V., Langor, D. W. (2008). A comparison of heat pulse velocity and lesion lengths for assessing the relative virulence of mountain pine beetle-associated fungi on jack pine. *Forest Pathology*, 38, 257-262.

Roe A. D., Rice A. V., Bromilow S. E., Cooke J. E. K., Sperling F. A. H. (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.

Roe A. D., Rice A. V., Coltman D. W., Cooke J. E. K., Sperling F. A. H. (2011). Comparative phylogeography, genetic differentiation, and contrasting reproductive modes in three fungal symbionts of a multipartite bark beetle symbiosis. *Molecular Ecology*, 20, 584–600.

RStudio Team (2015). *RStudio: Integrated development for R*. RStudio, Inc., Boston, MA.

Shapiro, S. S., Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591.

Snoeijers, S. S., Perez-Garcia, A., Joosten, M. H. A. J., De Wit, P. G. M. (2000). The effect of nitrogen on disease development and gene expression in bacterial and fungal plant pathogens. *European Journal of Plant Pathology*, 106, 493-506.

Stamp, N. (2003). Out of the quagmire of plant defense hypotheses. *The Quarterly Review of Biology*, 78, 23–55.

Toca, A., Oliet, J. A., Villar-Salvador, P., Maroto, J., Jacobs, D. F. (2017). Species ecology determines the role of nitrogen nutrition in the frost tolerance of pine seedlings. *Tree Physiology*, 38, 96–108.

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 5, 99-114.

Vadeboncoeur, M. A. (2010). Meta-analysis of fertilization experiments indicates multiple limiting nutrients in northeastern deciduous forests. *Canadian Journal of Forest Research,* 40, 1766-1780.

Van den Ackerveken, G. F. J. M., Dunn, R. M., Cozijnsen, T. J., Vossen, P., Van den Broek, H. W. J., De Wit, P. J. G. M. (1994). Nitrogen limitation induces expression of the avirulence gene *Avr9* in the tomato pathogen *Cladosporium fulvum*. *Molecular and General Genetics*, 243, 277–285.

Wainhouse, D., Ashburner, R., Ward, E., Rose, J. (1998). The effect of variation in light and nitrogen on growth and defense in young Sitka spruce. *Functional Ecology*, 12, 561-572.

Wang, Y., Lim, L., Diguistini, S., Robertson, G., Bohlmann, J., Breuil, C. (2012). A specialized ABC efflux transporter GcABC-G1 confers monoterpene resistance to *Grosmannia clavigera*, a bark beetle-associated fungal pathogen of pine trees. *New Phytologist*, 197, 886–898.

Wong, B. L., Berryman, A. A. (1977). Host resistance to the fir engraver beetle. 3. Lesion development and containment of infection by resistant *Abies grandis* inoculated with *Trichosporium symbioticum*. *Canadian Journal of Botany*, 55, 2358–2365.

Zabel, R. A., Morell, J. J. (1992). Wood stains and discolorations. In Zabel R. A., Morell J. J. (Eds.), *Wood microbiology: Decay and its prevention*. Academic Press, San Diego, California, United States. pp. 326–343.

Zimmerman, C. F., Keefe, C. W., Bashe, J. (1997). *Method 440.0 Determination of carbon and nitrogen in sediments and particulates of estuarine/coastal waters using elemental analysis*. U.S. Environmental Protection Agency, Washington, District of Columbia, United States, EPA/600/R-15/009, pp 365.5-1 – 365.5-9.

4.0 Chapter 4: Expression profiling lodgepole pine defense mechanisms in response to *Grosmannia clavigera* inoculation and varying levels of nitrogen fertilization

4.1 Introduction

The mountain pine beetle (MPB; *Dendroctonus ponderosae* Hopkins) is currently causing one of the largest outbreaks of bark beetles ever recorded in North America, decimating over 18 million hectares of pine forests across western Canada, including over one million hectares in Alberta alone (Hodge *et al.* 2017). Lodgepole pine (*Pinus contorta* Douglas ex Loudon *var. latifolia*) has a range that overlaps extensively with that of MPB (Safranyik *et al.* 2010), sharing an evolutionary history with the pest as one of its main hosts (Raffa and Berryman, 1987). Climate change has caused a north- and eastward shift in MPB range, and MPB has spread from British Columbia across northern Alberta (Cullingham *et al.* 2011, 2012). The resulting mass mortality of lodgepole pine is troubling for ecosystems as well as the timber, paper and outdoor recreation industries (Corbett *et al.* 2016; Morris *et al.* 2018).

MPB vectors a microbial complex that is introduced into the phloem of the tree during host colonization. Some of the most pathogenic MPB-vectored fungi belong to the family *Ophiostomatales,* including the necrotrophic blue-stain fungus *Grosmannia clavigera* (Robinson-Jeffrey and Davidson) which can rapidly colonize the phloem and sapwood of the tree (Ballard *et al.* 1984). Larvae and emerging new generation of beetles feed on blue-stain fungi mycelium that is

abundant in the pupal chambers (Paine *et al.* 1997). *G. clavigera* aids the beetle by detoxifying defense chemicals in tree resin meant to prevent the MPB from successfully borrowing further into the bark (DiGuistini *et al.* 2011). Fungal colonization results in the swelling of phloem polyphenolic parenchyma cells (Arango-Velez *et al.* 2014) and in the production of xylem parenchyma cell ingrowths called tyloses (Ballard *et al.* 1982). While tyloses prevent the axial spread of the fungus, they ultimately disrupt water transport resulting in tree mortality (Ballard *et al.* 1982; Arango-Velez *et al.* 2016).

Lodgepole pine exhibit both constitutive and inducible defense mechanisms to confine the *G. clavigera* infection and actively cause hyphal mortality (Franceschi *et al.* 2005). Constitutive (preformed) defenses provide immediate barriers to an invasion, either through physical features, such as stone cells, or through chemical properties involving a multitude of secondary metabolites (Franceschi *et al.* 2005). Products of the phenylpropanoid pathway (phenolics) and isoprenoid pathway (terpenoids) can have potent effects on pathogens (Franceschi *et al.* 2005). An invasion also activates inducible defenses associated with cells surrounding the inoculation site, including the formation of new traumatic resin ducts and a hypersensitive response associated with the accumulation of specialized phenolics and terpenoids along with lignin, which serves as a physical barrier to the spread of infection (Franceschi *et al.* 2005). Inducible defenses also include the production of defense proteins, such as chitinases (Liu *et al.* 2005), peroxidases (Fossdal *et al.* 2001) and thaumatins

(Selitrennikoff 2001). Previous experiments have shown alterations in tree defense shortly after wounding and inoculation with *G. clavigera* (Arango-Velez *et al.* 2016).

Identification of signaling mechanisms and hormones is necessary to understand the regulatory networks associated with induced defense processes in lodgepole pine to *G. clavigera*. Plants detect pathogens by identifying pathogen-associated molecular patterns which triggers a variety of defense responses, such as mitogen-activated protein kinase (MAPK) cascades, that transduce the signal to transcription factors, ultimately altering gene regulation (Fagard *et al.* 2014; Taiz *et al.* 2015). Jasmonic acid (JA) and ethylene are plant hormones commonly associated with the induced defense response, both of which can recruit a significant portion of a plant's defense network after herbivore or pathogen attack (Broekgaarden *et al.* 2015; Ruan *et al.* 2019). Increased levels of JA are associated with response to necrotrophic pathogens like *G. clavigera* (Glazebrook 2005; Zhang *et al.* 2017). In conifers, production of ethylene has been associated with the biosynthesis of monoterpenes in response to fungal contamination (Popp *et al.*1995).

Nitrogen (N) availability has been shown to affect both partitioning of resources into growth vs. defense processes, and allocation of resources into carbon (C)-based secondary metabolites and N-based protein defenses (Blodgett *et al.* 2005; Massad *et al.* 2012; Cook *et al.* 2015; Sun *et al.* 2020). The relative proportion of resources that are invested in C-intensive defenses vs. N-intensive

defenses may be influenced by the relative amount of N availability and could potentially alter the efficacy of lodgepole pine defense against *G. clavigera* (Bryant *et al.* 1983; Tuomi *et al.* 1988; Tuomi *et al.* 1991). Several theories have been presented regarding allocation of nutrients in plants. The carbon-nutrient balance (CNB) theory assumes that C is first allocated to growth, with any remaining going towards defense or storage (Bryant *et al.* 1983). The CNB theory states the sizes of internal (assimilated) C and nutrient pools relative to each other determines the chemical quality of defense metabolites (Matyssek *et al.* 2002; Matyssek *et al.* 2005). The growth-differentiation balance (GDB) hypothesis states that growth and defense-related metabolism respond in terms of a trade-off, and that while defense is favored at low resource availability at the expense of growth, growth is favored at high resource availability when defense is low (Herms and Mattson 1992). The optimal defense (OD) hypothesis predicts that the construction and maintenance of constitutive defenses against herbivores and pathogens are costly because they divert resources away from growth (Stamp 2003). Constitutive defense elements are therefore concentrated, and a timely induced defensive strategy is strongest, within the most vulnerable tissues (Herms and Mattson 1992; Meldau *et al.* 2012).

While there is a considerable body of literature examining the effect of N availability on plant defenses using physiological or ecological approaches, few studies have investigated N effects on defense using molecular or genomic approaches. Most of these studies have been carried out with annual plants,

particularly crop plants, while relatively few have been carried out with forest trees, and even fewer have examined the effect of N availability on molecular defense responses in lodgepole pine. Here, we have used a transcriptomics approach to test the hypotheses that N availability: (1) affects well-characterized components of lodgepole pine defense against *G. clavigera,* such as monoterpene synthesis, (2) modulates expression of genes thought to be important in mediating *G. clavigera*-elicited responses in lodgepole pine, and (3) alters the ratio of N-based to C-based defense-related genes that are up-regulated in response to *G. clavigera* inoculation, with a greater proportion of N-based defense genes up-regulated in response to higher N availability. To this end, ribonucleic acid sequencing (RNA-Seq) was carried out on secondary xylem (wood) or secondary phloem (bark) harvested from lodgepole pine seedlings fertilized with either 0.3 mM or 10 mM ammonium nitrate (NH_4NO_3) and subjected to either mock (wound) or *G. clavigera* (fungus) inoculation, using a full factorial experimental design. RNA-Seq data were then mined using a variety of bioinformatics approaches, including Venn diagrams (Kestler *et al.* 2004), functional category enrichment (Thimm *et al.* 2004; Young *et al.* 2010) and weighted correlation network analyses (Langfelder and Horvath 2008), to assess whether differential gene expression patterns provided support for any of the N-defense theories outlined above.

4.2 Materials and methods

4.2.1 Plant material and experimental design

A growth chamber experiment, described in detail in Section 2.2.1, was performed to test the responses of lodgepole pine to *G. clavigera* inoculation under three levels of N fertilization. One-year old lodgepole pine seedlings supplied by PRT Armstrong Nursery (Armstrong, British Columbia, Canada) were transplanted into 3.78 L pots and grown in a complete randomized block design under 19 °C constant temperature, 35% relative humidity, 15 h day / 9 h night photoperiod, and approximately 200-250 μmol photosynthetically active radiation (PAR) light intensity. Given that light levels in greenhouses and outdoors tend to be higher than the levels used for our study, our trees were most likely light-limited. One week after repotting, 0.5 g/L 15:30:15 (N:P:K) fertilizer solution (Plant Products Ltd, Brampton, Ontario, Canada) was applied weekly to seedlings until soil saturation for two weeks in total. This fertilizer application was followed by two weeks of 0.5 g/L 20:20:20 (N:P:K) fertilizer solution (Plant Products Ltd, Brampton, Ontario, Canada) for another two weeks. At six weeks after repotting, seedlings were fertilized weekly with Hocking's complete nutrient solution (Hocking 1971) containing one of three levels of NH_4NO_3, 0 mM (no N), 0.3 mM (low N) or 10 mM (high N), as part of the experiment's multifactorial design. Fertilization treatments continued until the conclusion of the experiment. Eleven weeks after repotting, wounding plus water (mock-inoculated) and wounding plus *G. clavigera* inoculation treatments were applied as described by Arango-Velez *et al.* (2016) with minor modifications. A spore suspension (~140 spores μL^{-1}) of *G. clavigera* isolate M001-03-03-07-UC04DL09, initially described by Roe *et al.*

(2010, 2011), was prepared according to Arango-Velez *et al*. (2016). Trees were wounded three times at 4-5 cm intervals along both the first and second season's growth using a syringe needle (23G1 PrecisionGlide, Becton, Dickinson and Company, Mississauga, Ontario, Canada). Fungal-inoculated trees received 10 µL of *G. clavigera* spores suspended in Milli-Q water (Merck Millipore, Burlington, Massachusetts, United States) in each wound. The mock-inoculated group received 10 µL Milli-Q water without the fungal inoculum. Foliar, phloem and xylem tissues from unwounded, wounded plus mock-inoculated and wounded plus fungal-inoculated trees were collected at 0, 1, 7, 14 and 28 days post inoculation (dpi). They were immediately flash frozen with liquid N and stored at -80°C until processing.

4.2.2 Monoterpene quantification and statistical analysis

Volatile collection

Volatile organic compounds (VOCs) released from seedlings were collected by Inka Lusebrink at 0, 1, 7, 14, and 28 dpi as previously reported by Lusebrink *et al*. (2011). In brief, trees were encased in an oven bag and a Porapak Q absorbent tube (Agilent Technologies, Santa Clara, California, United States) was inserted in the bag at the top of the seedling, affixed with Parafilm® (Bemis Company, Inc., Neenah, Wisconsin, United States). Volatile emissions were collected for 4 hours at a constant flow rate of 200 mL/min. After collection, the absorbent tubes were

capped and stored at -40 °C until extraction. Porapak Q tubes were extracted with 1 mL of dichloromethane spiked with 0.01% (v/v) tridecane as surrogate standard by Mehvash Malik and subsequently stored at -40 °C before gas chromatography/mass spectrometry (GC/MS) analysis.

Extracted VOC samples (1 µL) were injected in an Agilent 7890A/5062C Gas Chromatograph/Mass spectrometer (Agilent Technologies, Santa Clara, California, United States) with the splitless injection port set at 250. Helium was used as a carrier gas with a flow of 1 mL/min through an HP Innowax column (I.D. 0.25 mm, length 30 m; Agilent Technologies, Santa Clara, California, United States). The oven temperature program was set at 50 °C for 2 minutes, increased by 1 °C/min to 60 °C for 1 minute and then ramped up to 250 °C by 20 °C increments.

Peaks were identified using the following standards: γ-terpinene (Sigma-Aldrich, St. Louis, Missouri, United States), 3-carene, terpinolene, (-)-α-pinene, (-)-β-pinene, (*S*)-(-)-limonene, myrcene, camphene, *p*-cymene (Fluka, Sigma-Aldrich, Buchs, Switzerland), bornyl acetate, *cis*-ocimene, α-phellandrene (SAFC Supply Solutions, St. Louis, Missouri, United States) and β-phellandrene (Glidco, Inc., Jacksonville, Florida, United States). Calibration with these standards allowed for analysis of quantitative differences of volatile samples among treatments.

Tissue Extracts

Seedling were harvested at 0, 7, 14 and 28 dpi. Secondary phloem and foliar samples were rapidly frozen in liquid N and stored at -80 °C prior to extraction. Tissue extraction performed by Mehvash Malik followed the protocol of Lusebrink *et al.* (2011). Extracts were stored in amber GC vials at -40 °C before GC/MS analysis.

Extracts of tissue samples (3 µL) were injected at a split ratio of 20:1 in an Agilent 7890A/5062C Gas Chromatograph/Mass spectrometer (Agilent Technologies, Santa Clara, California, United States) with an HP-Chiral-20B column (I.D. 0.25 mm, length 30 m; Agilent Technologies, Santa Clara, California, United States), helium carrier gas flow at 1.1 mL/min and oven temperature at 75 °C for 15 min, increased to 230 °C by 5 °C increments. Calibration was performed with the standards used in VOC GC/MS-analysis along with (+)-α-pinene, (+)-β-pinene, and (*R*)-(+)-limonene (Fluka, Sigma-Aldrich, Buchs, Switzerland), which allowed for the quantification of tissue chemical content as well as the analysis of differences in stereoisomer composition of the differently treated seedlings.

Statistical analysis

In order to address how seedling volatile emission and monoterpene production changed due to the N, inoculation and day treatments, a multivariate analysis of variance (MANOVA; Pillai 2014) was performed by Inka Lusebrink

using the R statistical language (R Core Team 2017). To compliment the MANOVA, redundancy analysis (RDA; Legendre and Legendre 1998) was carried out on the monoterpene data using the vegan v2.5-6 R package (Oksanen *et al.* 2019). RDA extracted and summarized the variation between monoterpene production and volatile emission that was explained by each treatment combination. The results were visualized using a principal component analysis (PCA) biplot generated with vegan. For both analyses, monoterpene data were transformed by log(x+1) and analyzed for all tissues separately.

4.2.3 Foliar N content quantification and N concentration statistical analysis

Foliar samples collected 1, 7, 14 and 28 dpi were chopped and lyophilized with the VirTis Freezemobile FM25-XL freeze drier (SP Scientific, Inc., Warminster, Pennsylvania, United States) and ground with the Mixer Mill 301 (Retsch GmbH, Haan, Germany) by Ekaterina Stolnikova to obtain a fine powder. Three mg of each sample were used for dry combustion at the Biogeochemical Analytical Service Laboratory at the University of Alberta (Edmonton, Alberta, Canada), where total C and N content were detected by thermal conductivity using the CE-440 Elemental Analyzer (Exeter Analytical, Inc., North Chelmsford, Massachusetts, United States) and the US EPA Test Method 440.0 protocol (Zimmerman *et al.* 1997). The positive control, which also served as the calibration

standard, was 99.9% acetanilide. The negative control was no sample or acetanilide.

Foliar N concentration was calculated by dividing foliar N content by dry weight of the needles, and the N concentration data were analyzed with the R packages outliers v0.14 (Komsta 2011), car v3.0-7 (Fox and Weisberg 2019), multcompView v0.1-8 (Graves *et al.* 2019) and emmeans v1.4.5 (Lenth 2020) in R v3.6.1 (R Core Team 2017; RStudio team 2015). Data were tested for normality using Shapiro-Wilk's test (Shapiro and Wilk 1965) and homogeneity of variance using Levene's test (Levene 1960). Statistical significance was tested using a three-way analysis of variance (ANOVA; Fisher 1934) with a significance threshold of 0.05. Tukey's multiple comparison test (Tukey 1949) was leveraged to detect significant differences in means at an α value of 0.05.

4.2.4 Stem tissue processing and RNA extraction

RNA-Seq was carried out on 32 tissue samples: 4 biological replicates per each of 8 treatments = 1 day (7 dpi) × 2 tissues (secondary phloem or secondary xylem) × 2 levels of N availability (0.3 mM or 10 mM NH_4NO_3) × 2 inoculation treatments (mock- or fungal-inoculated). Secondary phloem and secondary xylem were ground separately using a Geno/Grinder 2010 (SPEX SamplePrep, Metuchen, New Jersey, United States) or by hand using a mortar and pestle, as described in Section 2.2.2. Total ribonucleic acid (RNA) extractions were

performed according to Pavy *et al.* (2008), as detailed in Section 2.2.3. RNA quantity was checked with the NanoQuant Spectrophotometer (Tecan, Männedorf, Switzerland) following the manufacturer's instructions, and quality was checked with the Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, California, United States) following manufacturer's instructions. High quantity was specified as an RNA concentration between 700 and 800 ng/μL. High quality was specified as an A260/A280 absorbance ratio between 2 and 2.2 and an RNA integrity number above 7.5.

4.2.5 cDNA library preparation and NGS

Complementary deoxyribonucleic acid (cDNA) libraries were constructed from the 32 RNA samples described above using the Illumina TruSeq Stranded mRNA Low Sample (LS) Protocol (Illumina, Inc. 2013). This process is described in detail in Section 2.2.4. Note that deoxynucleotide uridine triphosphate (dUTP) was used rather than deoxynucleotide thymidine triphosphate (dTTP) when generating the second strand of cDNA *in vivo* so that only the antisense strand (corresponding to the gene) would be amplified later. This method produced "stranded" cDNA where the orientation of the expressed gene was maintained. The 63 base pair (bp) TruSeq Indexed Adapters (Illumina, Inc., San Diego, California, United States) were then ligated to the 5' ends of the double-stranded cDNA. Polymerase chain reaction (PCR) followed so that the cDNA strands were amplified in anticipation of sequencing. Libraries were validated for quality using

the Agilent DNA 1000 kit (Agilent Technologies, Inc., Santa Clara, California, United States).

The 32 cDNA libraries were prepared using the NextSeq System Denature and Dilute Libraries Protocol (Illumina, Inc. 2015) and were combined into four pools, with each pool containing one biological replicate for each of the eight treatments. The Illumina NextSeq 500 platform (Illumina, Inc., San Diego, California, United States) was used for deep sequencing of each pool from both 5′ and 3′ ends for 151 bp reads following the manufacturer's system guide (Illumina, Inc. 2018) at the Molecular Biology Service Unit of the University of Alberta's Department of Biological Sciences (Edmonton, Alberta, Canada). The sequencing process is detailed in Section 2.2.5. Sequence data were converted to FASTQ files (Cock *et al.* 2010) using the Illumina BaseSpace Sequence Hub (Illumina, Inc., San Diego, California, United States). The raw Illumina 151 bp pair-end sequences were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive under accession number PRJNA524866.

4.2.6 Production of an annotated reference transcriptome

In the absence of a sequenced genome for lodgepole pine, a master transcriptome was constructed by Dr. Rhiannon Peery for use as a reference in RNA-Seq analysis (Peery *et al.* submitted). The lodgepole pine master transcriptome was deposited in the NCBI BioProject archive under accession

number PRJNA524866. Reads from the 32 sequenced cDNA libraries described above and an additional 16 lodgepole pine seedling libraries from a separate experiment were trimmed using trimmomatic v0.36 (Bolger *et al.* 2014) and used to create a *de novo* assembly with Trans-ABySS v1.5.5 (Robertson *et al.* 2010) following the steps outlined in the flowchart depicted in Figure 4.1. The additional 16 libraries included in the master transcriptome were constructed from stems of 2- to 3-month-old lodgepole pine seedlings, mock-inoculated or inoculated with *Cronartium harknessii* (E. Meinecke), the fungus causing western gall rust, and collected at either 1 week or 3 weeks post inoculation.

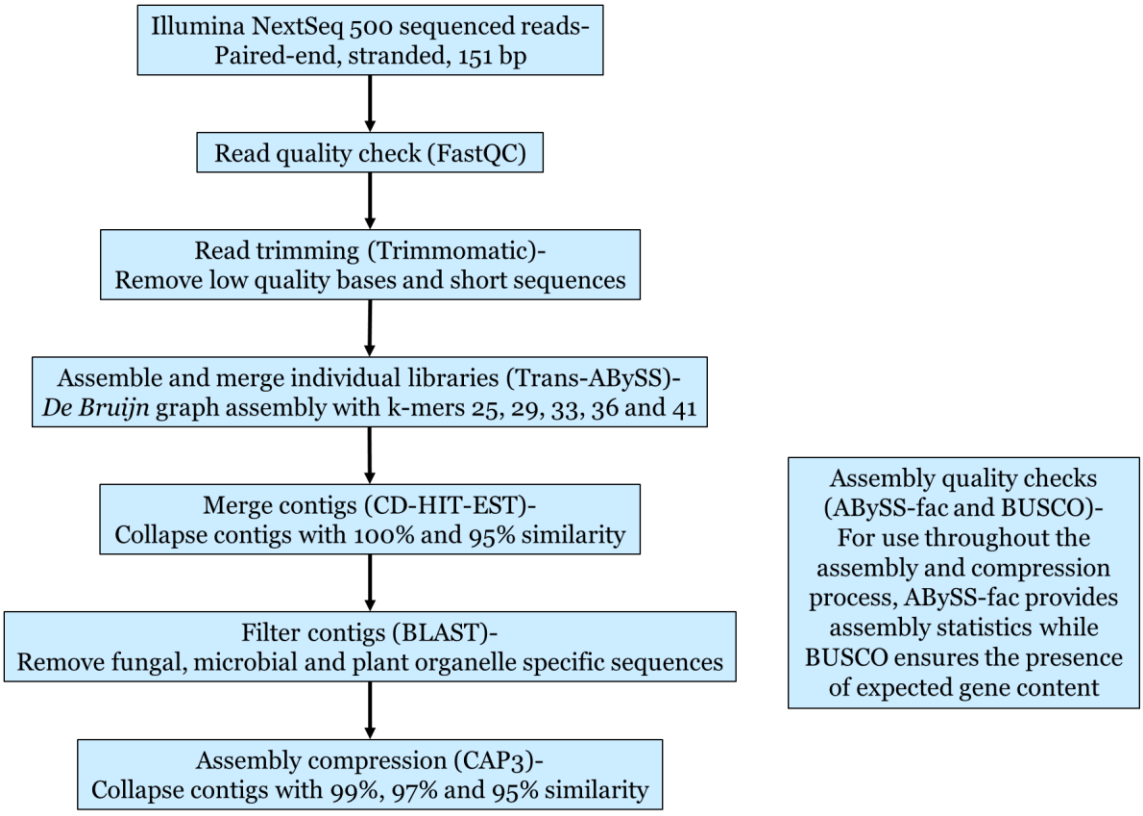


Figure 4.1. *De novo* assembly of a lodgepole pine master transcriptome. 151 base pair paired-end reads were assessed for quality, trimmed, and assembled *de novo* into contiguous sequences (contigs). Compression techniques based on sequence similarity collapsed allelic variants in the *de novo* assembly to reduce redundancy. Filtering for fungal and microbial sequences helped ensure a cleaner assembly composed of pine sequences only. Open source software implemented in this pipeline include trimmomatic (Bolger *et al.* 2014), Trans-ABySS (Robertson *et al.* 2010), CD-HIT-EST (Huang *et al.* 2010), BLAST (Altschul *et al.* 1990) and CAP3 (Huang *et al.* 1999). Quality checks were performed through-out using fastQC (Andrews 2014), the abyss-fac function (Robertson *et al.* 2010) and BUSCO (Simão *et al.* 2015). This pipeline was developed and implemented by Dr. Rhiannon Peery.

Contiguous sequences (contigs) were annotated against a custom in-house database that included all annotated conifer transcriptome data on Dendrome (Wegrzyn *et al.* 2008), ConGenie (Nystedt *et al.* 2013) and the conifer expressed sequence tag (EST) database in GenBank (Benson *et al.* 2011). TransDecoder v5.0.2 (Haas *et al.* 2013) was used to detect the longest open reading frames (ORFs). For contigs that were found to have more than one ORF, the TransDecoder output file included duplicate contig identifiers (IDs). For downstream analyses, the contig IDs with the smallest p-values following differential expression analysis were selected to eliminate duplicates.

The TransDecoder output file was modified by Rhiannon Peery to generate the general transfer format (gtf) file required by *featureCounts* (Rsubread R package v1.24.2; Liao *et al.* 2014) for transcript abundance enumeration (Shi and Liao 2016). ORFs were assigned the appropriate annotations using the customized conifer database. Contigs were additionally annotated using NCBI RefSeq non-redundant protein database (Pruitt *et al.* 2004) with a *Viridiplantae* GenInfo

Identifier (GI) list constraint (*https://www.ncbi.nlm.nih.gov /protein/?term=Viridiplantae*). Gene Ontology (GO) terms (The Gene Ontology Consortium 2015) and Interpro (Apweiler *et al.* 2001) classifications were assigned using the TRAPID tool (Van Bel *et al.* 2013). TAIR10 *Arabidopsis thaliana* protein sequence annotations (Lamesch *et al.* 2011) along with MapMan categories (Thimm *et al.* 2004) were provided by the Mercator tool (Lohse *et al.* 2014). MapMan categorized genes using the TAIR10 annotations in conjunction with Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa and Goto 2000) and GO databases (The Gene Ontology Consortium 2015) to create a mapping file that designated each gene as belonging to one of 31 possible categories (Thimm *et al.* 2004). The MapMan categories, along with all annotations mentioned above, were used when data mining genes in Section 4.2.10.

In this thesis, the term "gene" is used to refer to a distinct contig of the master transcriptome or reads from RNA-Seq corresponding to a contig. In the absence of a sequenced lodgepole pine genome, we recognize that some closely related or non-overlapping sequences likely represent the same gene. This means that there is some degree of redundancy in the master transcriptome and the ensuing RNA-Seq and differentially expressed (DE) gene set.

4.2.7 Read trimming and transcript abundance enumeration

Sequence processing for differential expression gene analyses followed a standard pipeline using both commercial and open-source software (Figure 4.2). Sequenced libraries were individually checked for quality and trimmed using the CLC Genomics Workbench v9.5.2 (*www.qiagenbioinformatics.com*). Optimized trim parameters included a limit of two adjacent ambiguous nucleotides, a minimum quality score cutoff of p=0.01 (Phred score=20) and a 51 and 151 bp minimum and maximum read length, respectively. The parameters were obtained according to Section 2.2.7 and described in Section 2.3.2. Trimmed reads were aligned to the master transcriptome by the CLC Genomics Workbench (*www.qiagenbioinformatics.com*), using all default mapping parameters, to create sequence alignment map (sam) files as described in Section 2.2.10.

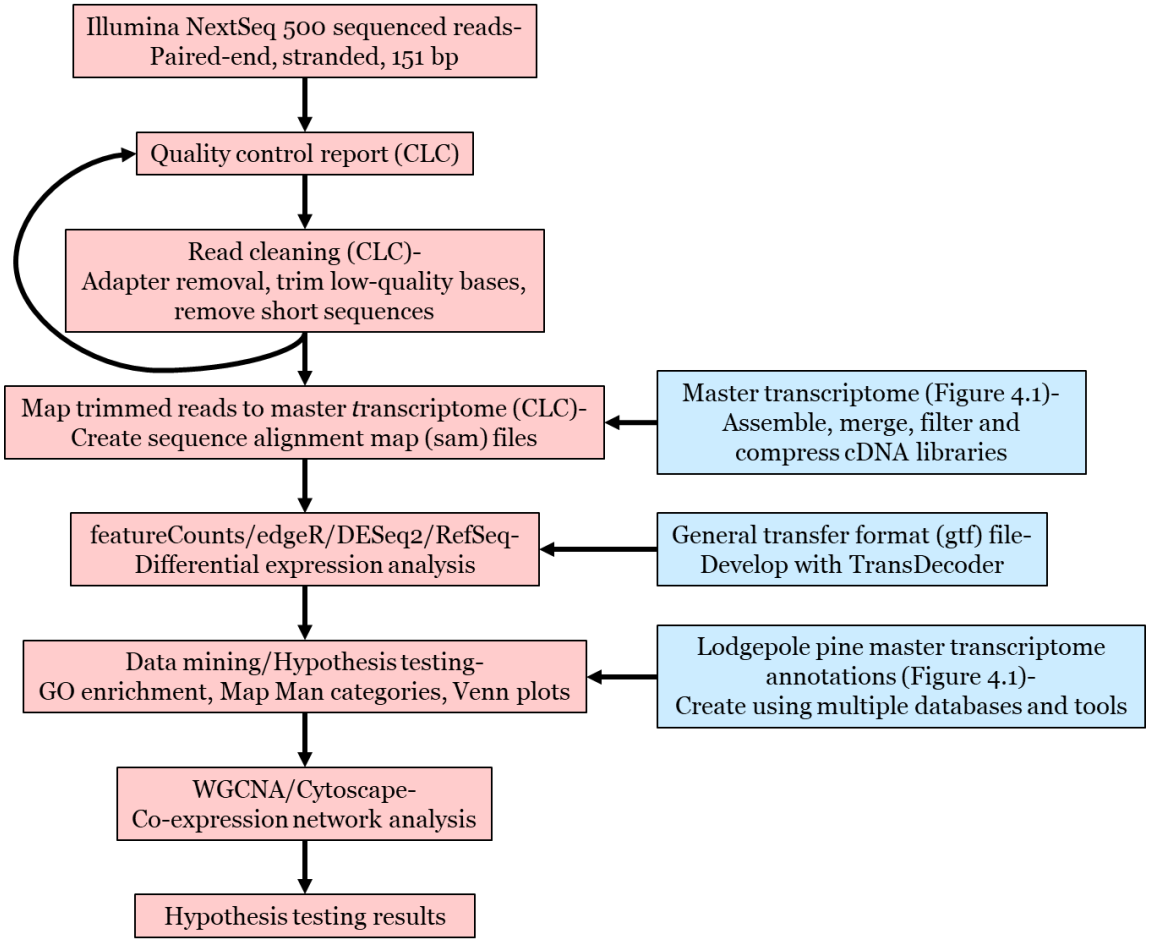


Figure 4.2. Differential expression analysis pipeline used with the CLC Genomics Workbench and open-source resources. The commercial software the CLC Genomics Workbench v9.5.2 (*www.qiagenbioinformatics.com*) along with the open-source bioinformatic tools, *featureCounts* from the Rsubread R package v1.24.2 (Liao *et al.* 2014), edgeR v3.20.9 (Robinson *et al.* 2010), DESeq2 v1.24.0 (Love *et al.* 2014), NCBI RefSeq non-redundant protein database (Pruitt *et al.* 2004), TransDecoder v5.0.2 (Haas *et al.* 2013), Gene Ontology (The Gene Ontology Consortium 2015), MapMan (Thimm *et al.* 2004), WGCNA v1.68 (Langfelder and Horvath 2008) and Cytoscape v3.7.2 (Shannon *et al.* 2003), were employed for differential expression analysis. Blue boxes correspond with the production of the annotated master transcriptome as detailed in Section 4.2.6.

The *featureCounts* tool from the Rsubread R package v1.24.2 (Liao *et al.* 2014) used the sam files and the gtf file developed from the master transcriptome

to produce raw count tables. The *featureCounts* script is included in Appendix B. The following criteria were used when counting the number of trimmed paired-end reads per library that mapped to the master transcriptome: mapping of both paired ends was not required for a read to be counted, the overlap of multiple reads was not permitted, and only the reads that had the largest overlap with the master transcriptome were chosen. Note that strand specificity was considered when counting the number of transcripts that aligned with the master transcriptome for all for all 32 sequenced cDNA libraries.

Expression values of RNA-Seq libraries were transformed prior to ordination using the DESeq2 v1.24.0 *varianceStabilizingTransformation* function with no specific parameters (Love *et al.* 2014), and RDA was performed on the normalized count data using the R package vegan v2.5-6 (Oksanen *et al.* 2019).

4.2.8 Differential expression analysis with edgeR

Count tables were used as input for edgeR v3.20.9 (Robinson *et al.* 2010), conducted in R v3.6.1 (R Core Team 2017; Figure 4.2). The edgeR script is included in Appendix B. Transcripts with low levels of abundance were removed from further analysis by applying a filter based on minimum read depth. Transcripts were retained for analysis when one count(s) per million (cpm) reads were present in at least four libraries (Anders *et al.* 2013). A more rigorous filter of two cpm was

used when combining multiple treatment groups for expression comparisons. Count normalization was carried out using the trimmed mean method (Robinson and Oshlak 2010). Normalized count data were fit into a gene-wise negative binomial generalized linear model following dispersion assessment in edgeR. A general linear model allowed for the combination of multiple control groups and experimental groups when detecting differential expression. A false discovery rate was applied using an adjusted p-value cutoff of 0.01 with the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995). The edgeR output was a list of DE sequences and accompanying statistics, such as the fold changes in expression levels on a log base two scale. A gene's fold change was the difference in the normalized experimental group expression value and the normalized control group expression value, divided by the normalized control group expression value. Note that each differential expression contrast resulted in a separate output.

4.2.9 Differential expression analysis with DESeq2

Count tables were also used as input for DESeq2 v1.24.0 (Love *et al.* 2014), conducted in R v3.6.1 (R Core Team 2017). The DESeq2 script is included in Appendix B. DESeq2 calculated size factors for each sample using the median ratio method described by Equation five in Anders and Huber (2010). It then fit the count data into a negative binomial general linearized model following dispersion assessment. P-values were calculated using the likelihood ratio test to identify significantly DEGs and were adjusted using the Benjamini-Hochberg procedure

(Benjamini and Hochberg 1995). The DESeq2 output was a list of DE sequences with accompanying statistics, such as the fold changes in expression levels on a log base two scale and the adjusted p-values for significance testing. Each differential expression contrast resulted in a separate output.

4.2.10 Data mining DEGs with defense-related categories

Statistically significant DEGs were determined using an adjusted p-value<0.01 and an absolute log base two fold change (log_2FC) of at least 2 ($|log_2FC| \geq 2$) for both edgeR and DESeq2 analyses. Only those genes that were identified as significantly DE in at least one differential expression contrast using both edgeR and DESeq2 were used for further analyses. A Venn diagram was produced using Venny v2.1 (Oliveros 2015) to discover lists of shared and distinct significantly DEGs between treatment groups. Duplicate contig IDs were removed prior to completing the Venn diagram analysis. Furthermore, heatmaps were produced using the gplots v3.0.3 R-package (Warnes *et al.* 2020) to compare gene expression patterns amongst genes of interest. Terpene synthase and chitinase gene annotations were clarified using BLAST (Altschul *et al.* 1990) prior to heatmap construction. Phylogenetic analysis was not utilized to classify chitinase genes so as to limit the annotation methods to those applied to the annotation of the master transcriptome, described in Section 4.2.6.

When annotations were assigned to the master transcriptome, a master annotation file was produced. This file was composed of the following columns: contig IDs, Interpro classifications, GO terms, the RefSeq annotation with the lowest e-value, MapMan category with accompanying annotations and pathway information, TAIR annotation and finally the annotation assigned by the customized in-house conifer database described in Section 4.2.6. Data tables of significantly DEGs with accompanying edgeR and DESeq2 statistics were assigned annotations in R v3.6.1 (RStudio team 2015; R Core Team 2017) by merging the contig IDs in the differential expression data tables with the contig IDs in the master annotation file. These annotations were then utilized to assign the following categories to each DEG: defense-related enzymes synthesizing C-based products (CD), N-based defense proteins, including enzymes (ND), transcription factors (TF), other regulators of genetic activity (R), hormone-related genes (H), cell wall-related genes (CW), signaling genes (S), N transport-related genes (NT) and N metabolism-related genes (NM). Categories were further resolved into subcategories related to putative gene family or gene function and associated with the lodgepole pine defense response to *G. clavigera* inoculation (Table 4.1).

Table 4.1. Gene subcategories for data mining DEGs. Annotations of genes determined to be significantly DE by both edgeR and DESeq2 were used to assign genes to one or more of the following categories: defense-related enzymes synthesizing C-based products (CD), N-based defense proteins, including enzymes (ND), cell wall-related genes (CW), hormone-related genes (H), transcription factors (TF), regulators of genetic activity (R), signaling genes (S), N transport-related genes (NT) and N metabolism-related genes (NM). Subcategories

associated with the lodgepole pine defense response to *G. clavigera* inoculation were assigned to the annotated DEGs if the subcategory was included in at least one of the associated annotations. Subcategories were organized by category. Transcription factor genes that did not fall into one of the included subcategories were considered "other". Abbreviations include abscisic acid (ABA), basic helix-loop-helix (bHLH), basic leucine zipper domain (bZIP), ethylene response factor (ERF), jasmonate (zinc-finger expressed in inflorescence meristem)-domain (JAZ), myeloblastosis oncogenes (MYB), no apical meristem, *Arabidopsis* transcription activation factor and cup-shaped cotyledon transcription factor (NAC), transcription factor with a WRKY amino acid sequence deoxyribonucleic acid binding domain at the N-terminus (WRKY), deoxyribonucleic acid (DNA), ribonucleic acid (RNA), mitogen-activated protein kinase (MAPK), mildew resistance locus O (MLO), nucleotide binding adaptor shared by nucleotide-binding oligomerization domain with leucine-rich repeat proteins, apoptotic protease activating factor 1, resistance proteins and cell-death protein 4 (NB-ARC), nucleotide-binding site - leucine-rich repeat gene (NBS-LRR), resistance gene (R-gene), nitrate (NO_3^-) and ammonium (NH_4^+).

CD	ND	CW	H	TF
alkaloid	β-1,3-glucanase	degradation	ABA	bHLH
chalcone	chitinase	modification	auxin	bZIP
cytochrome P450	dirigent	proteins	brassinosteroids	ERF
flavonoid	osmotin-like	synthesis	cytokinin	JAZ
isoprenoid	thaumatin		ethylene	MYB
lignin	pathogenesis-related		gibberelin	NAC
phenol	peroxidase		jasmonic acid	WRKY
phenylpropanoid			salicylic acid	other
phytoalexin				
pinoresinol				
stilbene				
sulfur-containing				
terpenoid				

R	S	NT	NM
chromatin	histidine kinase	NO_3^- transport	NO_3^- metabolism
DNA-related	MAPK	NH_4^+ transport	NH_4^+ metabolism
histone	MLO		
reverse transcriptase	NB-ARC		
RNA-related	NBS-LRR		
	receptor kinase		
	respiratory burst		
	R-gene		
	signal transduction		

To achieve the categorization and subcategorization of DEGs, the following steps were taken: First, the annotated differential expression data tables, one for

each differential expression contrast, were downloaded as comma separated value (csv) files and opened in Microsoft Excel (2019). For each DEG data table, five new columns named “Category_1” through “Category_5” were created along with five new columns named “Subcategory_1” through “Subcategory_5”. The MapMan annotation column was then used to sort each data table. This column was chosen first because MapMan has a well curated hierarchical classification structure that involves the use of numbered bins and sub bins. Sorting the data using this column ensured that DEGs with similar annotations and functions were adjacent to one another, an important factor to help reduce typographical errors. A word search within the sorted MapMan column was then employed for each subcategory term listed in Table 4.1, going one category at a time. For example, a search was performed within the sorted MapMan annotation column for the term “alkaloid”. Every time “alkaloid” was detected, the term “alkaloid” was typed in the accompanying “Subcategory_1” cell and a “C” was typed in the accompanying “Category_1” cell. If more than one subcategory term was detected for a DEG in the MapMan category column, the annotation was reviewed further to ensure that the proper categorization and subcategorization occurred. This technique was then repeated for the following annotation columns: Interpro classifications (“Category_2”), the RefSeq annotation with the lowest e-value (“Category_3”), TAIR annotation (“Category_4”) and finally the annotation assigned by the customized in-house conifer database described in Section 4.2.6 (“Category_5”). It is very important to note that if a category or subcategory was already assigned

to a DEG, it was not repeated, and an "NA" was typed instead. The GO term column was not used for the categorization process because the tree graph topology of assigned GO terms often did not provide enough resolution to accurately categorize DEGs. Finally, the categories and subcategories were tallied for each differential expression data set (i.e. for each differential expression contrast), and the proportion of putative enzymes synthesizing C-based products relative to N-rich defense proteins, from this point called the CD:ND ratio, was calculated for all differential expression contrasts.

4.2.11 Correlated gene expression network analysis and module enrichment

Network construction using the Weighted Gene Correlation Network Analysis (WGCNA) v1.68 R package (Langfelder and Horvath 2008) was performed on the same RNA-Seq data used as the input for featureCounts described in Section 4.2.7. The authors of WGCNA (Langfelder and Horvath 2008) predicted that gene co-expression networks follow a scale-free power-law degree distribution as is seen in protein-protein interactions and metabolic networks (Maslov and Sneppen 2006). In order to achieve the scale-free topology assumption, WGCNA network construction is designed to be an unsupervised and unbiased method for clustering genes based on their expression profiles across all samples, rather than their differential expression patterns between samples

(Zhang and Horvath 2005). Therefore, all contigs were entered into the network analyses, those significantly DE and those not significantly DE.

Expression values were transformed using the DESeq2 *varianceStabilizingTransformation* function (Anders and Huber 2010) with no specific parameters, and hierarchical clustering based on average linkage was used to detect outlier samples. Pairwise correlation values between genes were calculated using biweight midcorrelation (Langfelder and Horvath 2012) with a 10% maximum outlier detection parameter, and correlation values were weighted by a soft-threshold power of 15 to achieve a scale-free network topology (Langfelder and Horvath 2008). A correlation dendrogram was generated by average linkage hierarchical clustering using the *dynamicTreeCut* function (Langfelder and Horvath 2008), and densely interconnected genes were clustered into modules using average sensitivity and p-value ratio of 1E-8 for reassigning genes to a closer module. Module eigengenes, the first principal component (PC1) of each module, were used to merge modules with a pairwise biweight midcorrelation value greater than 0.35 using the function *mergeCloseModules* with a dissimilarity cut-height of 0.65. The WGCNA (Langfelder and Horvath 2008) script is included in Appendix B.

Genes in the color-coded modules were annotated using MapMan hierarchical functional category bins (Thimm *et al.* 2004), and over-representation analysis was performed with a hypergeometric test. This analysis was designed to determine for each module the functional MapMan categories that

exhibited a significantly greater proportion of sequences relative to the number of sequences in that category in the total data set. Significant enrichment was determined by $p<0.05$. Sequences in the four largest modules were also assigned gene categories for gene-level comparisons. Differential expression results produced with edgeR as described above were used to compare gene expression patterns between the four largest modules.

4.2.12 Network visualization

Hub genes were identified using the WGCNA-assigned intramodular connectivity and module membership values. The intramodular connectivity measure was calculated by taking the sum of the pairwise biweight midcorrelation values of each gene to its module counterparts. Module membership was calculated by correlating each gene and its associated module eigengene using biweight midcorrelation (Langfelder and Horvath 2012). The module genes with the highest intramodular connectivity along with a sufficiently high module membership value (>0.80) were categorized as being hub genes. The WGCNA v1.68 (Langfelder and Horvath 2008) function *exportNetworkToCytoscape* was used to export the edge file for the most connected genes in the largest color-coded module, orange, using a network threshold of 0.96 out of 1.00. The edge file was imported into Cytoscape v3.7.2 (Shannon *et al.* 2003) for visualization with the graphical user interface. A subnetwork composed of all significantly DE orange module genes connected to four selected hub genes were extracted from the edge

file in R (R Core Team 2017) and visualized using Cytoscape v3.7.2 (Shannon *et al.* 2003).

4.3 Results

4.3.1 Foliar N concentration

Foliar N concentration is commonly used as a measure of a plant's N status (Muñoz-Huerta *et al.* 2013). Accordingly, total N was measured in foliage to determine whether the N concentration was significantly impacted by the fertilization treatments used in this study (Figure 4.3). Three-way ANOVA indicated that level of N fertilization was significant ($p<0.01$), as was the interaction of N fertilization level with dpi ($p<0.01$) and the effect of inoculation ($p=0.04$). However, post hoc comparisons by least squares means indicated that although total foliar N of 10 mM NH_4NO_3-treated seedlings was higher than that of the 0 and 0.3 mM NH_4NO_3-treated seedings within an inoculation treatment (control, wound plus mock inoculation, wound plus fungal inoculation) at all time points with two exceptions, the pairwise comparisons were only significant at 28 dpi.

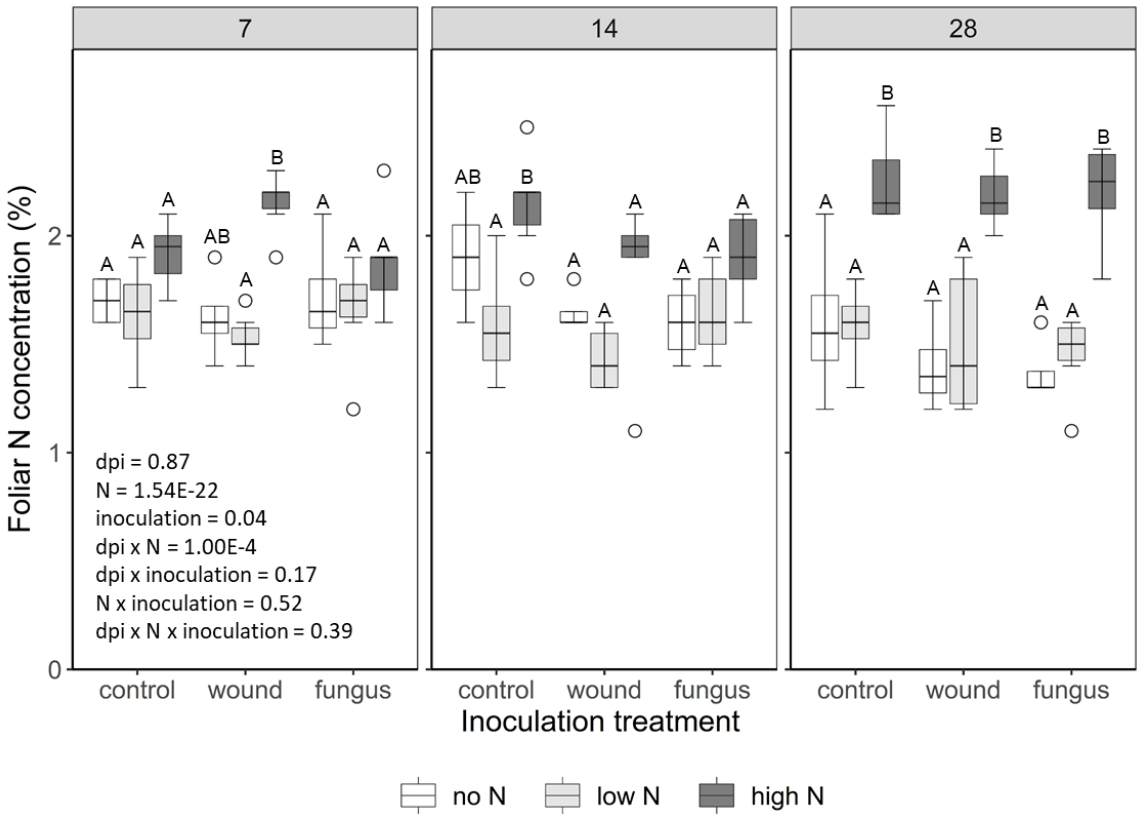


Figure 4.3. Lodgepole pine foliar N concentration as a function of fertilization and inoculation treatment Lodgepole pine seedlings treated with Hocking's complete nutrient solution supplemented with 0 mM (no N), 0.3 mM (low N) or 10 mM (high N) NH_4NO_3 were subjected to one of three inoculation treatments: no wound (control), mock-inoculated (wound), and *G. clavigera*-inoculated (fungus). N concentration (total N content/tissue dry weight) was determined for foliage collected at 7, 14 or 28 dpi, n=4-7. Three-way ANOVA p-values are listed in the bottom left of the boxplot. Letters above each box represent significantly different means between fertilization treatments for each inoculation type within each time point, α=0.05. There were no pair-wise significant differences between inoculation type for each fertilization treatment within each time point.

4.3.2 Monoterpene analysis

Monoterpenes are not only important for defense of pines against MPB and their fungal associates, they are also essential for MPB detection of appropriate,

high quality hosts (Raffa 2013). Tree monoterpenes serve as the substrate for MPB pheromone biosynthesis, several of which are part of the beetle's communication arsenal, vital to the mass attack strategy that MPB uses to overcome plant defenses to allow for successful tree colonization (Raffa 2013; Chiu *et al.* 2019). Accordingly, we quantified monoterpenes to provide a snapshot of whether N availability and/or *G. clavigera* inoculation modulated secondary phloem or needle monoterpene concentrations or of emitted (volatile) monoterpenes. The MANOVA revealed only a few statistically significant differences in monoterpene concentrations between control and *G. clavigera* treatments (Table 4.2). In phloem tissues, only β-phellandrene concentrations were significantly higher in *G. clavigera*-inoculated vs. control low N-treated seedlings, and only at 28 dpi. Needle tissues of 7 dpi low N-treated seedlings showed significant differences in β-phellandrene and (+)-α-pinene concentrations between *G. clavigera*-inoculated vs. control low N-treated seedlings, but both β-phellandrene and (+)-α-pinene concentrations were significantly lower in inoculated vs. control seedlings. Needles from high N-treated seedlings showed significantly decreased (-)-β-pinene concentrations between 28 dpi inoculated vs. control seedlings, and significantly increased camphene concentrations between 14 dpi inoculated vs. control seedlings. Volatile monoterpene profiles showed a greater number of significant differences, but only for low N-treated seedlings: both α-pinene and β-pinene exhibited significantly higher volatile concentrations for *G. clavigera* inoculated

vs. control seedlings at 14 and 28 dpi. Camphene also showed significantly greater volatile concentrations for *G. clavigera* vs. control seedlings at 28 dpi.

Table 4.2. Quantified monoterpene MANOVA results. Dpi refers to the number of days following inoculation that monoterpene samples were collected, and fertilization refers to the level of NH_4NO_3 (Low N or High N) used to fertilize the plants. MANOVA determined if the mean level of each monoterpene was significantly impacted by each dpi and fertilization level combination. Only p-values are shown, and the p-value significance threshold is 0.05. (↑) Monoterpene production was higher in fungal-inoculated seedlings compared to controls, (↓) monoterpene production was lower in fungal-inoculated seedlings compared to controls.

	Time point	Fertilization	Monoterpene	p-value	
Phloem	28 dpi	High N	β-phellandrene	0.037	↑
Needles	7 dpi	Low N	(+)-α-pinene	0.027	↓
			β-phellandrene	0.039	↓
	14 dpi	High N	camphene	0.027	↑
	28 dpi	High N	(-)-β-pinene	0.048	↓
Volatiles	14 dpi	Low N	α-pinene	0.048	↑
			β-pinene	0.038	↑
	28 dpi	Low N	α-pinene	0.004	↑
			β-pinene	0.01	↑
			camphene	0.032	↑

RDA was also used to examine monoterpene profiles in VOCs. RDA revealed some separation of control and *G. clavigera*-inoculated samples along PC1, with several monoterpenes contributing to the variance between these groups (Figure 4.4). Low N- and high N-treated samples showed slight separation along the second principal component (PC2), with α-pinene and β-pinene contributing to

these differences. RDA was also performed on phloem monoterpenes collected at 7 and 14 dpi (Figure 4.5). No obvious separation occurred along PC1 or PC2.

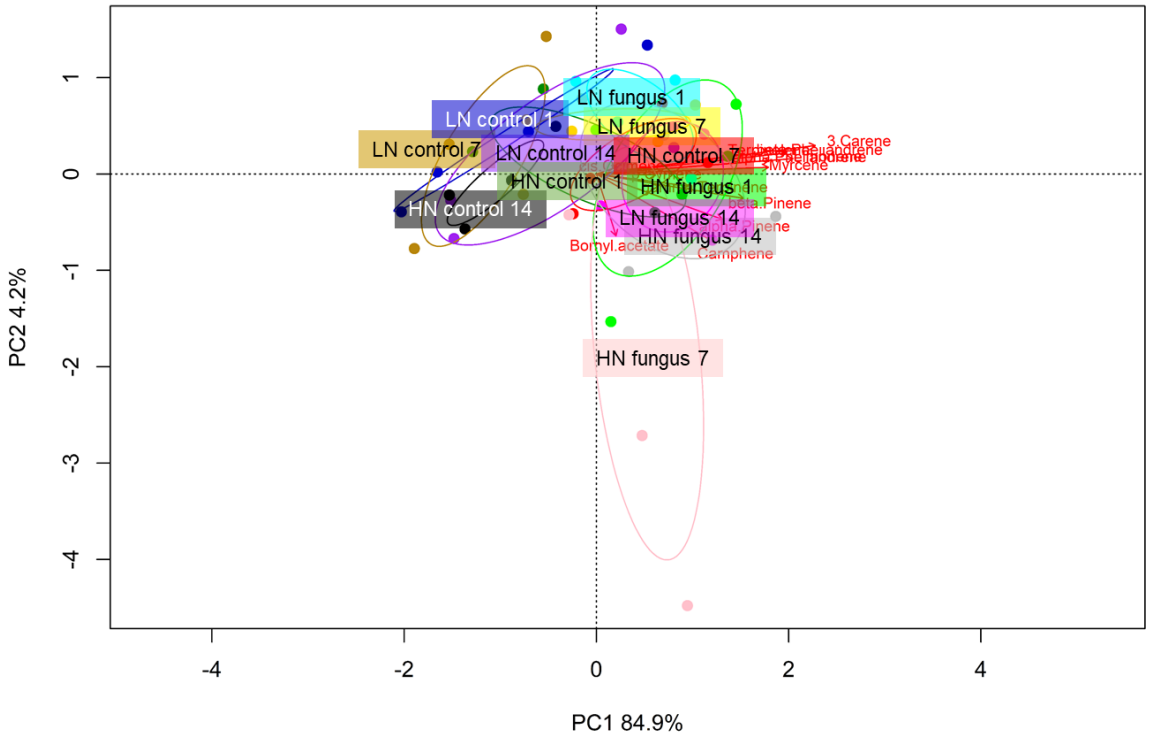


Figure 4.4. PCA of VOCs emitted from *G. clavigera*-inoculated or control low N- and high N-treated lodgepole pine showed some separation of fungal-inoculated and control samples along PC1. Lodgepole pine seedlings grown under low N (LN; 0.3 mM NH_4NO_3) or high N (HN; 10 mM NH_4NO_3) conditions were either inoculated with *G. clavigera* (fungus) or unwounded (control). VOCs collected at 7 or 14 dpi were analyzed by GC/MS, and the data analyzed by RDA with the R package vegan, n=4. The 95% confidence ellipses are shown, and labels are displayed in the center of each ellipse.

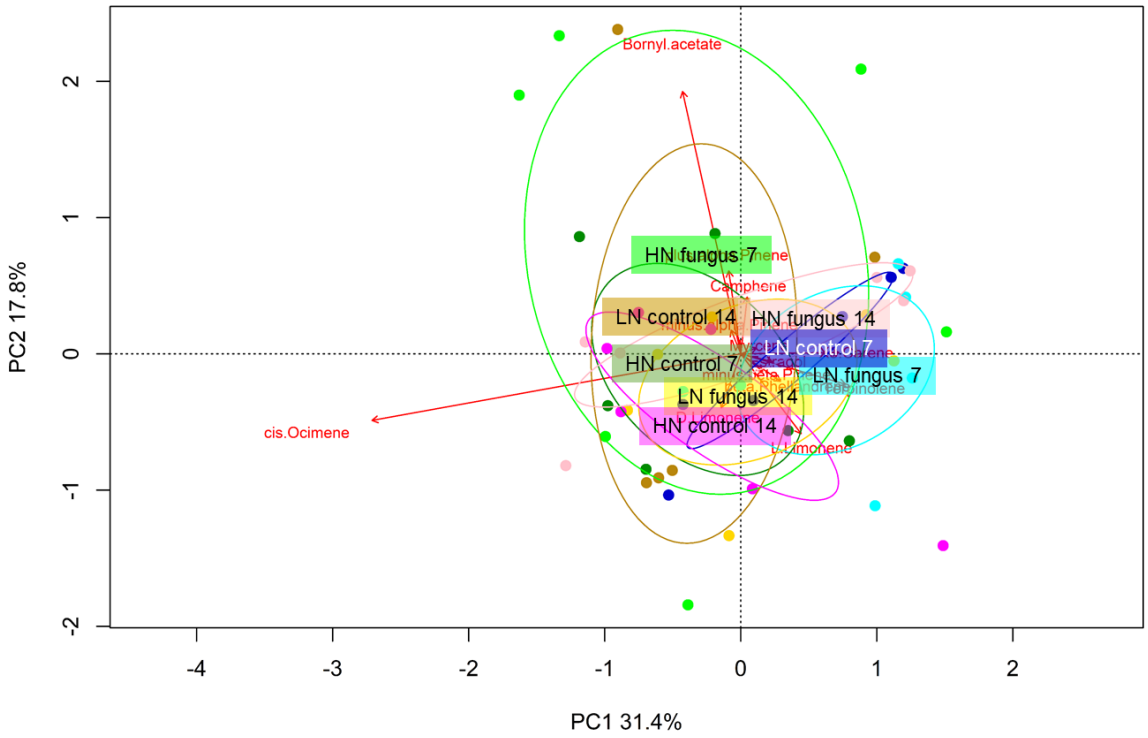


Figure 4.5. PCA of phloem monoterpenes from *G. clavigera*-inoculated or control low N- and high N-treated lodgepole pine revealed little separation along PC1 or PC2. Lodgepole pine seedlings grown under low N (LN; 0.3 mM NH_4NO_3) or high N (HN; 10 mM NH_4NO_3) conditions were either inoculated with *G. clavigera* (fungus) or unwounded (control). Phloem monoterpenes collected at 7 or 14 dpi were analyzed by GC/MS, and the data analyzed by RDA with the R package vegan, n=6. The 95% confidence ellipses are shown, and labels are displayed in the center of each ellipse.

4.3.3 Illumina sequence data processing and alignment to reference transcriptome

The 32 sequenced cDNA libraries produced ~12 GB of paired-end read data per library, yielding a combined data set of 1.51 TB when uncompressed. Sequencing statistics are detailed in Table 4.3. A complete table of library information and sequencing statistics can be found in Appendix A (Table 6.1).

Libraries had an average read depth of 147,736,562 ± 42,033,517 prior to processing, and 128,435,263 ± 36,785,104 reads post trimming. Optimized trim parameters, obtained according to Section 2.2.7 and described in Section 2.3.2, were used when preparing all 32 libraries for alignment to the master reference transcriptome for differential expression analysis. Reads had an average length of 128 bp across all libraries, while trimmed reads had an average read length of 118 bp across all libraries. A full table of trim results can be found in Appendix A (Table 6.4).

Table 4.3. Illumina sequencing statistics. The 32 cDNA libraries were sequenced using the Illumina NextSeq 500 system, described in detail in Section 2.2.5. Sequencing by synthesis generated 151 bp reads from each cluster of cloned deoxyribonucleic acid (DNA) fragments on the Illumina flow cell surface. The resulting FASTQ files were gzip compressed.

Sample	Number of reads	Number of clusters	Compressed size (GB)
LowN.Wound.2P.1	180,372,518	90,186,259	14.74
LowN.Wound.2P.2	37,755,380	18,877,690	3.16
LowN.Wound.2P.3	137,492,696	68,746,348	11.40
LowN.Wound.2P.4	217,216,226	108,608,113	19.08
LowN.Wound.2X.1	151,378,244	75,689,122	13.25
LowN.Wound.2X.2	164,866,450	82,433,225	13.78
LowN.Wound.2X.3	72,752,224	36,376,112	6.12
LowN.Wound.2X.4	167,470,418	83,735,209	13.96
LowN.Fungus.2P.1	184,711,540	92,355,770	16.30
LowN.Fungus.2P.2	131,092,962	65,546,481	11.37
LowN.Fungus.2P.3	182,365,750	91,182,875	15.75
LowN.Fungus.2P.4	101,208,170	50,604,085	8.11
LowN.Fungus.2X.1	90,826,876	45,413,438	7.18
LowN.Fungus.2X.2	191,671,236	95,835,618	16.68
LowN.Fungus.2X.3	155,327,508	77,663,754	13.59

Table 4.3. Illumina sequencing statistics. Continued.

Sample	Number of reads	Number of clusters	Compressed size (GB)
LowN.Fungus.2X.4	148,323,014	74,161,507	12.34
HighN.Wound.2P.1	141,089,824	70,544,912	12.12
HighN.Wound.2P.2	107,467,012	53,733,506	8.36
HighN.Wound.2P.3	174,066,564	87,033,282	14.36
HighN.Wound.2P.4	152,908,486	76,454,243	13.51
HighN.Wound.2X.1	150,927,944	75,463,972	13.44
HighN.Wound.2X.2	157,993,480	78,996,740	13.62
HighN.Wound.2X.3	64,955,740	32,477,870	4.91
HighN.Wound.2X.4	236,424,162	118,212,081	18.09
HighN.Fungus.2P.1	171,576,144	85,788,072	14.52
HighN.Fungus.2P.2	143,078,948	71,539,474	12.22
HighN.Fungus.2P.3	175,546,304	87,773,152	14.57
HighN.Fungus.2P.4	159,472,904	79,736,452	12.31
HighN.Fungus.2X.1	134,323,488	67,161,744	10.63
HighN.Fungus.2X.2	172,612,450	86,306,225	15.02
HighN.Fungus.2X.3	123,530,906	61,765,453	10.99
HighN.Fungus.2X.4	147,004,298	73,502,149	12.79

Trimmed reads were mapped to the master transcriptome using the CLC Genomics Workbench (*www.qiagenbioinformatics.com*). The master transcriptome, developed by Dr. Rhiannon Peery, contained 375,632 contigs, with a minimum contig length of 500 bp, an N50 length of 2001 bp and an average contig length of 719 bp (Peery *et al.* submitted). The average percentage of mapped reads per library was 90.66%, and the average percentage of mapped unbroken read pairs was 85.93%. A complete table of mapping results can be found in Appendix A (Table 6.5). The sam files were exported and used for differential expression analysis.

4.3.4 Transcript abundance enumeration

Coding sequence features were detected by TransDecoder (Haas *et al.* 2013) and 154,398 coding sequences (41.1% of contigs) were included in the gtf file utilized for transcript enumeration. The average percentage of reads assigned to features was 63.06% per library. The mean library coverage was 38,136,887 ± 11,33,132 bp and ranged from a minimum of 10,079,404 bp to a maximum of 66,940,510 bp. A full table of *featureCounts* transcript enumeration results can be found in Appendix A (Table 6.6). RDA carried out for xylem and phloem data sets separately demonstrated that for the most part biological replicates clustered closely together on PC1 and PC2 (Figure 4.6). PC1, which respectively explained 32.3% and 43.7% of the variance amongst phloem and xylem samples, separated mock-inoculated samples from *G. clavigera*-inoculated samples. PC2, explaining 9.1% of the variance amongst phloem samples, separated low N and high N phloem samples. Separation by N availability along PC2 was not evident for xylem samples.

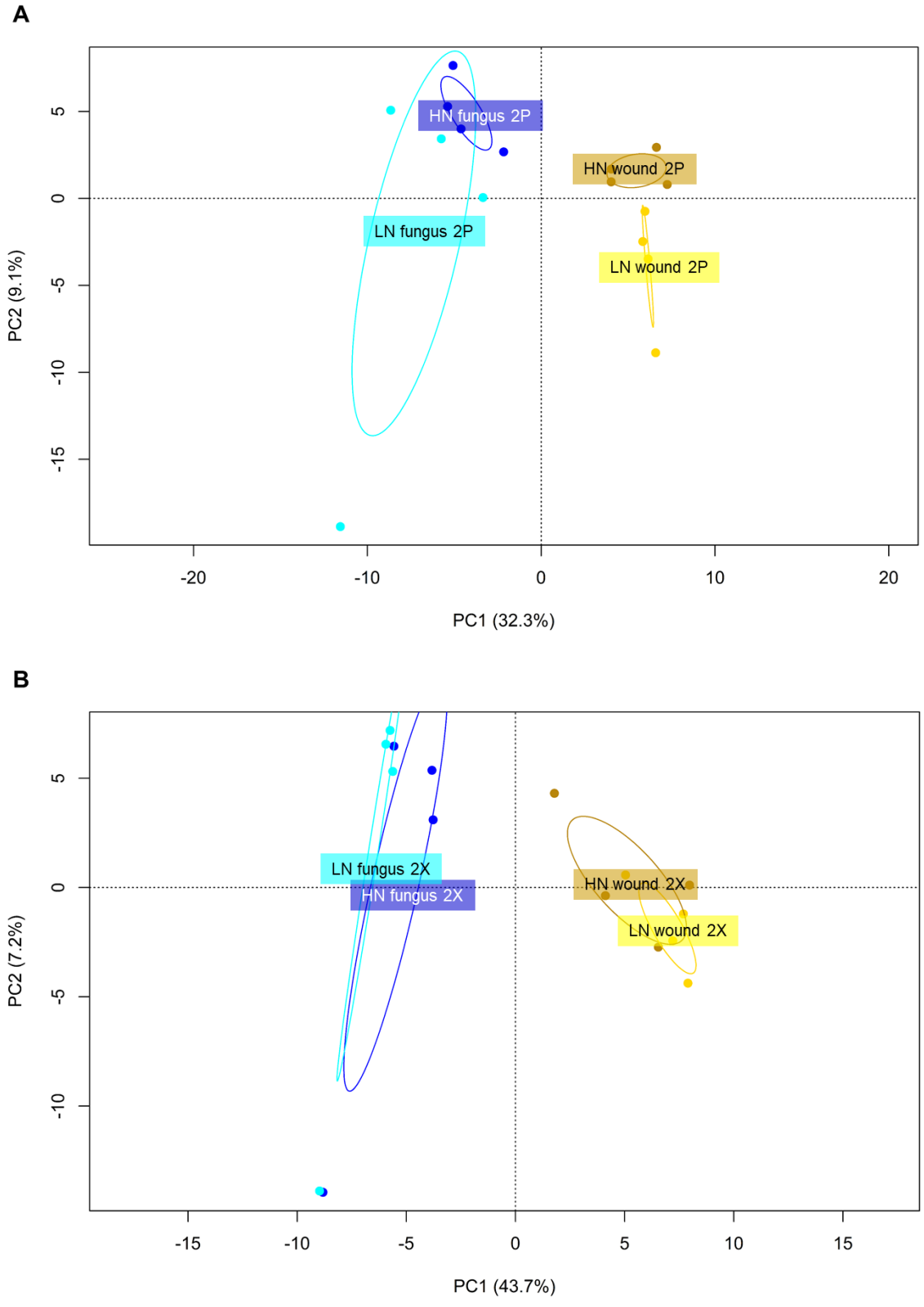

A
HN fungus 2P
HN wound 2P
LN fungus 2P
LN wound 2P
PC2 (9.1%)
PC1 (32.3%)
B
LN fungus 2X
HN fungus 2X
HN wound 2X
LN wound 2X
PC2 (7.2%)
PC1 (43.7%)

Figure 4.6. Two-dimensional variance in expression levels of RNA-Seq data following DESeq2 normalization. RNA-Seq data represent (**A**) secondary phloem (2P) and (**B**) secondary xylem (2X) harvested from lodgepole pine seedlings grown under 0.3 mM NH_4NO_3 (LN) or 10 mM NH_4NO_3 (HN) and either inoculated with *G. clavigera* (fungus) or mock-inoculated (wound). There were four biological replicates for each fertilization/inoculation treatment combination. Expression data produced by *featureCounts* was normalized using a DESeq2 variance transformation prior to ordination. RDA was performed using the R package vegan, and labels are displayed in the center of each 95% confidence ellipse.

4.3.5 Differential expression analysis with edgeR and DESeq2

A general linear model was applied by both edgeR (Robinson *et al.* 2010) and DESeq2 (Love *et al.* 2014) to make 12 different comparisons between the 12 treatments (2 tissues x 2 N levels x 3 inoculation treatments) used in this study (Table 4.4). The intersection of genes identified as significantly DE in at least one differential expression contrast at an adjusted p-value<0.01 by both edgeR and DESeq2 and filtered to include only those genes that showed$|\log_2 FC| \geq 2$ and also exhibited a transcript abundance of at least one cpm reads in at least four libraries will be referred to as the core DEGs. Volcano plots were used to visualize differential expression results (Figures 4.7-4.10).

Table 4.4. Differential expression contrasts for edgeR and DESeq2. Secondary phloem (2P) and secondary xylem (2X) were sampled from lodgepole pine seedlings grown under one of two nitrogen regimes (LowN or HighN) and either inoculated with *G. clavigera* (Fungus) or mock-inoculated (Wound). There were four biological replicates for each fertilization and inoculation treatment combination. Genes identified were significantly DE at an adjusted p-value<0.01 and filtered to include only those genes that showed an absolute log base two fold change of at least 2 ($|\log_2 FC| \geq 2$) and also exhibited a transcript abundance of at least one cpm reads in at least four libraries. The intersection of significantly DEGs detected by both methods are also tabulated and are termed the core DEGs.

	edgeR		DESeq2		Core DEGs	
Differential expression contrast	**Up regulated**	**Down regulated**	**Up regulated**	**Down regulated**	**Up regulated**	**Down regulated**
LowN 2P Fungus vs. Wound	3489	2290	4803	3134	2254	1027
LowN 2X Fungus vs. Wound	4483	2804	6303	4499	3114	1233
HighN 2P Fungus vs. Wound	1389	808	2887	2756	1034	400
HighN 2X Fungus vs. Wound	2716	1580	3498	2450	1843	736
Wound 2P HighN vs. LowN	9	3	143	179	7	0
Wound 2X HighN vs. LowN	75	22	122	161	21	17
Fungus 2P HighN vs. LowN	0	2	50	75	0	1
Fungus 2X HighN vs. LowN	1	0	10	10	0	0
LowN and HighN 2P Fungus vs. Wound	4193	3883	6271	5353	1989	823
LowN and HighN 2X Fungus vs. Wound	5900	4406	7118	6439	2601	999
Wound and Fungus 2P HighN vs. LowN	5	15	29	58	3	5
Wound and Fungus 2X HighN vs. LowN	21	3	14	5	6	0

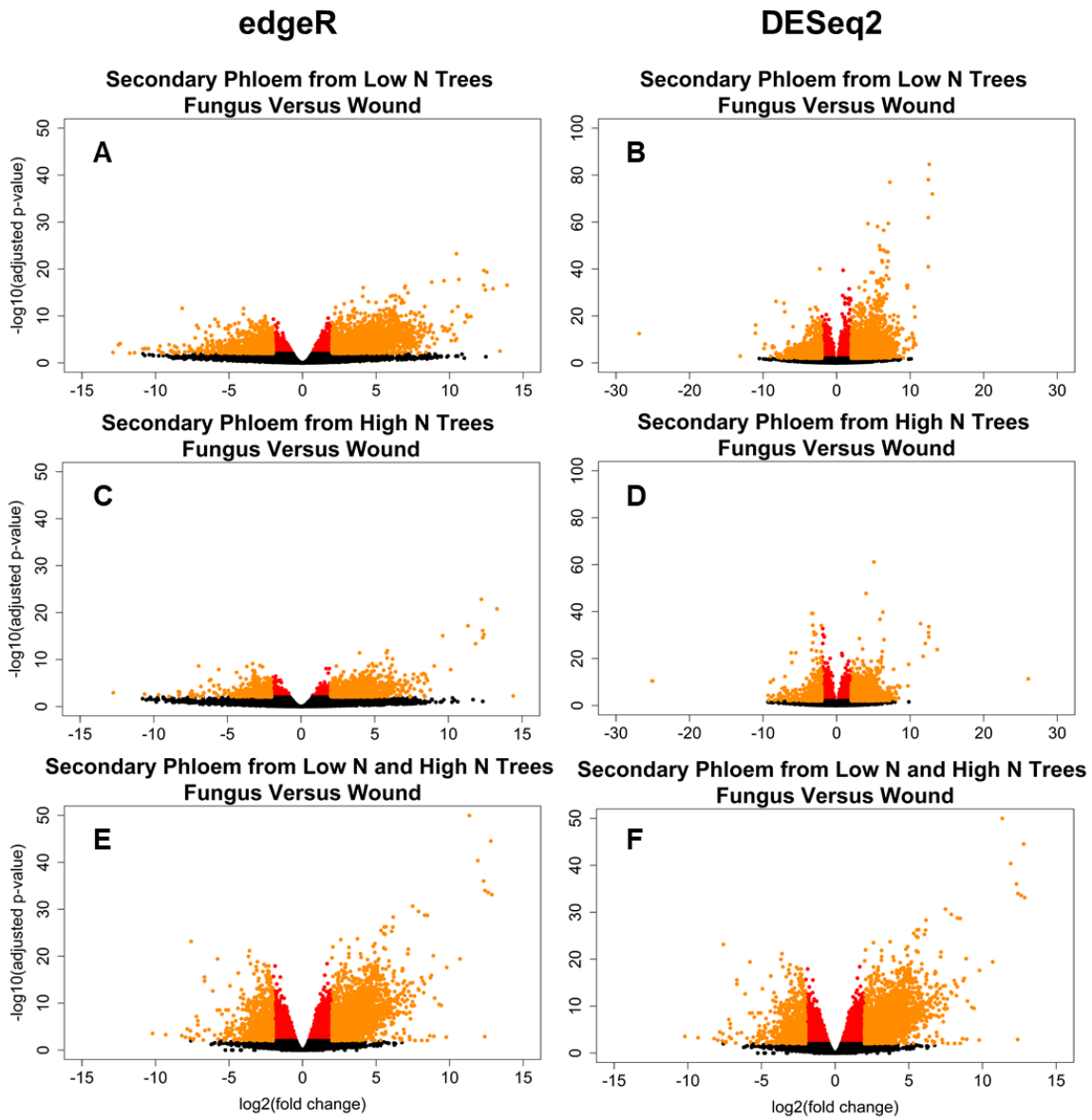


Figure 4.7. Volcano plots depicting DEGs determined by edgeR or DESeq2 in secondary phloem of *G. clavigera*- vs. mock-inoculated lodgepole pine grown under 0.3 mM or 10 mM NH_4NO_3. Fold change values were calculated with edgeR (**A**, **C**, **E**) or DESeq2 (**B**, **D**, **F**). Observed fold changes (log base two) are plotted against the Benjamini-Hochberg adjusted p-values (-log base 10) for each gene. Black points were not significantly DE. Red points were significantly DE (adjusted p-value<0.01). Orange points were significantly DE (adjusted p-value<0.01) and exhibited an absolute log base two fold change of at least 2 ($|\log_2 FC| \geq 2$). Data denoted by orange were used for subsequent analyses and data mining.

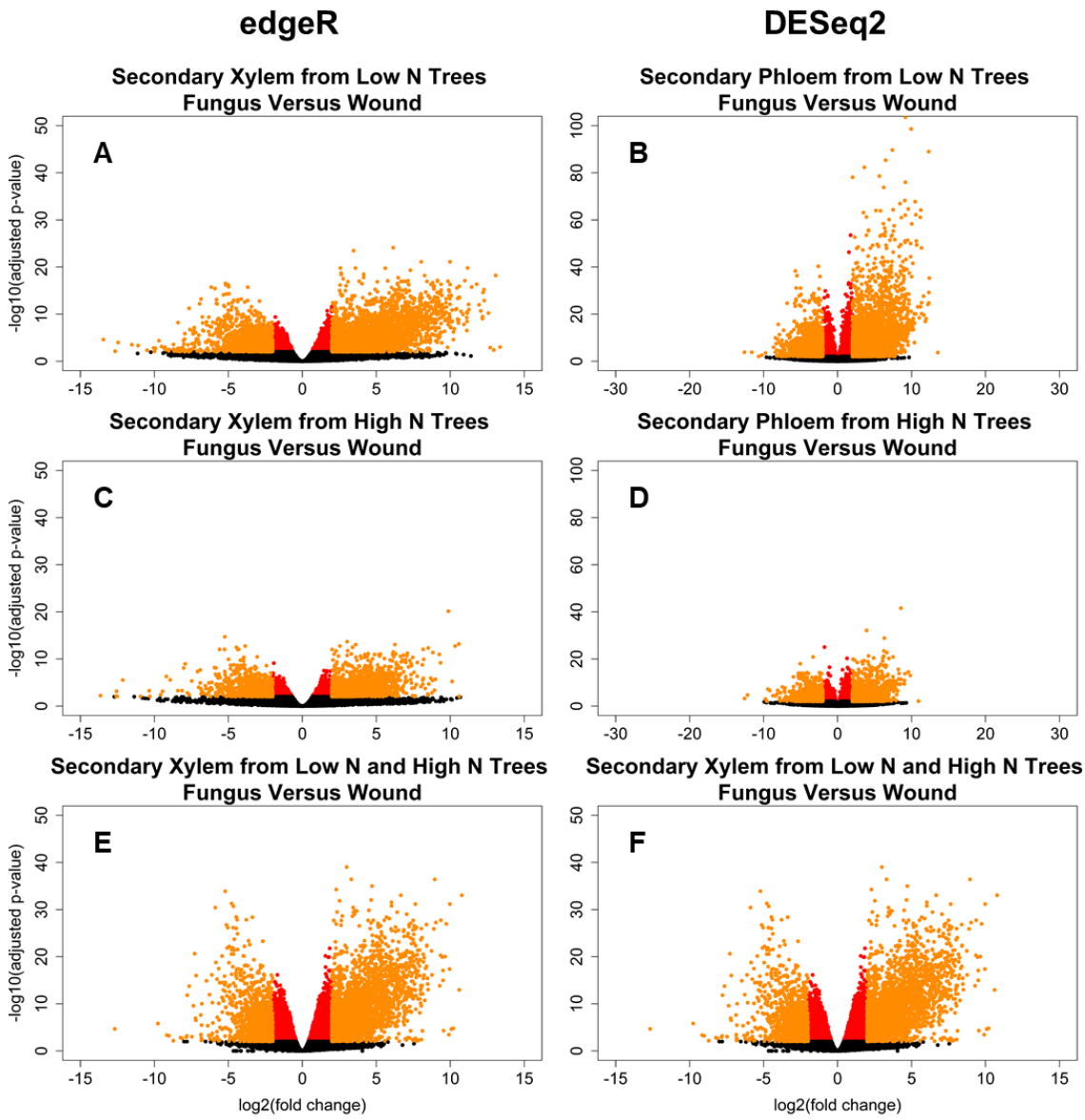


Figure 4.8. Volcano plots depicting DEGs determined by edgeR or DESeq2 in secondary xylem of *G. clavigera*- vs. mock-inoculated lodgepole pine grown under 0.3 mM or 10 mM NH_4NO_3. Fold change values were calculated with edgeR (**A**, **C**, **E**) or DESeq2 (**B**, **D**, **F**). Observed fold changes (log base two) are plotted against the Benjamini-Hochberg p-values (-log base 10) for each gene. Black points were not significantly DE. Red points were significantly DE (adjusted p-value<0.01). Orange points were significantly DE (adjusted p-value<0.01) and exhibited an absolute log base two fold change of at least 2 ($|\log_2 FC| \geq 2$). Data denoted by orange were used for subsequent analyses and data mining.

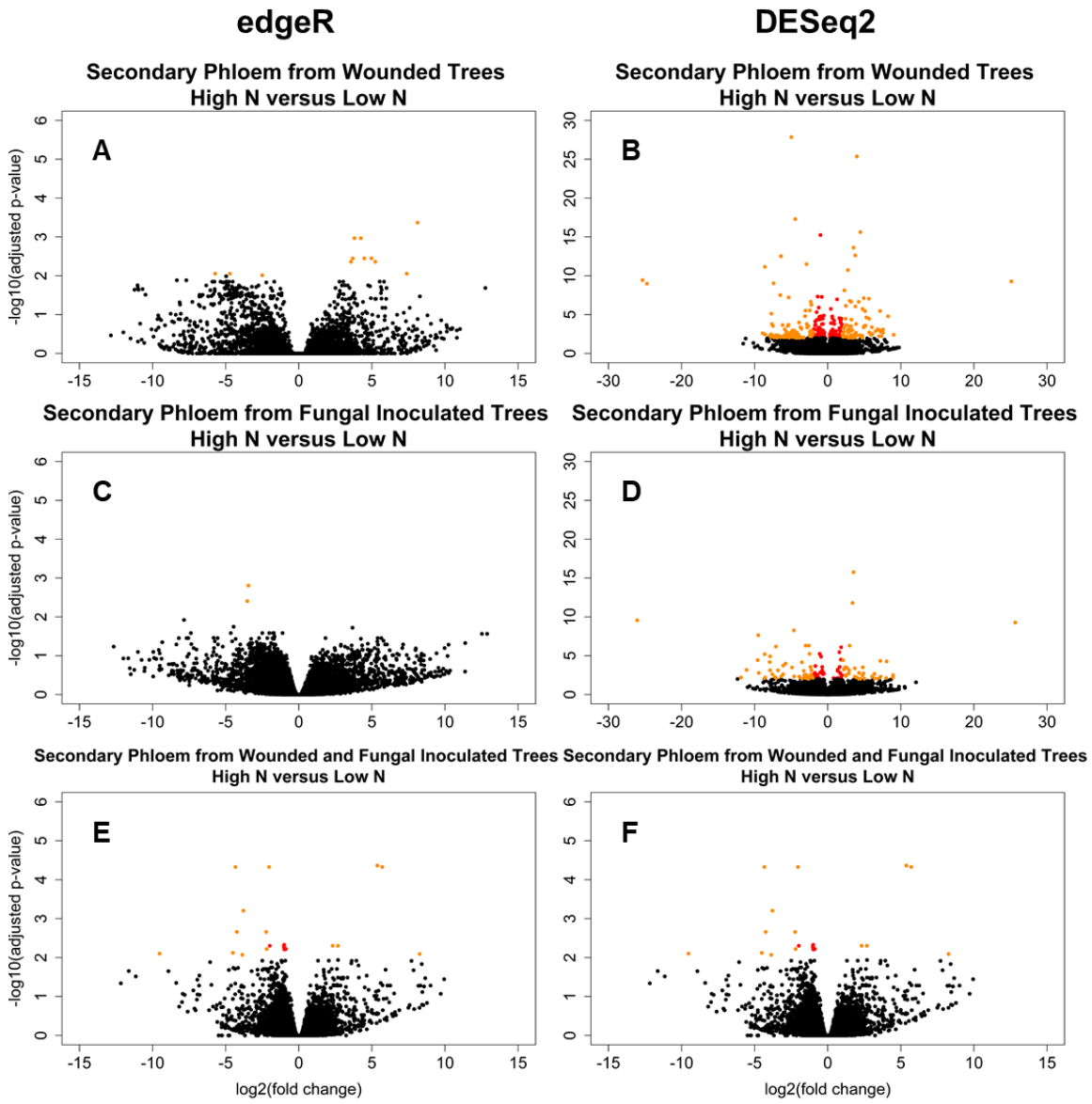


Figure 4.9. Volcano plots depicting DEGs determined by edgeR or DESeq2 in secondary phloem of lodgepole pine grown under 0.3 mM vs. 10 mM NH_4NO_3 and either mock-inoculated or inoculated with *G. clavigera*. Fold change values were calculated with edgeR (**A**, **C**, **E**) or DESeq2 (**B**, **D**, **F**). Observed fold changes (log base two) are plotted against the Benjamini-Hochberg adjusted p-values (-log base 10) for each gene. Black points were not significantly DE. Red points were significantly DE (adjusted p-value<0.01). Orange points were significantly DE (adjusted p-value<0.01) and exhibited an absolute log base two fold change of at least 2 ($|\log_2 FC| \geq 2$).

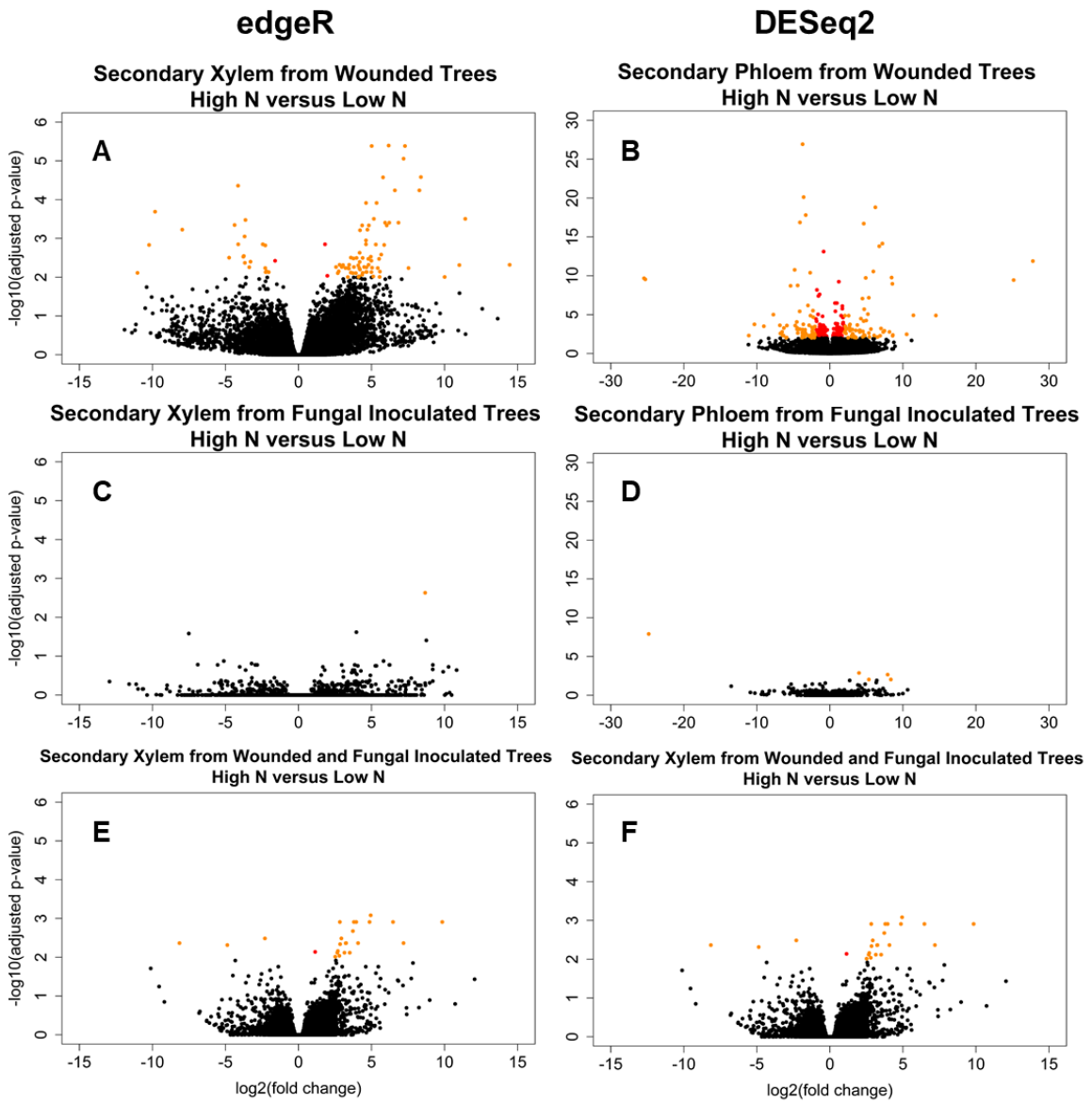
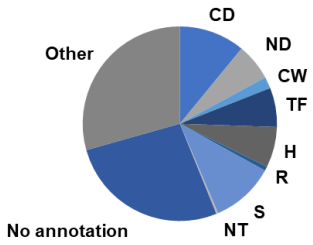


Figure 4.10. Volcano plots depicting DEGs determined by edgeR or DESeq2 in secondary xylem of lodgepole pine grown under 0.3 mM vs. 10 mM NH_4NO_3 and either mock-inoculated or inoculated with *G. clavigera*. Fold change values were calculated with edgeR (**A**, **C**, **E**) or DESeq2 (**B**, **D**, **F**). Observed fold changes (log base two) are plotted against the Benjamini-Hochberg adjusted p-values (-log base 10) for each gene. Black points were not significantly DE. Red points were significantly DE (adjusted p-value<0.01). Orange points were significantly DE (adjusted p-value<0.01) and exhibited an absolute log base two fold change of at least 2 ($|\log_2 FC| \geq 2$).

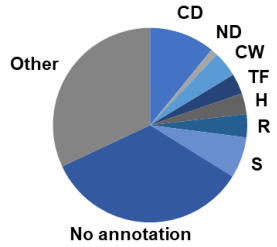
4.3.6 Transcript annotation and categorization

The categories and associated subcategories found in Table 4.1 were chosen to perform a comprehensive assessment of the defense responses induced by *G. clavigera* challenge. The proportion of core DEGs annotated as belonging to each category was similar for all N treatment and tissue type combinations (Figure 4.11; Table 4.5). Differences in expression profiles were more apparent for secondary phloem, where the high N application increased both the number of up-regulated N-based defense-associated genes and the number of down-regulated defense-associated C-based precursor biosynthesizing genes when contrasted with the low N treatment. The CD and ND categories were leveraged to test the CNB hypothesis. The CD:ND ratio for up-regulated genes was smaller in the high N-treated tissues compared with the low N-treated tissues.

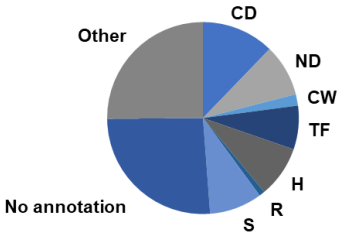
A


LOW N 2P FUNGUS VS WOUND
UP-REGULATED
CD:ND=1.8
CD
ND
CW
TF
H
R
S
NT
No annotation
Other

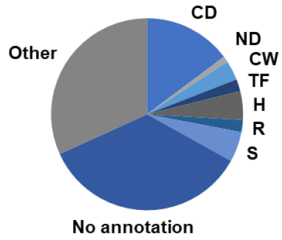
B


LOW N 2P FUNGUS VS WOUND
DOWN-REGULATED
CD:ND=9.2
CD
ND
CW
TF
H
R
S
No annotation
Other

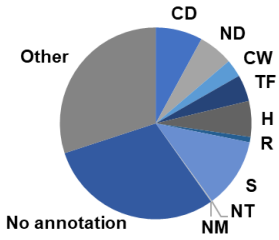
C


HIGH N 2P FUNGUS VS WOUND
UP-REGULATED
CD:ND=1.4
CD
ND
CW
TF
H
R
S
No annotation
Other

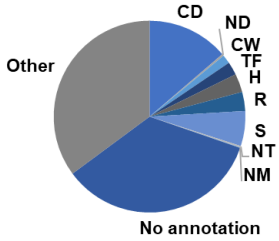
D


HIGH N 2P FUNGUS VS WOUND
DOWN-REGULATED
CD:ND=14.5
CD
ND
CW
TF
H
R
S
No annotation
Other

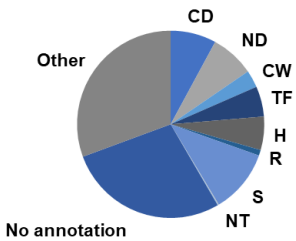
E


LOW N 2X FUNGUS VS WOUND
UP-REGULATED
CD:ND=1.3
CD
ND
CW
TF
H
R
S
NT
NM
No annotation
Other

F


LOW N 2X FUNGUS VS WOUND
DOWN-REGULATED
CD:ND=41.8
CD
ND
CW
TF
H
R
S
NT
NM
No annotation
Other

G


HIGH N 2X FUNGUS VS WOUND
UP-REGULATED
CD:ND=1
CD
ND
CW
TF
H
R
S
NT
No annotation
Other

H

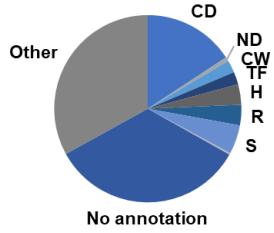

HIGH N 2X FUNGUS VS WOUND
DOWN-REGULATED
CD:ND=23.2
CD
ND
CW
TF
H
R
S
No annotation
Other

Figure 4.11. Functional categorization of core DEGs in *G. clavigera*-inoculated lodgepole pine seedlings grown under 0.3 mM or 10 mM NH_4NO_3. Secondary phloem (2P, **A**-**D**) and secondary xylem (2X, **E**-**H**) were sampled from lodgepole pine seedlings grown under low N (**A**, **B**, **E**, **F**) or high N conditions (**C**, **D**, **G**, **H**) and either inoculated with *G. clavigera* (Fungus) or mock-inoculated (Wound). Pie charts depict the relative proportion of core DEGs for each fungus vs. wound contrast that were annotated to the following categories: defense-related enzymes synthesizing C-based products (CD), N-based defense proteins, including enzymes (ND), cell wall-related genes (CW), transcription factors (TF), hormone-related genes (H), other regulators of genetic activity (R), signaling genes (S), genes involved in N transport (NT) and genes involved in N metabolism (NM). The CD:ND ratio is included with each pie chart.

Table 4.5. Functional categorization of core DEGs in lodgepole pine seedlings in response to *G. clavigera* inoculation, expressed for each contrast as a percentage of all core DEGs. Values in this table are depicted in the pie charts of Figure 4.11. The four sets of columns represent core DEGs that exhibited either significantly higher (Up) or lower transcript abundance (Down) in secondary phloem (2P) or secondary xylem (2X) of *G. clavigera*- vs. mock-inoculated seedlings grown under either low or high N conditions. Values indicate the percentage of all up- or down-regulated DEGs in a given contrast that were assigned to the following categories based on their annotations: defense-related enzymes synthesizing C-based products (CD), N-based defense proteins, including enzymes (ND), cell wall-related genes (CW), transcription factors (TF), hormone-related genes (H), other regulators (R), signaling genes (S), genes involved in N transport (NT), genes involved in N metabolism (NM).

	Low N 2P		High N 2P		Low N 2X		High N 2X	
Category	Up	Down	Up	Down	Up	Down	Up	Down
CD	11.0	10.9	12.2	14.8	7.9	13.7	7.9	15.8
ND	6.2	1.2	8.8	1.0	5.9	0.3	7.6	0.7
CW	1.9	4.3	1.9	3.3	3.0	1.5	3.0	2.2
TF	6.5	3.4	7.4	2.0	4.5	2.2	5.2	2.2
H	6.8	3.5	8.7	4.8	6.0	3.1	5.8	3.5
R	0.7	3.9	1.0	2.0	0.9	3.2	1.0	3.5
S	10.6	6.8	8.9	5.1	11.8	5.9	11.0	5.0
NT	0.3	0.0	0.0	0.0	0.2	0.2	0.1	0.3
NM	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0
No annotation	26.8	34.2	26.1	35.1	29.8	34.6	27.7	33.8
Other	29.4	32.3	25.1	31.8	30.0	35.1	30.6	33.0
Total genes	2291	1010	1047	393	3194	1218	1838	736

4.3.7 Expression patterns of genes implicated in the defense response

A Venn diagram of core DEGs was used to make tissue level comparisons between different N treatments (Figure 4.12). To explore the relative number of genes that were significantly DE for each tissue type exclusively in one N fertilization level and not the other, all Venn diagram core DEGs falling into the low N group or high N group, but not both, were tabulated (Table 4.6). When compared with the high N-treated trees, a greater number of genes were DE in the low N-treated trees. The 488 genes significantly DE in all treatments included many key players implicated in fungal pathogen-induced defense (Table 4.7). While most functional subcategories represented in this shared response were up-regulated, cytochrome P450s, which play critical roles in many biosynthetic pathways (Xu *et al.* 2015), showed overall down-regulation.

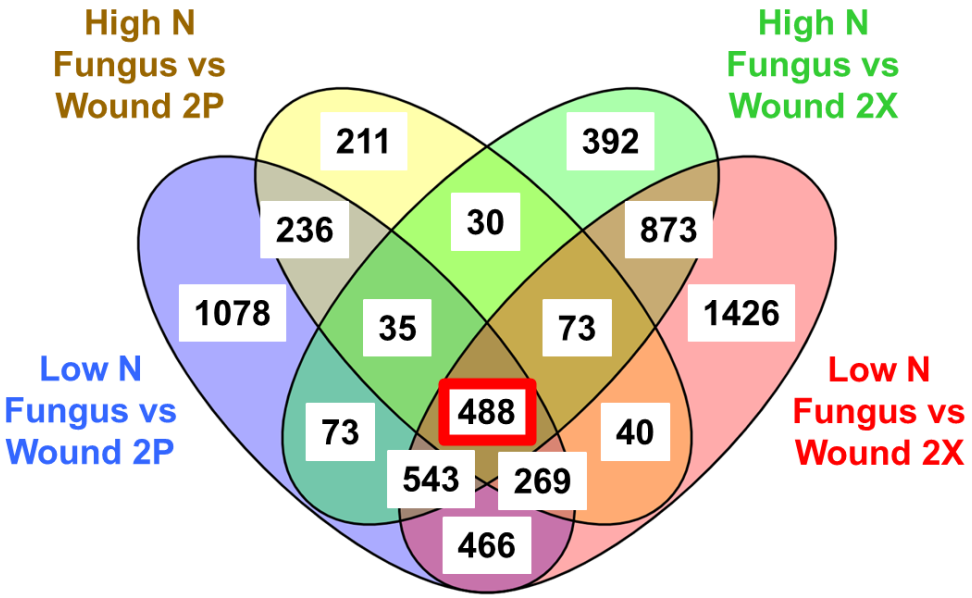


Figure 4.12. Changes in gene expression showed distinctive and overlapping patterns in response to *G. clavigera* and N availability. The Venn plot indicates the number of core DEGs unique to and shared between the transcriptomic responses of secondary phloem (2P) and secondary xylem (2X) under both N levels. The most central cell, highlighted in red, contains a common set of 488 genes DE in both 2X and 2P regardless of N availability. Prior to differential expression testing, genes were filtered using a transcript abundance cutoff of at least one cpm reads in at least four libraries. The cutoff for designation as a core DEG, and inclusion in this Venn diagram, was a Benjamini-Hochberg adjusted p-value<0.01 and $|\log_2 FC| \geq 2$ for both edgeR and DESeq2.

Table 4.6. N availability impacted the number of DEGs up- and down-regulated following fungal inoculation. Genes included in the Venn diagram that belonged exclusively to one N fertilization level for each tissue type were tabulated. The four columns represent core DEGs that exhibited either significantly higher (Up) or lower (Down) transcript abundance in secondary phloem (2P) or secondary xylem (2X) of *G. clavigera*- vs. mock-inoculated seedlings grown under either 0.3 mM (Low N) or 10 mM (High N) NH_4NO_3 conditions. Genes enumerated were significantly DE in either low N- or high N-treated trees, but not both.

Tissue	Fertilization	Total	Up	Down
2P	Low N	2159	1378	781
	High N	354	162	192
2X	Low N	2201	1463	738
	High N	738	262	268

Table 4.7. Functional subcategories of putative defense-associated genes represented in the common response of lodgepole pine to *G. clavigera* inoculation. Select classifications were assigned to the 488 genes found to be significantly DE in response to *G. clavigera* in both secondary phloem and secondary xylem under either low or high N availability. Classifications are the subcategories of the major functional categories described in Table 4.1, many of which have known roles in plant defense.

Up-regulated	flavonoid terpenoid phenylpropanoid sulfur-containing β-1,3-glucanase chitinase osmotin-like thaumatin pathogenesis-related protein peroxidase MYB ERF JAZ WRKY NAC other transcription factors other genetic regulators signal transduction MAPK NBS-LRR receptor kinase signaling response to fungus cell wall degradation cell wall response to chitin
Down-regulated	cytochrome P450

Terpene synthases and chitinases are well characterized components of conifer defense against pests and pathogens (Kovalchuk *et al.* 2013; Kolosova *et al.* 2014; Keefover-Ring *et al.* 2015; West *et al.* 2016). Predictably, sequences annotated as terpene synthases and chitinases were well represented in the 488 core DEGs common to all treatments (Table 4.7). Of the genes used to create the Venn diagram, multiple mono-, di- and sesquiterpene synthases were found to be significantly DE following fungal attack, with the 0.3 mM NH_4NO_3-fertilized seedlings presenting stronger patterns of up-regulation, as indicated by larger

log_2FC values, when compared with the 10 mM NH_4NO_3-fertilized seedlings (Figure 4.13). Eleven terpene synthase genes included in the heatmaps belonged to the 488 genes highlighted in Figure 4.12 and described in Table 4.7. In response to fungal inoculation in both xylem and phloem tissues, β-phellandrene synthase genes were down-regulated. This runs counter to the metabolite monoterpene analysis, which found that β-phellandrene levels were increased in *G. clavigera*-inoculated phloem samples compared with control phloem samples (Table 4.2).

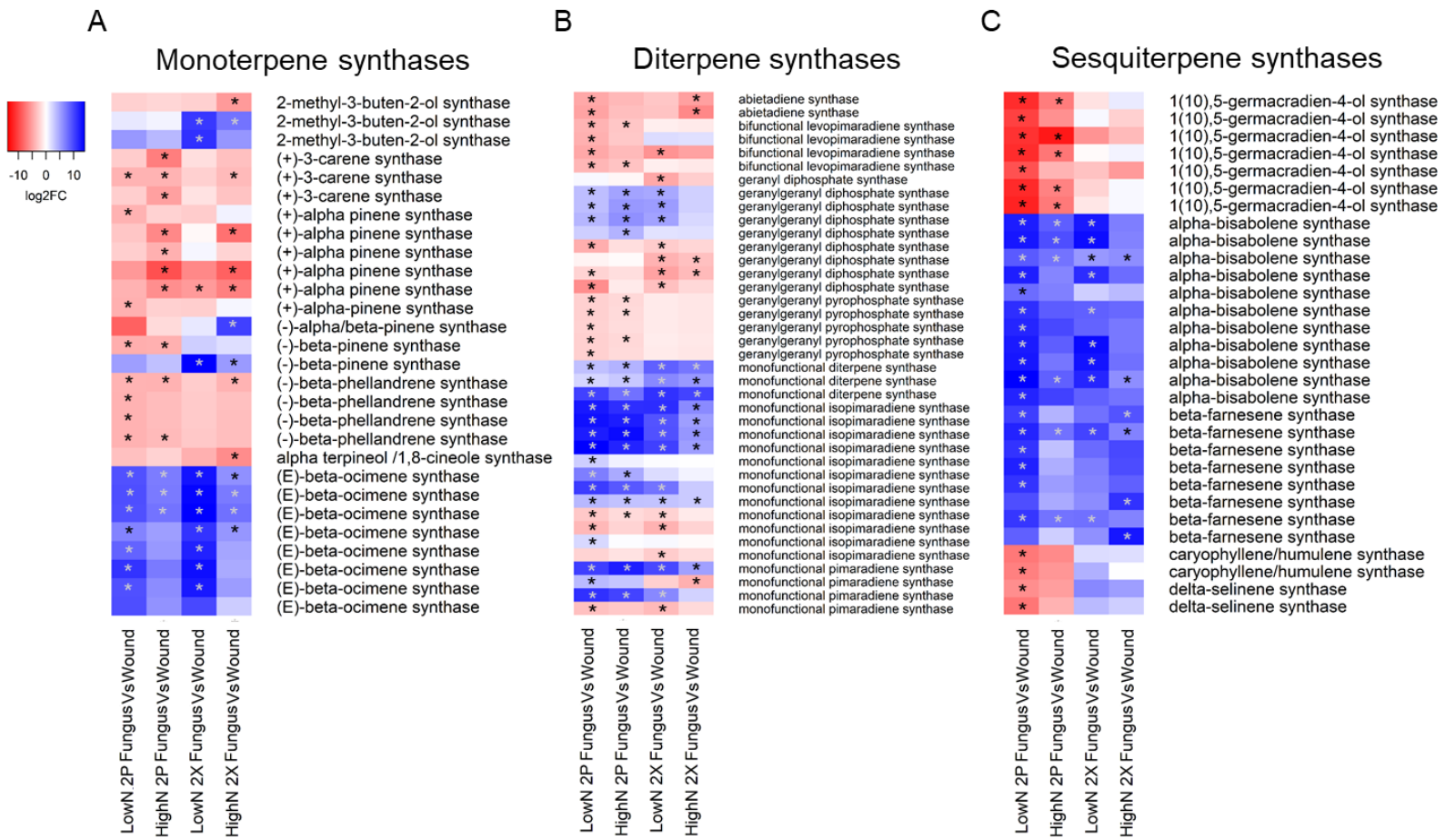
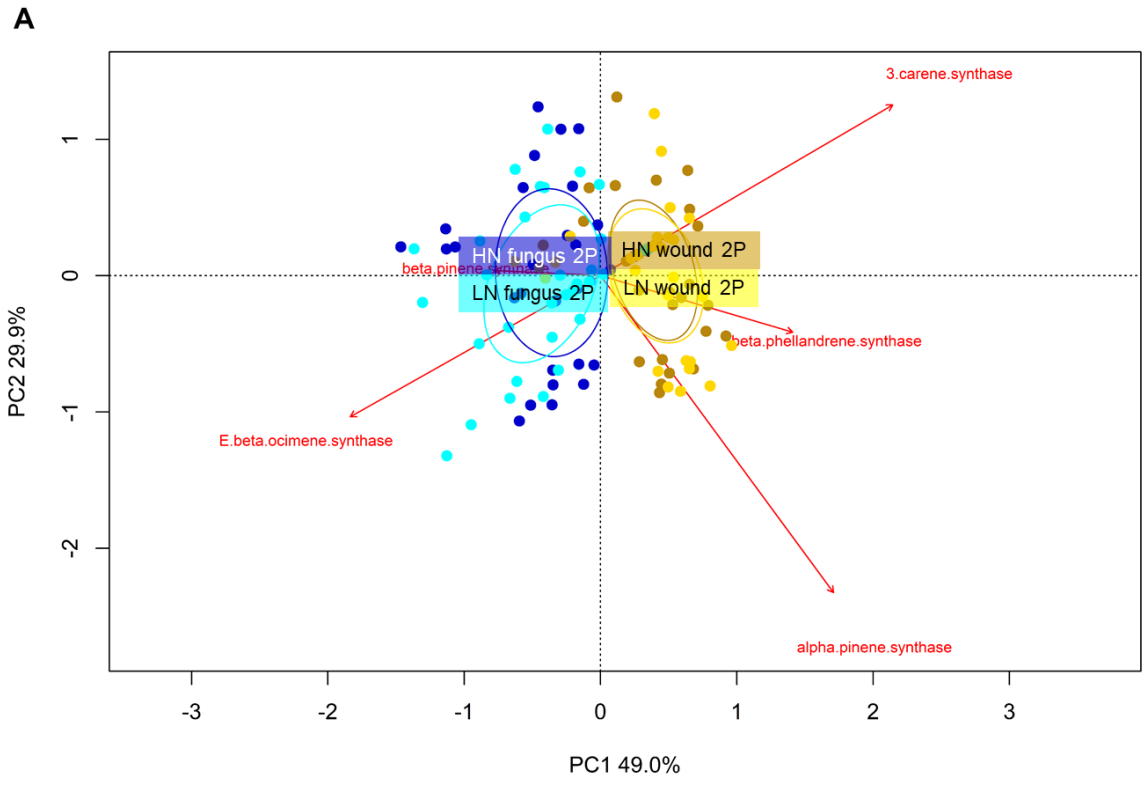


Figure 4.13. Expression levels of terpene synthases significantly increased in both xylem and phloem of *G. clavigera*-inoculated lodgepole pine seedlings relative to mock-inoculated seedlings, with fold-changes modulated by N availability. Mono- (**A**), di- (**B**) and sesquiterpene (**C**) synthase genes were DE in response to fungal inoculation. Differential gene expression was detected using edgeR. Blue indicates up-regulation and red indicates down-regulation. Color intensity indicates the degree of fold change in transcript abundance between *G. clavigera*- and mock-inoculated samples. Genes determined to be significantly DE with an adjusted p-value<0.01 and |log2FC|≥2 are marked with an asterisk.

RDA was also used to examine putative monoterpene synthase gene expression profiles in phloem (Figure 4.14A) and xylem (Figure 4.14B) tissues. In both figures, RDA revealed separation of mock-inoculated and *G. clavigera*-inoculated samples along PC1, with several monoterpene synthase genes contributing to these differences. For secondary phloem, low N- and high N-treated samples showed no separation along the PC2. In secondary xylem, low N- and high N-treated samples showed slight separation along the PC2, with 3-carene synthase, (+)-α-pinene synthase and β-phellandrene synthase contributing to these differences. All genes used for RDA appear in the monoterpene synthase heatmap (Figure 4.13 A). Two (E)-β-ocimene synthase genes belonged to the 488 genes highlighted in Figure 4.12 and Table 4.7.


A
3.carene.synthase
HN fungus 2P
HN wound 2P
beta.pinene.synthase
LN fungus 2P
LN wound 2P
beta.phellandrene.synthase
E.beta.ocimene.synthase
alpha.pinene.synthase
PC2 29.9%
PC1 49.0%

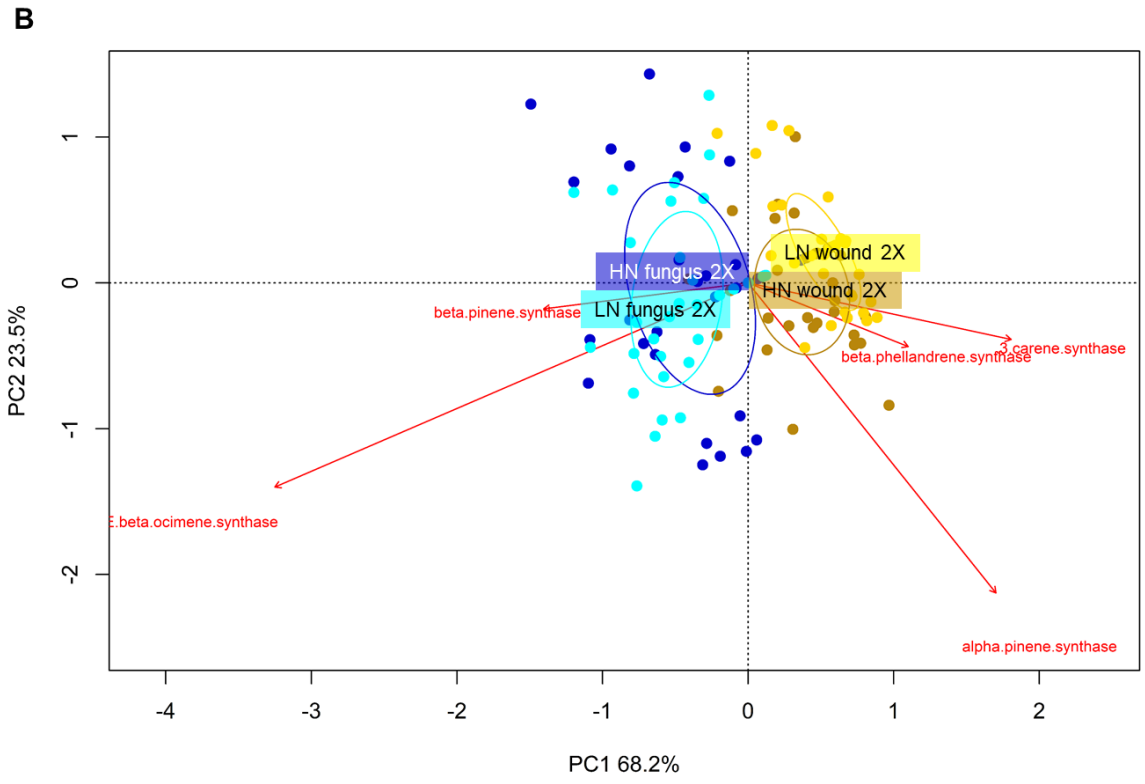

B
LN wound 2X
HN fungus 2X
HN wound 2X
beta.pinene.synthase
LN fungus 2X
3.carene.synthase
beta.phellandrene.synthase
E.beta.ocimene.synthase
alpha.pinene.synthase
PC2 23.5%
PC1 68.2%

Figure 4.14. PCA of phloem and xylem monoterpene synthase expression levels of RNA-Seq data following DESeq2 normalization showed separation of fungal- and mock-inoculated samples along PC1. Lodgepole pine seedlings grown under low N (LN; 0.3 mM NH_4NO_3) or high N (HN; 10 mM NH_4NO_3) conditions were either *G. clavigera*-inoculated (fungus) or mock-inoculated (wound). Monoterpene-related gene expression from (**A**) secondary phloem (2P) and (**B**) secondary xylem (2X) collected at 7 dpi were analyzed by RDA with the R package vegan, n=4. The 95% confidence ellipses are shown, and labels are displayed in the center of each ellipse.

G. clavigera inoculation also resulted in significantly greater transcript abundances for several chitinase genes, mainly belonging to classes I, IV and VII. Chitinase differential expression was consistently elicited by fungal inoculation across all classes detected aside from class II (Figure 4.15). Their expression profiles were also modulated by N availability, with low N trees presenting stronger patterns of up-regulation, indicated by larger log_2FC inductions, in both xylem and phloem when compared with the high N concentration treated trees.

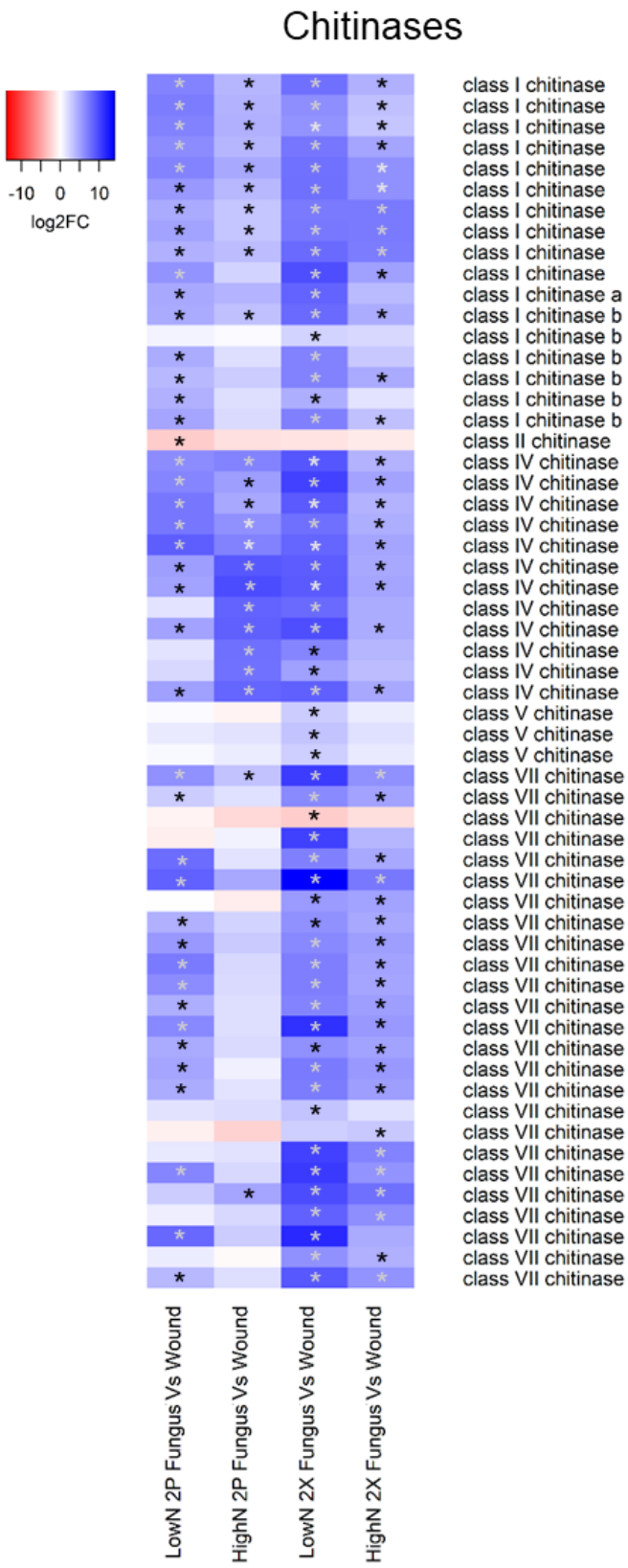


Figure 4.15. Expression levels of several chitinases significantly increased in both xylem and phloem of *G. clavigera*-inoculated lodgepole pine seedlings relative to mock-inoculated seedlings, with fold-changes modulated by N availability. Differential gene expression was detected using edgeR. Blue indicates up-regulation and red indicates down-regulation. Color intensity indicates the degree of fold change in transcript abundance between *G. clavigera*- and mock-inoculated samples. Genes determined to be significantly DE with an adjusted p-value<0.01 and $|\log_2 FC| \geq 2$ are marked with an asterisk.

4.3.8 Correlated gene expression network analysis

A network analysis conducted using the R package WGCNA (Langfelder and Horvath 2008) was used to characterize the coordinated responses of young lodgepole pine trees to *G. clavigera* inoculation (i.e. identify groups of genes with similar expression profiles across treatments). Average linkage sample clustering identified a high N-treated fungal-inoculated secondary xylem sample to be an outlier based on visual inspection, and these data were removed from the network analysis (Figure 4.16). The remaining 31 data sets were used for network analysis by WGCNA (Langfelder and Horvath 2008). All 154,398 contigs were correlated by similar patterns of expression and connected by edges weighted by the biweight midcorrelation value. Contigs were consolidated into distinct color-coded modules which maximized the interconnectedness (i.e. biweight midcorrelation values) between genes. The network results were displayed in an interconnectedness dendrogram, where genes with the highest intramodular connectivity proportional to the number of neighbors that a pair of genes shared in common were located at the tip of each branch (Figure 4.17). Application of the *dynamicTreeCut* function detected 22 modules. Modules were merged using the *mergeCloseModules* function if their module eigengenes had a pairwise biweight midcorrelation value greater than 0.35. This process resulted in nine highly discorrelated modules (Figure 4.18). Visualization of DEG fold change data associated with each of the merged modules revealed that within the orange module, most DEGs were up-regulated in response to G. *clavigera* inoculation (Figure 4.18). Genes making up

the dark cyan module showed both up- and down-regulation in response to *G. clavigera*, while the dark magenta and dark slate blue modules were mainly composed of genes down-regulated by *G. clavigera* infection.

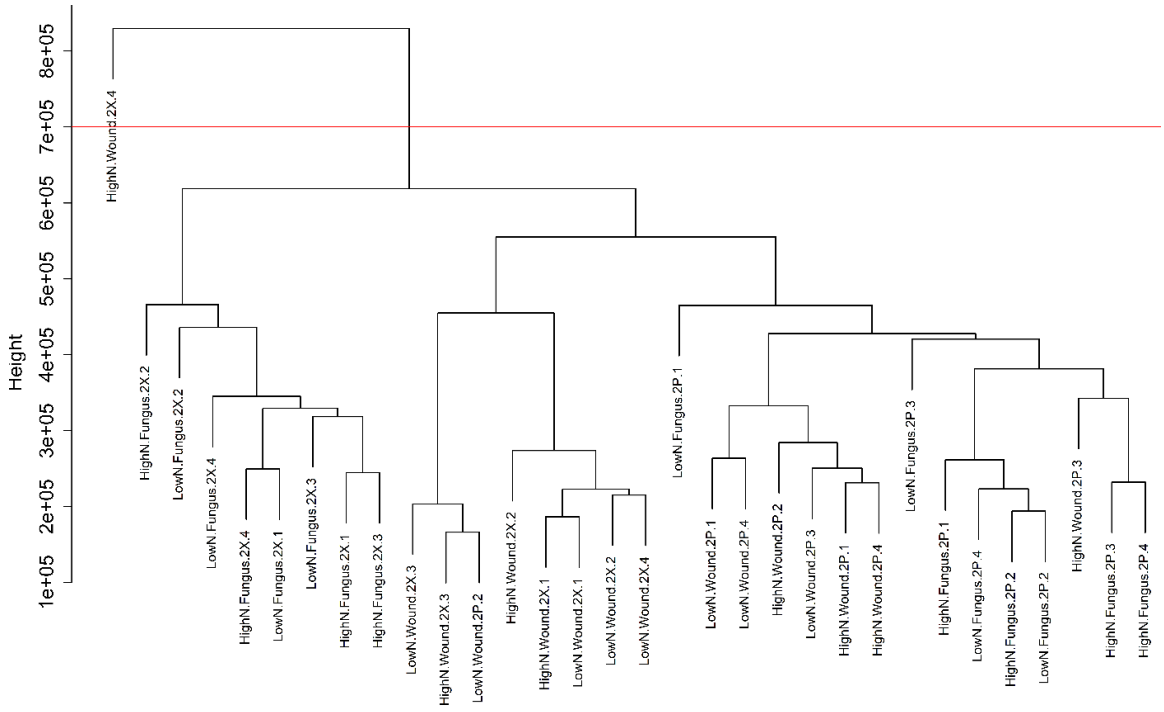


Figure 4.16. Hierarchical clustering of samples for the detection of outliers. Filtered expression data for 32 lodgepole pine cDNA libraries were produced by *featureCounts* and normalized with a DESeq2 variance stabilizing transformation. Average linkage sample clustering detected outlier sample HighN.Fungus.2X.4. The red line cut height was chosen based on visual inspection to identify the outlier sample.

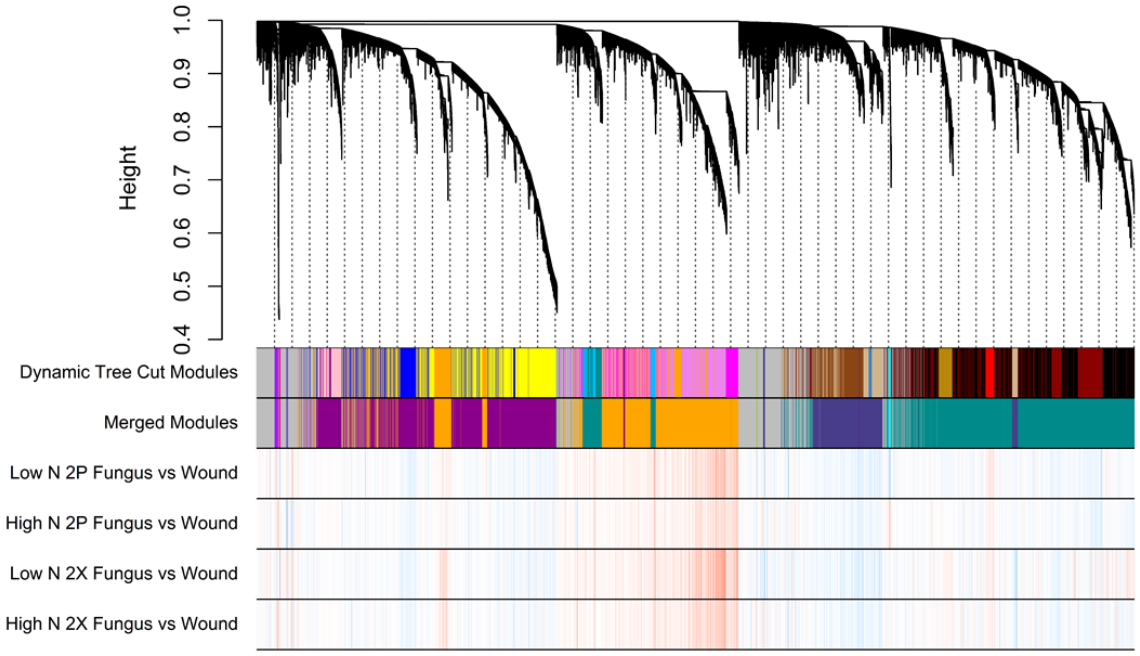


Figure 4.17. Gene expression correlation dendrogram with module identification, overlaid with edgeR differential expression data. Genes were clustered using biweight midcorrelation and assigned to color-coded modules by WGCNA. The dendrogram depicts genes that fall into each module, and genes are oriented across the x-axis by the strength of their co-expression as measured by biweight midcorrelation. The height measure orients genes based on how disconnected they are to the module eigengene, where genes with the highest intramodular connectivity are located at the tip of each module branch. The dynamic tree cut method imposed no module merge functionality, resulting in 22 modules. Module eigengenes were used to combine modules with a biweight midcorrelation value greater than 0.35, leading to nine merged modules. Log_2FC values for each gene in the dendrogram are depicted in the heatmap below the modules, as calculated by edgeR for each of the four treatment contrasts. In the heatmap, red indicates up-regulation while blue indicates down-regulation. Color intensity indicates the degree of fold change in transcript abundance between *G. clavigera*- and mock-inoculated samples.

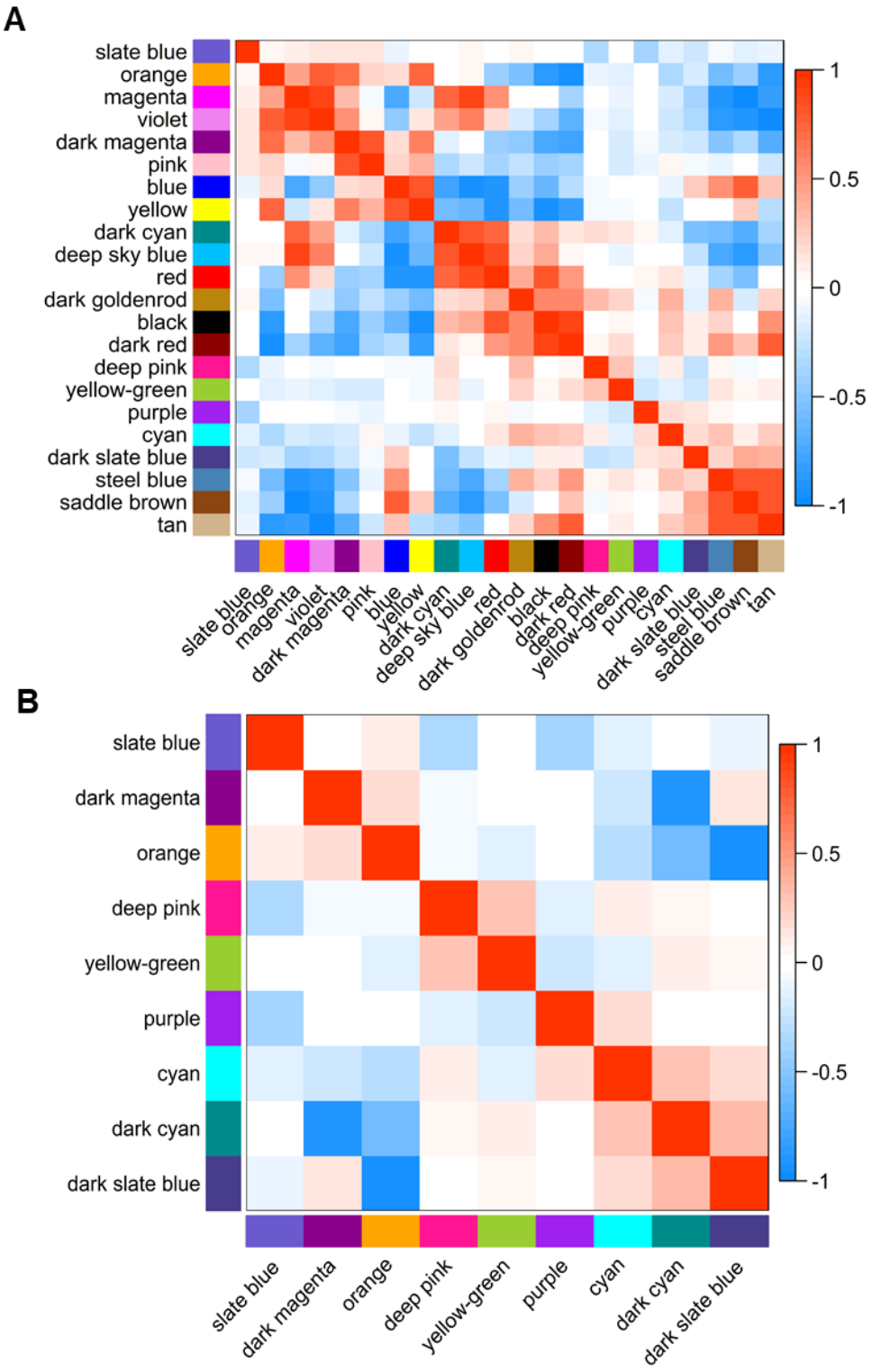

A
slate blue
orange
magenta
violet
dark magenta
pink
blue
yellow
dark cyan
deep sky blue
red
dark goldenrod
black
dark red
deep pink
yellow-green
purple
cyan
dark slate blue
steel blue
saddle brown
tan
1
0.5
0
-0.5
-1
B
slate blue
dark magenta
orange
deep pink
yellow-green
purple
cyan
dark cyan
dark slate blue
1
0.5
0
-0.5
-1

Figure 4.18. Module eigengene correlation with WGCNA dynamic tree cut modules and merged modules. (**A**) Modules produced by the *dynamicTreeCut* function. (**B**) Merged modules produced with the *mergeCloseModules* function. Genes were clustered using biweight midcorrelation and assigned to color-coded modules by WGCNA. Module eigengenes were correlated using biweight midcorrelation. Modules produced by *dynamicTreeCut* with a correlation value greater than 0.35 were merged, resulting in nine merged modules that were used in subsequent analyses.

We next determined whether any of these four major WGCNA (Langfelder and Horvath 2008) modules exhibited statistical overrepresentation of core DEGs belonging to functional categories associated with defense. Enrichment analysis by hypergeometric test was carried out using MapMan functional categories (bins; Thimm *et al.* 2004). The signaling, transportation and miscellaneous bins showed significant enrichment in all four modules (Figure 4.19). The orange module contained the most significantly enriched MapMan bins, including signaling, hormone metabolism and cell wall. Surprisingly, the MapMan bin stress, which included stress caused by biotic factors such as *G. clavigera*, was not significantly enriched for any module.

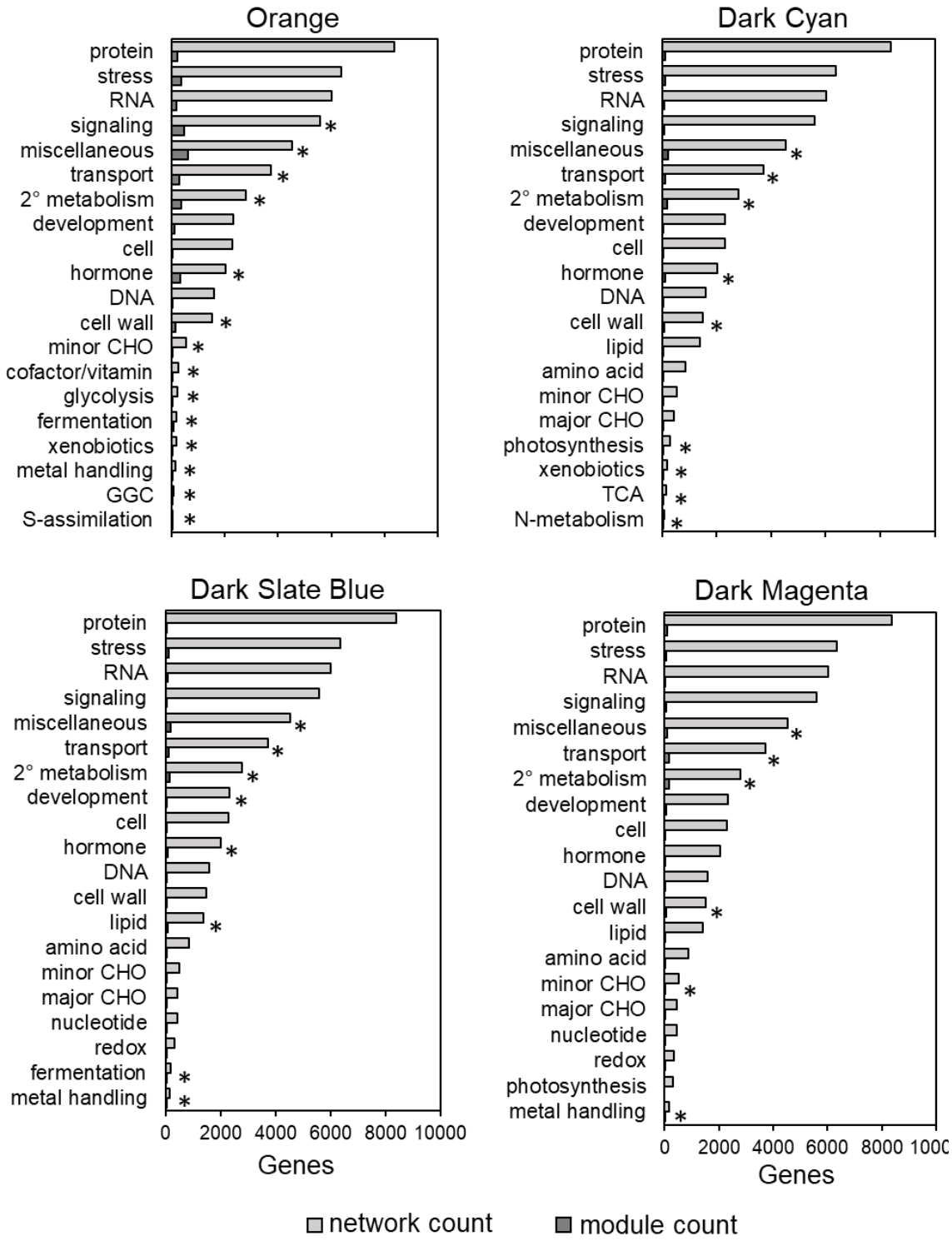

Orange
protein
stress
RNA
signaling
miscellaneous
transport
2° metabolism
development
cell
hormone
DNA
cell wall
minor CHO
cofactor/vitamin
glycolysis
fermentation
xenobiotics
metal handling
GGC
S-assimilation
Dark Cyan
protein
stress
RNA
signaling
miscellaneous
transport
2° metabolism
development
cell
hormone
DNA
cell wall
lipid
amino acid
minor CHO
major CHO
photosynthesis
xenobiotics
TCA
N-metabolism
Dark Slate Blue
protein
stress
RNA
signaling
miscellaneous
transport
2° metabolism
development
cell
hormone
DNA
cell wall
lipid
amino acid
minor CHO
major CHO
nucleotide
redox
fermentation
metal handling
0 2000 4000 6000 8000 10000
Genes
Dark Magenta
protein
stress
RNA
signaling
miscellaneous
transport
2° metabolism
development
cell
hormone
DNA
cell wall
lipid
amino acid
minor CHO
major CHO
nucleotide
redox
photosynthesis
metal handling
0 2000 4000 6000 8000 1000
Genes
network count
module count

Figure 4.19. WGCNA modules contained significantly enriched MapMan bins that highlighted primary and secondary metabolic processes. DEGs were annotated according to MapMan hierarchical functional categories (bins), and hypergeometric tests were performed for the set of core DEGs associated with each of the four largest WGCNA modules. For each module, the abundance of core DEGs within each functional category (module count) was compared to the total number of genes within that functional category (network count). Statistically overrepresented functional categories (p-value<0.05) are indicated with asterisks. Abbreviations include secondary metabolism (2° metabolism), cell organization (cell), hormone metabolism (hormone), amino acid metabolism (amino acid), lipid metabolism (lipid), tricarboxylic acid cycle/organic acid transformation (TCA), gluconeogenesis/glyoxylate cycle (GGC), carbohydrate metabolism (CHO), nucleotide metabolism (nucleotide), biodegradation of xenobiotics (xenobiotics) and redox regulation (redox).

As a means of examining expression patterns within the context of plant defense theories, we next looked at the representation of core DE putative defense-associated genes within these four WGCNA (Langfelder and Horvath 2008) modules using the subcategories described in Table 4.1 (Table 4.8). The orange module contained the largest number of N-based defense-related genes that were either up- or down-regulated in response to *G. clavigera* inoculation, exhibiting a CD:ND ratio of 1.5. In comparison, the dark cyan, dark magenta and dark slate blue modules had CD:ND ratios of 4.3, 9.0 and 30.3, respectively.

Table 4.8. Profiling the lodgepole pine defense response in the four largest WGCNA modules. Genes significantly DE in response to *G. clavigera* for at least one of the four differential expression contrasts depicted in Figure 4.18 were assigned to annotation categories and subcategories as described in Section 4.2.10. Values represent the number of core DEGs within each module assigned to annotation categories and subcategories. Fungal-induced (up-regulated) and fungal-repressed (down-regulated) genes are included.

	Gene	Orange	Dark cyan	Dark slate blue	Dark magenta
C-based defenses	alkaloid	0	1	0	0
	cytochrome P450	31	18	30	8
	chalcone	9	3	0	0
	stilbene	9	2	0	0
	flavonoid	97	70	46	5
	isoprenoid	10	5	7	2
	terpenoid	40	28	9	16
	lignin	60	28	24	41
	phenylpropanoid	21	1	2	1
	phytoalexin	0	0	0	0
	pinoresinol	0	0	0	3
	phenol	5	3	0	5
	sulfur-containing	21	12	3	0
	total	303	171	121	81
N-based defenses	β-1,3-glucanase	32	2	0	0
	chitinase	74	1	0	1
	dirigent	12	0	0	6
	osmotin-like	13	1	0	0
	thaumatin	2	2	0	0
	pathogenesis-related	20	3	0	0
	peroxidase	52	31	4	2
	total	205	40	4	9
Transcription factors	MYB	24	15	3	3
	ERF	26	13	5	1
	JAZ	26	5	0	0
	WRKY	18	6	0	0
	bHLH	28	1	5	0
	bZIP	13	0	2	0
	NAC	0	0	0	0
	other	37	16	13	5
	total	172	56	28	9
Signaling genes	receptor kinase	336	48	35	17
	histidine kinase	2	1	6	3
	MAPK	24	2	0	0
	signal transduction	2	0	2	0
	respiratory burst	4	0	6	6
	R-gene	34	10	21	0
	NBS-LRR	7	6	10	0
	NB-ARC	5	17	4	0
	total	421	84	80	26

We further explored the effect of N availability on log_2FC values for core DEGs included in Table 4.8 for the orange and dark cyan modules (Figures 4.20 and 4.21, respectively). Within either xylem or phloem, all C-based and N-based defense associated core DEGs in the orange module displayed higher median log_2FC values in response to *G. clavigera* inoculation in low N compared to high N seedlings. Genes that encode the enzymes that synthesize phenolic compounds in phloem showed a strong association with the low N treatment, which was starkly contrasted with the low expression levels of these genes in the phloem of high N-treated trees. Most orange module signaling genes and transcription factors followed the same pattern for both tissue types. Putative pathogen recognition nucleotide binding adaptor shared by nucleotide-binding oligomerization domain with leucine-rich repeat proteins, apoptotic protease activating factor 1, resistance proteins and cell-death protein 4 (NB-ARC) genes (McHale *et al.* 2006) in phloem and putative jasmonate (zinc-finger expressed in inflorescence meristem)-domain transcription factors (JAZs) (Thines *et al.* 2007) in xylem were the only two gene categories that displayed increased expression levels in high N-treated trees compared with low N-treated trees.

In keeping with the observation from Figure 4.18 that genes within the dark cyan module exhibited up- or down-regulation, and also showed less consistency of log_2FC values across the four treatment contrasts, the line graphs produced from putative defense-associated gene categories showed greater diversity in expression profiles between tissue type and N availability than that of the orange module.

Osmotin-like proteins (Hakim *et al.* 2018) displayed distinctive increased expression in the phloem and decreased expression in the xylem of high N-treated trees compared with low N-treated trees. JAZs in phloem tissue were similarly up-regulated under both N treatments, while JAZs in xylem displayed increased expression levels in high N-treated trees compared with low N-treated trees. Histidine kinase (Nongpiur *et al.* 2012) gene expression in xylem was intensified with increased N availability. Phloem genes that encode the enzymes that synthesize phenolic compounds had lower median levels of differential expression in high compared to low N treatments.

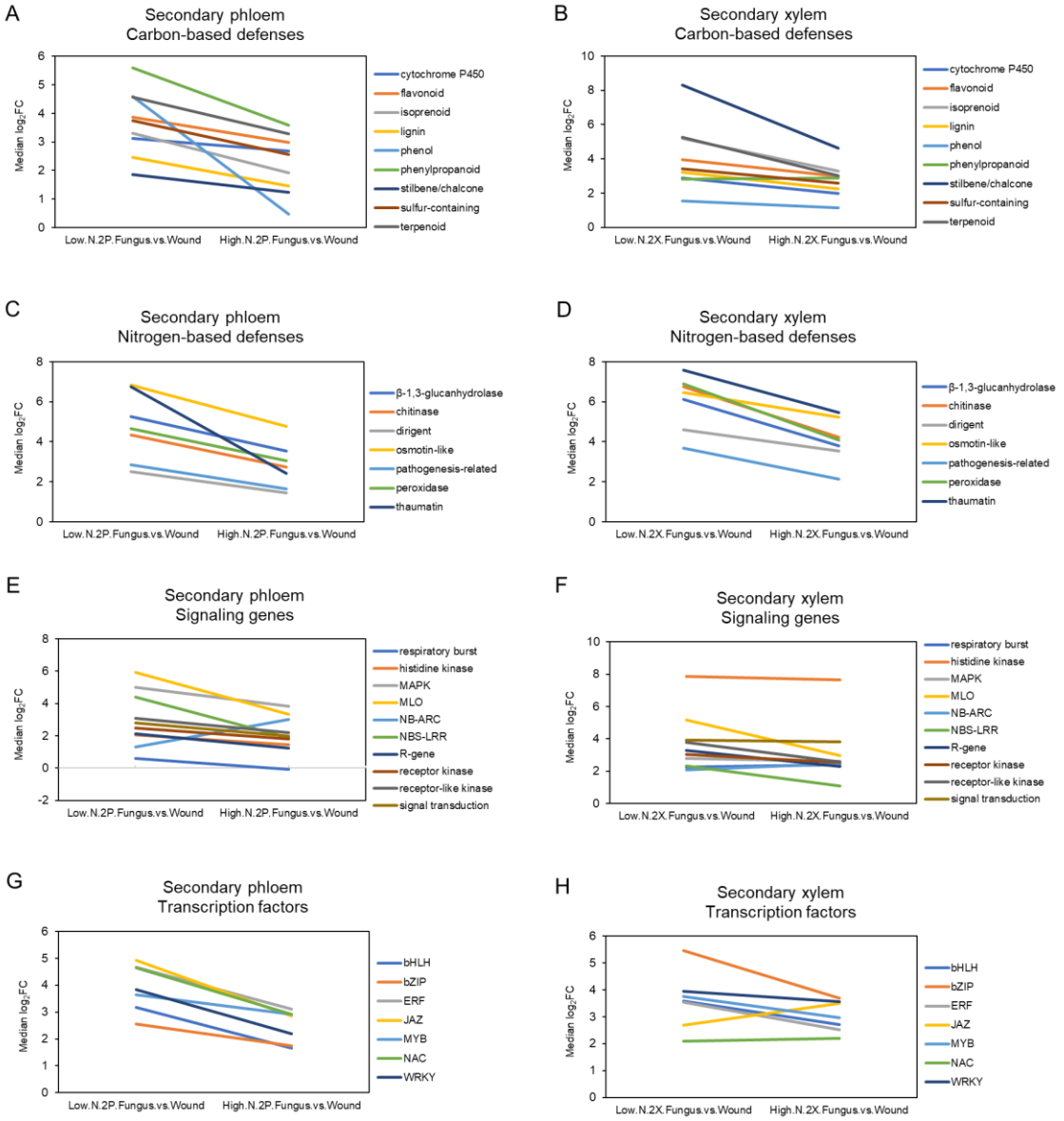


Figure 4.20. Effect of N availability on median log$_2$FC values of orange module putative defense-associated gene subcategories. Genes of the orange module identified as *G. clavigera*-responsive core DEGs in secondary phloem (**A**, **C**, **E**, **G**) and secondary xylem (**B**, **D**, **F**, **H**) were annotated according to the functional subcategories outlined in Table 4.1 and enumerated in Table 4.8 for each of the following: (**A**, **B**) C-based defenses, (**C**, **D**) N-based defenses, (**E**, **F**) defense-related signaling genes and (**G**, **H**) transcription factors. The median log$_2$FC values for core DEGs assigned to each functional subcategory are plotted for each low N – high N contrast.

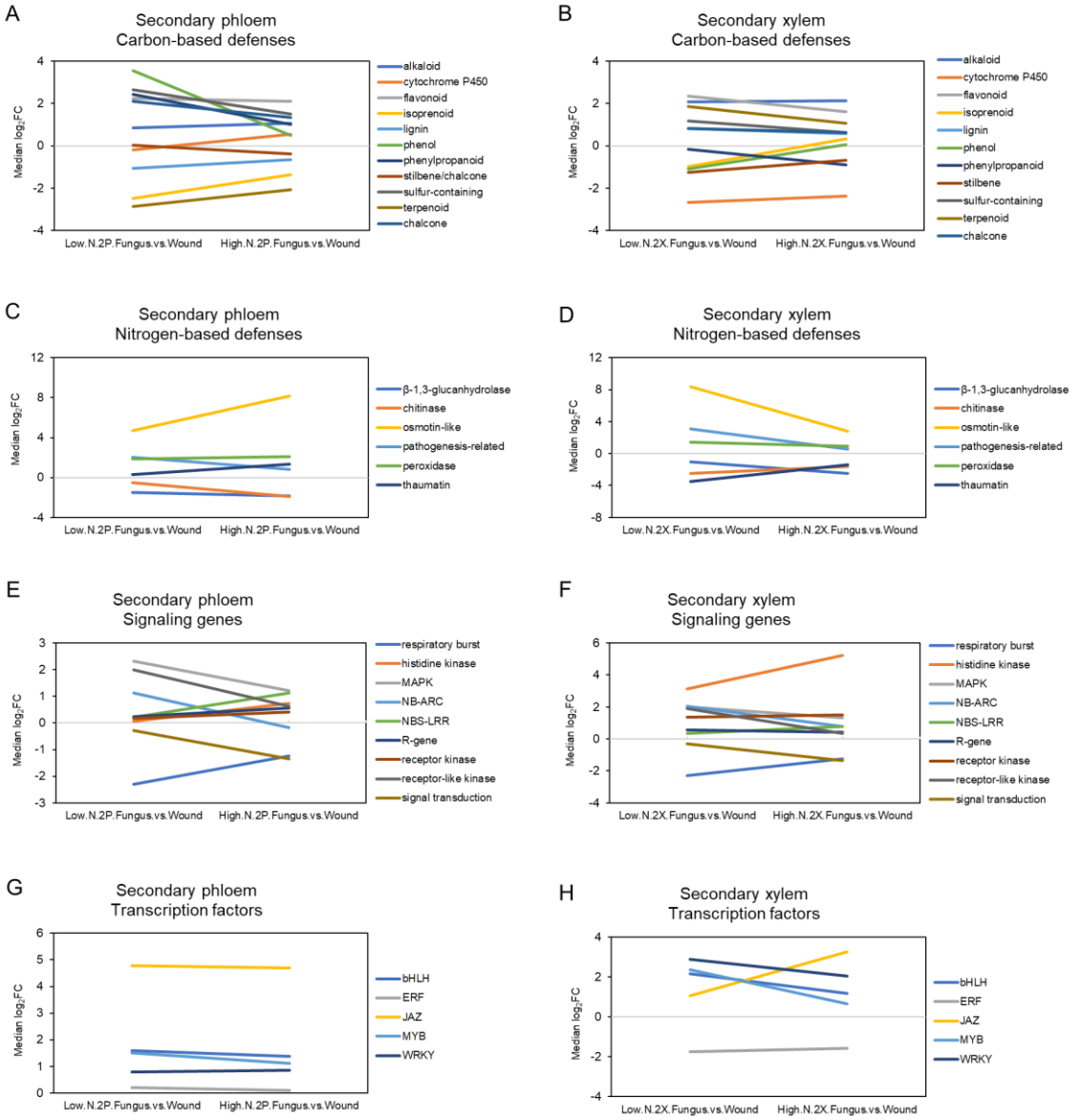


Figure 4.21. Effect of N availability on median log$_2$FC values of dark cyan module putative defense-associated gene subcategories. Genes of the orange module identified as *G. clavigera*-responsive core DEGs in secondary phloem (**A**, **C**, **E**, **G**) and secondary xylem (**B**, **D**, **F**, **H**) were annotated according to the functional subcategories outlined in Table 4.1 and enumerated in Table 4.8 for each of the following: (**A**, **B**) C-based defenses, (**C**, **D**) N-based defenses, (**E**, **F**) defense-related signaling genes and (**G**, **H**) transcription factors. The median log$_2$FC values for core DEGs assigned to each functional subcategory are plotted for each low N – high N contrast.

Expression heatmaps were also generated from the core DEGs included in Table 4.8 for the orange (Figure 4.22) and dark cyan (Figure 4.23) modules. These DEGs were showcased in Figures 4.20 and 4.21, respectively. Several core DEGs in the orange module heatmap showed a uniform pattern of greater fold-change induction in 0.3 mM NH_4NO_3-treated seedlings when compared with the 10 mM NH_4NO_3-treated seedlings. In the dark cyan module, core DEG expression was modulated by N, though expression patterns were less uniform.

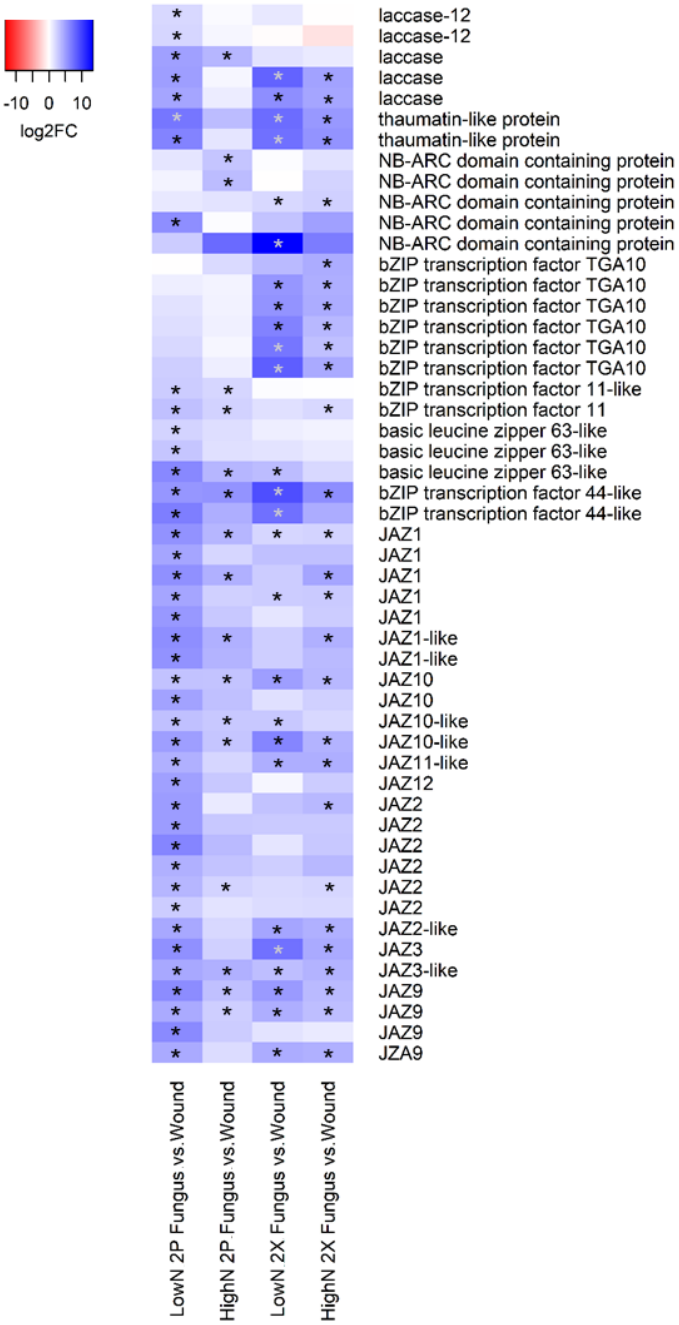


Figure 4.22. The orange WGCNA module contained significantly DE putative defense-related genes. Core DEGs in the orange module, represented in Table 4.8 and showcased in Figure 4.21, were determined using edgeR and DESeq2, though only edgeR log_2FC values are displayed. The heatmap compares log_2FC values of secondary phloem (2P) and secondary xylem (2X) of *G. clavigera*- (Fungus) vs. mock-inoculated (Wound) lodgepole pine grown under low or high N availability. Blue indicates up-regulation and red indicates down-regulation. Color intensity indicates the degree of fold change in transcript abundance between *G. clavigera*- and mock-inoculated samples. DEGs found to be significantly DE by both edgeR and DESeq2 (adjusted p-value<0.01 and $|log_2FC| \geq 2$) are marked with asterisks.

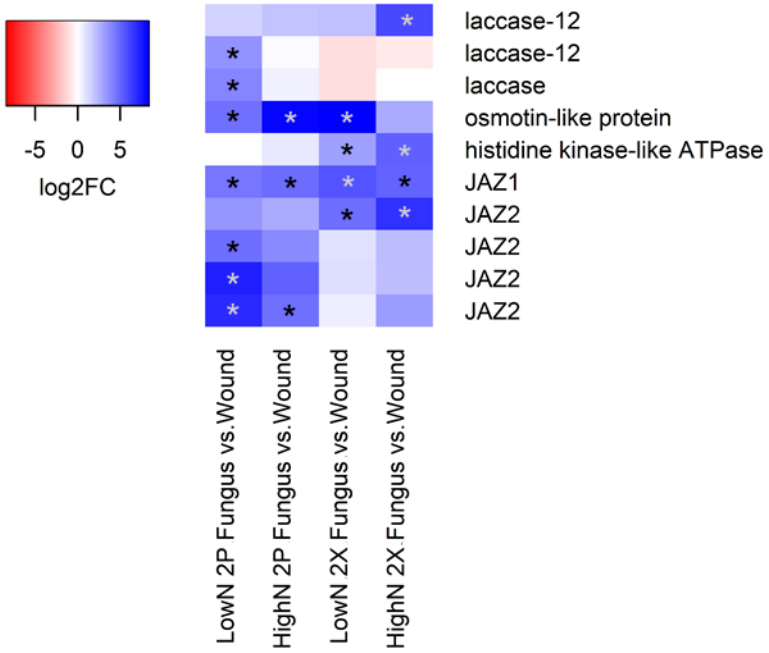


Figure 4.23. The dark cyan WGCNA module significantly contained DE putative defense-related genes. Core DEGs in the dark cyan module, represented in Table 4.8 and showcased in Figure 4.22, were determined using edgeR and DESeq2, though only edgeR log_2FC values are displayed. The heatmap compares log_2FC values of secondary phloem (2P) and secondary xylem (2X) of *G. clavigera*- (Fungus) vs. mock -inoculated (Wound) lodgepole pine grown under low or high N availability. Blue indicates up-regulation and red indicates down-regulation. Color intensity indicates the degree of fold change in transcript abundance between *G. clavigera*- and mock-inoculated samples. DEGs found to be significantly DE by both edgeR and DESeq2 (adjusted p-value<0.01 and $|log_2FC| \geq 2$) are marked with asterisks.

4.3.9 Hub gene identification and network visualization

Network analyses of transcriptome data allow for the identification of putative regulons (Mentzen and Wurtele 2008; McClure *et al*. 2016). Regulons are groups of genes whose expression are coordinately regulated in response to environmental or developmental cues (Mentzen and Wurtele 2008). Hub genes – genes showing the greatest degree of connectivity with other genes within a given network – can be used to identify putative regulons (Blais and Dynlacht 2005). Hub genes that represent transcription factors or other regulators of gene expression can offer new insights into the regulation of these putative regulons, and by extension, the plant's response to the environmental or developmental cue under investigation (Blais and Dynlacht 2005).

Accordingly, a network approach was used to identify hub genes and their putative regulons invoked in lodgepole pine's response to *G. clavigera* inoculation under low vs. high N availability. Orange module and dark cyan module hub genes were detected using intramodular connectivity and module membership values, described in Section 4.2.12. A total of 118 hub genes and 207 hub genes were found in the orange and dark cyan modules, respectively. The top 20 hub genes with the largest intramodular connectivity values (i.e. were co-expressed with the greatest number of module genes) in the orange and dark cyan modules are given in Tables 4.9 and 4.10, respectively. The 20 orange module hub genes in this table were significantly DE in response to *G. clavigera* inoculation in at least one tissue, and many had annotations consistent with direct roles in defense (Table 4.9). In

contrast, none of the dark cyan hub genes were significantly DE in response to *G. clavigera* inoculation, and few of the annotations suggested direct roles in defense (Table 4.10). Consequently, only significantly DE hub genes of the orange module were investigated further.

Table 4.9. Orange module hub genes identified with WGCNA correlation statistics. Detected orange module hub genes had the highest intramodular connectivity to all other orange module genes and a sufficiently high module membership value (>0.80). Hub genes were annotated with the NCBI RefSeq database and according to the functional subcategories outlined in Table 4.1. Differential expression was detected using edgeR and the cutoff for designation as significantly DE was a Benjamini-Hochberg adjusted p-value<0.01 and |log2FC|≥2.

Gene	Contig ID	Category	Key word	Intramodular connectivity	Module membership	LowN 2P Fungus vs Wound	HighN 2P Fungus vs Wound	LowN 2X Fungus vs Wound	HighN 2X Fungus vs Wound
cellulose synthase A7	A0.A1.A6.R10268425	CW	synthesis	2045.61	0.82	up-regulated	not significant	not significant	not significant
WRKY 31	A0.A0.A22.R10730181	TF	WRKY	1457.98	0.98	up-regulated	up-regulated	up-regulated	up-regulated
alcohol dehydrogenase 1	A0.A1.A16.R7477978	Other	NA	1424.99	0.97	up-regulated	not significant	up-regulated	up-regulated
ammonium transporter 2	A0.A0.A19.R15384791	N-transport	ammonium	1419.60	0.97	up-regulated	not significant	up-regulated	up-regulated
JAZ 9	A0.A1.A45.R16247440	TF	JAZ	1401.80	0.98	up-regulated	up-regulated	up-regulated	up-regulated
(E)-β-ocimene synthase	A0.A1.A49.R9298939	C	terpenoid	1391.49	0.97	up-regulated	up-regulated	up-regulated	up-regulated
C3HC4-type RING finger	A0.A1.A30.R8666338	Other	NA	1374.57	0.97	not significant	not significant	up-regulated	up-regulated
WRKY 31	A0.A1.A55.R13668950	TF	WRKY	1373.83	0.98	up-regulated	up-regulated	up-regulated	up-regulated
(E)-β-ocimene synthase	A0.A1.A21.R18311935	C	terpenoid	1369.86	0.95	up-regulated	up-regulated	up-regulated	up-regulated
late embryogenesis abundant	A0.A1.A55.R13674086	Other	NA	1369.76	0.97	up-regulated	up-regulated	up-regulated	up-regulated
HXXXD-type acyl-transferase	A0.A1.A31.R7898913	Other	NA	1367.65	0.95	up-regulated	up-regulated	up-regulated	up-regulated
(E)-β-ocimene synthase	A0.A0.A18.R16758902	C	terpenoid	1364.61	0.954	up-regulated	up-regulated	up-regulated	up-regulated
plant cadium resistance 2	A0.A0.A62.R9736383	Other	NA	1364.39	0.95	up-regulated	up-regulated	up-regulated	up-regulated
receptor kinase with LRR domain	A0.A1.A62.R9782082	S	receptor kinase	1361.91	0.98	up-regulated	up-regulated	up-regulated	up-regulated
glucose-methanol-choline oxidoreductase	A0.A0.A52.R15193491	Other	NA	1361.29	0.97	up-regulated	up-regulated	up-regulated	up-regulated
receptor kinase with LRR domain	A0.A0.A16.R21420069	S	receptor kinase	1361.22	0.97	up-regulated	up-regulated	up-regulated	up-regulated
oxophytodienoate-reductase 3	A0.A1.A27.R11540461	H	JA	1356.54	0.97	up-regulated	not significant	up-regulated	up-regulated
C3HC4-type RING finger	A0.A1.A20.R20696961	Other	NA	1354.11	0.97	up-regulated	not significant	up-regulated	up-regulated
IAA carboxylmethyltransferase 1	A0.A1.A15.J11080264	H	auxin	1351.22	0.96	up-regulated	not significant	up-regulated	not significant
phosphoinositide phosphatase	A0.A1.A21.R18348468	Other	NA	1349.84	0.97	up-regulated	not significant	up-regulated	up-regulated

Table 4.10. Dark cyan module hub genes identified with WGCNA correlation statistics. Detected dark cyan module hub genes had the highest intramodular connectivity to all other dark cyan module genes and a sufficiently high module membership value (>0.80). Hub genes were annotated with the NCBI RefSeq database and according to the functional subcategories outlined in Table 4.1. Differential expression was detected using edgeR and the cutoff for designation as significantly DE was a Benjamini-Hochberg adjusted p-value<0.01 and |log2FC|≥2.

Gene	Contig ID	Category	Key word	Intramodular connectivity	Module membership	LowN 2P Fungus vs Wound	HighN 2P Fungus vs Wound	LowN 2X Fungus vs Wound	HighN 2X Fungus vs Wound
pectin lyase	A0.A0.A13.R20740520	CW	degradation	1984.42	0.96	not significant	not significant	not significant	not significant
pectin lyase	A0.A1.A59.R10150375	CW	degradation	1977.14	0.96	not significant	not significant	not significant	not significant
pectin lyase	A1.A14.R21488629	CW	degradation	1932.26	0.96	not significant	not significant	not significant	not significant
heat shock protein 90.1	A0.A0.A54.R11679805	Other	NA	1911.99	0.96	not significant	not significant	not significant	not significant
late embryogenesis abundant hydroxyproline-rich glycoprotein	A0.A0.A8.R14957266	Other	NA	1902.18	0.95	not significant	not significant	not significant	not significant
peroxidase	A0.A0.A35.R8236277	N	peroxidase	1894.79	0.97	not significant	not significant	not significant	not significant
xyloglucan endotransglucosylase/hydrolase 10	A0.A1.A1.J14945512	CW	modification	1889.60	0.95	not significant	not significant	not significant	not significant
thaumatin	A0.A0.A6.R2618409	N	thaumatin	1889.02	0.95	not significant	not significant	not significant	not significant
xyloglucan endotransglucosylase/hydrolase 9	A1.A23.R15164872	CW	modification	1877.04	0.95	not significant	not significant	not significant	not significant
sucrose synthase 3	A0.A1.A24.R16271572	Other	NA	1873.61	0.95	not significant	not significant	not significant	not significant
glycosyl hydrolase 9A4	A0.A1.A60.R12852418	CW	degradation	1871.29	0.95	not significant	not significant	not significant	not significant
peptide transporter 2	A1.A31.R20529385	Other	NA	1870.63	0.96	not significant	not significant	not significant	not significant
heat shock protein 3	A1.A39.R13363549	Other	NA	1865.92	0.95	not significant	not significant	not significant	not significant
SKU5	A1.A38.R11321379	CW	NA	1865.85	0.96	not significant	not significant	not significant	not significant
laccase 12	A1.A20.R16464529	C	lignin	1862.51	0.95	not significant	not significant	not significant	not significant
actin-11	A0.A0.A27.S10677287	Other	NA	1857.24	0.95	not significant	not significant	not significant	not significant
cotton fibre expressed protein	A0.A0.A0.R14350897	Other	NA	1856.97	0.96	not significant	not significant	not significant	not significant
gamma-tocopherol methyltransferase	A1.A4.R19726908	C	isoprenoid	1852.21	0.97	not significant	not significant	not significant	not significant
Rab GTPase homolog B18	A0.A1.A49.J9370194	Other	NA	1850.49	0.96	not significant	not significant	not significant	not significant
CYP704A	A0.A0.A39.R5874899	C	P450	1845.59	0.96	not significant	not significant	not significant	not significant

Cytoscape (Shannon *et al.* 2003) was used to visualize the 3,408 significantly DEGs in the orange module that met a stringent correlation threshold of 0.96 out of 1.00 (Figure 4.24). A subnetwork composed of nodes connected to four hub genes annotated as (E)-β-ocimene synthase, JAZ9 and two WRKY31 transcripts - contig IDs A0.A1.A21.R18311935, A0.A1.A45.R16247440, A0.A1.A55.R13668950, A0.A0.A22.R10730181, respectively - was produced to identify additional members of putative regulons associated with these contigs (Figure 4.25). The subnetwork contained regulators of gene expression, including other transcription factors, C-based defense-associated genes, sugar transporters and defense-related signaling genes. The expression profile for each subnetwork gene was displayed in a heatmap (Figure 4.26). The four hub genes were significantly DE in both xylem and phloem in response to *G. clavigera* inoculation under both 0.3 mM and 10 mM NH_4NO_3 conditions, but log_2FC values were greater in low N than high N conditions for all four hub genes. A similar trend existed for the other subnetwork genes, in that the low N treatment elicited larger log_2FC values when compared with the high N treatment. Many subnetwork genes were solely significantly DE under 0.3 mM NH_4NO_3 conditions.

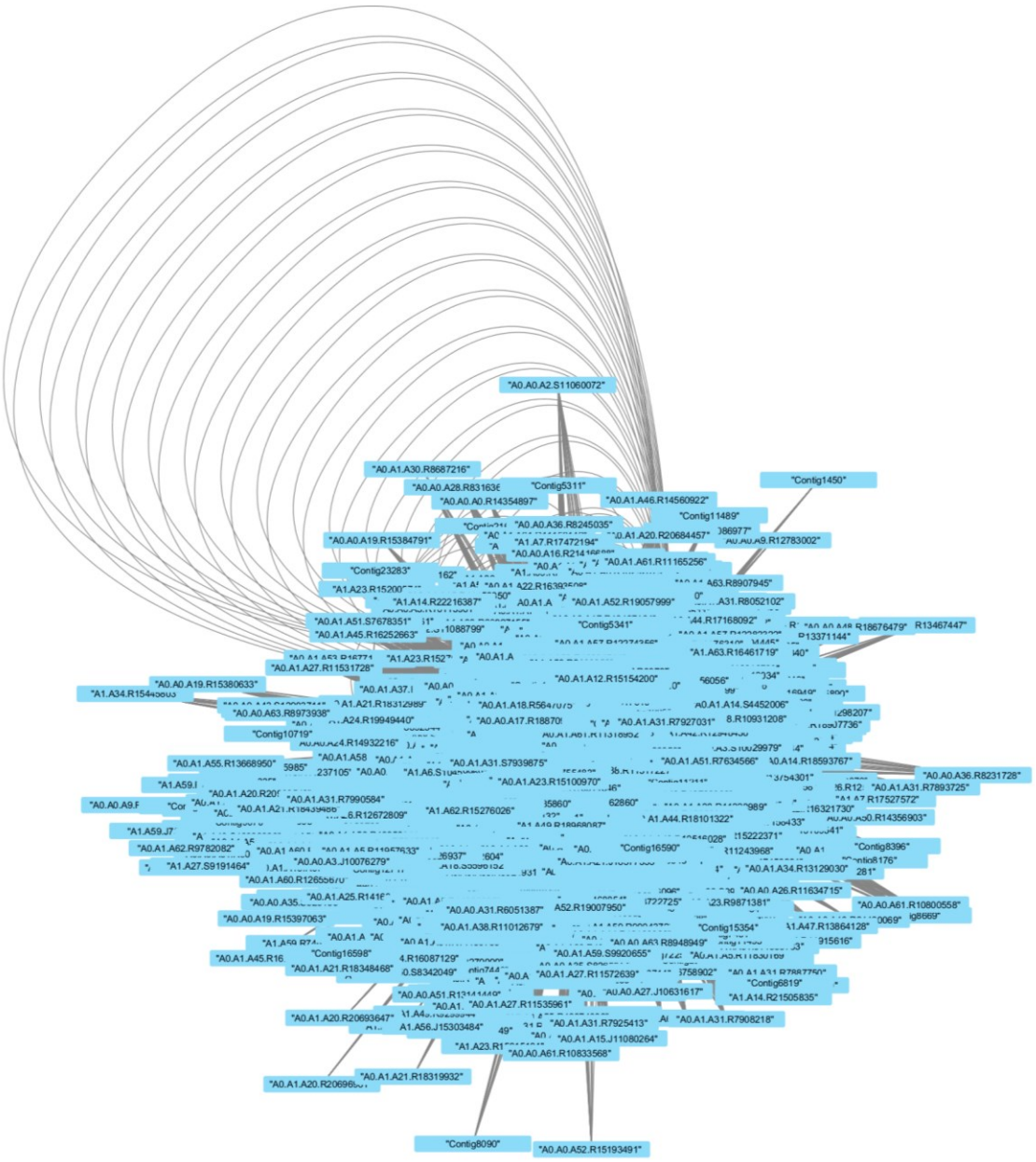


Figure 4.24. Orange module co-expression network. Cytoscape allowed for visualization of 795,816 edges connecting 3,048 co-expressed significantly DEGs. Nodes were connected by WGCNA using biweight midcorrelation. Edges displayed met a stringent correlation threshold of 0.96.

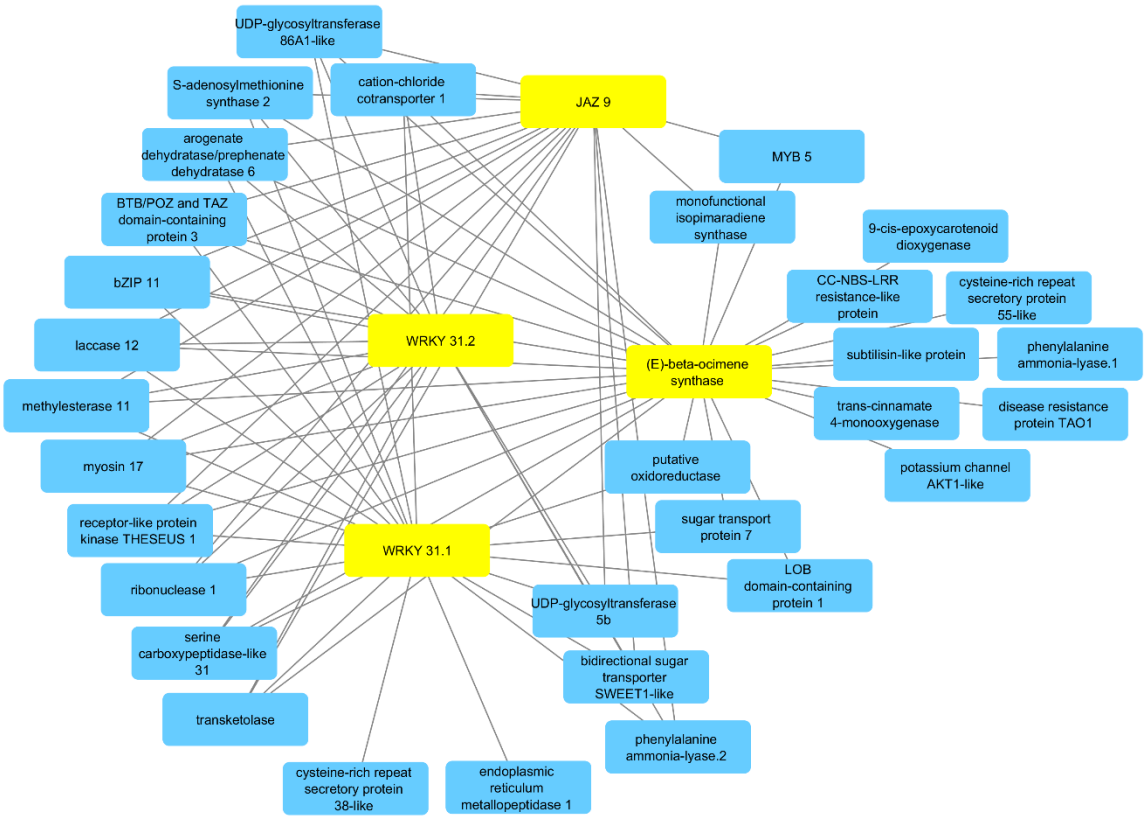


Figure 4.25. Orange module co-expression subnetwork with four hub genes. One (E)-β-ocimene synthase transcript (contig ID A0.A1.A21.R18311935), one JAZ9 transcript (contig ID A0.A1.A45.R16247440) and two WRKY31 transcripts termed WRKY 31.1 (contig ID A0.A1.A55.R13668950) and WRKY 31.2 (contig ID A0.A0.A22.R10730181) were identified as hub genes by WGCNA and are highlighted in yellow. Cytoscape allowed for visualization of 35 co-expressed genes associated with these hub genes. Edges displayed met a stringent correlation threshold of 0.96.

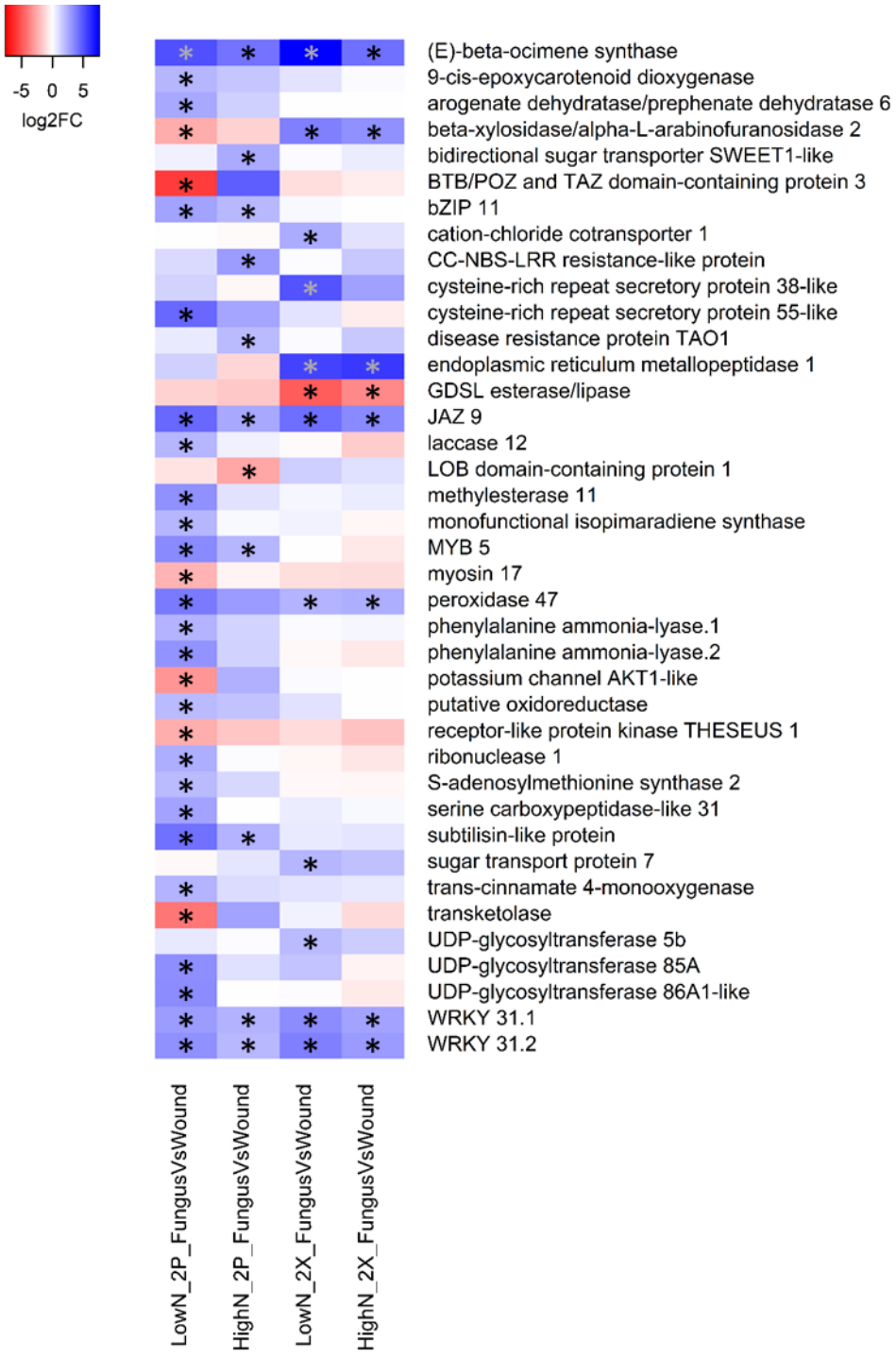


Figure 4.26. Orange module subnetwork genes were significantly DE in both xylem and phloem of *G. clavigera*-inoculated lodgepole pine seedlings relative to mock-inoculated controls. Genes in the orange module subnetwork

showcased in Figure 4.26 were identified as core DEGs using edgeR and DESeq2, though only edgeR log_2FC values are displayed. The heatmap compares log_2FC values of secondary phloem (2P) and secondary xylem (2X) of *G. clavigera*- (Fungus) vs. mock -inoculated (Wound) lodgepole pine grown under low or high N availability. Blue indicates significantly greater transcript abundance in *G. clavigera*-inoculated seedlings relative to mock-inoculated seedlings, with color intensity representing the degree at which differential expression occurred as measured by log_2FC. DEGs found to be significantly DE by both edgeR and DESeq2 (adjusted p-value<0.01 and $|log_2FC| \geq 2$) are marked with asterisks.

4.4 Discussion

While there have been a plethora of ecological studies with many plant species examining the relationship between N availability and plant defense in the context of plant defense theories, there have been relatively few investigations of how N availability affects defense responses at the level of gene expression in either angiosperms or gymnosperms. Accordingly, the objectives of this study were: (1) to use transcriptomics to determine whether contrasting levels of N availability altered lodgepole pine patterns of gene expression in response to inoculation with *G. clavigera*, with an emphasis on defense-associated genes, (2) to relate these changes in gene expression with quantification of monoterpenes, a well-characterized class of defense compounds, and (3) to determine whether the documented N-associated shifts to *G. clavigera*-responsive gene expression were consistent with one or more plant defense theories. In particular, we investigated whether genomic-scale patterns of gene expression provided evidence for

increased levels of N availability leading to (1) reduced allocation of resources towards defense, and (2) reduced ratios of C-based to N-based defenses.

4.4.1 Assessing the physiological relevance of the N fertilization treatments chosen for this study

Conifers are known for their conservative uptake of soil N (Mekonnen *et al.* 2019) and for prioritizing N storage over generalized growth enhancement (Millard and Gretlet 2010; Palma *et al.* 2020). Conifers grow successfully in nutrient poor and nutrient rich soils (Brockley 2001). The 33.3-fold difference in NH_4NO_3 fertilization levels relied upon in our study was consistent with the variation of soil N levels that may be found in typical field sites in western Canada (Prescott and Preston 1994; Köchy and Wilson 2005). In higher plants like conifers, N is mainly stored in the photosynthetic protein ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) (Peoples and Gifford 1990). Consequently, foliar N levels tend to be higher than in other tissues and also tend to be more responsive to soil N availability (Peoples and Gifford 1990). Measuring foliar N content is therefore an often-used indicator of a plant's N status (Muñoz-Huerta *et al.* 2013). N remobilization is seasonally programmed (Millard and Gretlet 2010), and N-rich storage proteins in stem tissues may need to be considered when comparing N concentrations later in the season (Wetzel *et al.* 1989). However, given the phenological stage of our trees, foliar N measurements were most appropriate. Foliar N levels were significantly different between the 0

mM, 0.3 mM and 10 mM NH_4NO_3-treated plants, indicating that these treatments were biologically relevant. Although foliar N levels were higher in 10 mM NH_4NO_3-treated plants than either 0 mM or 0.3 mM NH_4NO_3 at every time point, for the most part the pairwise differences were only statistically significant at 28 dpi. This finding differs with similarly-conducted experiments – such as the experiment described in Chapter 3 – where contrasting N fertilization levels led to significantly different foliar N levels at earlier time points. Slight differences between the two experimental designs, detailed in Sections 3.2.1-3.2.2.2 and 4.2.1, along with a relatively low number of biological replicates, may have influenced the discrepancy between the foliar N concentration data. Physiologically, significant differences were detected in foliar monoterpene concentrations as early as 7 dpi, indicating that defense-related physiological and biochemical processes were impacted by the contrasting N treatments at this time point.

Several factors likely contributed to our foliar N results. First, our trees were light-limited in the growth chambers used for this study. Light controls the activity of nitrate reductase and, directly or indirectly, provides the reducing power necessary for the reductive incorporation of nitrate into amino groups (Kaiser *et al.* 1999). Photosynthetic and respiratory carbon metabolism is also required to generate the carbon skeletons necessary for amino acid synthesis (Champigny 1995). Therefore, suboptimal levels of photosynthesis may have reduced the seedlings ability to assimilate inorganic N found in our fertilization treatments. The foliar N results also suggest that it may have been prudent to treat the

seedlings with contrasting N fertilization treatments for a longer period before inoculating the seedlings with *G. clavigera*. However, there were several factors that needed to be balanced in choosing an appropriate inoculation date. For example, to ensure that any observed N effect was not due to differential seedling establishment following replanting, seedling regrowth following storage and repotting needed to be established prior to application of the low and high N treatment. The presumed N status of the plants was also balanced against the phenological stage of the seedlings. Considerable physiological and development changes occur to perennials during the course of the growing season, and these can affect the defense response (Larisch *et al.* 2012; Galindo-González *et al.* 2012). Given that the effects of phenology on defense responses is largely unknown, our standard practice is to inoculate seedlings with *G. clavigera* when they are at the same phenological stage as mature trees during the MPB attack window. Using this metric, if plants had been fertilized with contrasting N levels for nine weeks rather than five weeks before inoculation, they would have been past the phenologically-relevant window of MPB attack in the field. Another factor that needed to be considered was the time point following inoculation when plants were harvested for RNA-Seq. In order to capture expression of genes associated with the acute induced defense response rather than genes involved in the more sustained containment response, previous gene expression analyses of *G. clavigera*-inoculated lodgepole pine seedlings using microarray data and reverse transcription quantitative polymerase chain reaction (RT-qPCR) suggested that 7

dpi was a more appropriate time point for gene expression analyses than either 14 or 28 dpi (Arango-Velez *et al*. in prep).

4.4.2 Lodgepole pine responds to *G. clavigera* inoculation using C- and N-based defense mechanisms

This study hinged upon the application of differential expression analysis to provide a global picture of how N availability affected lodgepole pine gene expression patterns in response to *G. clavigera* attack. The number of DEGs detected in either 0.3 mM vs. 10 mM NH_4NO_3 *G. clavigera*-inoculated or 0.3 mM vs. 10 mM NH_4NO_3 mock-inoculated comparisons was very low in both xylem and phloem, particularly when the data were analyzed using edgeR. This is consistent with the detectable but not significant differences in foliar N concentration between 0.3 mM NH_4NO_3- and 10 mM NH_4NO_3-treated seedlings. However, the Venn diagram analysis comparing all four *G. clavigera*- vs. mock-inoculated treatments revealed both shared and distinct DEGs in 0.3 mM NH_4NO_3- and 10 mM NH_4NO_3-treated plants. Therefore, even though the N fertilization treatments used in this study were insufficient to drive significant differences in foliar N concentration at 7 dpi – values that in large part reflect differential accumulation of RuBisCO – they were sufficient to modulate seedling responses to *G. clavigera* inoculation.

Mining annotations of these shared and distinct DEGs revealed several DEGs classically associated with defense responses of conifers. The up-regulation

of C-based defense genes putatively involved in biosynthesis of phenylpropanoids, flavonoids, terpenoids and sulfur-containing defense compounds was accompanied by the induction of N-based defense genes, such as β-1,3-glucanases, chitinases, peroxidases, osmotin-like proteins, thaumatins and other pathogenesis-related (PR) proteins (Franceschi *et al.* 2005; Van Loon *et al.* 2006). Genes encoding putative transcription factors and other regulators of plant defense responses were also identified as DEGs.

The finding that expression profiles of some putative C-defense associated genes were significantly altered by N availability, while expression profiles of other putative C-defense associated genes were not is not inconsistent with published literature investigating the influence of N availability on defense-associated gene expression and secondary metabolite synthesis in angiosperms and gymnosperms (Fagard *et al.* 2014; Sun *et al.* 2020). Fertilization has been shown to influence the production of C-based secondary metabolites in conifers, in some cases resulting in an increase of concentrations (Björkman *et al.* 1991) and in other cases a decrease of concentrations (Koricheva and Keinänen 1998; Tomova *et al.* 2005). For example, Tomova *et al.* (2005) discovered a decrease of concentrations of phenolic compounds in the roots of Norway spruce (*Picea abies* Miller) in response to NH_4NO_3 fertilization with increasing concentrations. Alternatively, Björkman *et al.* (1991) showed that fertilization increased resin concentrations in the foliage of field grown Scots pine (*Pinus sylvestris* Loureiro) when compared with untreated trees.

Differential N availability also has been shown to impact defense protein production (Sun *et al.* 2020). Dietrich *et al.* (2004) found that low N conditions reduced the production of *Arabidopsis thaliana* Linnaeus peroxidases and chitinases. Xuehui *et al.* (2014) tested the effects of four distinct levels of urea fertilization on *Solanum tuberosum* (Peoppig ex Walpers) exposed to the fungal pathogen *Phytophthora infestans* (Mont.) de Bary. They found that the highest fertilization application induced the largest increase in chitinase activity 1, 2 and 3 days post infection (Xuehui *et al.* 2014). However the third highest fertilization application resulted in the largest increase in chitinase activity by day 4, indicating that there was an optimal level of fertilization that was neither too high, nor too low to best counter *P. infestans* with N-based defenses (Xuehui *et al.* 2014).

4.4.3 Terpene synthase and chitinase genes implicated in lodgepole pine defense against *G. clavigera*

DEGs encoding putative C-defense compound synthesizing terpene synthases and N-defense chitinases figured prominently in the Venn diagram and WGCNA (Langfelder and Horvath 2008) network data mining of the RNA-Seq data. Many of these DEGs were significantly up-regulated in response to *G. clavigera* in both tissues under both N treatments. Given the important and well-characterized roles that terpenoid compounds and chitinases play in conifer defense against pests and pathogens (Kovalchuk *et al.* 2013; Kolosova *et al.* 2014; Keefover-Ring *et al.* 2015; West *et al.* 2016), DEGs annotated as belonging to these

two functional categories were investigated further. Relationships between putative terpene synthases and the monoterpenes that were measured as part of this experiment were also examined.

4.4.3.1 Terpene synthases and relationships to monoterpene profiles

Terpenoids form the largest class of plant metabolites (Bohlmann and Keeling 2008) and possess antimicrobial and anti-feedant properties (Keeling and Bohlmann 2006; Block *et al.* 2019). Diterpenes, also known as resin acids (Kovalchuk *et al.* 2013), and sesquiterpenes differ from volatile monoterpenes by the number of C atoms. Where monoterpenes have 10 Cs, diterpenes have 20 Cs and sesquiterpenes have 15 Cs (Steele *et al.* 1998). All three terpenoid classes are found in oleoresin, an essential chemical and physical conifer defense strategy that is both constitutively produced and induced in response to intruding MPB and vectored pathogenic fungi (Franceschi *et al.* 2005; Celedon and Bohlmann 2019). Volatile monoterpenes additionally serve as olfactory cues for MPB, mediating host-herbivore interactions (Chiu *et al.* 2017; Chiu *et al.* 2019).

DEGs annotated as 21 different terpene synthases were DE in response to *G. clavigera* inoculation in both phloem and xylem of 0.3 mM and 10 mM NH_4NO_3-treated trees, including mono-, di- and sesquiterpene synthases. Monoterpene synthases are the best studied of the terpene synthases. A number of monoterpenes synthases putatively involved in synthesizing monoterpenes known to play a role in MPB-pine interactions were DE in response to *G. clavigera*

inoculation, including sequences encoding for putative (+)-3-carene synthase, (+)-α-pinene synthase, (-)-β-pinene synthase, (-)-β-phellandrene synthase, alpha terpineol/1,8-cineole synthase and (E)-β-ocimene synthase (Chiu *et al.* 2017). While most of these monoterpene synthases showed significant down-regulation in at least one of the four comparisons in response to *G. clavigera* inoculation, (E)-β-ocimene synthase was significantly up-regulated, most strongly in 0.3 mM NH_4NO_3-treated seedlings. (E)-β-ocimene synthase was also identified as a hub gene in the WGCNA (Langfelder and Horvath 2008) analyses, indicating that this expression profile is shared with a comparatively large number of genes.

Overall, in phloem, significant differences in monoterpene synthases expression profiles were not generally matched with significant differences in monoterpene metabolite concentrations, with the exception of β-phellandrene synthase/β-phellandrene. However, there is partial or good agreement between gene expression patterns and trends in metabolite concentrations for α-pinene synthase/(-)-α-pinene, β-pinene synthase/(-)-β-pinene, 3-carene synthase/3-carene and (E)-β-ocimene synthase/cis-ocimene, although phloem metabolite concentration differences between *G. clavigera*- and mock-inoculated samples for these compounds were not significant. This likely reflects the high variability between biological replicates that was observed in monoterpene concentrations. It may also reflect the relatively low abundance of compounds like 3-carene and cis-ocimene relative to compounds like beta-phellandrene (Smith 2000), where

variability at the low range of concentrations make it more challenging to detect significant differences.

In a study testing the toxicity of different monoterpenes to MPB, Chiu *et al.* (2017) demonstrated that β-phellandrene, (-)-α-pinene, (-)-β-pinene, and 3-carene synthase are moderately toxic relative to the most toxic (-)-limonene and the least toxic terpinolene. Because of this moderate toxicity and overall lack of statistical significance between monoterpene profiles of 0.3 mM NH_4NO_3- and 10 mM NH_4NO_3-treated trees, we predict that the degree of N-mediated changes in phloem monoterpene concentrations elicited in this study are unlikely to affect lodgepole pine defense against MPB and its fungal associate *G. clavigera*.

Perhaps because cis-ocimene is often present in low concentrations in conifers (Smith 2000), the effect of this monoterpene on MBP-pine interactions has received less attention than other monoterpenes. However, Huber *et al.* (2000) showed that cis-ocimene can act as a kairomone, a pine-produced chemical that can be detected by MPB via their antennae. In *A. thaliana*, exposure to allo-ocimene has been found to trigger plant defense responses via the signaling pathways of salicylic acid, JA, and ethylene (Kishimoto *et al.* 2006). Kishimoto *et al.* (2006) found that allo-ocimene applied to *A. thaliana* leaves inhibited *Botrytis cinereal* hyphal growth by inducing accumulation of antifungal substances like phytoalexins and increased lignification of infected tissues. β-Ocimene can also influence plant-pest interactions, such as restricting the feeding behavior of aphids (*Myzus persicae* Sulzer) on Chinese cabbage (*Brassica pekinensis* Skeels) (Kang

et al. 2018). Our results indicate that further investigation is warranted for conifer species to fully understand how ocimene affects the growth of pathogenic fungi and influences bark beetle behavior.

A greater number of foliar and volatile monoterpenes showed significant differences between treatments than phloem monoterpenes, with α-pinene, β-pinene, camphene, and β-phellandrene all showing significant differences between *G. clavigera*-inoculated and control samples for at least one time point. This leads to the speculation that N fertilization has a greater impact on defenses in foliar tissues than in stem tissues. Interestingly, concentrations of volatile α-pinene, β-pinene and camphene all significantly increased in *G. clavigera*-inoculated vs. control seedlings, and only in the 0.3 mM NH_4NO_3-treated seedlings. In contrast, α-pinene, β-pinene and β-phellandrene concentrations were all significantly lower in foliage of *G. clavigera*-inoculated vs. control seedlings, while camphene concentrations significantly increased. Therefore, the volatile mix of monoterpenes, which largely arise from the foliage, do not necessarily reflect the mix of monoterpenes within the foliage itself. VOCs, including volatile monoterpenes, are particularly important for MPB-pine interactions, including the influence of MPB behavior when dispersed beetles are seeking new hosts (Chiu *et al.* 2019). Baradat and Yazdani (1988) found that α- and β-pinenes were more affected by environmental factors than other monoterpenes. Furthermore, (+)-α-pinene is a precursor chemical that female MPB utilize to make the aggregation pheromone (-)-trans-verbenol, which attracts dispersed beetles to an attacked tree

(Bentz *et al.* 2005; Chiu *et al.* 2019). This aggregation is essential for successful mass attack and subsequent colonization of the pine host (Bentz *et al.* 2005; Chiu *et al.* 2019). Therefore, we predict that increasing N availability may reduce α-pinene in VOCs emitted by lodgepole pine, perhaps reducing the efficacy of MPB mass attack of these trees. Field experiments can address these predictions.

Monoterpene VOCs were also detected in phloem tissue, and redundancy analysis did not result in a separation between phloem samples along PC1 or PC2. Putative monoterpene coding in phloem and xylem tissue, however, revealed that (+)-α-pinene had a greater association with mock-inoculated trees, and (E)-β-ocimene had a strong affiliation with *G. clavigera*-inoculated trees. This resulted in the up-regulation of (E)-β-ocimene synthase and the down-regulation of (+)-α-pinene synthase in response to fungal inoculation. Significant increases in the amount α-pinene produced in stem tissue is characteristic of lodgepole pine after a mechanical injury (Marpeau *et al.* 1989). Alternatively, local increases in Scots pine (+)-α-pinene in response to wounding was significantly lower than in response to a fungal complex that included the pathogenic blue-stain fungus *Ophiostoma brunneociliatum* Mathiesen-Käärik, indicating that (+)-α-pinene was more associated with fungal inoculation (Villari *et al.* 2012). Contradicting evidence in literature shows that reliance on monoterpene mechanisms for defense are host and pathogen specific.

Several diterpene and sesquiterpene synthases were also DE in fungal-inoculated lodgepole pine seedlings, with DEGs exhibiting both significant up- and

down-regulation. As major components of resin, this finding suggests that the diterpene and sesquiterpene composition of the plant's resin may have been altered in response to *G. clavigera* infection. As our study quantified only monoterpenes, no metabolite data are available to validate this speculation. Interestingly, more of these diterpene and sesquiterpene synthases were significantly DE in phloem (bark) – the tree's defensive barrier against stem-invading pests and pathogens – than in xylem, and more of them were significantly DE in the 0.3 mM NH_4NO_3-treated seedlings than the 10 mM NH_4NO_3-treated seedlings. This suggests that *G. clavigera*-induced changes in bark resin composition were greater in the low N-treated seedlings than the high N-treated seedlings if these gene expression patterns are reflected in metabolite profiles of di- and sesquiterpenes.

DE diterpene synthases included monofunctional diterpene synthase, which possesses one enzymatic active site that produces a cyclic isomer of the substrate geranylgeranyl diphosphate (Hall *et al.* 2013). Monofunctional isopimaradiene synthase, which also was significantly DE in at-least one *G. clavigera*-inoculated vs. mock-inoculated comparison, is a specific type of monofunctional diterpene synthase that provides the hydrocarbon precursors for most other resin acids (Keeling *et al.* 2008). Arango-Velez *et al.* (2014) found that lodgepole pine β-farnesene synthase was significantly up-regulated in secondary phloem in response to inoculation with *G. clavigera* under well-watered conditions. Similarly, our study found that β-farnesene synthase was significantly

up-regulated in secondary phloem by inoculation with *G. clavigera* under low N conditions. Our results also included the significant up-regulation of α-bisabolene synthase in the secondary phloem of inoculated trees grown under low N conditions. α-Bisabolene synthase is induced by wounding in grand fir (*Abies Grandis* Douglas ex D. Don) stem tissue (Bohlmann *et al.* 1998).

4.4.3.2 Chitinases

Plant chitinases differ in their biological and biochemical properties and are divided into seven classes- classes I through VII - based on deoxyribonucleic acid (DNA) sequence and structural motifs (Neuhaus 1999). A recent rigorous analysis by Peery *et al.* (in review) of the chitinase superfamily in the lodgepole pine master transcriptome, used as the reference for RNA-Seq in this study, identified 134 putatively distinct chitinases belonging to classes I-V and VII. In this study, we identified 56 class I, IV, V and VII chitinase sequences that were significantly induced by *G. clavigera* inoculation in lodgepole pine seedlings. More chitinases were significantly DE in xylem than phloem, and more were DE in response to 0.3 mM NH_4NO_3 than 10 mM NH_4NO_3. Of these 56 significantly DEGs, 55 were DE in xylem of 0.3 mM NH_4NO_3-treated seedlings, 43 were DE in xylem of 10 mM NH_4NO_3-treated seedlings, 40 were DE in phloem of 0.3 mM NH_4NO_3-treated seedlings, and 24 in phloem of 10 mM NH_4NO_3-treated seedlings. The vast majority of these DEGs were defense-associated class I, IV and VII chitinases. Class V chitinases were only significantly DE in 0.3 mM NH_4NO_3-treated

seedlings, while very few class VII chitinases were significantly DE in 10 mM NH_4NO_3-treated seedlings. Although a subset of these sequences is likely to be redundant (i.e. represent the same gene), these numbers illustrate the degree to which chitinases are invoked during the conifer defense response. The greater number of lodgepole pine chitinases significantly DE under low N conditions contrasts with the findings of Sun *et al.* (2020), who linked increased N availability with increased chitinase synthesis in crops. A proteomics analyses of these samples would be an appropriate means by which to validate the extent to which this wholesale up-regulation of chitinase transcripts in response to *G. clavigera* translates into increased chitinase protein accumulation in these tissues.

4.4.4. Using transcriptomics data to evaluate plant defense theories

Upon attack by a pest or pathogen, plants have a finite pool of resources to allocate towards induced defenses (Stamp 2003). This cost of constitutive and induced defenses is drawn from the plant's total resource economy, and thus resources that are directed towards defense are typically unavailable for other processes, such as growth (Stamp 2003). As outlined in the introduction to this chapter, a number of defense theories have been proposed to explain observed patterns of resource allocation to growth vs. defense processes (Stamp 2003). Given that N is often a limiting nutrient to optimal plant growth, the impact of N on growth vs. defense has received particular attention (Stamp 2003). These theories have been tested mainly at the ecological level. The RNA-Seq data set

generated for this study provided a unique opportunity to explore how broad-scale patterns of gene expression can be used to assess to what degree these theories are upheld at the molecular level. Accordingly, annotations were used to classify core DEGs as defense-related enzymes synthesizing C-based products, N-based defense proteins and several other categories relevant to assessing these plant defense theories.

In some instances, data mining identified DEGs that were annotated the same and showed similar expression patterns. Multiple transcripts that have a high percentage of overlap and only differ where an intron was retained in one transcript and not the other can result in distinct contigs (Hölzer and Marz 2019). Therefore, a certain amount of redundancy in the differential expression analysis results likely resulted from different isoforms of the same gene assembled into multiple contigs within the master transcriptome. Though this undoubtedly occurred, the probability of occurrence in one category of genes versus another is unknown. We therefore proceeded with the data as it was when addressing the following plant defense hypotheses.

4.4.4.1 The CNB hypothesis

The CNB hypothesis states that there is both a genetic and environmental component to resource allocation towards growth, defense and other processes (Bryant *et al.* 1983; Tuomi *et al.* 1988; Tuomi *et al.* 1991). This hypothesis generally considers the genetic component of resource allocation to be comparatively fixed,

whereas the environmental component of resource allocation is flexible. Based on this foundation, the CNB hypothesis states that the flexible component of resource allocation to growth vs. defense is influenced by environmental factors that impact the sizes of C and nutrient resource pools relative to each other (Matyssek *et al.* 2002; Matyssek *et al.* 2005). Placed more specifically in the context of N resource availability, the CNB hypothesis predicts that environmental factors that decrease the plant's internal (assimilated) C:N ratio will result in an increased proportion of assimilated N available for non-growth processes, such as storage and defense (Stamp 2003). In contrast, environmental factors that increase the plant's internal C:N ratio will result in C, and potentially N, resources in excess of those needed to support growth processes being allocated towards defense and storage (Herms and Mattson 1992).

Under the CNB's central tenet, the phenotypic plasticity of any given plant species will determine the extent to which resource allocation is fixed (genetically determined) vs. flexible (environmentally determined). Therefore, a species' phenotypic plasticity determines the degree to which the plant's assimilated C:N ratio influences the allocation of C and N resources towards defense, potentially impacting the CD:ND ratio. Put another way, the C:N ratio of species with a higher degree of phenotypic plasticity will exert a larger influence over allocation of C and N resources towards defense than the C:N ratio of species with a lower degree of phenotypic plasticity.

The patterns of lodgepole pine resource allocation to CD vs. ND in this study suggest that there is an overall increase in the proportion of *G. clavigera*-induced DEGs classified as ND in both xylem and phloem of 10 mM NH_4NO_3-treated seedlings compared to 0.3 mM NH_4NO_3-treated seedlings. In other words, a decreased C:N ratio was correlated with an increased allocation of N resources towards N-based defenses. In phloem, this was accompanied by a more modest increase in *G. clavigera*-induced DEGs classified as CD. Interestingly, 10 mM NH_4NO_3 conditions concomitantly led to an increased proportion of down-regulated (i.e. *G. clavigera*-repressed) significantly DEGs classified as CD in both xylem and phloem. In the orange module, all C-based and N-based defense associated genes had higher median levels of differential expression (log_2FC) in 0.3 mM compared to 10 mM NH_4NO_3 treatments, suggesting that under low N conditions, trees preferentially allocated resources towards defense, whereas under high N conditions, trees preferentially allocated resources towards growth. These fold-change values support the notion of a higher C:N ratio leading to a greater allocation of resources towards defense.

Taken together, the proportion and fold-change values for DEGs categorized as CD or ND and the overall patterns of gene expression within the RNA-Seq data set are in general agreement with the CNB hypothesis. There is also agreement of volatile monoterpene profiles with the CNB hypothesis, where concentrations of α-pinene, camphene and β-pinene were significantly greater in *G. clavigera*- vs. mock-inoculated 0.3 mM NH_4NO_3-treated seedlings. Given that

the shifts of defense-associated gene expression and metabolite patterns are not strong, we can also infer that lodgepole pine exhibits a comparatively low degree of phenotypic plasticity in resource allocation to defense relative to other species that show stronger trends. This inference is consistent with the extended period of time – nine weeks – required for the differential N treatments to effect significant differences in foliar N concentration values in lodgepole pine in this experiment. By comparison, treatment of *Populus trichocarpa* x *deltoides* saplings with comparable levels of NH_4NO_3-supplemented Hockings nutrient solution led to significant differences in total foliar N after only one week (Cooke *et al.* 2005)

Despite this overall agreement with the CNB hypothesis at the level of global gene expression, a finer-scale inspection of gene expression patterns reveals that profiles of some DEGs do not necessarily align with CNB predictions. For example, DE terpene synthases do not show a uniform pattern of greater fold-change induction in 0.3 mM NH_4NO_3-treated seedlings, and monoterpene profiles in needles and phloem largely do not agree with the CNB hypothesis. Similarly, functional subcategories within the CD and ND categories of the dark cyan module do not show greater fold changes in 0.3 mM NH_4NO_3-treated seedlings than in 10 mM NH_4NO_3-treated seedlings. Taken together, these observations imply that while lodgepole pine's transcriptome response to *G. clavigera* inoculation largely aligns with the CNB hypothesis, the CNB hypothesis does not take into account one or more key factors that shape lodgepole pine's transcriptomic responses to *G. clavigera* inoculation. It must also be noted that data are not presently available

to assess the relative allocation of resources towards growth by 0.3 mM NH_4NO_3- vs. 10 mM NH_4NO_3-treated seedlings. Experiments to compare growth of stems from 1 mM NH_4NO_3- vs. 10 mM NH_4NO_3-treated seedlings are presently underway.

4.4.4.2 The GDB hypothesis

The GDB hypothesis predicts that differentiation processes like cell maturation, which includes synthesis of defense and storage compounds, are favored at low resource availability at the expense of growth, and growth is favored at high resource availability when the need for cellular differentiation processes is low (Herms and Mattson 1992). The lines of evidence presented above in favor of the CNB hypothesis also provide support for the GDB hypothesis. This again implies that the 0.3 mM NH_4NO_3-treated trees exhibited a strengthened defense response to *G. clavigera*.

Relative to the CNB hypothesis, the GDB hypothesis better accounts for the seasonal growth habit of woody perennials like lodgepole pine because this hypothesis considers maturation processes (Lorio 1986). Over the course of the growing season, secondary xylem and secondary phloem of woody perennials undergo a transition from early season to late season growth (Cartenì *et al.* 2018). Within late season xylem, fewer cell divisions occur, tracheary elements mature more quickly and have thicker cell walls, and ray cells accumulate greater levels of secondary metabolites and storage compounds (Larisch *et al.* 2012; Cartenì *et al.*

2018). As the tree transitions to growth cessation, synthesis of secondary metabolites and storage compounds slows, as does cell division, which eventually stops to signal the onset of growth cessation (Cooke *et al.* 2005; Cooke *et al.* 2012). There is a body of evidence suggesting that defense proteins are also synthesized in greater quantities during the transition to growth cessation and dormancy acquisition. For example, chitinase synthesis in conifer needle and stem tissues during the transition to growth cessation and dormancy has been well-documented (Jarząbek *et al.* 2009; Galindo-González *et al.* 2012; Galindo-González *et al.* 2015). Chitinase expression is seasonally regulated in Douglas-fir (*Pseudotsuga menziesii* (Mirbel) Franco) (Zamini *et al.* 2003). Galindo-González *et al.* (2015) found that chitinases accumulated during the white spruce (*Pinus glauca* Moench) transition from active growth to dormancy as constitutive defense proteins.

In addition to serving as defense proteins, chitinases have been demonstrated to serve as vegetative storage proteins in angiosperm species, such as banana (Peumans *et al.* 2002), alfalfa (Meuriot *et al.* 2004) and tamarind (Rao and Gowda 2008). More recently, chitinases have been proposed to serve as vegetative storage proteins in white spruce (Galindo-González *et al.* 2015). Vegetative storage proteins act as temporary reservoirs of amino-acids (Cantón *et al.* 2005; Cooke et al. 2005), accumulating starting in late summer/early autumn (Wetzel *et al.* 1989) and are utilized by trees when they reenter active growth the following spring (Cantón *et al.* 2005).

If we speculate that reduced N availability in the 0.3 mM NH_4NO_3-treated seedlings resulted in a quicker transition to late season growth and growth cessation, then we might expect that the higher fold-change inductions of chitinase expression in response to *G. clavigera* in 0.3 mM NH_4NO_3-treated seedlings relative to 10 mM NH_4NO_3-treated seedlings may have been accentuated by enhanced synthesis of chitinases that we observe in conifers during growth cessation. In addition to enhanced accumulation of chitinases and other defense-associated proteins during the transition to growth cessation, Galindo-González *et al.* (2012) also found that pathways associated with terpenoid production were down-regulated. If 0.3 mM NH_4NO_3-treated seedlings transitioned more quickly to growth cessation than 10 mM NH_4NO_3-treated seedlings, then perhaps the accelerated phenological transition exerted an attenuating effect on *G. clavigera*-induced synthesis of monoterpenes in the 0.3 mM NH_4NO_3-treated seedlings, resulting in the modest changes that we observed in this experiment. To this end, the differential expression results comparing 0.3 mM and 10 mM NH_4NO_3 treatments could be inspected to determine whether there are significant differences in chitinase and monoterpene synthase transcript abundances at 7 dpi. Given that DESeq2 (Love *et al.* 2014) was better able to detect differential expression in the high N vs. low N contrasts, data mining may be better suited for the DESeq2 data rather than relying on the core set of DEGs utilized for this thesis project.

To further test these hypotheses, microscopy samples collected from a duplicate experiment (Chapter 3) are being used to compare cambial zone activity, relative rates of secondary xylem and secondary phloem maturation, and progression towards late season cellular characteristics between 1 mM NH_4NO_3- and 10 mM NH_4NO_3-treated seedlings. Results of these analyses should provide evidence for the developmental trajectories of the *G. clavigera*- and mock-inoculated seedlings grown under 0.3 mM and 10 mM NH_4NO_3. Proteomic analysis of chitinase concentrations in seedlings from these treatments through the full 28 dpi time course would provide evidence as to whether seasonally-associated accumulation of defense proteins was occurring earlier in 0.3 mM NH_4NO_3. Furthermore, metabolomic analysis of monoterpenes could verify if their production was attenuated in the low N-treated seedlings as a result of an earlier transition towards growth cessation. Should these results indicate that N did not impact growth, seasonally-programmed N storage in chitinases or monoterpene production, new plant defense models would need to be explored.

4.4.4.3 The optimal defense hypothesis

The optimal defense (OD) hypothesis states that resource allocation to defend a particular plant part is based on the value of the tissue, the benefit of defense and the probability of attack (Stamp 2003). As they divert resources from growth and reproduction, defensive elements are costly to construct and maintain (Stamp 2003). According to the OD hypothesis, phenotypic plasticity allows plants

to off-set the cost of defensive elements by optimizing the expression of both constitutive and inducible defensive traits in the tissues most vulnerable to pest and pathogen attack (Stamp 2003; Bakhtiari *et al.* 2019). Though we suggest that lodgepole pine has a relatively low degree of phenotypic plasticity in resource allocation to defense, the application of 0.3 mM vs. 10 mM NH_4NO_3 fertilization was sufficient to impact tissue-specific expression patterns of enzymes that synthesize C-based defense-associated compounds and N-based defense proteins and enzymes in fungal-inoculated vs. mock-inoculated seedlings.

We observed that more β-farnesene and α-bisabolene synthases were induced by fungal inoculation in phloem than xylem of low N-treated trees. Metabolomic analysis is needed to assess if the up-regulation of these sesquiterpene synthases resulted in an increase in the production of the sesquiterpene metabolites. Our results may indicate that β-farnesene and α-bisabolene were worth the C cost of production in the phloem tissue, which is the point of entry for fungal invasion (Ballard *et al.* 1982). DiGuistini *et al.* (2011) provided evidence that *G. clavigera* detoxifies mono- and diterpenes in oleoresin, though they did not explore the fungus' ability to tolerate treatment with the sesquiterpenes whose synthases were DE in our system. Though sesquiterpenes are found in smaller amounts in constitutive and induced resin compared with mono- and diterpenes (Keeling *et al.* 2008), research exploring β-farnesene and α-bisabolene toxicity against *G. clavigers* may be warranted.

We detected a greater number of class VII chitinases significantly DE in xylem when compared with phloem under both N regimes. Proteomic analysis is needed to assess if the up-regulation of these chitinase genes resulted in an increase in the production of chitinase proteins. In accordance with the OD hypothesis, we can postulate that *G. clavigera*-inoculated seedlings allocated more N to defense in the xylem compared with the phloem because it is the main tissue colonized by the fungus (Ballard *et al.* 1982). Functioning xylem tissue is vital for water transport throughout the entire plant (Taiz *et al.* 2015), and fungal colonization of xylem tissue can result in host mortality (Ballard *et al.* 1982; Arango-Velez *et al.* 2016). We therefore speculate that the benefit of N allocation to the xylem tissue far exceeds the N cost of the production of class VII chitinases.

4.4.5 Network analysis to identify hub genes reveals several defense-associated genes, including transcription factors

Network analysis with WGCNA (Langfelder and Horvath 2008) was used to identify hub genes. The most highly connected genes with >0.80 module membership, as assigned by WGCNA (Langfelder and Horvath 2008), were deemed hub genes. They showed the greatest degree of expression profile similarity to other genes within their given module. Whereas all of the top 20 most connected hub genes (i.e. were co-expressed with the greatest number of orange module genes) identified in Table 4.9 were DE in at least one of the four *G. clavigera* vs. mock inoculation comparisons, none of the hub genes that were

identified using the same criteria in the dark cyan module were DE. Of the 11 hub genes identified in Table 4.9 that were up-regulated under all four *G. clavigera* vs. mock inoculation comparisons (i.e. belonged to the set of 488 core DEGs identified through the Venn diagram analysis), eight had annotations consistent with roles in plant defense. These included the putative (E)-β-ocimene synthase discussed above, as well as the transcription factor with a WRKY amino acid sequence DNA binding domain at the N-terminus (WRKY; Rushton *et al.* 2010), JAZ transcription factor (Ruan *et al.* 2019), receptor kinase with leucine-rich repeat (LRR) domain (Chakraborty *et al.* 2019), and oxophytodienoate-reductase 3 (Wasternack and Hause 2019).

Hub genes that are DE in response to *G. clavigera* inoculation and are annotated as putative transcription factors or other putative regulators of gene expression are of particular interest, as these represent potential regulators of lodgepole pine's molecular response to *G. clavigera*. The transcription factor JAZ9 was co-expressed with the (E)-β-ocimene synthase hub gene, which was strongly up-regulated by fungal inoculation in all treatment and tissue types. JAZs are involved in JA-mediated signaling (Ruan *et al.* 2019), and JA is biosynthesized in plants in response to abiotic and biotic stresses (Ruan *et al.* 2019). *In vivo* levels of JA and JA-isoleucine (JA-Ile) increased in lodgepole pine in response to *G. clavigera* (Arango-Velez *et al.* 2016), providing evidence that *G. clavigera* is a necrotrophic pathogen (Glazebrook 2005; Zhang *et al.* 2017). JA initiates the degradation of JAZs, freeing ethylene response factor 1 (ERF1) to activate

expression of JA-dependent pathogen-responsive genes (Pieterse *et al.* 2012; Broekgaarden *et al.* 2015). Pine homolog class I and class IV chitinases are DE in response to foliar and root applications of JA (Davis *et al.* 2002), suggesting JAZ transcription factor involvement in the production of these PR proteins. Furthermore, the application of methyl-jasmonate (MeJA), a volatile JA derivative, resulted in massive up-regulation of terpene synthase genes and the accumulation of terpenoids in the stems of Norway and Sitka spruce (*Picea sitchensis* Bongard) (Martin *et al.* 2002; Miller *et al.* 2005).

Two WRKY31 hub genes were also co-expressed with the JAZ9 and the (E)-β-ocimene hub gene. WRKYs are involved in senescence, seed development, dormancy and germination and other developmental processes in angiosperms (Rushton *et al.* 2010). WRKYs regulate the production of phenylpropanoids, alkaloids, and terpenes in a myriad of plant species (Schluttenhofer and Yuan 2015). *Oryza sativa* (Linnaeus) WRKY4 was rapidly and strongly induced by the necrotrophic fungus *Rhizoctonia solani* J.G. Kühn (Wang *et al.* 2015). The up-regulation of WRKY4 was concurrent with elevated expression of JA- and ethylene-responsive PR genes 1, 5 and 10 (Wang *et al.* 2015). WRKY31 was up-regulated in rice seedlings by the hemibiotrophic fungus *Magnaporthe grisea* (T.T. Hebert) M.E. Barr, and overexpression of the transcription factor was associated with the blockade of pathogen invasion (Zhang *et al.* 2008).

Though they were not used for subnetwork construction, two receptor kinases with LRR domains, both of which belonged to the set of 488 core DEGs

identified through the Venn diagram analysis, were identified as two of the top 20 most connected hub genes in the orange module. A typical plasma membrane-localized receptor kinase contains a unique extracellular domain, such as an LRR domain, a transmembrane domain and an intracellular signaling domain (Chakraborty *et al.* 2019). Some receptor kinases with LRR domains identify plant- or pathogen-derived peptides, which can activate down-stream signaling cascades that modulate the expression of transcription factors and regulators (Liang and Zhou 2018). Huffaker *et al.* (2006) suggested that the *A. thaliana*-derived peptide *At*Pep1, which is detected by the LRR domain containing receptor kinase PEPR1 (Yamaguchi *et al.* 2006), activated transcription of a defense-associated defensin gene. The peptide precursor gene PROPEP1 was induced by wounding and exogenous application of MeJA and ethylene in *A. thaliana*, and its overexpression produced a phenotype with enhanced resistance toward the root pathogen *Pythium irregulare* Buisman (Huffaker *et al.* 2006).

An oxophytodienoate reductase 3 was identified as one of the top 20 most connected hub genes in the orange module, though it was not used for subnetwork construction. The gene was significantly DE in all fungal- vs. mock-inoculated contrasts excluding phloem from the high N-treated trees. The metabolite *cis*-12-oxophytodienoic acid is reduced by 12 oxophytodienoate reductase 3 in the peroxisome as one step in the biosynthesis of JA (Wasternack and Hause 2019). Miller *et al.* (2005) detected the accumulation of 12-oxophytodienoate reductase transcripts in pine weevil-attacked Sitka spruce. Their results suggested that pine

weevil attack initiated JA-mediated defense signaling that may have resulted the induction of JA-responsive genes, such as terpene synthases (Miller *et al.* 2005).

4.5 Conclusion

Processes involved in lodgepole pine defense responses to *G. clavigera* invasion under two different levels of soil N supply were revealed by a combination of monoterpene metabolite and transcriptional profiling in conjunction with a network analysis approach. Quantification of monoterpene levels in VOCs, foliage and phloem determined that several monoterpenes were significantly impacted by both N fertilization level and *G. clavigera* inoculation, including the VOC α-pinene, validating hypothesis (1). Our findings provided preliminary support for the GDB hypothesis (Stamp 2003). In support of hypothesis (2), we found that more resources were allocated towards defense in low N- vs. high N-treated trees. Significantly DE defense-associated genes, such as mono- and sesquiterpene synthases and chitinases, had larger $\log_2$FC values in low N-treated trees compared with adequately fertilized trees. This may have been due to the low N-treated trees' earlier transition into growth cessation. It is possible that more resources were allocated to growth in high N-treated trees resulting in a lower intensity defense response, though without the appropriate growth measurements, we can only speculate. Validating hypothesis (3), the CD:ND ratio for up-regulated genes was smaller (i.e. a greater proportion of ND were up-regulated compared with the up-

regulated CD) in the high N-treated tissues compared with the low N-treated tissues. In accordance with the CNB hypothesis (Stamp 2003), the high N fertilization treatment reduced the internal C:N ratio and correlated with an increased allocation of N resources towards N-based defenses. Our results also opposed the predictions of the CNB hypothesis since a uniform pattern of greater fold-change induction in 0.3 mM NH_4NO_3-treated seedlings for C-based defense associated genes, including terpene synthases, was not detected. Our results partially aligned with the assumptions and predictions of the OD hypothesis (Stamp 2003) in that more β-farnesene and α-bisabolene synthases were up-regulated in the phloem (fungus' entry point) and more class VII chitinases were up-regulated in the xylem (main tissue inhabited by the fungus), when compared with the opposing tissue. Finally, network analysis identified JAZ and WRKY transcription factors as hub genes that represent potential regulators of lodgepole pine's molecular response to G. clavigera. The analyses carried out in this study demonstrated the intersection between N use and defense in lodgepole pine seedlings challenged by G. clavigera and grown under low and high NH_4NO3 applications in growth chamber conditions.

4.6 References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403-410.

Anders, S., Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, 11, 1-12.

Anders, S., McCarthy, D. J., Chen, Y. S., Okoniewski, M., Smyth, G. K., Huber, W., Robinson, M. D. (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature Protocols*, 8, 1765-1786.

Andrews, S. (2014). FastQC a quality control tool for high throughput sequence data. *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*. (Last accessed February 14, 2018).

Apweiler, R. Attwood, T. K., Bairoch A., Bateman A., Birney E., Biswas M., Bucher P., Cerutti L., Corpet F., Croning M. D., Durbin R., Falquet L., Fleischmann W., Gouzy J., Hermjakob H., Hulo N., Jonassen I., Kahn D., Kanapin A., Karavidopoulou Y., Lopez R., Marx B., Mulder N. J., Oinn T. M., Pagni M., Servant F., Sigrist C. J., Zdobnov E. M. (2001). The InterPro database, An integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Research,* 29, 37-40.

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

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

Bakhtiari, M., Formenti, L., Caggía, V., Glauser, G., Rasmann, S. (2019). Variable effects on growth and defence traits for plant ecotypic differentiation and phenotypic plasticity along elevation gradients. *Ecology and Evolution*, 9, 3740-3755.

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.

Baradat, P., Yazdani, R. (1988). Genetic expression for monoterpenes in clones of *Pinus sylvestris* grown on different sites. *Scandinavian Journal of Forest Research,* 3, 25-36.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *Series B* (*Methodological*), 57, 289-300.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Sayers, E. W. (2011). GenBank. *Nucleic Acids Research*, 39, D32-D37.

Bentz, B. J., Kegley, S., Gibson, K., Thier, R. (2005). A test of high-dose verbenone for stand-level protection of lodgepole and whitebark pine from mountain pine beetle (*Coleoptera: Curculionidae: Scolytinae*) attacks. *Journal of Economic Entomology*, 98, 1614-1621.

Björkman, C., Larsson, S., Gref, R. (1991). Effects of nitrogen fertilization on pine needle chemistry and sawfly performance. *Oecologia*, 86, 202-209.

Blais, A., Dynlacht, B. D. (2005). Constructing transcriptional regulatory networks. *Genes and Development, 19*, 1499-1511.

Block, A. K., Vaughan, M. M., Schetz, E. A., Chrisyensen, S. A. (2019). Biosynthesis and function of terpenoid defense compounds in maize (*Zea mays*). *Planta*, 249, 21–30.

Blodgett, J. T., Herms, D. A., Bonello, P. (2005). Effects of fertilization on red pine defense chemistry and resistance to *Sphaeropsis sapinea*. *Forest Ecology and Management*, 208, 373-382.

Bohlmann, J., Crock, J., Jetter, R., Croteau, R. (1998). Terpenoid-based defenses in conifers: cDNA cloning, characterization, and functional expression of wound-inducible (E)-α-bisabolene synthase from grand fir (*Abies grandis*). *Proceedings of the National Academy of Sciences*, 95, 6756-6761.

Bohlmann, J., Keeling, C. I. (2008). Terpenoid biomaterials. *The Plant Journal,* 54, 656–669.

Bolger, A. M., Lohse, M., Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114-2120.

Brockley, R. P. (2001). Fertilization of lodgepole pine in western Canada. In Bamsey, C. (Ed.), *Enhanced forest management: Fertilization and economics conference*. March 1-2, 2001, Edmonton, Alberta, Canada. pp. 44-55.

Broekgaarden, C., Caarls, L., Vos, I. A., Pieterse, C. M., Wees, S. C. (2015). Ethylene: Traffic controller on hormonal crossroads to defense. *Plant Physiology*, 169, 2371-2379.

Bryant, J. P., Chapin, F. S. III, Klein, D. R. (1983). Carbon/nutrient balance of boreal plants in relation to vertebrate herbivory. *Oikos*, 40, 357–368.

Cantón, F. R., Suárez1, M. F., Cánovas, F. M. (2005). Molecular aspects of nitrogen mobilization and recycling in trees. *Photosynthesis Research*, 83, 265–278.

Cartenì, F., Deslauriers, A., Rossi, S., Morin, H., Micco, V. D., Mazzoleni, S., Giannino, F. (2018). The physiological mechanisms behind the earlywood-to-latewood transition: A process-based modeling approach. *Frontiers in Plant Science,* 9, 1-12.

Celedon, J. M., Bohlmann, J. (2019). Oleoresin defenses in conifers: Chemical diversity, terpene synthases and limitations of oleoresin defense under climate change. *New Phytologist*, 224, 1444–1463.

Chakraborty, S., Nguyen, B., Wasti, S. D., Xu, G. (2019). Plant leucine-rich repeat receptor kinase (LRR-RK): Structure, ligand perception, and activation mechanism. *Molecules,* 24, 1-37.

Champigny, M. L. (1995). Integration of photosynthetic carbon and nitrogen metabolism in higher plants. *Photosynthesis Research*, 46, 117-127.

Chiu, C. C., Keeling, C. I., Bohlmann, J. (2019). The cytochrome P450 CYP6DE1 catalyzes the conversion of α-pinene into the mountain pine beetle aggregation pheromone trans-verbenol. *Scientific Reports*, 9, 1-10.

Chiu, C. C., Keeling C. K., Bohlmann, J. (2017). Toxicity of pine monoterpenes to mountain pine beetle. *Scientific Reports*, 7, 1-8.

Cock, P. J., Fields, C. J., Goto, N., Heuer, M. L., Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, 38, 1767-1771.

Cook, S., Carroll, A., Kimsey, M., Shaw, T. (2015). Changes in a primary resistance parameter of lodgepole pine to bark beetle attack one year following fertilization and thinning. *Forests*, 6, 280-292.

Cooke, J. E. K., Martin, T. J., Davis, J. M. (2005). Short-term physiological and developmental responses to nitrogen availability in hybrid poplar. *New Phytologist*, 167, 41-52.

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.

Cullingham, C. I., Roe, A. D., Sperling, F. A. H., Coltman, D. W. (2012). Phylogeographic insights into an irruptive pest outbreak. *Ecology and Evolution*, 2, 908–919.

Davis, J. M., Wu, H., 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.

Dietrich, R., Ploss, K., Heil, M. (2004). Constitutive and induced resistance to pathogens in Arabidopsis thaliana depends on nitrogen supply. *Plant, Cell and Environment*, *27*, 896-906.

DiGuistini, S., Wang, Y., Liao, N. Y., Taylor, G., Tanguay, P., Feau, N., Henrissat, B., Chan, S. K., Hesse-Orce, U., Alamouti, S. M., Tsui, C. K. M., Docking, R. T., Levasseu, A., Haridas, S., Robertson, G., Birol, I., Holt, R. A., Marra, M. A., Hamelin, R. C., Hirst, M., Jones, S. J. M., Bohlman, J., Breuil, C. (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*, 108, 2504–2509.

Fagard, M., Launay, A., Clement, G., Courtial, J., Dellagi, A., Farjad, M., Krapp, A., Soulié, M. C., Masclaux-Daubresse, C. (2014). Nitrogen metabolism meets phytopathology. *Journal of Experimental Botany*, 65, 5643-5656.

Fisher, R. A. (1934). *Statistical methods for research workers* (5th ed.). Oliver and Boyd, Edinburgh, Scotland.

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.

Fox, J., Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage Publications, Inc., Thousand Oaks, California, United States.

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, 353-375.

Galindo-González, L. M., Kayal, W. E., Morris, J. S., Cooke, J. E. K. (2015). Diverse chitinases are invoked during the activity-dormancy transition in spruce. *Tree Genetics and Genomes*, 11, 3-21.

Galindo-González, L. M., Kayal, W. E., Ju, C. J., Allen, C. C., King-Jones, S., Cooke, J. E. K. (2012). Integrated transcriptomic and proteomic profiling of white spruce stems during the transition from active growth to dormancy. *Plant, Cell and Environment,* 35, 682-701.

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

Graves, S., Piepho, H.-P., Selzer, L. Dorai-Raj, S. (2019). multcompView: Visualizations of paired comparisons. R package version 0.1-8. *https://CRAN.R-project.org/package=multcompView.*

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, N., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N., Regev, A. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8, 1494-1512.

Hakim, Ullah, A., Hussain, A., Shaban, M., Khan, A. H., Alariqi, M., Gulb, S., Juna, Z., Lina, S., Lia, J., Jina, S., Munis, M. F. H. (2018). Osmotin: A plant defense tool against biotic and abiotic stresses. *Plant Physiology and Biochemistry*, 123, 149-159.

Hall, D. E., Yuen, M. M., Jancsik, S., Quesada, A., Dullat, H. K., Li, M., Henderson, H., Arango-Velez, A., Liao, N. Y., Docking, R. T., Chan, S. K., Cooke, J. E. K., Breuil, C., Jones, S. J. M., Keeling, C. I., Bohlmann, J. (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, 1-14.

Herms, D. A., Mattson, W. J. (1992). The dilemma of plants: To grow or defend. *The Quarterly Review of Biology*, 67, 283–335.

Hocking, D. (1971). Preparation and use of a nutrient solution for culturing seedlings of lodgepole pine and white spruce, with selected bibliography. *Northern Forest Research Centre Information Report Nor-X-1*. Canadian Forest Service, Department of the Environment, Edmonton, Alberta, Canada.

Hodge, J., Cooke, B., McIntosh, R. (2017). *A strategic approach to slow the spread of mountain pine beetle across Canada*. Canadian Council of Forest Ministers, Forest Pest Working Group.

Hölzer, M., Marz, M. (2019). De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. *GigaScience*, 8, 1-16.

Huang, X. (1999). CAP3: A DNA sequence assembly program. *Genome Research*, *9*, 868-877.

Huang, Y., Niu, B., Gao, Y., Fu, L., Li, W. (2010). CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics,* 26, 680-682.

Huber, D. P. W., Gries, R., Borden, J. H., Pierce, H. D. (2000). A survey of antennal response by five species of coniferophagous bark beetles (Coleoptera: Scolytidae) to bark volatiles of six species of angiosperm trees. *Chemoecology*, 10, 103-113.

Huffaker, A., Pearce, G., Ryan, C. A. (2006). An endogenous peptide signal in Arabidopsis activates components of the innate immune response. *Proceedings of the National Academy of Sciences,* 103, 10098-10103.

Illumina, Inc. (2013). *Illumina TruSeq stranded mRNA low sample (LS) protocol*. Author, San Diego, California, United States.

Illumina, Inc. (2015). *NextSeq system denature and dilute libraries protocol*. Author, San Diego, California, United States.

Illumina, Inc. (2018). *NextSeq 500 system guide*. Author, San Diego, California, United States.

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.

Kaiser, W. M., Weiner, H., Huber, S. C. (1999). Nitrate reductase in higher plants: A case study for transduction of environmental stimuli into control of catalytic activity. *Physiologia Plantarum*, 105, 384-389.

Kanehisa, M., Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28, 27-30.

Kang, Z., Liu, F., Zhang, Z., Tian, H., Liu, T. (2018). Volatile β-ocimene can regulate developmental performance of peach aphid *Myzus persicae* through activation of defense responses in Chinese cabbage *Brassica pekinensis*. *Frontiers in Plant Science,* 9, 1-12.

Keefover-Ring, K., Trowbridge, A., Mason, C. J., Raffa, K. F. (2015). Rapid induction of multiple terpenoid groups by ponderosa pine in response to bark beetle-associated fungi. *Journal of Chemical Ecology*, 42, 1-12.

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

Keeling, C., Weisshaar, S., Lin, R. P. C., Bohlmann, J. (2008). Functional plasticity of paralogous diterpene synthases involved in conifer defense. *PNAS*, 105, 1085–1090.

Kishimoto, K., Matsui, K., Ozawa, R., Takabayashi, J. (2006). Analysis of defensive responses activated by volatile allo-ocimene treatment in *Arabidopsis thaliana*. *Phytochemistry*, 67, 1520-1529.

Köchy, M., Wilson, S. D. (2005). Variation in nitrogen deposition and available soil nitrogen in a forest–grassland ecotone in Canada. *Landscape Ecology*, 20, 191-202.

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

Komsta, L. (2011). Outliers: Tests for outliers. R package version 0.14. *https://CRAN.R-project.org/package=outliers.*

Koricheva, J., M., Keinanen, M. (1998). Regulation of woody plant secondary metabolism by resource availability: Hypothesis testing by means of meta-analysis. *Oikos*, 83, 212-226.

Kovalchuk, A., Keriö, S., Oghenekaro, A. O., Jaber, E., Raffaello, T., Asiegbu, F. O. (2013). Antimicrobial defenses and resistance in forest trees: Challenges and perspectives in a genomic era. *Annual Review of Phytopathology*, 51, 221–244.

Kestler, H. A., Muller, A., Gress, T. M., Buchholz, M. (2004). Generalized Venn diagrams: A new method of visualizing complex genetic set relations. *Bioinformatics*, 21, 1592-1595.

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D. L., Garcia-Hernadez, M., Karthikeyan, A. S., Lee, C. H., Nelson, W. D., Ploetz, L., Singh, S., Wensel, A., Huala, E. (2011). The *Arabidopsis* Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Research*, 40, D1202-D1210.

Langfelder, P., Horvath, S. (2012). Fast R functions for robust correlations and hierarchical clustering. *Journal of Statistical Software*, 46, 1-17.

Langfelder, P., Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, 1-13.

Larisch, C., Dittrich, M., Wildhagen, H., Lautner, S., Fromm, J., Polle, A., Hedrich, R., Rennenberg, H., Müller, T., Ache, P. (2012). Poplar wood rays are involved in seasonal remodeling of tree physiology. *Plant Physiology*, 160, 1515-1529.

Legendre, P, Legendre, L. (1998). *Numerical ecology* (2nd ed). Elsevier, Amsterdam, Netherlands. pp. 853.

Lenth, R. (2020). emmeans: Estimated marginal means, aka least-squares means. R package version 1.4.5. *https://CRAN.R-project.org/package=emmeans*.

Levene, H. (1960). *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (Olkin, I., Ghurye, S. G., Hoeffding, W., Madow, W. G., Mann, H. B, Eds.). Stanford University Press, Palo Alto, California, United States. pp. 278-292.

Liang, X., Zhou, J. (2018). Receptor-like cytoplasmic kinases: Central players in plant receptor kinase–mediated signaling. *Annual Review of Plant Biology*, 69, 267-299.

Liu, J.-J., Ekramoddoullah, A. K. M., 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.

Lohse, M., Nagel, A., Herter, T., May, P., Schroda, M., Zrenner, R., Tohge, T., Fernie, A. R., Stitt, M., Usadel, B. (2014). Mercator: A fast and simple web server for genome scale functional annotation of plant sequence data. *Plant, Cell and Environment*, 37, 1250–1258.

Lorio, P. L. (1986). Growth-differentiation balance: A basis for understanding southern pine beetle-tree interactions. *Forest Ecology and Management*, 14, 259-273.

Love, M. I., Huber, W., Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*, 1-21.

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

Marpeau, A., Walter, J., Launay, J., Charon, J., Baradat, P., Gleizes, M. (1989). Effects of wounds on the terpene content of twigs of maritime pine (*Pinus pinaster* Ait.). *Trees*, 3, 220-226.

Martin, D., Tholl, D., Gershenzon, J., Bohlmann, 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.

Maslov, S., Sneppen, K. (2006). Chapter 3 Large-Scale Topological Properties of Molecular Networks. In Koonin, E. V., Wolf, Y. I., Karev, G. P. (Eds.), *Power Laws, Scale-Free Networks and Genome Biology*. Springer Science+Business Media, Berlin, Germany. pp. 25-39.

Massad, T. J., Trumbore, S. E., Ganbat, G., Reichelt, M., Unsicker, S., Boeckler, A., Gleixner, G., Gershenzon, J., Ruehlow, S. (2014). An optimal defense strategy for phenolic glycoside production in *Populus trichocarpa*- Isotope labeling demonstrates secondary metabolite production in growing leaves. *New Phytologist*, 203, 607-619.

Matyssek, R., Agerer, R., Ernst, D., Munch, J.-C., Osswald, W., Pretzsch, H., Priesack, E., Schnyder, H., Treutter, D. (2005). The plant's capacity in regulating resource demand. *Plant Biology*, 7, 560–580.

Matyssek, R., Schnyder, H., Elstner, E.-F., Munch, J.-C., Pretzsch, H., Sandermann, H. (2002). Growth and parasite defence in plants: The balance between resource sequestration and retention. *Plant Biology*, 4, 133–136.

McClure, R. S., Overall, C. C., McDermott, J. E., Hill, E. A., Markillie, L. M., McCue, L. A., Beliaev, A. S. (2016). Network analysis of transcriptomics expands regulatory landscapes in *Synechococcus* sp. PCC 7002. *Nucleic Acids Research, 44*, 8810-8825.

McHale, L., Tan, X., Koehl, P., Michelmore, R. W. (2006). Plant NBS-LRR proteins: Adaptable guards. *Genome Biology*, 7, 212.1-212.11.

Mekonnen, Z. A., Riley, W. J., Randerson, J. T., Grant, R. F., Rogers, B. M. (2019). Expansion of high-latitude deciduous forests driven by interactions between climate warming and fire. *Nature Plants*, 5, 952–958.

Meldau, S., Erb, M., Baldwin, I. T. (2012). Defence on demand: Mechanisms behind optimal defence patterns. *Annals of Botany*, 110, 1503-1514.

Mentzen, W. I., Wurtele, E. (2008). Regulon organization of Arabidopsis. *BMC Plant Biology, 44*, 8810-8825.

Meuriot, F., Noquet, C., Avice, J.-C., Volenec, J. J., Cunningham, S. M., Sors, T. G., Caillot, S., Ourry, A. (2004). Methyl jasmonate alters N partitioning, N reserves accumulation and induces gene expression of a 32-kDa vegetative storage protein that possesses chitinase activity in *Medicago sativa* taproots. *Plant Physiology*, 120, 113–123.

Millard, P., Gretlet, G.-A. (2010). Nitrogen storage and remobilization by trees: Ecophysiological relevance in a changing world. *Tree Physiology*, 30, 1083–1095.

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.

Muñoz-Huerta, R., Guevara-Gonzalez, R., Contreras-Medina, L., Torres-Pacheco, I., Prado-Olivarez, J., Ocampo-Velazquez, R. (2013). A review of methods for sensing the nitrogen status in plants: Advantages, disadvantages and recent advances. *Sensors*, 13, 10823–10843.

Neuhaus, J. M. (1999). Plant chitinases (PR-3, PR-4, PR-8, PR-11). In Datta, S. K., Muthukrishnan, S. (Eds.), *Pathogenesis-related proteins in plants*. CRC Press, Boca Ranton, Florida, United States. pp. 77–105.

Nongpiur, R., Soni, P., Karan, R., Singla-Pareek, S. L., Pareek, A. (2012). Histidine kinases in plants. *Plant Signaling and Behavior*, 7, 1230-1237.

Nystedt, B., Street, N., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D. G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., Vicedomini, R., Sahlin, K., Sherwood, E., Elfstrand, M., Gramzow, L., Holmberg, K., Hällman, J., Keech, O., Klasson, L., Koriabine, M., Kucukoglu, M., Käller, M., Luthman, J., Lysholm, F., Niittylä, T., Olson, A., Rilakovic, N., Ritland, C., Rosselló, J. A., Sena, J., Svensson, T., Talavera-López, C., Theißen, G., Tuominen, H., Vanneste, K., Wu, Z.-Q., Zhang, B., Zerbe, P., Arvestad, L., Bhalerao, R., Bohlmann, J., Bousquet, J., Gil, R. G., Hvidsten, T. R., de Jong, P., MacKay, J., Morgante, M., Ritland, K., Sundberg, B., Thompson, S. L., Van de Peer, Y., Andersson, B., Ingvarsson, P. K., Lundeberg, J., Jansson, S. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, 497, 579–584.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H. (2019). vegan: Community ecology package. R package version 2.5-6. *https://CRAN.R-project.org/package=vegan*.

Oliveros, J. C. (2015) Venny: An interactive tool for comparing lists with Venn's diagrams. https://bioinfogp.cnb.csic.es/tools/venny/index.html.

Paine, T. D., Raffa, K. F., Harrington, T. C. (1997). Interactions among scolytid bark beetles, their associated fungi, and live host conifers. *Annual Review of Entomology*, 42, 179–206.

Palma, A. C., Winter, K., Aranda, J., Dalling, J. W., Cheesman, A. W., Turner, B. L., Cernusak, L. A. (2020). Why are tropical conifers disadvantaged in fertile soils? Comparison of *Podocarpus guatemalensis* with an angiosperm pioneer *Ficus insipida*. *Tree Physiology,* 40, 810-821.

Pavy, N., Boyle, B., Nelson, C., Paule, C., Giguère, I., Caron, S., Parsons, L. S., Dallaire, N., Bedon, F., Bérubé, H., Cooke, J., Mackay, J. (2008). Identification of conserved core xylem gene stets: Conifer cDNA microarray development, transcript profiling and computational analyses. *New Phytology*, 180, 766–786.

Peery, R. M., McAllister, C. H., Cullingham, C. I., Mahon, E. L., Arango-Velez, A., Cooke, J. E. K. Comparative genomics of the chitinase gene family in lodgepole and jack pines: Contrasting responses to biotic threats and landscape level investigation of genetic differentiation. Submitted to *Botany*, July 5, 2020, cjb-2020-0125.

Peoples M. B., Gifford R. M. (1990). Long-distance transport of nitrogen and carbon from sources to sinks in higher plants. In Dennis, D. T., Turpin, D. H. (Eds.), *Plant Physiology Biochemistry and Molecular Biology*. Longman, Harlow, Essex, England. pp. 434–447.

Peumans, W. J., Proost, P., Swennen, R. L., Van Damme, E. J. M. (2002). The abundant class III chitinase homolog in young developing banana fruits behaves as a transient vegetative storage protein and most probably serves as an important supply of amino acids for the synthesis of ripening-associated proteins. *Plant Physiology*, 130, 1063–1072.

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

Pillai, K. C. (2004). Multivariate analysis of variance (MANOVA). In Kotz, S., Read, C. B., Balakrishnan, N., Vidakovic, B., Johnson, N. L. (Eds.), *Encyclopedia of statistical sciences*. John Wiley & Sons, Inc., Hoboken, New Jersey, United States.

Prescott, C. E., Preston, C. M. (1994). Nitrogen mineralization and decomposition in forest floors in adjacent plantations of western red cedar, western hemlock, and Douglas-fir. *Canadian Journal of Forest Research,* 24, 2424-2431.

Pruitt, K. D., Tatusova, T., Maglott, D. R. (2004). NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 33, D501–D504.

Qiagen, CLCBio. *www.qiagenbioinformatics.com*. (Last accessed March 16, 2020).

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *https://www.R-project.org*.

Raffa, K. F. (2013). Terpenes tell different tales at different scales: Glimpses into the chemical ecology of conifer - bark beetle - microbial interactions. *Journal of Chemical Ecology,* 40, 1-20.

Raffa, K. F., Berryman, A. A. (1987). Interacting selective pressures in conifer-bark beetle systems: A basis for reciprocal adaptations? *The American Naturalist*, 129, 234-262.

Rao, D. H., Gowda, L. R. (2008). Abundant class III acidic chitinase homologue in tamarind (*Tamarindus indica*) seed serves as the major storage protein. *Journal of Agricultural and Food Chemistry*, 56, 2175–2182.

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., Mungall, K., Lee, S., Okada, H., Qian, J., Griffith, M., Raymond, A., Thiessen, N., Cezard, T., Butterfield, Y., Newsome, R., Chan, S., She, R., Varhol, R., Birol, I., Birol, I. (2010). *De novo* assembly and analysis of RNA-seq data. *Nature Methods*, 7, U909-U962.

Robinson, M. D., McCarthy, D. J., Smyth, G. K. (2010). EdgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139-140.

Robinson, M. D., Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11, PAGES.

Roe A. D., Rice A. V., Bromilow S. E., Cooke J. E. K., Sperling F. A. H. (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.

Roe A. D., Rice A. V., Coltman D. W., Cooke J. E. K., Sperling F. A. H. (2011). Comparative phylogeography, genetic differentiation, and contrasting reproductive modes in three fungal symbionts of a multipartite bark beetle symbiosis. *Molecular Ecology*, 20, 584–600.

RStudio Team (2015). *RStudio: Integrated development for R*. RStudio, Inc., Boston, MA.

Ruan, J., Zhou, Y., Zhou, M., Yan, J., Khurshid, M., Weng, W., Cheng, Z., Zhang, K. (2019). Jasmonic acid signaling pathway in plants. *International Journal of Molecular Sciences*, 20, 2479.

Rushton, P. J., Somssich, I. E., Ringler, P., She. Q. J. (2010). WRKY transcription factors. *Trends in Plant Science*, 15, 247-258.

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.

Schluttenhofer, C., Yuan, L. (2015). Regulation of specialized metabolism by WRKY transcription factors. *Plant Physiology*, 167, 295–306.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13, 2498-2504.

Shapiro, S. S., Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591.

Shi, W., Liao, Y. (2016). *Subread/Rsubread users guide*. Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, The University of Melbourne, Melbourne, Australia.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 3, 3210–3212.

Smith, Richard H. (2000). *Xylem monoterpenes of pines: distribution, variation, genetics, function*. Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture, Albany, California, United States. General Technical Report PSWGTR- 177.

Stamp, N. (2003). Out of the quagmire of plant defense hypotheses. *The Quarterly Review of Biology*, 78, 23–55.

Steele, C. L., Katoh, S., Bohlmann, J., Croteau, R. (1998). Regulation of oleoresinosis in grand fir (*Abies grandis*): Differential transcriptional control of monoterpene, sesquiterpene, and diterpene synthase genes in response to wounding. *Plant Physiology*, 116, 1497–1504.

Sun, Y., Wang, M., Mur, L. A. J., Shen, Q., Guo, S. (2020). Unravelling the roles of nitrogen nutrition in plant disease defences. *International Journal of Molecular Sciences*, 21, 1–20.

The Gene Ontology Consortium (2015). Gene ontology consortium: Going forward. *Nucleic Acids Research*, 43, D1049-D1056.

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Muller, L. A., Rhee, S. Y., Stitt, M. (2004). MapMan: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal*, 37, 914-939.

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

Tomova, L., Braun, S., Fluckiger, W. (2005). The effect of nitrogen fertilization on fungistatic phenolic compounds in roots of beech (*Fagus sylvatica*) and Norway spruce (*Picea abies*). *Forest Pathology*, 35, 262–276.

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 5, 99-114.

Tuomi, J., Fagerstrom, T., Niemela, P. (1991). Carbon allocation, phenotypic plasticity, and induced defenses. In Tallamy, D. W., Raupp, M. J. (Eds.), *Phytochemical Induction by Herbivores*. Wiley, New York, New York, United States. pp. 85–104.

Tuomi, J., Niemelä, P., Chapin, III, F. S., Bryant, J. P., Sirén, S. (1988). Defensive responses of trees in relation to their carbon/nutrient balance. In Mattson, W. J. *et al.* (Eds.), *Mechanisms of Woody Plant Defenses Against Insects: Search for Pattern*. Springer, New York, New York, United States. pp. 57–72.

Van Bel, M., Proost, S., Neste, C. V., Deforce, D., Peer, Y. V., Vandepoele, K. (2013). TRAPID: An efficient online tool for the functional and comparative analysis of *de novo* RNA-Seq transcriptomes. *Genome Biology*, 14, 1-10.

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

Villari, C., Battisti, A., Chakraborty, S., Michelozzi, M., Bonello, P., Faccoli, M. (2012). Nutritional and pathogenic fungi associated with the pine engraver beetle trigger comparable defenses in Scots pine. *Tree Physiology*, 32, 867-879.

Wang, H., Meng, J., Peng, X., Tang, X., Zhou, P., Xiang, J., Deng, X. (2015). Rice WRKY4 acts as a transcriptional activator mediating defense response toward *Rhizoctonia solani*, the causing agent of rice sheath blight. *Plant Molecular Biology,* 89, 157-171.

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., Venables, B. (2020). gplots: Various R programming tools for plotting data. R package version 3.0.3. *https://CRAN.R-project.org/package=gplots*.

Wasternack, C., Hause, B. (2019). The missing link in jasmonic acid biosynthesis. *Nature Plants,* 5, 776-777.

Wegrzyn, J. L., Lee, J. M., Tearse, B. R., Neale, D. B. (2008). TreeGenes: A forest tree genome database. *International Journal of Plant Genomics*, 2008, 1-7.

West, D., Bernklau, E., Bjostad, L., Jacobi, W. (2016). Host defense mechanisms against bark beetle attack differ between ponderosa and lodgepole pines. *Forests, 7*, 1-11.

Wetzel, S., Demmers, C., Greenwood, J. S. (1989). Seasonally fluctuating bark proteins are a potential form of nitrogen storage in three temperate hardwoods. *Planta*, 178, 275-281.

Xu, J., Wang, X., Guo, W. (2015). The cytochrome P450 superfamily: Key players in plant development and defense. *Journal of Integrative Agriculture*, 14, 1673–1686.

Xuehui, J., Na, H., Feng, J., Yang, Y., Dan, W., Chunying, X., Ruichang, Z. (2014). The effect of nitrogen supply on potato yield, tuber size and pathogen resistance in *Solanum tuberosum* exposed to *Phytophthora infestans*. *African Journal of Agricultural Research,* 9, 2657-2663.

Yamaguchi, Y., Pearce, G., Ryan, C. A. (2006). The cell surface leucine-rich repeat receptor for *At*Pep1, an endogenous peptide elicitor in *Arabidopsis*, is functional in transgenic tobacco cells. *Proceedings of the National Academy of Sciences,* 103, 10104-10109.

Young, M. D., Wakefield, M. J., Smyth, G. K., Oshlack, A. (2010). Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biology, 11*, PAGES.

Zamani, B. A., Sturrock, R., Ekramoddoullah, A. K. M., Wiseman, S. B. (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.

Zhang, B., Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology,* 4, 1-43.

Zhang, L., Zhang, F., Melotto, M., Yao, J., He, S. Y. (2017). Jasmonate signaling and manipulation by pathogens and insects. *Journal of Experimental Botany,* 68, 1371-1385.

Zhang, J., Peng, Y., Guo, Z. (2008). Constitutive expression of pathogen-inducible OsWRKY31 enhances disease resistance and affects root growth and auxin response in transgenic rice plants. *Cell Research*, 18, 508-521.

Zimmerman, C. F., Keefe, C. W., Bashe, J. (1997). *Method 440.0 Determination of carbon and nitrogen in sediments and particulates of estuarine/coastal waters using elemental analysis*. U.S. Environmental Protection Agency, Washington, District of Columbia, United States, EPA/600/R-15/009, pp 365.5-1 – 365.5-9.

5.0 Chapter 5: Conclusion

In western Canada alone, the current mountain pine beetle (MPB) outbreak has devastated nearly 19 million hectares of pine forest (Hodge *et al.* 2017). The beetles and their fungal associates can kill entire stands of pine, leaving characteristic yellow and red foliage in place of green starting one year after attack (Wulder *et al.* 2006; Page *et al.* 2012). This MPB epidemic negatively impacts communities, industries and ecosystems (Shore *et al.* 2006; Kurz *et al.* 2008; Corbett *et al.* 2016). Scientific research is a critical component of the response effort to minimize this persistent threat (Hodge *et al.* 2017). A thorough understanding of the underlying mechanisms that contribute to host susceptibility is crucial to track the spread risk of MPB (Cullingham *et al.* 2019). Reforestation efforts can benefit from a deeper understanding of nitrogen (N)-based fertilization on pine trees under attack by bark beetles and their microbial cohorts (Government of British Columbia 2006). We can use powerful transcriptomic tools to explore molecular processes that impact the lodgepole pine response to the MPB-vectored *Grosmannia clavigera* under varying levels of N availability (Wang *et al.* 2009).

For the temperate trees of western Canada, N is the most growth limiting nutrient (Brockley 2001). Trees utilize N for both primary and secondary metabolism in greater quantities than any other essential nutrient (Taiz *et al.* 2015). Limiting this nutrient destabilizes tree biochemistry and reduces N

availability for protein production (Taiz *et al.* 2015; Tang *et al.* 2019). Plant N saturation is reached when the availability of inorganic N exceeds the plant's nutritional demand (Wilson and Skeffington 1994). An excess of N availability can result in an increase in soil acidity and N leaching (Wilson and Skeffington 1994). N fertilization increases vegetative growth by stimulating the replication and differentiation of meristematic cells (Hacke *et al.* 2010; Taiz *et al.* 2015). Because they provide higher quality food for larvae and greater surface areas for brood development, larger hosts have been linked with greater MPB reproductive success (Nelson *et al.* 2018). The fungal associate *G. clavigera* redistributes N from the tree phloem tissue to the MPB pupal chambers (Goodsman *et al.* 2012). There, brood and emerging beetles feed on the pathogenic fungi (Paine *et al.* 1997).

Research on the effect of varied N fertilization regimes on conifer defense responses to necrotrophic fungi is lacking, especially studies that implement a transcriptomics approach. Vega *et al.* (2015) found that the susceptibility of tomato (*Solanum lycopersicum* Linnaeus) to the necrotrophic pathogen *Botrytis cinerea* (Persoon) was modulated by alterations in nitrate ($NO3^-$) fertilization concentrations. They correlated differential expression results with disease incident measurements and showed that $NO3^-$-limited conditions resulted in the decreased production of defense-related transcripts and increased susceptibility to fungal colonization (Vega *et al.* 2015). Disease incident measurements included lesion length and percentage of leaf area with disease symptoms. Vega *et al.* (2015) employed co-expression network analysis to find the key transcription factors

involved in the defense-related hormone signaling pathways that were expressed in response to fungal inoculation and modulated by NO_3^- availability (Vega *et al.* 2015). Similarly, my research utilized a transcriptomics approach to study the lodgepole pine-*G. clavigera* pathosystem.

This master's project relied upon the implementation of bioinformatic techniques, including sequenced read pre-processing, differential expression analysis using a *de novo* master transcriptome assembly, the annotation and data mining of contiguous sequences (contigs), and co-expression network analysis. Chapter 2 emphasized the importance of choice for both tools and parameters when producing an effective and reproducible bioinformatics pipeline. The CLC Genomics Workbench v9.5.2 (*www.qiagenbioinformatics.com*) and Trans-ABySS v1.5.5 (Robertson *et al.* 2010) were analyzed head-to-head as *de novo* assemblers of short reads (Table 5.1). The CLC Genomics Workbench (*www.qiagenbioinformatics.com*) is ideal for users that lack command line experience, though it must be purchased to reap the benefits. The CLC Genomics Workbench (*www.qiagenbioinformatics.com*) is highly effective in generating *de novo* assemblies when time is limited. However, the commercial software cannot assemble reads from multiple sequenced libraries combined to produce a master reference transcriptome. The application of the open-source assembler Trans-ABySS (Robertson *et al.* 2010) requires a knowledge of Unix command line and the up-stream implementation of alternative read pre-processing tools. Trans-ABySS (Robertson *et al.* 2010) is time-use intensive but results in a more accurate

and comprehensive assembly. Following our RNA-Seq experiment, I suggest using Trans-ABySS (Robertson *et al.* 2010) for constructing reference transcriptomes out of Illumina NGS data from non-model species.

Table 5.1. Summary of the CLC Genomics Workbench vs. Trans-ABySS for *de novo* assembly. The CLC Genomics Workbench and Trans-ABySS were implemented for the assembly of 151 bp paired-end lodgepole pine sequences. A comprehensive comparison of the two tools revealed the importance of choosing the assembler most appropriate for the research goals and the availability of financial, temporal and computational resources. Abbreviations include United States dollar (USD) and graphical user interface (GUI).

	Trans-ABySS	**CLC Genomics Workbench**
Read trimming available	No	Yes
Cost	Free	$5,500 USD *
Code availability	Open-source	Proprietary
# lodgepole pine libraries / assembly	≥ 48	< 4
Ease of use	Command line	Straight forward GUI with detailed manual
Run time	Long	Short
Assembly quality	High	Modest

* Obtained from Smith (2014).

Chapter 3 explored the impact of N availability on lesion length and foliar N in lodgepole pine. Seedlings grown under low (1 mM) or high (10 mM) ammonium nitrate (NH_4NO_3) conditions were either inoculated with *G. clavigera* or wounded and mock-inoculated with water. Fertilization levels chosen for the study outlined in Chapters 3 and 4 were sufficient to elicit changes in foliar N concentration. This was most likely due to the buildup of ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) as a carboxylase/oxygenase enzyme and a

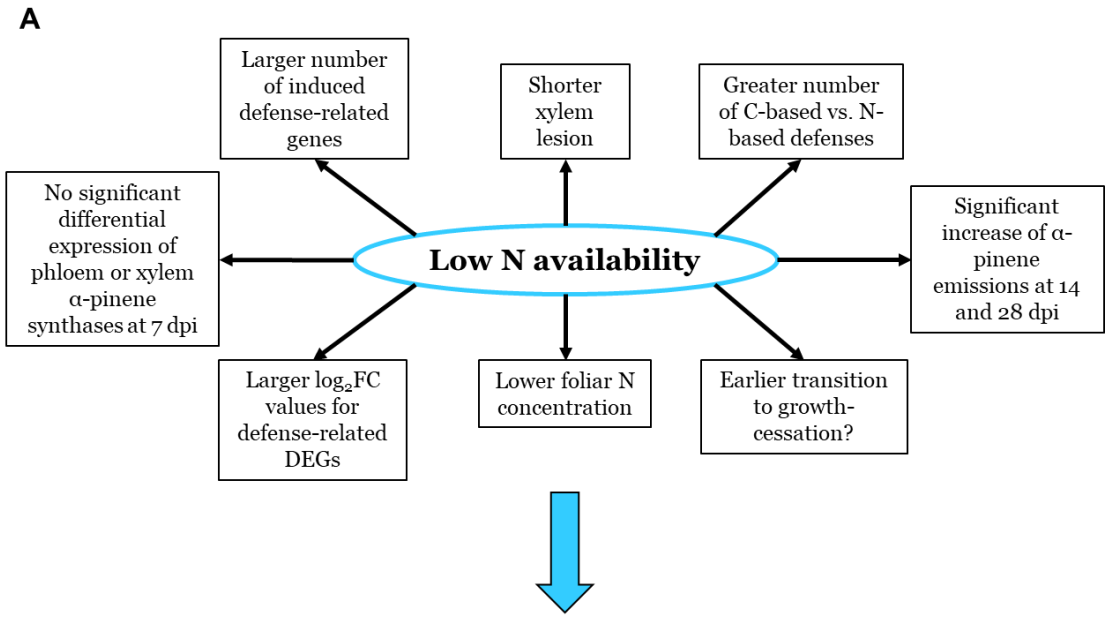
potential storage protein (Cheng and Fuchigami 2000). Furthermore, increased fertilization seemed to stimulate a stronger defense response through the creation of a longer lesion, possibly in response to increased N-stimulated fungal growth (Lahr and Krokene 2013; Mur *et al.* 2016; McAllister *et al.* 2018). Chapter 4 expanded upon these physiological discoveries by revealing that varied concentrations of N availability significantly impacted volatile organic compound (VOC) α-pinene production at 14 and 28 days post inoculation (dpi) with *G. clavigera*, resulting in increased production of the VOC in the low N-treated trees. In phloem, though α-pinene metabolite concentrations were not significantly different at any time point, α-pinene synthases were significantly down-regulated in high N-treated seedlings at 7 dpi. Given that MPB use phloem α-pinene as an aggregate pheromone precursor (Chiu *et al.* 2019), higher concentrations of NH_4NO_3-based fertilizers may decrease the likelihood of MPB mass-attack, while trees that received lower concentrations of N may be more susceptible to the mass-attack strategy.

Chapter 4 used RNA-Seq to analyze the impact of N on the lodgepole pine defense responses to *G. clavigera*. Seedlings grown under low (0.3 mM NH_4NO_3) or high (10 mM NH_4NO_3) N conditions were either wounded and inoculated with *G. clavigera* or wounded and mock-inoculated with water. Based on gene expression patterns, the results suggested that the composition and quality of lodgepole pine defenses against *G. clavigera* were altered by the level of N supply. When compared with the 0.3 mM NH_4NO_3-treated trees, the availability of 10 mM

NH_4NO_3 fertilizer resulted in fewer defense-associated differentially expressed genes (DEGs), smaller log base two fold change (log_2FC) values and a smaller carbon (C)- to N-based defenses ratio (CD:ND) for *G. clavigera*-induced DEGs. As the relationships between the terpene synthase gene expression profiles and the monoterpene metabolite analyses demonstrated, significant changes in transcript abundance of genes encoding enzymes or modifiers of biosynthetic pathways did not necessarily translate into significant differences in metabolite profiles. Therefore, we can use the results of the transcriptome analyses as predictors of how biochemical, physiological and cellular processes are involved in lodgepole pine response to *G. clavigera* inoculation, and as hypothesis generators for further in-depth studies at these levels.

In lodgepole pine seedlings challenged by *G. clavigera,* the results in Chapters 3 and 4 facilitate an understanding of the intersection between N availability and defense, and also raise questions about tradeoffs with growth. High N- vs. low N-treated trees yielded opposing results across a number of metrics and analyses (Figure 5.1). Because transcriptomic analyses suggested that tree defenses were less robust when N nutrition was more available, we predict that high N conditions are advantageous for fungal colonization, as has been observed in other plant-fungal pathogen systems (Lahr and Krokene 2013; Mur *et al.* 2016). We speculate that these trees maintained consistent growth, which provided more cellular components for fungus nutrition. Low N conditions were more advantageous for the lodgepole pine because the transcriptomics analyses

suggested that defense responses were enhanced and N availability for the fungus was low.



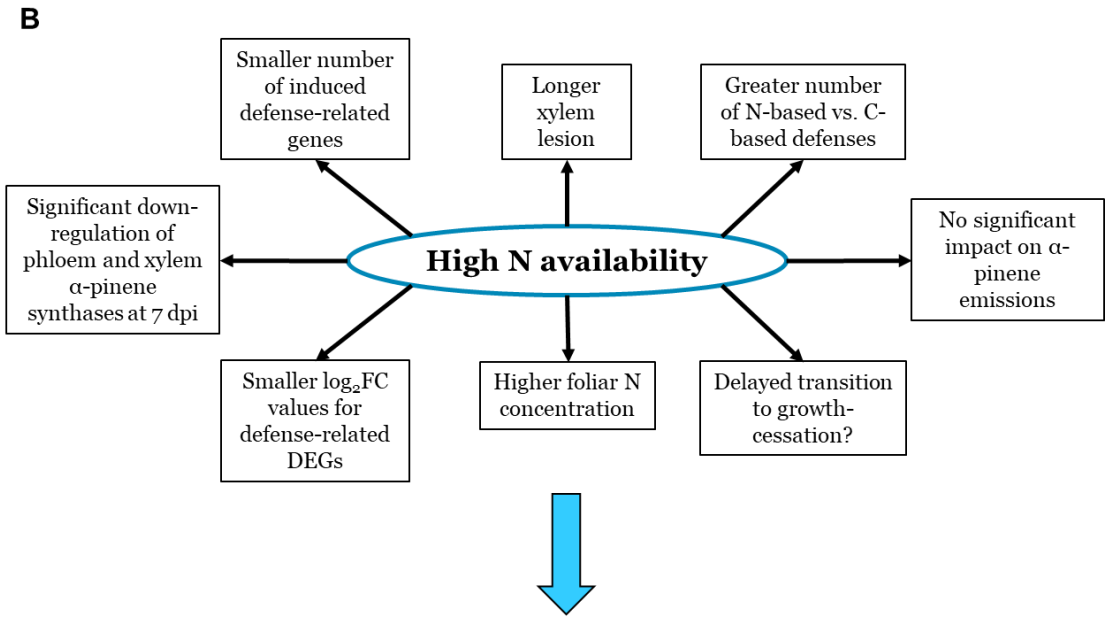


Figure 5.1. N availability impacts lodgepole pine response to *G. clavigera*. Lodgepole pine seedlings were supplied with 0.3 mM (Low; **A**) or 10 mM (High; **B**) NH_4NO_3 fertilization and inoculated with *G. clavigera*. N availability impacted physiology and defense strategies, resulting in differences in host quality. Low N-treated trees may have a lower propensity to fungal colonization compared to high N-treated trees. When contrasted with the 10 mM NH_4NO_3-treated trees, seedlings given the lower concentration fertilization may be more susceptible to MPB mass attack. Abbreviations include log base two fold change (log_2FC), differentially expressed genes (DEGs) and days post inoculation (dpi).

Plant growth/defense tradeoff theories can offer explanatory frameworks for these patterns of lodgepole pine-*G. clavigera* interactions across resource gradients. Note that to truly test if these hypotheses explain our results, microscopy with stem cross-sections is needed to compare the degree of cambial tissue expansion between N treatment types as an assessment of N-induced growth. The carbon-nutrient balance (CNB) hypothesis predicts that environmental factors that influence the plant's internal (assimilated) C:N ratio will result in an altered proportion of assimilated N available for non-growth processes, such as storage and defense (Stamp 2003). The CNB hypothesis states that a species' phenotypic plasticity determines the degree to which the plant's assimilated C:N ratio influences the allocation of C and N resources towards defense, potentially impacting the CD:ND ratio (Tuomi *et al.* 1988; Tuomi *et al.* 1991). As predicted by the CNB hypothesis, we found that N availability modulated the ratio of C- to N-based defenses. High compared to low N-treated seedlings had smaller CD:ND values for up-regulated genes in both phloem and xylem tissue. Our results also opposed the predictions of the CNB hypothesis since a uniform

pattern of greater fold-change induction in low N-treated seedlings for C-based defense-associated genes, including terpene synthases, was not detected.

Rather than being restricted to immediate effects on the C to nutrient balance within the plant, the growth-differentiation balance (GDB) hypothesis is specific in predicting the ways different environmental factors effect tradeoffs between growth and differentiation (Stamp 2003). Differentiation includes the production and maintenance of defense mechanisms and storage structures (Stamp 2004). In response to *G. clavigera* inoculation, the orange module revealed low N- compared to high N-treated trees resulted in a larger and more intense induction of several defense-related DEGs as indicated by larger $\log_2$FC values. I suggest that, in accordance with the GDB hypothesis, our high N- compared to low N-treated seedlings allocated more resources to growth and assimilated more N in RuBisCO at the expense of the defense responses to fungal colonization.

The optimal defense (OD) hypothesis states that, since they divert resources from growth and reproduction, defensive elements are costly to construct and maintain (Stamp 2003). Constitutive defense elements are therefore concentrated, and a timely induced defensive strategy is strongest, within the most vulnerable tissues (Herms and Mattson 1992; Meldau et al. 2012). Our results partially aligned with the assumptions and predictions of the OD hypothesis in that more sesquiterpene synthases and class VII chitinases were up-regulated in the phloem (fungus' entry point) and xylem (main tissue inhabited by the fungus), respectively,

than in the opposing tissue. The benefit of defending fungal-inoculated tissue may have exceeded the C cost of β-farnesene and α-bisabolene production in phloem and the N cost of class VII chitinase production in xylem, especially in the low N-treated trees. The OD hypothesis predicts that resource limitation results in weakened defense responses (Stamp 2003), however our low N-treated trees were not N-limited, so we cannot address this prediction.

To fill the void in the literature regarding the influence of N on lodgepole pine defense responses to *G. clavigera* challenge, our study used a transcriptomics approach. The field of bioinformatics includes other "omic" technologies, namely proteomics and metabolomics. Differential gene expression analysis is useful for the characterization of the lodgepole pine defense responses to pathogen challenge because RNA transcripts are proxies for proteins (Vogel and Marcotte 2012). Protein abundance can be determined in a high-throughput manner using next-generation proteomics, which provides a more direct measure of functional activity (Altelaar *et al.* 2013). Additionally, the detection and quantification of metabolites and other small molecules holds immense potential for clarifying complex interactions between lodgepole pine and MPB-vectored *G. clavigera*. Proteomic and metabolomic technologies include high-performance liquid chromatography to separate proteins or metabolites, respectively, which are then identified using mass spectrometry (MS; Turnbaugh and Gordon 2008; Altelaar *et al.* 2013). The implementation of multiple forms of "omic" data in an integrated framework

would result in an improved mechanistic model of lodgepole pine defenses against *G. clavigera* under varying levels of N fertilization.

My thesis project suggests that 0.3 mM compared with the 10 mM NH_4NO_3 fertilization bolstered the lodgepole pine defenses against *G. clavigera*. Fertilization can increase the quality and quantity of marketable timber (Government of British Columbia 2006), though silviculture intervention should not be undertaken without the appropriate experimentation. Reforestation programs in western Canada seek to replenish pine forests decimated by the MPB epidemic and wildfire (Government of British Columbia 2013). Before implementing soil enhancement measures as part of these efforts, the impact of varying levels of N availability on tree defense capacities should be tested, especially if the trees are at high risk for biotic threats. This includes exploring the impact of N fertilization on alternative MPB hosts and other conifer-pest-pathogen systems. With a sound scientific foundation, fertilization can benefit reforestation efforts that seek to mitigate the social, economic and ecological impacts of the MPB epidemic.

5.1 References

Altelaar, A. F., Munoz, J., Heck, A. J. (2013). Next generation proteomics: Towards an integrative view of proteome dynamics. *Nature Reviews Genetics*, 14, 35–48.

Brockley, R. P. (2001). Fertilization of lodgepole pine in western Canada. In Bamsey, C. (Ed.), *Enhanced Forest Management: Fertilization and Economics Conference*. March 1-2, 2001, Edmonton, Alberta, Canada. pp. 44-55.

Bryant, J. P., Chapin, F. S. III, Klein, D. R. (1983). Carbon/nutrient balance of boreal plants in relation to vertebrate herbivory. *Oikos*, 40, 357–368.

Cheng, L., Fuchigami, L. H. (2000). Rubisco activation state decreases with increasing nitrogen content in apple leaves. *Journal of Experimental Botany*, 51, 1687-1694.

Chiu, C. C., Keeling, C. I., Bohlmann, J. (2019). The cytochrome P450 CYP6DE1 catalyzes the conversion of α-pinene into the mountain pine beetle aggregation pheromone trans-verbenol. *Scientific Reports*, 9, 1-10.

Corbett, L. J., Withey, P., Lantz, V. A., Ochuodho, T. O. (2016). The economic impact of the mountain pine beetle infestation in British Columbia: Provincial estimates from a CGE analysis. *Forestry*, 89, 100-105.

Cullingham, C. I., Janes, J. K., Hamelin, R. C., James, P. M., Murray, B. W., Sperling, F. A. (2019). The contribution of genetics and genomics to understanding the ecology of the mountain pine beetle system. *Canadian Journal of Forest Research*, 49, 721-730.

Goodsman, D. W., Erbilgin, N., Lieffers, V. J. (2012). The impact of phloem nutrients on overwintering mountain pine beetles and their fungal symbionts. *Environmental Entomology*, 41, 478-486.

Government of British Columbia (2006). *Forest fertilization in British Columbia*. British Columbia Forest Service, Ministry of Forests and Range, Victoria, British Columbia, Canada.

Government of British Columbia (2013). *FFT current reforestation (CR) and timber supply mitigation (TSM) 5 year update and annual operating plan (AOP)*

guidance. British Columbia Forest Service, Forests, Lands and Natural Resource Operations, Victoria, British Columbia, Canada.

Hacke, U. G., Plavcova, L., Almeida-Rodriguez, A., King-Jones, S., Zhou, W., Cooke, J. E. K. (2010). Influence of nitrogen fertilization on xylem traits and aquaporin expression in stems of hybrid poplar. *Tree Physiology, 30*, 1016-1025.

Hodge, J., Cooke, B., McIntosh, R. (2017). *A strategic approach to slow the spread of mountain pine beetle across Canada*. Canadian Council of Forest Ministers, Forest Pest Working Group.

Kurz, W. A., Dymond, C. C., Stinson, G., Rampley, G. J., Neilson, E. T., Carroll, A. L., Ebata, T., Safranyik, L. (2008). Mountain pine beetle and forest carbon feedback to climate change. *Nature*, 452, 987-990.

Lorio, P. L. (1986). Growth-differentiation balance: A basis for understanding southern pine beetle-tree interactions. *Forest Ecology and Management*, 14, 259-273.

Matyssek, R., Schnyder, H., Elstner, E.-F., Munch, J.-C., Pretzsch, H., Sandermann, H. (2002). Growth and parasite defence in plants: The balance between resource sequestration and retention. *Plant Biology*, 4, 133–136.

McAllister, C. H., Fortier, C. E., Onge, K. R. S., Sacchi, B. M., Nawrot, M. J., Locke, T., Cooke, J. E. K. (2018). A novel application of RNase H2-dependent quantitative PCR for detection and quantification of *Grosmannia clavigera*, a mountain pine beetle fungal symbiont, in environmental samples. *Tree Physiology*, 38, 485–501.

Nelson, M. F., Murphy, J. T., Bone, C., Altaweel, M. (2018). Cyclic epidemics, population crashes, and irregular eruptions in simulated populations of the mountain pine beetle, *Dendroctonus ponderosae*. *Ecological Complexity*, 36, 218-229.

Page, W. G., Jenkins, M. J., Runyon, J. B. (2012). Mountain pine beetle attack alters the chemistry and flammability of lodgepole pine foliage. *Canadian Journal of Forest Research*, 42, 1631-1647.

Paine, T. D., Raffa, K. F., Harrington, T. C. (1997). Interactions among scolytid bark beetles, their associated fungi, and live host conifers. *Annual Review of Entomology*, 42, 179–206.

Qiagen, CLCBio. *www.qiagenbioinformatics.com*. (Last accessed March 16, 2020).

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., Mungall, K., Lee, S., Okada, H., Qian, J., Griffith, M., Raymond, A., Thiessen, N., Cezard, T., Butterfield, Y., Newsome, R., Chan, S., She, R., Varhol, R., Birol, I., Birol, I. (2010). *De novo* assembly and analysis of RNA-seq data. *Nature Methods*, 7, U909-U962.

Shore, T. L., Safranyik, L., Hawkes, B. C., Taylor, S. W. (2006). Chapter 3 Effects of the mountain pine beetle on lodgepole pine stand structure and dynamics. In Safranyik, L., Wilson, W.R. (Eds.). *The mountain pine beetle: a synthesis of biology, management, and impacts on lodgepole pine*. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia. pp. 95-114.

Smith, D. R. (2014). Buying in to bioinformatics: An introduction to commercial sequence analysis software. *Briefings in Bioinformatics*, 16, 700-709.

Stamp, N. (2003). Out of the quagmire of plant defense hypotheses. *The Quarterly Review of Biology*, 78, 23–55.

Stamp, N. (2004). Can the growth-differentiation balance hypothesis be tested rigorously? *Oikos*, 107, 439-448.

Taiz, L., Zeiger, E., Møller, I. M., Murphy, A. (Eds.). (2015). *Plant physiology and development* (6th ed.). Sinauer Associates, Inc., Publishers, Sunderland, Massachusetts, United States.

Tang, J., Sun, B., Cheng, R., Shi, Z., Luo, D., Liu, S., Centritto, M. (2019). Effects of soil nitrogen (N) deficiency on photosynthetic N-use efficiency in N-fixing and non-N-fixing tree seedlings in subtropical China. *Scientific Reports*, 9, 1-14.

Tuomi, J., Fagerstrom, T., Niemela, P. (1991). Carbon allocation, phenotypic plasticity, and induced defenses. In Tallamy, D. W., Raupp, M. J. (Eds.), *Phytochemical Induction by Herbivores*. Wiley, New York, New York, United States. pp. 85–104.

Tuomi, J., Niemelä, P., Chapin, III, F. S., Bryant, J. P., Sirén, S. (1988). Defensive responses of trees in relation to their carbon/nutrient balance. In Mattson, W. J. *et al.* (Eds.), *Mechanisms of Woody Plant Defenses Against Insects: Search for Pattern*. Springer, New York, New York, United States. pp. 57–72.

Turnbaugh, P. J., Gordon, J. I. (2008). An invitation to the marriage of metagenomics and metabolomics. *Cell*, 134, 708–713.

Vega, A., Canessa, P., Hoppe, G., Retamal, I., Moyano, T. C., Canales, J., Gutiérrez, R. A., Rubilar, J. (2015). Transcriptome analysis reveals regulatory networks underlying differential susceptibility to *Botrytis cinerea* in response to nitrogen availability in *Solanum lycopersicum*. *Frontiers in Plant Science*, 6, 1-17.

Vogel, C., Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, 13, 227-232.

Wang, Z., Gerstein, M., Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics,* 10, 57-63.

Wilson, E., Skeffington, R. (1994). The effects of excess nitrogen deposition on young Norway spruce trees. Part I the soil. *Environmental Pollution*, 86, 141-151.

Wulder, M. A., White, J. C., Bentz, B. J., Ebata, T. (2006). Augmenting the existing survey hierarchy for mountain pine beetle red-attack damage with satellite remotely sensed data. *The Forestry Chronicle*, 82, 187-202.

References

Addison, A., Powell, J. A., Bentz, B. J., Six, D. L. (2015). Integrating models to investigate critical phenological overlaps in complex ecological interactions: The mountain pine beetle-fungus symbiosis. *Journal of Theoretical Biology*, 368, 55–66.

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.

Agarwal, S., Macfarlan, T. S., Sartor, M. A., Iwase, S. (2015). Sequencing of first-strand cDNA library reveals full-length transcriptomes. *Nature Communications*, 6, 1-12.

Agrios, G. N. (2005). *Plant pathology*. (5th ed.). Elsevier Academic Press, New York, New York, United States.

Altelaar, A. F., Munoz, J., Heck, A. J. (2013). Next generation proteomics: Towards an integrative view of proteome dynamics. *Nature Reviews Genetics*, 14, 35–48.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403-410.

Alves, M., Dadalto, S., Gonçalves, A., Souza, G. D., Barros, V., Fietto, L. (2014). Transcription factor functional protein-protein interactions in plant defense responses. *Proteomes*, 2, 85-106.

Ambawat, S., Sharma, P., Yadav, N. R., Yadav, R. C. (2013). MYB transcription factor genes as regulators for plant responses: An overview. *Physiology and Molecular Biology of Plants*, 19, 307-321.

Amponsah, I. G., Lieffers, V. J., Comeau, P. G., Brockley, R. P. (2004). Growth response and sapwood hydraulic properties of young lodgepole pine following repeated fertilization. *Tree Physiology*, 24, 1099-1108.

Anders, S., Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, 11, 1-12.

Anders, S., McCarthy, D. J., Chen, Y. S., Okoniewski, M., Smyth, G. K., Huber, W., Robinson, M. D. (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature Protocols*, 8, 1765-1786.

Andrews, S. (2014). FastQC a quality control tool for high throughput sequence data. *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*. (Last accessed February 14, 2018).

Apweiler, R. Attwood, T. K., Bairoch A., Bateman A., Birney E., Biswas M., Bucher P., Cerutti L., Corpet F., Croning M. D., Durbin R., Falquet L., Fleischmann W., Gouzy J., Hermjakob H., Hulo N., Jonassen I., Kahn D., Kanapin A., Karavidopoulou Y., Lopez R., Marx B., Mulder N. J., Oinn T. M., Pagni M., Servant F., Sigrist C. J., Zdobnov E. M. (2001). The InterPro database, An integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Research,* 29, 37-40.

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

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

Asiegbu, F.O., M. Denekamp, G. Daniel and M. Johansson. 1995. Immunocytochemical localization of pathogenesis-related proteins in roots of

Norway spruce infected with *Heterobasidion annosum*. *European Journal of Forest Pathology*, 25, 169–178.

Ayres, M. P., Wilkens, R. T., Ruel, J. J., Lombardero, M. J., Vallery, E. (2000). Nitrogen budgets of phloem-feeding bark beetles with and without symbiotic fungi. *Ecology*, 81, 2198–2210.

Bakhtiari, M., Formenti, L., Caggía, V., Glauser, G., Rasmann, S. (2019). Variable effects on growth and defence traits for plant ecotypic differentiation and phenotypic plasticity along elevation gradients. *Ecology and Evolution*, 9, 3740-3755.

Balint-Kurti, P. (2019). The plant hypersensitive response: Concepts, control and consequences. *Molecular Plant Pathology*, 20, 1163-1178.

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.

Ballard, R. G., Walsh, M. A., Cole, W. E. (1984). The penetration and growth of blue-stain fungi in the sapwood of lodgepole pine attacked by mountain pine beetle. *Canadian Journal of Botany*, 62, 1724-1729.

Bancroft, B. (2008). *Fundamentals of natural lodgepole pine regeneration and drag scarification*. B.C. Ministry of Forests, Forest Renewal Section Silviculture Practices Branch, Victoria, British Columbia, Canada. pp. 1-30.

Baradat, P., Yazdani, R. (1988). Genetic expression for monoterpenes in clones of *Pinus sylvestris* grown on different sites. *Scandinavian Journal of Forest Research,* 3, 25-36.

Bedon, F., Bomal, C., Caron, S., Levasseur, C., Boyle, B., Mansfield, S. D., Schmidt, A., Gershenzon, J., Grima-Pettenati, J., Séguin, A., 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*, 61, 3847-3864.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *Series B* (*Methodological*), 57, 289-300.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Sayers, E. W. (2011). GenBank. *Nucleic Acids Research*, 39, D32-D37.

Bentley, D. R., Balasubramanian, S., Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456, 53-59.

Bentz, B. J., Kegley, S., Gibson, K., Thier, R. (2005). A test of high-dose verbenone for stand-level protection of lodgepole and whitebark pine from mountain pine beetle (*Coleoptera: Curculionidae: Scolytinae*) attacks. *Journal of Economic Entomology*, 98, 1614-1621.

Björkman, C., Larsson, S., Gref, R. (1991). Effects of nitrogen fertilization on pine needle chemistry and sawfly performance. *Oecologia*, 86, 202-209.

Blais, A., Dynlacht, B. D. (2005). Constructing transcriptional regulatory networks. *Genes and Development, 19*, 1499-1511.

Bleiker, K. P., Six, D. L. (2007). Dietary benefits of fungal associates to an eruptive herbivore: Potential implications of multiple associates on host population dynamics. *Environmental Entomology*, 36, 1384-1396.

Block, A. K., Vaughan, M. M., Schetz, E. A., Chrisyensen, S. A. (2019). Biosynthesis and function of terpenoid defense compounds in maize (*Zea mays*). *Planta*, 249, 21–30.

Blodgett, J. T., Herms, D. A., Bonello, P. (2005). Effects of fertilization on red pine defense chemistry and resistance to *Sphaeropsis sapinea*. *Forest Ecology and Management*, 208, 373-382.

Bohlmann, J., Crock, J., Jetter, R., Croteau, R. (1998). Terpenoid-based defenses in conifers: cDNA cloning, characterization, and functional expression of wound-inducible (E)-α-bisabolene synthase from grand fir (*Abies grandis*). *Proceedings of the National Academy of Sciences*, 95, 6756-6761.

Bohlmann, J., Keeling, C. I. (2008). Terpenoid biomaterials. *The Plant Journal,* 54, 656–669.

Bolger, A. M., Lohse, M., Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114-2120.

Boller, T. (1985). Induction of hydrolases as a defense reaction against pathogens. *UCLA Symposium on Molecular and Cellular Biology,* 22, 247-262.

Bossinger, G., Spokevicius, A. V. (2018). Sector analysis reveals patterns of cambium differentiation in poplar stems. *Journal of Experimental Botany*, 69, 4339-4348.

Box, G. E. P., Cox, D. R. (1964). An analysis of transformation. *Journal of the Royal Statistical Society, Series B (Methodological)*, 26, 211-252.

Brady, S. M., Long, T. A., Benfey, P. N. (2006). Unraveling the dynamic transcriptome. *The Plant Cell Online*, 18, 2101-2111.

Brockley, R. P. (2000). Using foliar variables to predict the response of lodgepole pine to nitrogen and sulphur fertilization. *Canadian Journal of Forest Research*, 30, 1389-1399.

Brockley, R. P. (2001). Fertilization of lodgepole pine in western Canada. In Bamsey, C. (Ed.), *Enhanced forest management: Fertilization and economics conference*. March 1-2, 2001, Edmonton, Alberta, Canada. pp. 44-55.

Broekgaarden, C., Caarls, L., Vos, I. A., Pieterse, C. M., Wees, S. C. (2015). Ethylene: Traffic controller on hormonal crossroads to defense. *Plant Physiology*, 169, 2371-2379.

Broschat, T. K. (2017). *Nutrient deficiency symptoms of woody ornamental plants in south Florida*. U. S. Department of Agriculture, UF/IFAS Extension, Gainesville, Florida, United States, ENH1098.

Bryant, J. P., Chapin, F. S. III, Klein, D. R. (1983). Carbon/nutrient balance of boreal plants in relation to vertebrate herbivory. *Oikos*, 40, 357–368.

Buchanan, B. B., Wolosiuk, R. A. (2015). Photosynthesis: The carbon reactions. In Taiz, L., Zeiger, E., Møller, I. M., Murphy, A. (Eds.), *Plant physiology and development* (6th ed.). Sinauer Associates, Inc., Publishers, Sunderland, Massachusetts, United States. pp. 203-241.

Cahais, V., Gayral, P., Tsagkogeorga, G., Melo-Ferreira, J., Ballenghien, M., Weinert, L., Chiari, Y., Belkhir, K., Ranwez, V., Galtier, N. (2012). Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Molecular Ecology Resources*, 12, 834-845.

Cantón, F. R., Suárez1, M. F., Cánovas, F. M. (2005). Molecular aspects of nitrogen mobilization and recycling in trees. *Photosynthesis Research*, 83, 265–278.

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.

Carroll, A., Taylor, S., Régnière, J., Safranyik, L. (2004). Effects of climate change on range expansion by the mountain pine beetle in British Columbia. In Shore, T. L., Brooks, J. E., Stone, J. E. (Eds.), *Mountain pine beetle symposium: Challenges and solutions*. October 30-31, 2003, Kelowna, British Columbia, Canada. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Information Report BC-X-399, Victoria, British Columbia, Canada. pp. 223-232.

Cartenì, F., Deslauriers, A., Rossi, S., Morin, H., Micco, V. D., Mazzoleni, S., Giannino, F. (2018). The physiological mechanisms behind the earlywood-to-latewood transition: A process-based modeling approach. *Frontiers in Plant Science,* 9, 1-12.

Celedon, J. M., Bohlmann, J. (2019). Oleoresin defenses in conifers: Chemical diversity, terpene synthases and limitations of oleoresin defense under climate change. *New Phytologist*, 224, 1444–1463.

Chakraborty, S., Nguyen, B., Wasti, S. D., Xu, G. (2019). Plant leucine-rich repeat receptor kinase (LRR-RK): Structure, ligand perception, and activation mechanism. *Molecules,* 24, 1-37.

Chaisson, M. J., Wilson, R. K., Eichler, E. E. (2015). Genetic variation and the *de novo* assembly of human genomes. *Nature Reviews Genetics*, 16, 627-640.

Chalot, M., Brun, A. (1998). Physiology of organic nitrogen acquisition by ectomycorrhizal fungi and ectomycorrhizas. *FEMS Microbiology Reviews*, 22, 21-44.

Champigny, M. L. (1995). Integration of photosynthetic carbon and nitrogen metabolism in higher plants. *Photosynthesis Research*, 46, 117-127.

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.

Chen, H., Walton, A. (2011). Mountain pine beetle dispersal: Spatiotemporal patterns and role in the spread and expansion of the present outbreak. *Ecosphere*, 2, 21-44.

Chen, Y., McCarthy, D., Ritchie, M., Robinson, M., Smyth, G. K. (2017, Dec 26). EdgeR: Differential expression analysis of digital gene expression data: User's Guide. Retrieved from *https://www.bioconductor.org/packages/devel/bioc/vignettes/edgeR/inst/doc/edgeRUsersGuide.pdf*

Cheng, L., Fuchigami, L. H. (2000). Rubisco activation state decreases with increasing nitrogen content in apple leaves. *Journal of Experimental Botany*, 51, 1687-1694.

Chiu, C. C., Keeling C. K., Bohlmann, J. (2017). Toxicity of pine monoterpenes to mountain pine beetle. *Scientific Reports*, 7, 1-8.

Chiu, C. C., Keeling, C. I., Bohlmann, J. (2019). The cytochrome P450 CYP6DE1 catalyzes the conversion of α-pinene into the mountain pine beetle aggregation pheromone trans-verbenol. *Scientific Reports*, 9, 1-10.

Chowdhury, S., Basu, A., Kundu, S. (2015). Cloning, characterization, and bacterial over-expression of an osmotin-like protein gene from *Solanum nigrum* L. with antifungal activity against three necrotrophic fungi. *Molecular Biotechnology*, 57, 371-381.

Christiansen, E., Krokene, P., Berryman, A. A., Franceschi, V. R., Krekling, T., Lieutier, F., Lönneborg, A., Solheim, H. (1999). Mechanical injury and fungal infection induce acquired resistance in Norway spruce. *Tree Physiology*, 19, 399-403.

CLC Bio (2013). *De novo assembly with CLC Assembly Cell 4.0: White paper executive summary*. Author, Cambridge, Massachusetts, United States.

Clérivet, A., Déon, V., Alami, I., Lopez, F., Geiger, J., Nicole, M. (2000). Tyloses and gels associated with cellulose accumulation in vessels are responses of plane tree seedlings (*Platanus* × *acerifolia*) to the vascular fungus *Ceratocystis fimbriata* f. sp *platani*. *Trees*, 15, 25-31.

Clérivet, A., El Modafar, C. (1994). Vascular modifications in *Platanus acerifolia* seedlings inoculated with *Ceratocystis fimbriata* f. sp. *platani*. *European Journal of Forest Pathology*, 24, 1–10.

Cock, P. J., Fields, C. J., Goto, N., Heuer, M. L., Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, 38, 1767-1771.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. J., Gaffney, D., Elo, L. L., Zhang, X., Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology,* 17, 1-19.

Cook, S. P., Shirley, B. M., Zambino, P. J. (2010). Nitrogen concentration in mountain pine beetle larvae reflects nitrogen status of the tree host and two fungal associates. *Environmental Entomology*, 39, 821-826.

Cook, S., Carroll, A., Kimsey, M., Shaw, T. (2015). Changes in a primary resistance parameter of lodgepole pine to bark beetle attack one year following fertilization and thinning. *Forests*, 6, 280-292.

Corbett, L. J., Withey, P., Lantz, V. A., Ochuodho, T. O. (2016). The economic impact of the mountain pine beetle infestation in British Columbia: Provincial estimates from a CGE analysis. *Forestry*, 89, 100-105.

Corley, S. M., Mackenzie, K. L., Beverdam, A., Roddam, L. F., Wilkins, M. R. (2017). Differentially expressed genes from RNA-Seq and functional enrichment results are affected by the choice of single-end versus paired-end reads and stranded versus non-stranded protocols. *BMC Genomics*, 18, 1-13.

Cornish-Bowden, A. (1985). Nomenclature for incompletely specified bases in nucleic-acid sequences - Recommendations 1984. *Nucleic Acids Research,* 13, 3021-3030.

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., 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.

Cullingham, C. I., James, P. M. A., Cooke, J. E. K., 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., Janes, J. K., Hamelin, R. C., James, P. M., Murray, B. W., Sperling, F. A. (2019). The contribution of genetics and genomics to understanding the ecology of the mountain pine beetle system. *Canadian Journal of Forest Research*, 49, 721-730.

Dalman, K., Wind, J. J., Nemesio-Gorriz, M., Hammerbacher, A., Lundén, K., Ezcurra, I., Elfstrand, M. (2017). Overexpression of PaNAC03, a stress induced NAC gene family transcription factor in Norway spruce leads to reduced flavonol biosynthesis and aberrant embryo development. *BMC Plant Biology*, 17, 1-17.

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

Davin, L.B., Lewis, N.G. (2000). Dirigent proteins and dirigent sites explain the mystery of specificity of radical precursor coupling in lignan and lignin biosynthesis. *Plant Physiology*, 123, 453-462.

De La Torre, A. R., Birol, I., Bousquet, J., Ingvarsson, P. K., Jansson, S., Jones, S. J., Keeling, C. I., MacKay, J., Nilsson, O., Ritland, K., Street, N., Yanchuk, A., Zerbe, P., Bohlmann, J. (2014). Insights into conifer giga-genomes. *Plant Physiology*, 166, 1724-1732.

Del Fabbro, C., Scalabrin, S., Morgante, M., Giorgi, F. M. (2013). An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PLoS ONE*, 8, 1-13.

Didion, J. P., Martin, M., Collins, F. S. (2017). Atropos: Specific, sensitive, and speedy trimming of sequencing reads. *PeerJ*, 5, 2-19.

Dietrich, R., Ploss, K., Heil, M. (2004). Constitutive and induced resistance to pathogens in *Arabidopsis thaliana* depends on nitrogen supply. *Plant, Cell and Environment*, 27, 896–906.

DiGuistini, S., Wang, Y., Liao, N. Y., Taylor, G., Tanguay, P., Feau, N., Henrissat, B., Chan, S. K., Hesse-Orce, U., Alamouti, S. M., Tsui, C. K. M., Docking, R. T., Levasseu, A., Haridas, S., Robertson, G., Birol, I., Holt, R. A., Marra, M. A., Hamelin, R. C., Hirst, M., Jones, S. J. M., Bohlman, J., Breuil, C. (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*, 108, 2504–2509.

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

Dougherty, P. M., Whitehead, D., Vose, J. M. (1994). Environmental influences on the phenology of pine. *Ecological Bulletins (Copenhagen)*, 43, 64-75.

Earl, D., Bradnam, K., John, J. S., Darling, A., Lin, D., Fass, J., Paten, B. (2011). Assemblathon 1: A competitive assessment of *de novo* short read assembly methods. *Genome Research*, 21, 2224-2241.

Ericsson, T. (1995). Growth and shoot-root ratio of seedlings in relation to nutrient availability. *Plant Soil*, 168, 205-214.

Evans J. R., Seemann J. R. (1989). The allocation of protein nitrogen in the photosynthetic apparatus: Cost, consequences and control. In Briggs, W. R. (Ed.), *Photosynthesis*. Alan R. Liss, Inc., New York, New York, United States. pp. 183–205.

Ewing, B., Green, P. (1998). Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Research*, 8, 186-194.

Ewing, B., Hillier, L., Wendl, M. C., Green, P. (1998). Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Research*, 8, 175-185.

Fagard, M., Launay, A., Clement, G., Courtial, J., Dellagi, A., Farjad, M., Krapp, A., Soulié, M. C., Masclaux-Daubresse, C. (2014). Nitrogen metabolism meets phytopathology. *Journal of Experimental Botany*, 65, 5643-5656.

Fasulo, D., Halpern, A., Dew, I., Mobarry, C. (2002). Efficiently detecting polymorphisms during the fragment assembly process. *Bioinformatics*, 18, S294–S302.

Ferrenberg, S., Kane, J. M., Mitton, J. B. (2014). Resin duct characteristics associated with tree resistance to bark beetles across lodgepole and limber pine. *Oecologia*, 174, 1283–1292.

Fettig, J. C., Klepzig, K. D., Billings, R. F., Munson, A. S., Nebeker, T. E., Negrón, J. F., Nowak, J. T. (2007). The effectiveness of vegetation management practices for prevention and control of bark beetle infestations in coniferous forests of the western and southern United States. *Forest Ecology and Management*, 238, 24–53.

Fisher, R. A. (1934). *Statistical methods for research workers* (5th ed.). Oliver and Boyd, Edinburgh, Scotland.

Flor, H. H. (1971). Current status of the gene-for-gene concept. *Annual Review of Phytopathology*, 9, 275-296.

Fossdal, C. G., Hietala, A. M., Kvaalen, H., Solheim, H. (2006). Changes in host chitinase isoforms in relation to wounding and colonization by *Heterobasidion annosum*: Early and strong defense response in 33-year-old resistant Norway spruce clone. *Tree Physiology*, 26, 169-177.

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.

Fox, J., Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage Publications, Inc., Thousand Oaks, California, United States.

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, 353-375.

Fuller M. P., Jellings A. J. (2003). Crop physiology. In Soffe, R. (Ed.), *Agriculture notebook*. Butterworths & Co Ltd, London, England.

Galindo-González, L. M., Kayal, W. E., Morris, J. S., Cooke, J. E. K. (2015). Diverse chitinases are invoked during the activity-dormancy transition in spruce. *Tree Genetics and Genomes*, 11, 3-21.

Galindo-González, L. M., Kayal, W. E., Ju, C. J., Allen, C. C., King-Jones, S., Cooke, J. E. K. (2012). Integrated transcriptomic and proteomic profiling of white spruce stems during the transition from active growth to dormancy. *Plant, Cell and Environment,* 35, 682-701.

Gibson, K., Kegley, S., Bentz, B. (2009). *Forest insect & disease leaflet 2*. U. S. Department of Agriculture, Forest Service, Pacific Northwest Region (R6), Portland, Oregon, United States.

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

Goodsman, D. W., Erbilgin, N., Lieffers, V. J. (2012). The impact of phloem nutrients on overwintering mountain pine beetles and their fungal symbionts. *Environmental Entomology*, 41, 478-486.

Government of Alberta (2010). *Mountain pine beetle & cold temperatures: The facts*. Ministry of Environment and Park, Environment and Sustainable Resource Development, Edmonton, Alberta, Canada.

Government of British Columbia (2006). *Forest fertilization in British Columbia*. British Columbia Forest Service, Ministry of Forests and Range, Victoria, British Columbia, Canada.

Government of British Columbia (2013). *FFT current reforestation (CR) and timber supply mitigation (TSM) 5 year update and annual operating plan (AOP) guidance*. British Columbia Forest Service, Forests, Lands and Natural Resource Operations, Victoria, British Columbia, Canada.

Graves, S., Piepho, H.-P., Selzer, L. Dorai-Raj, S. (2019). multcompView: Visualizations of paired comparisons. R package version 0.1-8. *https://CRAN.R-project.org/package=multcompView*.

Gruffman, L., Ishida, T., Nordin, A., Näsholm, T. (2012). Cultivation of Norway spruce and Scots pine on organic nitrogen improves seedling morphology and field performance. *Forest Ecology and Management*, 276, 118-124.

Guérard, N., Dreyer, E., Lieutier, F. (2000). Interactions between Scots pine, *Ips acuminatus* (Gyll.) and *Ophiostoma brunneociliatum* (Math.): Estimation of the critical thresholds of attack and inoculation densities and effects on hydraulic properties in the stem. *Annals of Forest Science, 57*, 681-690.

Guimarães, A. C., Meireles, L. M., Lemos, M. F., Guimarães, M. C., Endringer, D. C., Fronza, M., Scherer, R. (2019). Antibacterial activity of terpenes and terpenoids present in essential oils. *Molecules*, 24, 1-12.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, N., William, T., Dewey, C. N.,

Henschel, R., LeDuc, R. D., Friedman, N., Regev, A. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8, 1494-1512.

Hacke, U. G., Plavcova, L., Almeida-Rodriguez, A., King-Jones, S., Zhou, W., Cooke, J. E. K. (2010). Influence of nitrogen fertilization on xylem traits and aquaporin expression in stems of hybrid poplar. *Tree Physiology, 30*, 1016-1025.

Hakim, Ullah, A., Hussain, A., Shaban, M., Khan, A. H., Alariqi, M., Gulb, S., Juna, Z., Lina, S., Lia, J., Jina, S., Munis, M. F. H. (2018). Osmotin: A plant defense tool against biotic and abiotic stresses. *Plant Physiology and Biochemistry*, 123, 149-159.

Hall, D. E., Yuen, M. M., Jancsik, S., Quesada, A., Dullat, H. K., Li, M., Henderson, H., Arango-Velez, A., Liao, N. Y., Docking, R. T., Chan, S. K., Cooke, J. E. K., Breuil, C., Jones, S. J. M., Keeling, C. I., Bohlmann, J. (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, 1-14.

Haukioja, E., Ossipov, V., Koricheva, J., Honkanen, T., Larsson, S., Lempa, K. F. (1998). Biosynthetic origin of carbon-based secondary compounds: Cause of variable responses of woody plants to fertilization? *Chemoecology*, 8, 133–139.

Hawkins, B. J., Boukcim, H., Plassard, C. (2008). A comparison of ammonium, nitrate and proton net fluxes along seedling roots of Douglas-fir and lodgepole pine grown and measured with different inorganic nitrogen sources. *Plant, Cell and Environment*, 31, 278–287.

Heldt, H.-W., Piechulla, B. (2010). *Plant biochemistry* (4th ed.). Elsevier Academic Press, Cambridge, Massachusetts, United States.

Herbert, Z. T., Kershner, J. P., Butty, V. L., Thimmapuram, J., Choudhari, S., Alekseyev, Y. O., Fan, J., Podnar, J. W., Wilcox, E., Gipson, J., Gillaspy, A., Jepsen,

K., Splinter, S., Durant, B., Morris, K., Berkeley, M, LeClerc, A., Simpson, S. D., Sommerville, G., Grimmett, L., Adams, M., Levine, S. S. (2017). Cross-site comparison of ribosomal depletion kits for Illumina RNAseq library construction. *BMC Genomics*, 19, 2-10.

Herms, D. A., Mattson, W. J. (1992). The dilemma of plants: To grow or defend. *The Quarterly Review of Biology*, 67, 283–335.

Hocking, D. (1971). Preparation and use of a nutrient solution for culturing seedlings of lodgepole pine and white spruce, with selected bibliography. *Northern Forest Research Centre Information Report Nor-X-1*. Canadian Forest Service, Department of the Environment, Edmonton, Alberta, Canada.

Hodge, J., Cooke, B., McIntosh, R. (2017). *A strategic approach to slow the spread of mountain pine beetle across Canada*. Canadian Council of Forest Ministers, Forest Pest Working Group.

Hoffland, E., van Beusichem, M. L., Jeger, M. J. (1999). Nitrogen availability and susceptibility of tomato leaves to *Botrytis cinerea*. *Plant and Soil*, 210, 263–272.

Hoffland, E., van Beusichem, M. L., Jeger, M. J. (2000). Effect of nitrogen supply rate on disease resistance in tomato depends on the pathogen. *Plant and Soil*, 218, 239–247.

Högberg, P., Näsholm, T., Franklin, O., Högberg, M. N. (2017). Tamm review: On the nature of the nitrogen limitation to plant growth in Fennoscandian boreal forests. *Forest Ecology and Management*, 403, 161-185.

Hölzer, M., Marz, M. (2019). De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. *GigaScience*, 8, 1-16.

Huang, X. (1999). CAP3: A DNA sequence assembly program. *Genome Research*, 9, 868-877.

Huang, Y., Niu, B., Gao, Y., Fu, L., Li, W. (2010). CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics,* 26, 680-682.

Hubbard, R. M., Rhoades, C. C., Elder, K., Negron, J. (2013). Changes in transpiration and foliage growth in lodgepole pine trees following mountain pine beetle attack and mechanical girdling. *Forest Ecology and Management,* 289, 312-317.

Huber, D., Römheld, V., Weinmann, M. (2012). Relationship between nutrition, plant diseases and pests. In Marschner, H. (Ed.), *Marschner's mineral nutrition of higher plants*. Academic Press, Cambridge, Massachusetts, USA. pp. 283–298.

Huber, D. P. W., Gries, R., Borden, J. H., Pierce, H. D. (2000). A survey of antennal response by five species of coniferophagous bark beetles (Coleoptera: Scolytidae) to bark volatiles of six species of angiosperm trees. *Chemoecology*, 10, 103-113.

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., 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., 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.

Huffaker, A., Pearce, G., Ryan, C. A. (2006). An endogenous peptide signal in Arabidopsis activates components of the innate immune response. *Proceedings of the National Academy of Sciences,* 103, 10098-10103.

Huse, S. M., Huber, J. A., Morrison, H. G., Sogin, M. L., Welch, D. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*, 8, R143.1-R143.9.

Illumina, Inc. (2013). *Illumina TruSeq stranded mRNA low sample (LS) protocol*. Author, San Diego, California, United States.

Illumina, Inc. (2015). *NextSeq system denature and dilute libraries protocol*. Author, San Diego, California, United States.

Illumina, Inc. (2016). *Illumina TruSeq indexed adapters*. Author, San Diego, California, United States.

Illumina, Inc. (2018). *NextSeq 500 system guide*. Author, San Diego, California, United States.

Ishihama, N., Yoshioka, H. (2012). Post-translational regulation of WRKY transcription factors in plant immunity. *Current Opinion in Plant Biology*, 15, 431-437.

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

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., Parcy, F. (2002). BZIP transcription factors in *Arabidopsis*. *Trends in Plant Science*, 7, 106-111.

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.

Jensen, B., Munk, L. (1997). Nitrogen-induced changes in colony density and spore production of *Erysiphe graminis* f. sp. *hordei* on seedlings of six spring barley cultivars. *Plant Pathology*, 46, 191–202.

Jones, C. G., Hartley, S. E. (1999). A protein competition model of phenolic allocation. *Oikos*, 86, 27-44.

Jung, H., Yoon, B. H., Kim, W. J., Kim, D. W., Hurwood, D. A., Lyons, R. E., Salin, K. R., Kim, H.-S., Baek, I., Chand, V., Mather, P. B. (2016). Optimizing hybrid *de novo* transcriptome assembly and extending genomic resources for giant freshwater prawns (*Macrobrachium rosenbergii*): The identification of genes and markers associated with reproduction. *International Journal of Molecular Sciences,* 17, 1-21.

Jwa, N., Hwang, B. K. (2017). Convergent evolution of pathogen effectors toward reactive oxygen species signaling networks in plants. *Frontiers in Plant Science*, 8, 1-12.

Kaiser, W. M., Weiner, H., Huber, S. C. (1999). Nitrate reductase in higher plants: A case study for transduction of environmental stimuli into control of catalytic activity. *Physiologia Plantarum*, 105, 384-389.

Kane, J. M., Kolb, T. E. (2010). Importance of resin ducts in reducing Ponderosa pine mortality from bark beetle attack. *Oecologia,* 164, 601-609.

Kanehisa, M., Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28, 27-30.

Kang, Z., Liu, F., Zhang, Z., Tian, H., Liu, T. (2018). Volatile β-ocimene can regulate developmental performance of peach aphid *Myzus persicae* through

activation of defense responses in Chinese cabbage *Brassica pekinensis*. *Frontiers in Plant Science,* 9, 1-12.

Keefover-Ring, K., Trowbridge, A., Mason, C. J., Raffa, K. F. (2015). Rapid induction of multiple terpenoid groups by ponderosa pine in response to bark beetle-associated fungi. *Journal of Chemical Ecology*, 42, 1-12.

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

Keeling, C., Weisshaar, S., Lin, R. P. C., Bohlmann, J. (2008). Functional plasticity of paralogous diterpene synthases involved in conifer defense. *PNAS*, 105, 1085–1090.

Kishimoto, K., Matsui, K., Ozawa, R., Takabayashi, J. (2006). Analysis of defensive responses activated by volatile allo-ocimene treatment in *Arabidopsis thaliana*. *Phytochemistry*, 67, 1520-1529.

Klenner, W., Walton, R., Arsenault, A., Kremsater, L. (2008). Dry forests in the southern interior of British Columbia: Historic disturbances and implications for restoration and management. *Forest Ecology and Management,* 256, 1711-1722.

Klepzig, K. D., Robison, D. J., Fowler, G., Minchin, P. R., Hain, F. P., Allen, H. L. (2005). Effects of mass inoculation on induced oleoresin response in intensively managed loblolly pine. *Tree Physiology*, 25, 681–688.

Köchy, M., Wilson, S. D. (2005). Variation in nitrogen deposition and available soil nitrogen in a forest–grassland ecotone in Canada. *Landscape Ecology*, 20, 191-202.

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

Komsta, L. (2011). Outliers: Tests for outliers. R package version 0.14. *https://CRAN.R-project.org/package=outliers*.

Koricheva, J., M., Keinanen, M. (1998). Regulation of woody plant secondary metabolism by resource availability: Hypothesis testing by means of meta-analysis. *Oikos*, 83, 212-226.

Kovalchuk, A., Keriö, S., Oghenekaro, A. O., Jaber, E., Raffaello, T., Asiegbu, F. O. (2013). Antimicrobial defenses and resistance in forest trees: Challenges and perspectives in a genomic era. *Annual Review of Phytopathology*, 51, 221–244.

Kestler, H. A., Muller, A., Gress, T. M., Buchholz, M. (2004). Generalized Venn diagrams: A new method of visualizing complex genetic set relations. *Bioinformatics*, 21, 1592-1595.

Krokene, P., Nagy, N. E. (2012). 5. Anatomical aspects of resin-based defences in pine. In Fett-Neto, A. G., Rodrigues-Corrêa, K. C. S. (Eds.), *Pine Resin: Biology, Chemistry and Applications*. Research Signpost, Thiruvananthapuram, India. pp. 67-86.

Kukurba, K. R., Montgomery, S. B. (2015). RNA sequencing and analysis. *Cold Spring Harbor Protocols*, 2015, 951–969.

Kurz, W. A., Dymond, C. C., Stinson, G., Rampley, G. J., Neilson, E. T., Carroll, A. L., Ebata, T., Safranyik, L. (2008). Mountain pine beetle and forest carbon feedback to climate change. *Nature*, 452, 987-990.

Kvam, V. M., Liu, P., Si, Y. (2012). A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. *American Journal of Botany, 99*, 248-256.

Kytö, M., Niemelä, P., Annila, E. (1996). Vitality and bark beetle resistance of fertilized Norway spruce. *Forest Ecology and Management*, 84, 149–157.

Lacerda, A. F., Vasconcelos, Ã. A., Pelegrini, P. B., Sa, M. F. (2014). Antifungal defensins and their role in plant defense. *Frontiers in Microbiology, 5*, 1-10.

Laluk, K., Mengiste, T. (2010). Necrotroph attacks on plants: Wanton destruction or covert extortion? *The Arabidopsis Book*, 8, 1-24.

Lahr, E. C., Krokene, P. (2013). Conifer stored resources and resistance to a fungus associated with the spruce bark beetle *Ips typographus*. *PLoS ONE*, 8, 1-8.

Lamarre, S., Frasse, P., Zouine, M., Labourdette, D., Sainderichin, E., Hu, G., Berre-Anton, V. L., Bouzayen, M., Maza, E. (2018). Optimization of an RNA-Seq differential gene expression analysis depending on biological replicate number and library size. *Frontiers in Plant Science,* 9*,* 1-18.

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D. L., Garcia-Hernadez, M., Karthikeyan, A. S., Lee, C. H., Nelson, W. D., Ploetz, L., Singh, S., Wensel, A., Huala, E. (2011). The *Arabidopsis* Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Research*, 40, D1202-D1210.

Langfelder, P., Horvath, S. (2012). Fast R functions for robust correlations and hierarchical clustering. *Journal of Statistical Software*, 46, 1-17.

Langfelder, P., Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, 1-13.

Larisch, C., Dittrich, M., Wildhagen, H., Lautner, S., Fromm, J., Polle, A., Hedrich, R., Rennenberg, H., Müller, T., Ache, P. (2012). Poplar wood rays are involved in seasonal remodeling of tree physiology. *Plant Physiology*, 160, 1515-1529.

Lee, S., Kim, J., Breuil, C. (2006). Pathogenicity of *Leptographium longiclavatum* associated with *Dendroctonus ponderosae* to *Pinus contorta*. *Canadian Journal of Forest Research,* 36, 2864-2872.

Legendre, P, Legendre, L. (1998). *Numerical ecology* (2nd ed). Elsevier, Amsterdam, Netherlands. pp. 853.

Lehri, B., Seddon, A. M., Karlyshev, A. V. (2017). The hidden perils of read mapping as a quality assessment tool in genome sequencing. *Scientific Reports*, 7, 1-8.

Leser, C., Treutter, D. (2005). Effects of nitrogen supply on growth, contents of phenolic compounds and pathogen (scab) resistance of apple trees. *Physiologia Plantarum*, 123, 49–56.

Lenth, R. (2020). emmeans: Estimated marginal means, aka least-squares means. R package version 1.4.5. *https://CRAN.R-project.org/package=emmeans*.

Levene, H. (1960). *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (Olkin, I., Ghurye, S. G., Hoeffding, W., Madow, W. G., Mann, H. B, Eds.). Stanford University Press, Palo Alto, California, United States. pp. 278-292.

Liang, X., Zhou, J. (2018). Receptor-like cytoplasmic kinases: Central players in plant receptor kinase–mediated signaling. *Annual Review of Plant Biology*, 69, 267-299.

Liao, Y., Smyth, G. K., Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30, 923-930.

Lieutier, F. (2002). Mechanisms of resistance in conifers and bark beetle attack strategies. In Wagner, M. R., Clancy, K. M., Lieutier, F., Paine, T. D. (Eds.), *21st International congress of entomology*. Springer, Iguassu Falls, Brazil. pp. 105–130.

Lieutier, F., Yart, A., Salle, A. (2009). Stimulation of tree defenses by *Ophiostomatoid* fungi can explain attack success of bark beetles on conifers. *Annals of Forest Science,* 66, 801-801.

Little, E. L., Jr. (1971). *Atlas of United States trees. Volume 1: Conifers and important hardwoods*. US Department of Agriculture Forest Service, Washington, D.C., United States. Miscellaneous Publication Number 1146.

Liu, J.-J., Ekramoddoullah, A. K. M., 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, Y. W., Zhou, J., White, K. P. (2014). RNA-seq differential expression studies: More sequence or more replication? *Bioinformatics,* 30, 301-304.

Lohse, M., Nagel, A., Herter, T., May, P., Schroda, M., Zrenner, R., Tohge, T., Fernie, A. R., Stitt, M., Usadel, B. (2014). Mercator: A fast and simple web server for genome scale functional annotation of plant sequence data. *Plant, Cell and Environment*, 37, 1250–1258.

Lorio, P. L. (1986). Growth-differentiation balance: A basis for understanding southern pine beetle-tree interactions. *Forest Ecology and Management*, 14, 259-273.

Lotan, J. E., Critchfield, W. B. (1990). Lodgepole pine. In Burns, R. M., Barbara H. H. (Tech. Coords.), *Silvics of North America volume 1. Conifers. Agriculture handbook 654.* U.S. Department of Agriculture, Forest Service, Washington, DC, United States.

Lourenço, A., Rencoret, J., Chemetova, C., Gominho, J., Gutiérrez, A., Río, J. C., Pereira, H. (2016). Lignin composition and structure differs between xylem, phloem and phellem in *Quercus suber* L. *Frontiers in Plant Science*, 7, 1-14.

Love, M. I., Huber, W., Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*, 1-21.

Luo, J., Zhou, J.-J., Masclaux-Daubresse, C., Wang, N., Wang, H., Zheng, B. (2019). Morphological and physiological responses to contrasting nitrogen regimes in *Populus cathayana* is linked to resources allocation and carbon/nitrogen partition. *Environmental and Experimental Botany*, 162, 247–255.

Lupi, C. (2013). Role of soil nitrogen for the conifers of the boreal forest: A critical review. *International Journal of Plant Soil Science*, 2, 155-189.

Lusebrink, I., Erbilgin, N., Evenden, M. L. (2013). The lodgepole × Jack pine hybrid zone in Alberta, Canada: A stepping stone for the mountain pine bBeetle on its journey east across the boreal forest? *Journal of Chemical Ecology*, 39, 1209-1220.

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

MacManes, M. D. (2014). On the optimal trimming of high-throughput mRNA sequence data. *Frontiers in Genetics*, 5, 1-7.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J.-B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F., Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437, 376-380.

Marín, I. C., Loef, I., Bartetzko, L., Searle, I., Coupland, G., Stitt, M., Osuna, D. (2011). Nitrate regulates floral induction in *Arabidopsis*, acting independently of light, gibberellin and autonomous pathways. *Planta*, 233, 539–552.

Marpeau, A., Walter, J., Launay, J., Charon, J., Baradat, P., Gleizes, M. (1989). Effects of wounds on the terpene content of twigs of maritime pine (*Pinus pinaster* Ait.). *Trees*, 3, 220-226.

Martin, D., Tholl, D., Gershenzon, J., Bohlmann, 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.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17, 1-10.

Martin, J. A., Wang, Z. (2011). Next-generation transcriptome assembly. *Nature Reviews Genetics*, 12, 671-682.

Maslov, S., Sneppen, K. (2006). Chapter 3 Large-Scale Topological Properties of Molecular Networks. In Koonin, E. V., Wolf, Y. I., Karev, G. P. (Eds.), *Power Laws, Scale-Free Networks and Genome Biology*. Springer Science+Business Media, Berlin, Germany. pp. 25-39.

Massad, T. J., Trumbore, S. E., Ganbat, G., Reichelt, M., Unsicker, S., Boeckler, A., Gleixner, G., Gershenzon, J., Ruehlow, S. (2014). An optimal defense strategy for phenolic glycoside production in *Populus trichocarpa*- Isotope labeling demonstrates secondary metabolite production in growing leaves. *New Phytologist*, 203, 607-619.

Matyssek, R., Agerer, R., Ernst, D., Munch, J.-C., Osswald, W., Pretzsch, H., Priesack, E., Schnyder, H., Treutter, D. (2005). The plant's capacity in regulating resource demand. *Plant Biology*, 7, 560–580.

Matyssek, R., Schnyder, H., Elstner, E.-F., Munch, J.-C., Pretzsch, H., Sandermann, H. (2002). Growth and parasite defence in plants: The balance between resource sequestration and retention. *Plant Biology*, 4, 133–136.

McAllister, C. H., Fortier, C. E., Onge, K. R. S., Sacchi, B. M., Nawrot, M. J., Locke, T., Cooke, J. E. K. (2018). A novel application of RNase H2-dependent quantitative PCR for detection and quantification of *Grosmannia clavigera*, a mountain pine beetle fungal symbiont, in environmental samples. *Tree Physiology*, 38, 485–501.

McCarthy, D. J., Chen, Y., Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, 40, 4288-4297.

McClure, R. S., Overall, C. C., McDermott, J. E., Hill, E. A., Markillie, L. M., McCue, L. A., Beliaev, A. S. (2016). Network analysis of transcriptomics expands regulatory landscapes in *Synechococcus* sp. PCC 7002. *Nucleic Acids Research, 44*, 8810-8825.

McHale, L., Tan, X., Koehl, P., Michelmore, R. W. (2006). Plant NBS-LRR proteins: Adaptable guards. *Genome Biology*, 7, 212.1-212.11.

Mekonnen, Z. A., Riley, W. J., Randerson, J. T., Grant, R. F., Rogers, B. M. (2019). Expansion of high-latitude deciduous forests driven by interactions between climate warming and fire. *Nature Plants*, 5, 952–958.

Meldau, S., Erb, M., Baldwin, I. T. (2012). Defence on demand: Mechanisms behind optimal defence patterns. *Annals of Botany*, 110, 1503-1514.

Meng, X., Zhang, S. (2013). MAPK cascades in plant disease resistance signaling. *Annual Review of Phytopathology*, 51, 245–266.

Mentzen, W. I., Wurtele, E. (2008). Regulon organization of Arabidopsis. *BMC Plant Biology, 44*, 8810-8825.

Mercado, J. E., Hofstetter, R. W., Reboletti, D. M., Negrón, J. F. (2014). Phoretic symbionts of the mountain pine beetle (*Dendroctonus ponderosae* Hopkins). *Forest Science*, 60, 512-526.

Meuriot, F., Noquet, C., Avice, J.-C., Volenec, J. J., Cunningham, S. M., Sors, T. G., Caillot, S., Ourry, A. (2004). Methyl jasmonate alters N partitioning, N reserves accumulation and induces gene expression of a 32-kDa vegetative storage protein that possesses chitinase activity in *Medicago sativa* taproots. *Plant Physiology*, 120, 113–123.

Miedes, E., Vanholme, R., Boerjan, W., Molina, A. (2014). The role of the secondary cell wall in plant resistance to pathogens. *Frontiers in Plant Science*, 5, 1-13.

Mihaliak, A. C., Lincoln, D. E. (1985). Growth pattern and carbon allocation to volatile leaf terpenes under nitrogen-limiting conditions. *Journal of Chemical Ecology*, 13, 2059–2067.

Millard, P., Gretlet, G.-A. (2010). Nitrogen storage and remobilization by trees: Ecophysiological relevance in a changing world. *Tree Physiology*, 30, 1083–1095.

Miller, A. J., Cramer, M. D. (2004). Root nitrogen acquisition and assimilation. *Plant Soil*, 274, 1-36.

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.

Miller, J. R., Koren, S., Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics*, 95, 315-327.

Miron, D., Battisti, F., Silva, F. K., Lana, A. D., Pippi, B., Casanova, B., Gnoattoc, S., Fuentefriab, A., Mayorgad, P., Schapoval, E. E. (2014). Antifungal activity and

mechanism of action of monoterpenes against dermatophytes and yeasts. *Revista Brasileira de Farmacognosia*, 24, 660-667.

Molina-Mora, J. A., Campos-Sánchez, R., Rodríguez, C., Shi, L., García, F. (2020). High quality 3C *de novo* assembly and annotation of a multidrug resistant ST-111 Pseudomonas aeruginosa genome: Benchmark of hybrid and non-hybrid assemblers. *Scientific Reports*, 10, 1-16.

Monteiro, S., Barakat, M., Piçarra-Pereira, M. A., Teixeira, A. R., Ferreira, R. B. (2003). Osmotin and thaumatin from grape: A putative general defense mechanism against pathogenic fungi. *Phytopathology, 93*, 1505-1512.

Moreira, X., Zas, R., Sampedro, L. (2012). Differential allocation of constitutive and induced chemical defenses in pine tree juveniles: A test of the optimal defense theory. *PLoS ONE*, 7, 1-8.

Morris, H., Brodersen, C., Schwarze, F. W. M. R., Jansen, S. (2016). The parenchyma of secondary xylem and its critical role in tree defense against fungal decay in relation to the CODIT model. *Frontiers in Plant Science*, 7, 1-18.

Muñoz-Huerta, R., Guevara-Gonzalez, R., Contreras-Medina, L., Torres-Pacheco, I., Prado-Olivarez, J., Ocampo-Velazquez, R. (2013). A review of methods for sensing the nitrogen status in plants: Advantages, disadvantages and recent advances. *Sensors*, 13, 10823–10843.

Mur, L. A. J., Simpson, C., Kumari, A., Gupta, A. K., Gupta, K. J. (2016). Moving nitrogen to the centre of plant defense against pathogens. *Annals of Botany*, 119, 703–709.

Muzika, R. M., Pregitzer, K. S., Hanover, J. W. (1989). Changes in terpene production following nitrogen fertilization of grand fir (*Abies grandis* (Dougl.) Lindl.) seedlings. *Oecologia*, 80, 485-489.

Nagy, N. E., Franceschi, V. R., Solheim, H., Krekling, T., Christiansen, E. (2000). Wound-induced traumatic resin duct development in stems of Norway spruce (*Pinaceae*): Anatomy and cytochemical traits. *American Journal of Botany*, 87, 302-313.

Näsholm, T., Ekblad, A., Nordin, A., Giesler, R., Högberg, M., Högberg, P. (1998). Boreal forest plants take up organic nitrogen. *Nature*, 392, 914-916.

Natural Resources Canada (2004). *Mountain pine beetle management, A guide for small woodland operations*. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia, Canada.

Natural Resources Canada (2017a, February 21). Mountain pine beetle (factsheet). *https://www.nrcan.gc.ca/forests/fire-insects-disturbances/top-insects/13397*. (Last accessed April 13, 2020).

Natural Resources Canada (2017b, June 14). Forest classification. *https://www.nrcan.gc.ca/our-natural-resources/forests-forestry/sustainable-forest-management/measuring-reporting/forest-classification/13179*. (Last accessed April 13, 2020).

Natural Resources Canada (2020, May 20). How much forest does Canada have? *https://www.nrcan.gc.ca/our-natural-resources/forests-forestry/how-much-forest-does-canada-have/17601*. (Last accessed July 1, 2020).

Negrón, J. F., Cain, B. (2018). Mountain pine beetle in Colorado: A story of changing forests. *Journal of Forestry, 117*, 144-151.

Nelson, M. F., Murphy, J. T., Bone, C., Altaweel, M. (2018). Cyclic epidemics, population crashes, and irregular eruptions in simulated populations of the mountain pine beetle, *Dendroctonus ponderosae*. *Ecological Complexity*, 36, 218-229.

Neuhaus, J. M. (1999). Plant chitinases (PR-3, PR-4, PR-8, PR-11). In Datta, S. K., Muthukrishnan, S. (Eds.), *Pathogenesis-related proteins in plants*. CRC Press, Boca Ranton, Florida, United States. pp. 77–105.

Nguyen, N. T., Nakabayashi, K., Mohapatra, P. K., Thompson, J., Fujita, K. (2003). Effect of nitrogen deficiency on biomass production, photosynthesis, carbon partitioning, and nitrogen nutrition status of *Melaleuca* and *Eucalyptus* species. *Soil Science and Plant Nutrition*, 49, 99-109.

Nolan, T., Hands, R. E., Bustin, S. A. (2006). Quantification of mRNA using real-time RT-PCR. *Nature Protocols*, 1, 1559-1582.

Nongpiur, R., Soni, P., Karan, R., Singla-Pareek, S. L., Pareek, A. (2012). Histidine kinases in plants. *Plant Signaling and Behavior*, 7, 1230-1237.

Nystedt, B., Street, N., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D. G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., Vicedomini, R., Sahlin, K., Sherwood, E., Elfstrand, M., Gramzow, L., Holmberg, K., Hällman, J., Keech, O., Klasson, L., Koriabine, M., Kucukoglu, M., Käller, M., Luthman, J., Lysholm, F., Niittylä, T., Olson, A., Rilakovic, N., Ritland, C., Rosselló, J. A., Sena, J., Svensson, T., Talavera-López, C., Theißen, G., Tuominen, H., Vanneste, K., Wu, Z.-Q., Zhang, B., Zerbe, P., Arvestad, L., Bhalerao, R., Bohlmann, J., Bousquet, J., Gil, R. G., Hvidsten, T. R., de Jong, P., MacKay, J., Morgante, M., Ritland, K., Sundberg, B., Thompson, S. L., Van de Peer, Y., Andersson, B., Ingvarsson, P. K., Lundeberg, J., Jansson, S. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, 497, 579–584.

Ogawa, S., Miyamoto, K., Nemoto, K., Sawasaki, T., Yamane, H., Nojiri, H., Okada, K. (2017). OsMYC2, an essential factor for JA-inductive sakuranetin production in rice, interacts with MYC2-like proteins that enhance its transactivation ability. *Scientific Reports*, 7, 1-11.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H. (2019). vegan: Community ecology package. R package version 2.5-6. *https://CRAN.R-project.org/package=vegan*.

Oliveros, J. C. (2015) Venny: An interactive tool for comparing lists with Venn's diagrams. *https://bioinfogp.cnb.csic.es/tools/venny/index.html*.

Olli, S., Kirti, P. (2006). Cloning, characterization and antifungal activity of defensin Tfgd1 from *Trigonella foenum-graecum* L. *Journal of Biochemistry and Molecular Biology*, 39, 278-283.

Opitz, L., Salinas-Riester, G., Grade, M., Jung, K., Jo, P., Emons, G., Ghadimi, B. M., Beißbarth, T., Gaedcke, J. (2010). Impact of RNA degradation on gene expression profiling. *BMC Medical Genomics, 3*, 1-14.

Orenstein, Y., Shamir, R. (2013). Design of shortest double-stranded DNA sequences covering all k-mers with applications to protein-binding microarrays and synthetic enhancers. *Bioinformatics*, 29, 71-79.

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.

Ott, D. S., Yanchuk, A. D., Huber, D. P. W., Wallin, K. F. (2012). Genetic variation of lodgepole pine (*Pinus contorta* var. *latifolia*) chemical and physical defenses that affect mountain pine beetle, *Dendroctonus ponderosae,* attack and tree mortality. *Journal of Chemical Ecology*, 37, 1002–1012.

Pagán, I., García-Arenal, F. (2018). Tolerance to plant pathogens: Theory and experimental evidence. *International Journal of Molecular Sciences*, 19, 1-17.

Page, W. G., Jenkins, M. J., Runyon, J. B. (2012). Mountain pine beetle attack alters the chemistry and flammability of lodgepole pine foliage. *Canadian Journal of Forest Research*, 42, 1631-1647.

Paine, T. D., Raffa, K. F., Harrington, T. C. (1997). Interactions among scolytid bark beetles, their associated fungi, and live host conifers. *Annual Review of Entomology*, 42, 179–206.

Palma, A. C., Winter, K., Aranda, J., Dalling, J. W., Cheesman, A. W., Turner, B. L., Cernusak, L. A. (2020). Why are tropical conifers disadvantaged in fertile soils? Comparison of *Podocarpus guatemalensis* with an angiosperm pioneer *Ficus insipida*. *Tree Physiology,* 40, 810-821.

Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H., Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Research,* 37, 1-7.

Pascual, M. B., Cánovas, F. M., Ávila, C. (2015). The NAC transcription factor family in maritime pine (*Pinus Pinaster*): Molecular regulation of two genes involved in stress responses. *BMC Plant Biology*, 15, 2-15.

Patzlaff, A., Mcinnis, S., Courtenay, A., Surman, C., Newman, L. J., Smith, C., Bevan, M. W., Mansfield, S., Whetten, R. W., Sederoff, R. R., Campbell, M. M. (2003). Characterisation of a pine MYB that regulates lignification. *The Plant Journal*, 36, 743-754.

Pavy, N., Boyle, B., Nelson, C., Paule, C., Giguère, I., Caron, S., Parsons, L. S., Dallaire, N., Bedon, F., Bérubé, H., Cooke, J., Mackay, J. (2008). Identification of conserved core xylem gene stets: Conifer cDNA microarray development, transcript profiling and computational analyses. *New Phytology*, 180, 766–786.

Pearson, W. R., Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences*, 85, 2444-2448.

Peery, R. M., McAllister, C. H., Cullingham, C. I., Mahon, E. L., Arango-Velez, A., Cooke, J. E. K. Comparative genomics of the chitinase gene family in lodgepole and jack pines: Contrasting responses to biotic threats and landscape level investigation of genetic differentiation. Submitted to *Botany*, July 5, 2020, cjb-2020-0125.

Peoples M. B., Gifford R. M. (1990). Long-distance transport of nitrogen and carbon from sources to sinks in higher plants. In Dennis, D. T., Turpin, D. H. (Eds.), *Plant Physiology Biochemistry and Molecular Biology*. Longman, Harlow, Essex, England. pp. 434–447.

Peumans, W. J., Proost, P., Swennen, R. L., Van Damme, E. J. M. (2002). The abundant class III chitinase homolog in young developing banana fruits behaves as a transient vegetative storage protein and most probably serves as an important supply of amino acids for the synthesis of ripening-associated proteins. *Plant Physiology*, 130, 1063–1072.

Pevzner, P. A., Tang, H. X., Waterman, M. S. (2001). An Eulerian path approach to DNA fragment assembly. *Proceedings of the National Academy of Sciences of the United States of America,* 98, 9748-9753.

Pieterse, C. M., Does, D. V., Zamioudis, C., Leon-Reyes, A., Wees, S. C. (2012). Hormonal modulation of plant immunity. *Annual Review of Cell and Developmental Biology, 28*, 489-521.

Pieterse, C. M., Leon-Reyes, A., Ent, S. V., 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.

Pillai, K. C. (2004). Multivariate analysis of variance (MANOVA). In Kotz, S., Read, C. B., Balakrishnan, N., Vidakovic, B., Johnson, N. L. (Eds.), *Encyclopedia of statistical sciences*. John Wiley & Sons, Inc., Hoboken, New Jersey, United States.

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.

Pré, M., Atallah, M., Champion, A., Vos, M. D., Pieterse, C. M., Memelink, J. (2008). The AP2/ERF domain transcription factor ORA59 integrates jasmonic acid and ethylene signals in plant defense. *Plant Physiology*, 147, 1347-1357.

Prescott, C., de Montigny, L., Harper, G. (2019). Eighteen-year growth responses to thinning and fertilization of a height-repressed lodgepole pine stand in interior British Columbia. *The Forestry Chronicle*, 95, 207-221.

Prescott, C. E., Preston, C. M. (1994). Nitrogen mineralization and decomposition in forest floors in adjacent plantations of western red cedar, western hemlock, and Douglas-fir. *Canadian Journal of Forest Research,* 24, 2424-2431.

Pruitt, K. D., Tatusova, T., Maglott, D. R. (2004). NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 33, D501–D504.

Qiagen, CLCBio. *www.qiagenbioinformatics.com*. (Last accessed March 16, 2020).

Qiao, H., Shen, Z., Huang, S. C., Schmitz, R. J., Urich, M. A., Briggs, S. P., Ecker, J. R. (2012). Processing and subcellular trafficking of ER-tethered EIN2 control response to ethylene gas. *Science*, 338, 390-393.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *https://www.R-project.org*.

Radhika, V., Kost, C., Bartram, S., Heil, M., Boland, W. (2008). Testing the optimal defense hypothesis for two indirect defenses: Extrafloral nectar and volatile organic compounds. *Planta*, 228, 449-457.

Raffa, K. F. (2013). Terpenes tell different tales at different scales: Glimpses into the chemical ecology of conifer - bark beetle - microbial interactions. *Journal of Chemical Ecology,* 40, 1-20.

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., Berryman, A. A. (1987). Interacting selective pressures in conifer-bark beetle systems: A basis for reciprocal adaptations? *The American Naturalist*, 129, 234-262.

Raffa, K. F., Smalley, E. B. (1988). Seasonal and long-term responses of host trees to microbial associates of the pine engraver, *Ips pini*. *Canadian Journal of Forest Research*, 18, 1624–1634.

Ralph, S. G., Park, J., Bohlmann, J., Mansfield, S. D. (2006a). 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.

Ralph, S. G., Yueh, H., Friedmann, M., Aeschliman, D., Zeznik, J. A., Nelson, C. C., Butterfield, Y. S. N., Kirkpatrick, R., Liu, J., Jones, S. J. M., Marra, M. A., Douglas, C. J., Ritland, K., Bohlmann, J. (2006b). 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 and Environment*, 29, 1545-1570.

Rao, D. H., Gowda, L. R. (2008). Abundant class III acidic chitinase homologue in tamarind (*Tamarindus indica*) seed serves as the major storage protein. *Journal of Agricultural and Food Chemistry*, 56, 2175–2182.

Reed, D. E., Ewers, B. E., Pendall, E. (2014). Impact of mountain pine beetle induced mortality on forest carbon and water fluxes. *Environmental Research Letters*, 9, 1-12.

Rice, A. V., Langor, D. W. (2008). A comparison of heat pulse velocity and lesion lengths for assessing the relative virulence of mountain pine beetle-associated fungi on jack pine. *Forest Pathology*, 38, 257-262.

Rice, A. V., Langor, D. W. (2009). Mountain pine beetle-associated blue-stain fungi in lodgepole × jack pine hybrids near Grande Prairie, Alberta (Canada). *Forest Pathology*, 39, 323-334.

Rice A. V., Thormann M. N., Langor D. W. (2007a). Mountain pine beetle associated blue-stain fungi cause lesions on jack pine, lodgepole pine, and lodgepole x jack pine hybrids in Alberta. *Canadian Journal of Botany,* 85, 307-315.

Rice A. V., Thormann M. N., Langor D. W. (2007b). Virulence of, and interactions among, mountain pine beetle associated blue-stain fungi on two pine species and their hybrids in Alberta. *Canadian Journal of Botany*, 85, 316-323.

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

Robert, J. A., Madilao, L. L., White, R., Yanchuk, A., King, J., Bohlmann, J. (2010). Terpenoid metabolite profiling in Sitka spruce identifies association of dehydroabietic acid, (+)-3-carene, and terpinolene with resistance against white pine weevil. *Botany*, 88, 810-820.

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., Mungall, K., Lee, S., Okada, H., Qian, J., Griffith, M., Raymond, A., Thiessen, N., Cezard, T., Butterfield, Y., Newsome, R., Chan, S., She, R., Varhol, R., Birol, I., Birol, I. (2010). *De novo* assembly and analysis of RNA-seq data. *Nature Methods*, 7, U909-U962.

Robinson, M. D., McCarthy, D. J., Smyth, G. K. (2010). EdgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139-140.

Robinson, M. D., Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11, PAGES.

Roe A. D., Rice A. V., Bromilow S. E., Cooke J. E. K., Sperling F. A. H. (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.

Roe A. D., Rice A. V., Coltman D. W., Cooke J. E. K., Sperling F. A. H. (2011). Comparative phylogeography, genetic differentiation, and contrasting reproductive modes in three fungal symbionts of a multipartite bark beetle symbiosis. *Molecular Ecology*, 20, 584–600.

Rohde, A., Bhalerao, R. P. (2007). Plant dormancy in the perennial context. *Trends in Plant Science,* 12, 217-223.

Roy, S., Laframboise, W. A., Nikiforov, Y. E., Nikiforova, M. N., Routbort, M. J., Pfeifer, J., Nagarajan, R., Carter, A. B., Pantanowitz, L. (2016). Next-generation sequencing informatics: Challenges and strategies for implementation in a clinical environment. *Archives of Pathology Laboratory Medicine, 140*, 958-975.

RStudio Team (2015). *RStudio: Integrated development for R*. RStudio, Inc., Boston, MA.

Ruan, J., Zhou, Y., Zhou, M., Yan, J., Khurshid, M., Weng, W., Cheng, Z., Zhang, K. (2019). Jasmonic acid signaling pathway in plants. *International Journal of Molecular Sciences*, 20, 2479.

Rushton, P. J., Somssich, I. E., Ringler, P., She. Q. J. (2010). WRKY transcription factors. *Trends in Plant Science*, 15, 247-258.

Sadasivam, S., Thayumanayan, B. (Eds.) (2019). *Molecular host plant resistance to pests*. CRC Press, Boca Raton, Florida, United States.

Safranyik, L., Carroll, A. L. (2006). Chapter 1 The biology and epidemiology of the mountain pine beetle in lodgepole pine forests. In Safranyik, L., Wilson, B. (Eds.), *The mountain pine beetle, A synthesis of biology, management, and impacts on*

lodgepole pine. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia, Canada. pp. 3-66.

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.

Sambaraju, K. R., Carroll, A. L., Aukema, B. H. (2019). Multiyear weather anomalies associated with range shifts by the mountain pine beetle preceding large epidemics. *Forest Ecology and Management, 438*, 86-95.

Sanger, F., Nicklen, S., Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74, 5463-5467.

Schatz, M. C., Delcher, A. L., Salzberg, S. L. (2010). Assembly of large genomes using second-generation sequencing. *Genome Research*, 20, 1165-1173.

Schluttenhofer, C., Yuan, L. (2015). Regulation of specialized metabolism by WRKY transcription factors. *Plant Physiology*, 167, 295–306.

Selitrennikoff, C. P. (2001). Antifungal proteins. *Applied and Environmental Microbiology*, 67, 2883-2894.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13, 2498-2504.

Shapiro, S. S., Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591.

Shendure, J., Ji, H. L. (2008). Next-generation DNA sequencing. *Nature Biotechnology,* 26, 1135-1145.

Shi, W., Liao, Y. (2016). *Subread/Rsubread users guide*. Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, The University of Melbourne, Melbourne, Australia.

Shore, T. L., Safranyik, L., Hawkes, B. C., Taylor, S. W. (2006). Chapter 3 Effects of the mountain pine beetle on lodgepole pine stand structure and dynamics. In Safranyik, L., Wilson, W.R. (Eds.). *The mountain pine beetle: A synthesis of biology, management, and impacts on lodgepole pine*. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia. pp. 95-114.

Shudo, E., Iwasa, Y. (2002). Optimal defense strategy: Storage vs. new production. *Journal of Theoretical Biology*, 219, 309-323.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 3, 3210–3212.

Six, D. L. (2003). A comparison of mycangial and phoretic fungi of individual mountain pine beetles. *Canadian Journal of Forest Research*, 33, 1331–1334.

Smith, D. R. (2014). Buying in to bioinformatics: An introduction to commercial sequence analysis software. *Briefings in Bioinformatics*, 16, 700-709.

Snoeijers, S. S., Perez-Garcia, A., Joosten, M. H. A. J., De Wit, P. G. M. (2000). The effect of nitrogen on disease development and gene expression in bacterial and fungal plant pathogens. *European Journal of Plant Pathology*, 106, 493-506.

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 and Development*, 12, 3703-3714.

Soneson, C., Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, 14, 1-18.

SPEX SamplePrep. (2013). Geno/Grinder 2010 Operating Manual. Author, Metuchen, New Jersey, USA.

Stahl, K., Moore, R., Mckendry, I. (2006). Climatology of winter cold spells in relation to mountain pine beetle mortality in British Columbia, Canada. *Climate Research*, 32, 13-23.

Stamp, N. (2003). Out of the quagmire of plant defense hypotheses. *The Quarterly Review of Biology*, 78, 23–55.

Stamp, N. (2004). Can the growth-differentiation balance hypothesis be tested rigorously? *Oikos*, 107, 439-448.

Steele, C. L., Katoh, S., Bohlmann, J., Croteau, R. (1998). Regulation of oleoresinosis in grand fir (*Abies grandis*): Differential transcriptional control of monoterpene, sesquiterpene, and diterpene synthase genes in response to wounding. *Plant Physiology*, 116, 1497–1504.

Stergiopoulos, I., Wit, P. J. (2009). Fungal effector proteins. *Annual Review of Phytopathology*, 47, 233-263.

Stockfors, J., Linder, S. (1998). Effect of nitrogen on the seasonal course of growth and maintenance respiration in stems of Norway spruce trees. *Tree Physiology*, 18, 155-166.

Sturm, M., Schroeder, C., Bauer, P. (2016). SeqPurge: Highly-sensitive adapter trimming for paired-end NGS data. *BMC Bioinformatics*, 17, 1-7.

Sun, Y., Wang, M., Mur, L. A. J., Shen, Q., Guo, S. (2020). Unravelling the roles of nitrogen nutrition in plant disease defences. *International Journal of Molecular Sciences*, 21, 1–20.

Sundell, D., Mannapperuma, C., Netotea, S., Delhomme, N., Lin, Y., Sjödin, A., de Peer, Y. V., Jansson, S., Hvidsten, T. R., Street, N. R. (2015). The plant genome integrative explorer resource: PlantGenIE.org. *New Phytologist*, 208, 1149-1156.

Suren, H., Hodgins, K. A., Yeaman, S., Nurkowski, K. A., Smets, P., Rieseberg, L. H., Aitken, S. N., Holliday, J. A. (2016). Exome capture from the spruce and pine giga-genomes. *Molecular Ecology Resources*, 16, 1136-1146.

Surget-Groba, Y., Montoya-Burgos, J. I. (2010). Optimization of *de novo* transcriptome assembly from next-generation sequencing data. *Genome Research*, 20, 1432-1440.

Taiz, L., Zeiger, E., Møller, I. M., Murphy, A. (Eds.). (2015). *Plant physiology and development* (6th ed.). Sinauer Associates, Inc., Publishers, Sunderland, Massachusetts, United States.

Tan, K., Oliver, R. P., Solomon, P. S., Moffat, C. S. (2010). Proteinaceous necrotrophic effectors in fungal virulence. *Functional Plant Biology*, 37, 907.

Tan, K., Phan, H. T., Rybak, K., John, E., Chooi, Y. H., Solomon, P. S., Oliver, R. P. (2015). Functional redundancy of necrotrophic effectors – consequences for exploitation for breeding. *Frontiers in Plant Science, 6*, 1-9.

Tang, J., Sun, B., Cheng, R., Shi, Z., Luo, D., Liu, S., Centritto, M. (2019). Effects of soil nitrogen (N) deficiency on photosynthetic N-use efficiency in N-fixing and non-N-fixing tree seedlings in subtropical China. *Scientific Reports*, 9, 1-14.

Taylor, S. W., Carroll, A. L. (2004). Disturbance, forest age, and mountain pine beetle outbreak dynamics in BC: A historical perspective. In Shore, T. L., Brooks, J. E., Stone, J. E. (Eds.), *Mountain pine beetle symposium: Challenges and*

solutions. October 30-31, 2003, Kelowna, British Columbia, Canada. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Information Report BC-X-399, Victoria, British Columbia, Canada. pp. 41-51.

Tegeder, M., Masclaux-Daubresse, C. (2017). Source and sink mechanisms of nitrogen transport and use. *New Phytologist*, 217, 35-53.

Tena, G., Boudsocq, M., Sheen, J. (2011). Protein kinase signaling networks in plant innate immunity. *Current Opinion in Plant Biology*, 14, 519-529.

Tenhaken, R. (2015). Cell wall remodeling under abiotic stress. *Frontiers in Plant Science*, 5, 1-9.

The Gene Ontology Consortium (2015). Gene ontology consortium: Going forward. *Nucleic Acids Research*, 43, D1049-D1056.

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Muller, L. A., Rhee, S. Y., Stitt, M. (2004). MapMan: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal*, 37, 914-939.

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

Thomas, T., Gilbert, J., Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, 2, 1-12.

Thomma, B. P., Nürnberger, T., Joosten, M. H. (2011). Of PAMPs and effectors: The blurred PTI-ETI dichotomy. *The Plant Cell*, 23, 4-15.

Toca, A., Oliet, J. A., Villar-Salvador, P., Maroto, J., Jacobs, D. F. (2017). Species ecology determines the role of nitrogen nutrition in the frost tolerance of pine seedlings. *Tree Physiology*, 38, 96–108.

Tomova, L., Braun, S., Fluckiger, W. (2005). The effect of nitrogen fertilization on fungistatic phenolic compounds in roots of beech (*Fagus sylvatica*) and Norway spruce (*Picea abies*). *Forest Pathology*, 35, 262–276.

Torres, M. A., Jones, J. D., Dangl, J. L. (2006). Reactive oxygen species signaling in response to pathogens. *Plant Physiology*, 141, 373-378.

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 5, 99-114.

Tuomi, J., Fagerstrom, T., Niemela, P. (1991). Carbon allocation, phenotypic plasticity, and induced defenses. In Tallamy, D. W., Raupp, M. J. (Eds.), *Phytochemical Induction by Herbivores*. Wiley, New York, New York, United States. pp. 85–104.

Tuomi, J., Niemelä, P., Chapin, III, F. S., Bryant, J. P., Sirén, S. (1988). Defensive responses of trees in relation to their carbon/nutrient balance. In Mattson, W. J. *et al.* (Eds.), *Mechanisms of Woody Plant Defenses Against Insects: Search for Pattern*. Springer, New York, New York, United States. pp. 57–72.

Turnbaugh, P. J., Gordon, J. I. (2008). An invitation to the marriage of metagenomics and metabolomics. *Cell*, 134, 708–713.

Vadeboncoeur, M. A. (2010). Meta-analysis of fertilization experiments indicates multiple limiting nutrients in northeastern deciduous forests. *Canadian Journal of Forest Research,* 40, 1766-1780.

Van Bel, M., Proost, S., Neste, C. V., Deforce, D., Peer, Y. V., Vandepoele, K. (2013). TRAPID: An efficient online tool for the functional and comparative analysis of *de novo* RNA-Seq transcriptomes. *Genome Biology*, 14, 1-10.

Van den Ackerveken, G. F. J. M., Dunn, R. M., Cozijnsen, T. J., Vossen, P., Van den Broek, H. W. J., De Wit, P. J. G. M. (1994). Nitrogen limitation induces expression of the avirulence gene *Avr9* in the tomato pathogen *Cladosporium fulvum*. *Molecular and General Genetics*, 243, 277–285.

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

Van Loon, L. C., van Strien, E. A. (1999). The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. *Physiological and Molecular Plant Pathology*, 55, 85-97.

Vega, A., Canessa, P., Hoppe, G., Retamal, I., Moyano, T. C., Canales, J., Gutiérrez, R. A., Rubilar, J. (2015). Transcriptome analysis reveals regulatory networks underlying differential susceptibility to *Botrytis cinerea* in response to nitrogen availability in *Solanum lycopersicum*. *Frontiers in Plant Science*, 6, 1-17.

Verly, C., Djoman, A. C., Rigault, M., Giraud, F., Rajjou, L., Saint-Macary, M., Dellagi, A. (2020). Plant defense stimulator mediated defense activation is affected by nitrate fertilization and developmental stage in *Arabidopsis thaliana*. *Frontiers in Plant Science, 11*, 1-15.

Vijay, N., Poelstra, J. W., Künstner, A., Wolf, J. B. (2012). Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive *in silico* assessment of RNA-seq experiments. *Molecular Ecology*, 22, 620-634.

Villari, C., Battisti, A., Chakraborty, S., Michelozzi, M., Bonello, P., Faccoli, M. (2012). Nutritional and pathogenic fungi associated with the pine engraver beetle trigger comparable defenses in Scots pine. *Tree Physiology*, 32, 867-879.

Vogel, C., Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, 13, 227-232.

Vos, I. A., Pieterse, C. M., Wees, S. C. (2013). Costs and benefits of hormone-regulated plant defences. *Plant Pathology,* 62, 43-55.

Wainhouse, D., Ashburner, R., Ward, E., Rose, J. (1998). The effect of variation in light and nitrogen on growth and defense in young Sitka spruce. *Functional Ecology*, 12, 561-572.

Wang, H., Meng, J., Peng, X., Tang, X., Zhou, P., Xiang, J., Deng, X. (2015). Rice WRKY4 acts as a transcriptional activator mediating defense response toward *Rhizoctonia solani*, the causing agent of rice sheath blight. *Plant Molecular Biology,* 89, 157-171.

Wang, C., Tang, Y. (2019). Responses of plant phenology to nitrogen addition: A meta-analysis. *Oikos*, 128, 1243-1253.

Wang, L., Li, P., Brutnel. T. P. (2010). Exploring plant transcriptomes using ultra high-throughput sequencing. *Briefings in Functional Genomics*, 9, 118–128.

Wang, Y., Lim, L., Diguistini, S., Robertson, G., Bohlmann, J., Breuil, C. (2012). A specialized ABC efflux transporter GcABC-G1 confers monoterpene resistance to *Grosmannia clavigera*, a bark beetle-associated fungal pathogen of pine trees. *New Phytologist*, 197, 886–898.

Wang, Z., Gerstein, M., Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics,* 10, 57-63.

Waring, H. R., Pitman, G. B. (1985). Modifying lodgepole pine stands to change susceptibility to mountain pine beetle attack. *Ecology*, 66, 889–897.

Warinowski, T., Koutaniemi, S., Kärkönen, A., Sundberg, I., Toikka, M., Simola, L. K., Kilpeläinen, I., Teeri, T. H. (2016). Peroxidases bound to the growing lignin polymer produce natural like extracellular lignin in a cell culture of Norway spruce. *Frontiers in Plant Science,* 7, 1-12.

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., Venables, B. (2020). gplots: Various R programming tools for plotting data. R package version 3.0.3. *https://CRAN.R-project.org/package=gplots*.

Wasternack, C., Hause, B. (2019). The missing link in jasmonic acid biosynthesis. *Nature Plants,* 5, 776-777.

Wegrzyn, J. L., Lee, J. M., Tearse, B. R., Neale, D. B. (2008). TreeGenes: A forest tree genome database. *International Journal of Plant Genomics*, 2008, 1-7.

West, D., Bernklau, E., Bjostad, L., Jacobi, W. (2016). Host defense mechanisms against bark beetle attack differ between ponderosa and lodgepole pines. *Forests, 7*, 1-11.

Wetzel, S., Demmers, C., Greenwood, J. S. (1989). Seasonally fluctuating bark proteins are a potential form of nitrogen storage in three temperate hardwoods. *Planta*, 178, 275-281.

Williams, C. R., Baccarella, A., Parrish, J. Z., Kim, C. C. (2016). Trimming of sequence reads alters RNASeq gene expression estimates. *BMC Bioinformatics*, 17, 1-13.

Wilson, E., Skeffington, R. (1994). The effects of excess nitrogen deposition on young Norway spruce trees. Part I the soil. *Environmental Pollution*, 86, 141-151.

Wong, B. L., Berryman, A. A. (1977). Host resistance to the fir engraver beetle. 3. Lesion development and containment of infection by resistant *Abies grandis* inoculated with *Trichosporium symbioticum*. *Canadian Journal of Botany*, 55, 2358–2365.

Wulder, M. A., White, J. C., Bentz, B. J., Ebata, T. (2006). Augmenting the existing survey hierarchy for mountain pine beetle red-attack damage with satellite remotely sensed data. *The Forestry Chronicle*, 82, 187-202.

Xu, J., Wang, X., Guo, W. (2015). The cytochrome P450 superfamily: Key players in plant development and defense. *Journal of Integrative Agriculture*, 14, 1673–1686.

Xuehui, J., Na, H., Feng, J., Yang, Y., Dan, W., Chunying, X., Ruichang, Z. (2014). The effect of nitrogen supply on potato yield, tuber size and pathogen resistance in *Solanum tuberosum* exposed to *Phytophthora infestans*. *African Journal of Agricultural Research,* 9, 2657-2663.

Yamaguchi, Y., Pearce, G., Ryan, C. A. (2006). The cell surface leucine-rich repeat receptor for *At*Pep1, an endogenous peptide elicitor in *Arabidopsis*, is functional in transgenic tobacco cells. *Proceedings of the National Academy of Sciences,* 103, 10104-10109.

Yamaoka, Y., Hiratsuka, Y., Maruyama, P. J. (1995). The ability of *Ophiostoma clavigerum* to kill mature lodgepole pine trees. *Forest Pathology*, 25, 401-404.

Young, M. D., Wakefield, M. J., Smyth, G. K., Oshlack, A. (2010). Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biology, 11*, PAGES.

Zabel, R. A., Morell, J. J. (1992). Wood stains and discolorations. In Zabel R. A., Morell J. J. (Eds.), *Wood microbiology: Decay and its prevention*. Academic Press, San Diego, California, United States. pp. 326–343.

Zamani, B. A., Sturrock, R., Ekramoddoullah, A. K. M., Wiseman, S. B. (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.

Zamyatnin, A. A. (2015). Plant proteases involved in regulated cell death. *Biochemistry*, 80, 1701-1715.

Zerbino, D. R., Birney, E. (2008). Velvet: Algorithms for d*e novo* short read assembly using de Bruijn graphs. *Genome Research*, 18, 821-829.

Zhang, B., Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology,* 4, 1-43.

Zhang, L., Zhang, F., Melotto, M., Yao, J., He, S. Y. (2017). Jasmonate signaling and manipulation by pathogens and insects. *Journal of Experimental Botany,* 68, 1371-1385.

Zhang, J., Peng, Y., Guo, Z. (2008). Constitutive expression of pathogen-inducible OsWRKY31 enhances disease resistance and affects root growth and auxin response in transgenic rice plants. *Cell Research*, 18, 508-521.

Zhang, W. Y., Chen, J. J., Yang, Y., Tang, Y. F., Shang, J., Shen, B. R. (2011). A practical comparison of *de novo* genome assembly software tools for next-generation sequencing technologies. *PLoS ONE,* 6, 1-12.

Zhao, S., Erbilgin, N. (2019). Larger resin ducts are linked to the survival of lodgepole pine trees during mountain pine beetle outbreak. *Frontiers in Plant Science*, 10, 1-14.

Zhao, S. R., Zhang, Y., Gordon, W., Quan, J., Xi, H. L., Du, S., von Schack, D., Zhang, B. (2015). Comparison of stranded and non-stranded RNA-seq transcriptome profiling and investigation of gene overlap. *BMC Genomics,* 16, 1-14.

Zhong, R., Ye, Z. (2014). Secondary cell walls: Biosynthesis, patterned deposition and transcriptional regulation. *Plant and Cell Physiology,* 56, 195-214.

Zimmerman, C. F., Keefe, C. W., Bashe, J. (1997). *Method 440.0 Determination of carbon and nitrogen in sediments and particulates of estuarine/coastal waters using elemental analysis*. U.S. Environmental Protection Agency, Washington, District of Columbia, United States, EPA/600/R-15/009, pp 365.5-1 – 365.5-9.

Appendix A: Supplemental tables

Table 6.1. Illumina NextSeq 500 sequencing statistics and run information. The 32 cDNA libraries were constructed using the Illumina TruSeq Stranded mRNA Low Sample (LS) Protocol (Illumina, Inc. 2013) by Dr. Chandra McAllister. During their construction, each cDNA library was fragmented into sequences 300-500 bp long and fragments were flanked with TruSeq Index Adapters (Illumina, Inc., San Diego, California, United States). The TruSeq Index Adapters included index sequences to differentiate libraries during the sequencing process. Libraries were sequenced at the Molecular Biology Service Unit of the University of Alberta's Department of Biological Sciences (Edmonton, Alberta, Canada) by Sophie Dang according to the NextSeq 500 System Guide (Illumina, Inc. 2018). The size of each library's FASTQ file is gzip compressed.

Run 1									
	Sample	LowN.Wound.2P.1	LowN.Wound.2X.2	LowN.Fungus.2P.3	LowN.Fungus.2X.4	HighN.Wound.2P.1	HighN.Wound.2X.2	HighN.Fungus.2P.3	HighN.Fungus.2X.4
	Library	14-01-2P	14-02-2X	15-03-2P	15-06-2X	17-01-2P	17-02-2X	18-03-2P	18-06-2X
	Index	5	6	12	19	2	4	7	16
	Index sequence	ACAGTG	GCCAAT	CTTGTA	GTGAAA	CGATGT	TGACCA	CAGATC	CCGTCC
	Number of Clusters	90,186,259	82,433,225	91,182,875	74,161,507	70,544,912	78,996,740	87,773,152	73,502,149
	Number of Reads	180,372,518	164,866,450	182,365,750	148,323,014	141,089,824	157,993,480	175,546,304	147,004,298
	Read 1 length	151	151	151	151	151	151	151	151
	Read 2 length	151	151	151	151	151	151	151	151
	Size	14.74 GB	13.78 GB	15.75 GB	12.34 GB	12.12 GB	13.62 GB	14.57 GB	12.79 GB
Run 2									
	Sample	LowN.Wound.2P.3	LowN.Wound.2X.3	LowN.Fungus.2P.1	LowN.Fungus.2X.2	HighN.Wound.2P.3	HighN.Wound.2X.3	HighN.Fungus.2P.1	HighN.Fungus.2X.2
	Library	14-03-2P	14-04-2X	15-01-2P	15-03-2X	17-03-2P	17-05-2X	18-01-2P	18-04-2X
	Index	5	4	12	19	2	6	7	16
	Index sequence	ACAGTG	TGACCA	CTTGTA	GTGAAA	CGATGT	GCCAAT	CAGATC	CCGTCC
	Number of Clusters	68,746,348	36,376,112	92,355,770	95,835,618	87,033,282	32,477,870	85,788,072	86,306,225
	Number of Reads	137,492,696	72,752,224	184,711,540	191,671,236	174,066,564	64,955,740	171,576,144	172,612,450
	Read 1 length	151	151	151	151	151	151	151	151
	Read 2 length	151	151	151	151	151	151	151	151
	Size	11.40 GB	6.12 GB	16.30 GB	16.68 GB	14.36 GB	4.91 GB	14.52 GB	15.02 GB
Run 3									
	Sample	LowN.Wound.2P.4	LowN.Wound.2X.1	LowN.Fungus.2P.2	LowN.Fungus.2X.3	HighN.Wound.2X.1	HighN.Wound.2P.4	HighN.Fungus.2P.2	HighN.Fungus.2X.3
	Library	14-04-2P	14-01-2X	15-02-2P	15-05-2X	17-01-2X	17-04-2P	18-02-2P	18-05-2X
	Index	6	2	7	16	5	4	15	12
	Index sequence	GCCAAT	CGATGT	CAGATC	CCGTCC	ACAGTG	TGACCA	ATGTCA	CTTGTA
	Number of Clusters	108,608,113	75,689,122	65,546,481	77,663,754	75,463,972	76,454,243	71,539,474	61,765,453
	Number of Reads	217,216,226	151,378,244	131,092,962	155,327,508	150,927,944	152,908,486	143,078,948	123,530,906
	Read 1 length	151	151	151	151	151	151	151	151
	Read 2 length	151	151	151	151	151	151	151	151
	Size	19.08 GB	13.25 GB	11.37 GB	13.59 GB	13.44 GB	13.51 GB	12.22 GB	10.99 GB
Run 4									
	Sample	LowN.Wound.2P.2	LowN.Wound.2X.4	LowN.Fungus.2P.4	LowN.Fungus.2X.1	HighN.Wound.2P.2	HighN.Wound.2X.4	HighN.Fungus.2P.4	HighN.Fungus.2X.1
	Library	14-02-2P	14-05-2X	15-04-2P	15-02-2X	17-02-2P	17-06-2X	18-04-2P	18-02-2X
	Index	5	6	19	12	15	2	18	7
	Index sequence	ACAGTG	GCCAAT	GTGAAA	CTTGTA	ATGTCA	CGATGT	GTCCGC	CAGATC
	Number of Clusters	18,877,690	83,735,209	50,604,085	45,413,438	53,733,506	118,212,081	79,736,452	67,161,744
	Number of Reads	37,755,380	167,470,418	101,208,170	90,826,876	107,467,012	236,424,162	159,472,904	134,323,488
	Read 1 length	151	151	151	151	151	151	151	151
	Read 2 length	151	151	151	151	151	151	151	151
	Size	3.16 GB	13.93 GB	8.11 GB	7.18 GB	8.36 GB	18.09 GB	12.31 GB	10.63 GB

Table 6.2. The CLC Genomics Workbench per sequence quality report. Paired-end stranded 151 bp reads produced by the Illumina NextSeq 500 platform (Illumina, Inc., San Diego, California, United States) were tested for quality using the CLC Genomics Workbench v9.5.2 (*www.qiagenbioinformatics.com*). Quality control was performed on sequence data.

Sample	Number of sequences	Maximum sequence length (bp)	Minimum sequence length (bp)	Number with 0% ambiguous nucleotides	Percentage with 0% ambiguous nucleotides	Number with average Phred ≥ 30	Percentage with average Phred ≥ 30
LowN.Wound.2P.1	180,372,518	151	35	179,986,754	99.79%	149,871,525	83.09%
LowN.Wound.2P.2	37,755,380	151	35	37,697,593	99.85%	32,262,664	85.45%
LowN.Wound.2P.3	137,492,696	151	35	137,368,048	99.91%	112,180,251	81.59%
LowN.Wound.2P.4	217,216,226	151	35	217,110,414	99.95%	172,429,516	79.38%
LowN.Wound.2X.1	164,866,450	151	35	164,507,002	99.78%	133,493,017	80.97%
LowN.Wound.2X.2	164,866,450	151	35	164,507,002	99.78%	130,473,017	79.14%
LowN.Wound.2X.3	72,752,224	151	35	72,690,706	99.92%	59,798,120	82.19%
LowN.Wound.2X.4	167,470,418	151	35	167,230,766	99.86%	147,742,694	88.22%
LowN.Fungus.2P.1	184,711,540	151	35	184,552,178	99.91%	149,281,708	80.82%
LowN.Fungus.2P.2	131,092,962	151	35	131,026,071	99.95%	104,514,979	79.73%
LowN.Fungus.2P.3	182,365,750	151	35	181,989,036	99.79%	149,964,643	82.23%
LowN.Fungus.2P.4	101,208,170	151	35	101,051,752	99.85%	86,572,779	85.54%
LowN.Fungus.2X.1	90,826,876	151	35	90,696,508	99.86%	80,761,113	88.92%
LowN.Fungus.2X.2	191,671,236	151	35	191,514,082	99.92%	154,253,716	80.48%
LowN.Fungus.2X.3	155,327,508	151	35	155,244,806	99.95%	123,218,772	79.33%
LowN.Fungus.2X.4	148,323,014	151	35	148,010,746	99.79%	122,883,262	82.85%
HighN.Wound.2P.1	141,089,824	151	35	140,799,639	99.79%	112,661,906	79.85%
HighN.Wound.2P.2	107,467,012	151	35	107,299,716	99.84%	96,308,395	89.62%
HighN.Wound.2P.3	174,066,564	151	35	173,910,095	99.91%	141,269,104	81.16%
HighN.Wound.2P.4	152,908,486	151	35	152,829,795	99.95%	119,800,388	78.35%
HighN.Wound.2X.1	150,927,944	151	35	150,855,631	99.95%	118,365,415	78.43%
HighN.Wound.2X.2	157,993,480	151	35	157,668,200	99.79%	131,088,685	82.97%
HighN.Wound.2X.3	64,955,740	151	35	64,886,484	99.89%	53,496,103	82.36%
HighN.Wound.2X.4	236,424,162	151	35	236,062,410	99.85%	211,884,958	89.62%
HighN.Fungus.2P.1	171,576,144	151	35	171,421,014	99.91%	138,354,948	80.64%
HighN.Fungus.2P.2	143,078,948	151	35	143,006,238	99.95%	114,157,038	79.79%
HighN.Fungus.2P.3	175,546,304	151	35	175,174,099	99.79%	144,210,902	82.15%
HighN.Fungus.2P.4	159,472,904	151	35	159,224,122	99.84%	142,339,007	89.26%
HighN.Fungus.2X.1	134,323,488	151	35	134,123,143	99.85%	120,230,007	89.51%
HighN.Fungus.2X.2	172,612,450	151	35	172,469,130	99.92%	138,084,086	80.00%
HighN.Fungus.2X.3	123,530,906	151	35	123,470,853	99.95%	96,734,419	78.31%
HighN.Fungus.2X.4	147,004,298	151	35	146,706,913	99.80%	113,031,687	76.89%

Table 6.3. The CLC Genomics Workbench per base quality report. Paired-end stranded 151 bp reads produced using the Illumina NextSeq 500 platform (Illumina, Inc., San Diego, California, United States) were tested for quality using the CLC Genomics Workbench v9.5.2 (*www.qiagenbioinformatics.com*). Quality control was performed on nucleotide data.

Sample	# of bases	# of non-ambiguous bases	% of non-ambiguous bases	# with median Phred ≥ 30	% with median Phred ≥ 30
LowN.Wound.2P.1	23,180,866,101	22,179,452,685	95.68%	23,134,256,834	99.80%
LowN.Wound.2P.2	5,125,181,350	4,977,576,127	97.12%	5,112,287,337	99.75%
LowN.Wound.2P.3	17,483,216,193	17,231,457,880	98.56%	17,447,346,761	99.79%
LowN.Wound.2P.4	29,267,417,150	29,059,618,488	99.29%	29,202,691,472	99.78%
LowN.Wound.2X.1	21,589,064,580	21,435,782,221	99.29%	21,546,780,802	99.80%
LowN.Wound.2X.2	21,589,064,580	20,811,858,255	96.40%	21,544,433,174	99.79%
LowN.Wound.2X.3	9,360,319,739	9,191,833,984	98.20%	9,341,315,986	99.80%
LowN.Wound.2X.4	22,683,666,168	22,112,037,781	97.48%	22,628,806,214	99.76%
LowN.Fungus.2P.1	25,108,880,858	24,747,312,974	98.56%	25,049,081,732	99.76%
LowN.Fungus.2P.2	17,426,182,334	17,300,713,821	99.28%	17,366,383,208	99.66%
LowN.Fungus.2P.3	24,764,685,913	24,051,462,959	97.12%	24,700,933,068	99.74%
LowN.Fungus.2P.4	13,211,034,375	12,973,235,756	98.20%	12,944,321,414	97.98%
LowN.Fungus.2X.1	11,621,266,747	11,370,247,385	97.84%	11,599,397,548	99.81%
LowN.Fungus.2X.2	25,518,788,932	25,151,318,371	98.56%	25,457,172,948	99.76%
LowN.Fungus.2X.3	20,716,074,696	20,566,918,958	99.28%	20,669,098,906	99.77%
LowN.Fungus.2X.4	19,356,685,096	18,729,528,499	96.76%	19,319,532,468	99.81%
HighN.Wound.2P.1	19,059,513,667	18,579,213,923	97.48%	19,011,267,954	99.75%
HighN.Wound.2P.2	13,711,003,138	13,316,126,248	97.12%	13,683,769,059	99.80%
HighN.Wound.2P.3	21,967,606,514	21,651,272,980	98.56%	21,920,628,401	99.79%
HighN.Wound.2P.4	20,616,521,378	20,468,082,424	99.28%	20,570,790,926	99.78%
HighN.Wound.2X.1	20,550,474,537	20,402,511,120	99.28%	20,502,999,499	99.77%
HighN.Wound.2X.2	21,388,682,122	20,926,686,588	97.84%	21,338,539,635	99.77%
HighN.Wound.2X.3	7,471,178,370	7,256,008,433	97.12%	7,461,361,979	99.87%
HighN.Wound.2X.4	29,594,952,471	28,636,076,011	96.76%	29,536,933,507	99.80%
HighN.Fungus.2P.1	22,252,454,544	21,771,801,526	97.84%	22,203,726,942	99.78%
HighN.Fungus.2P.2	18,679,535,446	18,545,042,791	99.28%	18,643,130,551	99.81%
HighN.Fungus.2P.3	22,920,132,043	22,260,032,240	97.12%	22,873,645,343	99.80%
HighN.Fungus.2P.4	20,209,414,634	19,481,875,707	96.40%	20,174,096,762	99.83%
HighN.Fungus.2X.1	17,426,330,262	17,049,921,528	97.84%	17,388,878,957	99.79%
HighN.Fungus.2X.2	22,896,453,152	22,484,316,995	98.20%	22,844,691,957	99.77%
HighN.Fungus.2X.3	16,678,206,141	16,558,123,057	99.28%	16,640,938,641	99.78%
HighN.Fungus.2X.4	19,962,727,620	19,244,069,426	96.40%	19,913,205,315	99.75%

Table 6.4. The CLC Genomics Workbench trimming statistics. Paired-end stranded 151 bp reads were pre-processed using the CLC Genomics Workbench v9.5.2 (*www.qiagenbioinformatics.com*) in anticipation of differential expression analysis (Chapter 4). The optimized Trim Adapter List (described in Section 2.2.7) was employed using parameters detailed in Figure 2.9 of Section 2.4.2. Trimming is performed sequentially, beginning with the quality trim, then the ambiguity followed by the adapter trim, ending with the minimum and maximum length filters.

Sample	Number of reads	Average length (bp)	Number of reads after trim	Percentage trimmed	Average length after trim (bp)
LowN.Wound.2P.1	180,372,518	128.5	157,513,972	87.33%	118.3
LowN.Wound.2P.2	37,755,380	135.7	33,386,073	88.43%	127.4
LowN.Wound.2P.3	137,492,696	127.1	117,591,123	85.53%	115.9
LowN.Wound.2P.4	217,216,226	134.8	185,320,938	85.32%	122.1
LowN.Wound.2X.1	151,378,244	133.2	127,713,524	84.37%	121.1
LowN.Wound.2X.2	164,855,450	131.0	141,127,839	85.61%	119.8
LowN.Wound.2X.3	72,752,224	128.7	62,799,356	86.32%	117.6
LowN.Wound.2X.4	167,470,418	135.5	152,696,325	91.18%	127.6
LowN.Fungus.2P.1	184,711,540	135.9	158,993,773	86.08%	123.5
LowN.Fungus.2P.2	131,092,962	133.9	111,777,938	85.27%	120.4
LowN.Fungus.2P.3	182,365,750	135.5	159,934,694	87.70%	124.5
LowN.Fungus.2P.4	101,211,170	130.5	92,108,794	91.01%	123.4
LowN.Fungus.2X.1	90,826,876	127.9	82,773,744	91.13%	120.6
LowN.Fungus.2X.2	191,671,236	133.1	163,867,527	85.49%	120.5
LowN.Fungus.2X.3	155,327,508	133.3	131,654,388	84.76%	120.5
LowN.Fungus.2X.4	148,323,014	130.5	129,807,667	87.52%	119.8
HighN.Wound.2P.1	141,089,824	135.2	120,364,608	85.31%	122.9
HighN.Wound.2P.2	107,232,830	127.6	98,439,412	91.80%	121.2
HighN.Wound.2P.3	174,066,864	126.2	147,615,809	84.80%	114.7
HighN.Wound.2P.4	152,908,486	134.8	128,927,011	84.32%	121.8
HighN.Wound.2X.1	150,927,944	136.1	127,719,030	84.62%	123.8
HighN.Wound.2X.2	157,993,480	135.3	139,522,052	88.31%	124.1
HighN.Wound.2X.3	64,954,740	115.0	54,327,559	83.64%	105.5
HighN.Wound.2X.4	236,424,162	125.2	216,098,499	91.40%	118.7
HighN.Fungus.2P.1	171,576,144	129.7	145,672,145	84.90%	118.0
HighN.Fungus.2P.2	143,078,948	130.5	121,484,104	84.91%	118.3
HighN.Fungus.2P.3	175,546,304	130.5	152,329,685	86.77%	120.8

Table 6.4. The CLC Genomics Workbench trimming statistics. Continued.

Sample	Number of reads	Average length (bp)	Number of reads after trim	Percentage trimmed	Average length after trim (bp)
HighN.Fungus.2P.4	159,472,904	126.7	145,534,561	91.26%	120.4
HighN.Fungus.2X.1	134,326,488	129.7	123,196,072	91.71%	123.0
HighN.Fungus.2X.2	172,612,450	132.7	146,743,716	85.01%	119.1
HighN.Fungus.2X.3	123,530,906	135.0	104,451,128	84.55%	121.5
HighN.Fungus.2X.4	147,004,298	135.8	128,435,334	87.37%	122.8

Table 6.5. The CLC Genomics Workbench mapping statistics. The CLC Genomics Workbench v9.5.2 (*www.qiagenbioinformatics.com*) generated sequence alignment map (sam) files by aligning trimmed reads to the master transcriptome using the mapping parameters outlined in Table 2.5 in Section 2.2.10. The construction of the master transcriptome is detailed in Section 4.2.6. Reference count and average length refer to the number and average length of contigs in the master transcriptome, respectively. Upon completion, sam files were exported and used as inputs for the read enumeration tool *featureCounts* (Liao *et al.* 2014) (Chapter 4). Broken paired reads are paired-end reads where only one pair aligned to the reference.

	Sample LowN.Wound.2P.1			**Sample** LowN.Wound.2P.2		
	Count	**% of reads**	**Average length**	**Count**	**% of reads**	**Average length**
Reference	375,632		719.67	375,632		719.67
Mapped reads	133,159,990	90.66%	119.51	27,997,670	90.75%	128.42
Not mapped reads	13,715,900	9.34%	120.97	2,853,970	9.25%	128.98
Reads in pairs	126,209,502	85.93%	137.04	26,274,766	85.16%	149.94
Broken paired reads	6,950,488	4.73%	129.47	1,722,904	5.58%	134.46
Total reads	146,875,890	100.00%	119.64	30,851,640	100.00%	128.47
	Sample LowN.Wound.2P.3			**Sample** LowN.Wound.2P.4		
	Count	**% of reads**	**Average length**	**Count**	**% of reads**	**Average length**
Reference	375,632		719.67	375,632		719.67
Mapped reads	100,108,068	91.07%	117.3	157,634,182	90.99%	123.66
Not mapped reads	9,818,528	8.93%	118.03	15,608,582	9.01%	124.47
Reads in pairs	94,574,476	86.03%	135.51	148,116,742	85.50%	144.69
Broken paired reads	5,533,592	5.03%	126.03	9,517,440	5.49%	126.61
Total reads	109,926,596	100.00%	117.37	173,242,764	100.00%	123.74
	Sample LowN.Wound.2X.1			**Sample** LowN.Wound.2X.2		
	Count	**% of reads**	**Average length**	**Count**	**% of reads**	**Average length**
Reference	375,632		719.67	375,632		719.67
Mapped reads	107,415,155	90.62%	121.88	116,792,608	90.38%	121.18
Not mapped reads	11,119,389	9.38%	122.69	12,438,436	9.62%	122.65
Reads in pairs	101,650,826	85.76%	141.82	110,713,662	85.67%	138.99
Broken paired reads	5,764,329	4.86%	124.26	6,078,946	4.70%	127.7
Total reads	118,534,544	100.00%	121.96	129,231,044	100.00%	121.32

Table 6.5. The CLC Genomics Workbench mapping statistics. Continued.

	Sample LowN.Wound.2X.3			Sample LowN.Wound.2X.4		
	Count	**% of reads**	**Average length**	**Count**	**% of reads**	**Average length**
Reference	375,632		719.67	375,632		719.67
Mapped reads	53,810,834	91.22%	118.81	131,982,578	90.49%	128.59
Not mapped reads	5,178,780	8.78%	119.79	13,875,096	9.51%	129.06
Reads in pairs	51,533,806	87.36%	136.92	124,561,040	85.40%	150.73
Broken paired reads	2,277,028	3.86%	125.16	7,421,538	5.09%	134.36
Total reads	58,989,614	100.00%	118.89	145,857,674	100.00%	128.64
	Sample LowN.Fungus.2P.1			**Sample** LowN.Fungus.2P.2		
	Count	**% of reads**	**Average length**	**Count**	**% of reads**	**Average length**
Reference	375,632		719.67	375,632		719.67
Mapped reads	135,770,029	91.06%	125.11	95,799,869	91.70%	122.22
Not mapped reads	13,323,113	8.94%	125.71	8,668,697	8.30%	122.99
Reads in pairs	128,020,834	85.87%	149.96	90,740,990	86.86%	142.1
Broken paired reads	7,749,195	5.20%	127.08	5,058,879	4.84%	126.28
Total reads	149,093,142	100.00%	125.16	104,468,566	100.00%	122.28
	Sample LowN.Fungus.2P.3			**Sample** LowN.Fungus.2P.4		
	Count	**% of reads**	**Average length**	**Count**	**% of reads**	**Average length**
Reference	375,632		719.67	375,632		719.67
Mapped reads	137,053,986	91.90%	125.52	81,288,445	92.29%	124.16
Not mapped reads	12,078,530	8.10%	126.57	6,791,295	7.71%	124.62
Reads in pairs	127,040,218	85.19%	153.68	77,832,714	88.37%	141.86
Broken paired reads	10,013,768	6.71%	130.93	3,455,731	3.92%	133.48
Total reads	149,132,516	100.00%	125.61	88,079,740	100.00%	124.2
	Sample LowN.Fungus.2X.1			**Sample** LowN.Fungus.2X.2		
	Count	**% of reads**	**Average length**	**Count**	**% of reads**	**Average length**
Reference	375,632		719.67	375,632		719.67
Mapped reads	73,319,912	92.63%	121.73	142,042,805	92.60%	122.09
Not mapped reads	5,831,844	7.37%	121.87	11,347,125	7.40%	122.96
Reads in pairs	70,287,216	88.80%	135.4	133,475,952	87.02%	147.81
Broken paired reads	3,032,696	3.83%	133.08	8,566,853	5.59%	126.58
Total reads	79,151,756	100.00%	121.74	153,389,930	100.00%	122.15

Table 6.5. The CLC Genomics Workbench mapping statistics. Continued.

	Sample LowN.Fungus.2X.3			Sample LowN.Fungus.2X.4		
	Count	**% of reads**	**Average length**	**Count**	**% of reads**	**Average length**
Reference	375,632		719.67	375,632		719.67
Mapped reads	113,537,914	92.26%	122.22	110,346,715	91.04%	121.22
Not mapped reads	9,529,006	7.74%	123.17	10,863,391	8.96%	121.54
Reads in pairs	106,098,414	86.21%	145.05	104,313,968	86.06%	136.71
Broken paired reads	7,439,500	6.05%	126.12	6,032,747	4.98%	127.11
Total reads	123,066,920	100.00%	122.3	121,210,106	100.00%	121.25
	Sample HighN.Wound.2P.1			**Sample** HighN.Wound.2P.2		
	Count	**% of reads**	**Average length**	**Count**	**% of reads**	**Average length**
Reference	375,632		719.67	375,632		719.67
Mapped reads	99,039,632	90.72%	124.18	86,385,647	91.40%	121.93
Not mapped reads	10,132,028	9.28%	125.66	8,131,849	8.60%	122.64
Reads in pairs	91,016,962	83.37%	150.51	82,872,222	87.68%	137.36
Broken paired reads	8,022,670	7.35%	129.29	3,513,425	3.72%	133.85
Total reads	109,171,660	100.00%	124.32	94,517,496	100.00%	121.99
	Sample HighN.Wound.2P.3			**Sample** HighN.Wound.2P.4		
	Count	**% of reads**	**Average length**	**Count**	**% of reads**	**Average length**
Reference	375,632		719.67	375,632		719.67
Mapped reads	125,565,033	91.30%	116.41	109,400,467	91.23%	123.45
Not mapped reads	11,964,735	8.70%	117.68	10,514,677	8.77%	124.33
Reads in pairs	118,987,654	86.52%	135.29	103,661,296	86.45%	145.01
Broken paired reads	6,577,379	4.78%	126.05	5,739,171	4.79%	124.93
Total reads	137,529,768	100.00%	116.52	119,915,144	100.00%	123.53
	Sample HighN.Wound.2X.1			**Sample** HighN.Wound.2X.2		
	Count	**% of reads**	**Average length**	**Count**	**% of reads**	**Average length**
Reference	375,632		719.67	375,632		719.67
Mapped reads	107,940,872	91.15%	124.6	118,687,084	90.83%	125.38
Not mapped reads	10,483,992	8.85%	125.41	11,987,900	9.17%	126.21
Reads in pairs	101,382,152	85.61%	147.64	108,933,286	83.36%	148.93
Broken paired reads	6,558,720	5.54%	125.69	9,753,798	7.46%	131.55
Total reads	118,424,864	100.00%	124.67	130,674,984	100.00%	125.45

Table 6.5. The CLC Genomics Workbench mapping statistics. Continued.

	Sample HighN.Wound.2X.3			Sample HighN.Wound.2X.4		
	Count	**% of reads**	**Average length**	**Count**	**% of reads**	**Average length**
Reference	375,632		719.67	375,632		719.67
Mapped reads	44,630,763	88.68%	107.19	190,518,119	91.93%	119.48
Not mapped reads	5,699,489	11.32%	106.93	16,722,085	8.07%	119.72
Reads in pairs	42,856,182	85.15%	115.36	180,532,778	87.11%	134.18
Broken paired reads	1,774,581	3.53%	114.18	9,985,341	4.82%	134.63
Total reads	50,330,252	100.00%	107.16	207,240,204	100.00%	119.49
	Sample HighN.Fungus.2P.1			**Sample** HighN.Fungus.2P.2		
	Count	**% of reads**	**Average length**	**Count**	**% of reads**	**Average length**
Reference	375,632		719.67	375,632		719.67
Mapped reads	124,053,801	91.23%	119.4	103,686,230	91.50%	120.05
Not mapped reads	11,922,347	8.77%	120.18	9,630,040	8.50%	120.95
Reads in pairs	117,006,664	86.05%	139.12	98,642,804	87.05%	137.5
Broken paired reads	7,047,137	5.18%	126.03	5,043,426	4.45%	124.82
Total reads	135,976,148	100.00%	119.47	113,316,270	100.00%	120.13
	Sample HighN.Fungus.2P.3			**Sample** HighN.Fungus.2P.4		
	Count	**% of reads**	**Average length**	**Count**	**% of reads**	**Average length**
Reference	375,632		719.67	375,632		719.67
Mapped reads	128,719,210	91.06%	121.12	128,522,601	92.10%	121.03
Not mapped reads	12,639,996	8.94%	122.59	11,021,025	7.90%	121.83
Reads in pairs	122,088,232	86.37%	137.76	124,795,650	89.43%	132.72
Broken paired reads	6,630,978	4.69%	128.81	3,726,951	2.67%	131.94
Total reads	141,359,206	100.00%	121.25	139,543,626	100.00%	121.09
	Sample HighN.Fungus.2X.1			**Sample** HighN.Fungus.2X.2		
	Count	**% of reads**	**Average length**	**Count**	**% of reads**	**Average length**
Reference	375,632		719.67	375,632		719.67
Mapped reads	108,801,538	92.08%	123.84	126,896,026	92.66%	121.27
Not mapped reads	9,354,684	7.92%	124.68	10,045,690	7.34%	122.14
Reads in pairs	103,926,968	87.96%	141.42	119,514,370	87.27%	144.13
Broken paired reads	4,874,570	4.13%	135.03	7,381,656	5.39%	125.17
Total reads	118,156,222	100.00%	123.91	136,941,716	100.00%	121.33

Table 6.5. The CLC Genomics Workbench mapping statistics. Continued.

	Sample HighN.Fungus.2X.3			Sample HighN.Fungus.2X.4		
	Count	**% of reads**	**Average length**	**Count**	**% of reads**	**Average length**
Reference	375,632		719.67	375,632		719.67
Mapped reads	89,076,704	91.87%	122.99	110,248,543	92.36%	124.94
Not mapped reads	7,882,388	8.13%	123.88	9,123,035	7.64%	125.8
Reads in pairs	84,240,876	86.88%	145.33	102,039,776	85.48%	150.33
Broken paired reads	4,835,828	4.99%	124.35	8,208,767	6.88%	130.24
Total reads	96,959,092	100.00%	123.06	119,371,578	100.00%	125

Table 6.6. Summary of *featureCounts* results for 32 lodgepole pine RNA-Seq samples. The Rsubread v1.24.2 function *featureCounts* (Liao *et al.* 2014) tallied the number of reads that aligned to 154,398 metafeatures, in this case coding sequences, within the reference master transcriptome. The *featureCounts* tool utilized the CLC Genomics Workbench v9.5.2 (*www.qiagenbioinformatics.com*)-generated sequence alignment map (sam) files as input. It also used the TransDecoder v5.0.2 (Haas *et al.* 2013)-generated gene transfer format (gtf) file as input, which contained annotation information for the reference transcriptome along with coordinates of each feature within its respective contig. A read was assigned to a feature if the read had the longest overlap with that feature. A read was unassigned either because the read aligned maximally with two or more features (Unassigned ambiguity) or because the read did not align with any of the metafeatures (Unassigned no features). Tallies for each feature were used for differential expression analysis (Chapter 4).

Sample	Run time (min)	Percent assigned	Total reads	Assigned	Unassigned ambiguity	Unassigned no features
LowN.Wound.2P.1	6.17	65.59%	69,437,829	45,544,232	31,814	23,861,783
LowN.Wound.2P.2	1.2	36.00%	27,997,670	10,079,404	7,992	17,910,274
LowN.Wound.2P.3	4.21	35.04%	100,108,068	35,081,972	27,183	64,998,913
LowN.Wound.2P.4	7.26	66.93%	82,494,603	55,215,619	30,726	27,248,258
LowN.Wound.2X.1	5.14	66.25%	56,081,451	37,151,852	24,729	18,904,870
LowN.Wound.2X.2	5.48	65.13%	61,017,116	39,739,739	31,698	21,245,679
LowN.Wound.2X.3	2.54	65.96%	27,878,903	18,389,938	15,009	9,473,956
LowN.Wound.2X.4	6.21	65.18%	69,088,138	45,034,352	27,352	24,026,434
LowN.Fungus.2P.1	6.56	67.96%	71,114,519	48,328,269	23,957	22,762,293
LowN.Fungus.2P.2	4.55	68.87%	49,967,680	34,412,217	21,073	15,534,390
LowN.Fungus.2P.3	6.52	66.45%	72,758,945	48,344,841	27,729	24,386,375
LowN.Fungus.2P.4	3.77	67.93%	42,093,566	28,595,000	17,845	13,480,721
LowN.Fungus.2X.1	3.38	66.61%	37,921,577	25,258,613	18,618	12,644,346
LowN.Fungus.2X.2	6.75	64.28%	74,662,752	47,994,549	29,329	26,638,874
LowN.Fungus.2X.3	5.24	64.84%	59,779,636	38,761,832	29,154	20,988,650
LowN.Fungus.2X.4	5.16	60.00%	57,795,164	34,677,296	23,334	23,094,534
HighN.Wound.2P.1	4.71	66.51%	2,959,498	35,222,082	21,235	17,716,181
HighN.Wound.2P.2	3.92	67.93%	44,600,600	30,296,984	22,283	14,281,333
HighN.Wound.2P.3	5.93	66.92%	65,364,390	43,744,547	27,651	21,592,192
HighN.Wound.2P.4	5.24	66.93%	57,070,142	38,199,561	21,009	18,849,572
HighN.Wound.2X.1	5.21	66.90%	56,705,815	37,937,141	25,295	18,743,379
HighN.Wound.2X.2	5.64	66.04%	63,530,137	41,952,882	29,149	21,548,106
HighN.Wound.2X.3	2.02	53.09%	23,082,253	12,254,787	12,458	10,815,008
HighN.Wound.2X.4	8.72	67.62%	99,002,017	66,940,510	68,821	31,992,686
HighN.Fungus.2P.1	5.87	66.57%	64,851,417	43,171,976	25,407	21,654,034
HighN.Fungus.2P.2	4.39	35.40%	103,686,230	36,707,788	27,648	66,950,794

Table 6.6. Summary of *featureCounts* results for 32 lodgepole pine RNA-Seq samples. Continued.

Sample	Run time (min)	Percent assigned	Total reads	Assigned	Unassigned ambiguity	Unassigned no features
HighN.Fungus.2P.3	6.01	66.87%	67,119,478	44,882,602	28,039	22,208,837
HighN.Fungus.2P.4	5.85	68.35%	65,801,810	44,973,484	32,785	20,795,541
HighN.Fungus.2X.1	5.00	67.79%	56,396,043	38,229,713	28,996	18,137,334
HighN.Fungus.2X.2	6.05	65.20%	66,611,164	43,428,452	26,747	23,155,965
HighN.Fungus.2X.3	4.30	67.56%	46,567,769	31,459,197	17,917	15,090,655
HighN.Fungus.2X.4	5.23	65.31%	58,748,161	8,368,954	23,130	20,356,077

References

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, N., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N., Regev, A. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8, 1494-1512.

Illumina, Inc. (2013). *Illumina TruSeq stranded mRNA low sample (LS) protocol*. Author, San Diego, California, United States.

Illumina, Inc. (2018). *NextSeq 500 system guide*. Author, San Diego, California, United States.

Liao, Y., Smyth, G. K., Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30, 923-930.

Qiagen, CLCBio. *www.qiagenbioinformatics.com*. (Last accessed March 16, 2020).

Appendix B: Scripts

transabyss.pbs

```
#!/bin/bash
#PBS -S /bin/bash
#PBS -l procs=32
#PBS -l pmem=8190mb
#PBS -l walltime=60:00:00
#PBS -N transabyss
#PBS -m bea
#PBS -M normingt@ualberta.ca

module load application/Trans-ABySS/1.5.5
module load application/ABySS/1.5.2
module load application/gmap/2014-12-02
module load application/samtools/0.1.19
module load application/blat/3.5
module load application/python/2.7.3
module load library/igraph/0.7.1

kmer1=25
kmer2=29
kmer3=33
kmer4=36
kmer5=41

name1=14-01-2P.k${kmer1}
name2=14-01-2P.k${kmer2}
name3=14-01-2P.k${kmer3}
name4=14-01-2P.k${kmer4}
name5=14-01-2P.k${kmer5}

reads1=/data/normingt/M018_RNA-Seq/14-01-2P-transabyss/read=51/M018-
14-01-2P_CLC_Trimmed.fastq

assemblydir1=./${name1}
assemblydir2=./${name2}
assemblydir3=./${name3}
assemblydir4=./${name4}
assemblydir5=./${name5}

finalassembly1=${assemblydir1}/${name1}-final.fa
```

```
finalassembly2=${assemblydir2}/${name2}-final.fa
finalassembly3=${assemblydir3}/${name3}-final.fa
finalassembly4=${assemblydir4}/${name4}-final.fa
finalassembly5=${assemblydir5}/${name5}-final.fa

transabyss --pe ${reads1} --SS --outdir ${assemblydir1} --name ${name1} -k
${kmer1} --threads 16 --island 0 -s 50 -c 2

transabyss --pe ${reads1} --SS --outdir ${assemblydir2} --name ${name2} -k
${kmer2} --threads 16 --island 0 -s 58 -c 2

transabyss --pe ${reads1} --SS --outdir ${assemblydir3} --name ${name3} -k
${kmer3} --threads 16 --island 0 -s 66 -c 2

transabyss --pe ${reads1} --SS --outdir ${assemblydir4} --name ${name4} -k
${kmer4} --threads 16 --island 0 -s 72 -c 2

transabyss --pe ${reads1} --SS --outdir ${assemblydir5} --name ${name5} -k
${kmer5} --threads 16 --island 0 -s 82 -c 2
```

transabyss-merge.pbs

```
#PBS -S /bin/bash
#PBS -l procs=32
#PBS -l pmem=8190mb
#PBS -l walltime=60:00:00
#PBS -N transabyss-merge
#PBS -m bea
#PBS -M normingt@ualberta.ca

module load application/Trans-ABySS/1.5.5
module load application/gmap/2014-12-02
module load application/samtools/0.1.19
module load application/ABySS/1.5.2
module load application/blat/3.5
module load application/python/2.7.3
module load library/igraph/0.7.1

transabyss-merge 14-01-2P.k25-final.fa 14-01-2P.k29-final.fa 14-01-2P.k33-
final.fa 14-01-2P.k36-final.fa 14-01-2P.k41-final.fa --mink 25 --maxk 41 --threads
16 --SS --out ./14-01-2P.k25-41.merged.fa
```

abyss-fac.pbs

```
#!/bin/bash
#PBS -S /bin/bash
#PBS -l procs=1
#PBS -l pmem=256mb
#PBS -l walltime=00:02:00
#PBS -N abyss-fac
##PBS -m bea
##PBS -M normingt@ualberta.ca

$PBS_O_WORKDIR

module load application/ABySS/1.5.2
module load application/blat/3.5
module load application/python/2.7.3
module load library/igraph/0.7.1

abyss-fac 14-01-2P.k33-final.fa
abyss-fac 14-01-2P.k25-41.merged.fa
```

featureCounts.slm

```
#!/bin/bash
#SBATCH --account=def-jek4
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=1
#SBATCH --mem-per-cpu=64000mb
#SBATCH --time=1:00:00
#SBATCH --mail-type=ALL
#SBATCH --mail-user=normingt@ualberta.ca

module load gcc r-bundle-bioconductor # Load R

source("http://bioconductor.org/biocLite.R") # Install library Rsubread
biocLite("Rsubread")

R --vanilla <<featureCounts-R # Use vanilla R parameters (--no-site-file, --no-
init-file, --no-environ and --no-restore)

library(Rsubread) # Load library

# Assign objects for file names:
sam_file <- "sample.sam"
gtf_file <- "master.gtf"
count_file <- "sample_counts.txt"
stat_file <- "sample_stats.txt"
annotation_file <- “featureCounts_annotation.txt”

# Create featureCounts file (read table) and .out file (counts, annotations, stats):
fc <- featureCounts(files=sam_file, annot.ext=gtf_file,
        isGTFAnnotationFile=TRUE, GTF.featureType="CDS",
        GTF.attrType="gene_id", isPairedEnd=TRUE, minFragLength=51,
        maxFragLength=1000, useMetaFeatures=TRUE, autosort=TRUE,
        requireBothEndsMapped=FALSE, allowMultiOverlap=FALSE,
        fraction=FALSE, largestOverlap=TRUE, strandSpecific=1,
        splitOnly=FALSE, countMultiMappingReads=TRUE,
        countChimericFragments=TRUE, reportReads=TRUE)

# Export count, stat and annotation tables:
write.table(x=data.frame(fc$counts, stringsAsFactors=FALSE), file=count_file,
        quote=FALSE, row.names=FALSE)
write.table(x=data.frame(fc$stat, stringsAsFactors=FALSE), file=stat_file,
        quote=FALSE, row.names=FALSE)
```

```
write.table(x=data.frame(fc$annotation, stringsAsFactors=FALSE),
        file=annotation_file, quote=FALSE, row.names=FALSE) # This file is the
        same for all sam files and only needs to be created once.

featureCounts-R # Close the script
```

M018_edgeR.R

```
# Install and load the edgeR and limma packages:
source("http://bioconductor.org/biocLite.R")
biocLite(pkgs = "limma", "edgeR")
library("limma")
library("edgeR")

# ~~~~~Import and prepare data for differential expression analysis~~~~~

# Import and join counts across all samples (obtained via the featureCounts function from the Rsubread R package v1.24.2) to create the raw counts table: Columns are ordered alphabetically by treatment type (High N Fungus 2P, High N Fungus 2X, High N Wound 2P, High N Wound 2X, Low N Fungus 2P, Low N Fungus 2X, Low N Wound 2P, Low N Wound 2X):
GeneID <- read.csv("M018_featureCounts_Ant.csv", header=TRUE)[,2]
GeneID <- as.data.frame(GeneID)
colnames(GeneID)[1] <- "GeneID"
M018.18.01.2P <- read.csv("M018-18-01-2P_counts.csv", header=TRUE)[,2]
M018.18.02.2P <- read.csv("M018-18-02-2P_counts.csv", header=TRUE)[,2]
M018.18.03.2P <- read.csv("M018-18-03-2P_counts.csv", header=TRUE)[,2]
M018.18.04.2P <- read.csv("M018-18-04-2P_counts.csv", header=TRUE)[,2]
M018.18.02.2X <- read.csv("M018-18-02-2X_counts.csv", header=TRUE)[,2]
M018.18.04.2X <- read.csv("M018-18-04-2X_counts.csv", header=TRUE)[,2]
M018.18.05.2X <- read.csv("M018-18-05-2X_counts.csv", header=TRUE)[,2]
M018.18.06.2X <- read.csv("M018-18-06-2X_counts.csv", header=TRUE)[,2]
M018.17.01.2P <- read.csv("M018-17-01-2P_counts.csv", header=TRUE)[,2]
M018.17.02.2P <- read.csv("M018-17-02-2P_counts.csv", header=TRUE)[,2]
```

```
M018.17.03.2P <- read.csv("M018-17-03-2P_counts.csv", header=TRUE)[,2]
M018.17.04.2P <- read.csv("M018-17-04-2P_counts.csv", header=TRUE)[,2]
M018.17.01.2X <- read.csv("M018-17-01-2X_counts.csv", header=TRUE)[,2]
M018.17.02.2X <- read.csv("M018-17-02-2X_counts.csv", header=TRUE)[,2]
M018.17.05.2X <- read.csv("M018-17-05-2X_counts.csv", header=TRUE)[,2]
M018.17.06.2X <- read.csv("M018-17-06-2X_counts.csv", header=TRUE)[,2]
M018.15.01.2P <- read.csv("M018-15-01-2P_counts.csv", header=TRUE)[,2]
M018.15.02.2P <- read.csv("M018-15-02-2P_counts.csv", header=TRUE)[,2]
M018.15.03.2P <- read.csv("M018-15-03-2P_counts.csv", header=TRUE)[,2]
M018.15.04.2P <- read.csv("M018-15-04-2P_counts.csv", header=TRUE)[,2]
M018.15.02.2X <- read.csv("M018-15-02-2X_counts.csv", header=TRUE)[,2]
M018.15.03.2X <- read.csv("M018-15-03-2X_counts.csv", header=TRUE)[,2]
M018.15.05.2X <- read.csv("M018-15-05-2X_counts.csv", header=TRUE)[,2]
M018.15.06.2X <- read.csv("M018-15-06-2X_counts.csv", header=TRUE)[,2]
M018.14.01.2P <- read.csv("M018-14-01-2P_counts.csv", header=TRUE)[,2]
M018.14.02.2P <- read.csv("M018-14-02-2P_counts.csv", header=TRUE)[,2]
M018.14.03.2P <- read.csv("M018-14-03-2P_counts.csv", header=TRUE)[,2]
M018.14.04.2P <- read.csv("M018-14-04-2P_counts.csv", header=TRUE)[,2]
M018.14.01.2X <- read.csv("M018-14-01-2X_counts.csv", header=TRUE)[,2]
M018.14.02.2X <- read.csv("M018-14-02-2X_counts.csv", header=TRUE)[,2]
M018.14.04.2X <- read.csv("M018-14-04-2X_counts.csv", header=TRUE)[,2]
M018.14.05.2X <- read.csv("M018-14-05-2X_counts.csv", header=TRUE)[,2]


counts_raw <- data.frame(GeneID, M018.18.01.2P, M018.18.02.2P,
M018.18.03.2P, M018.18.04.2P, M018.18.02.2X, M018.18.04.2X,
M018.18.05.2X, M018.18.06.2X, M018.17.01.2P, M018.17.02.2P, M018.17.03.2P,
M018.17.04.2P,  M018.17.01.2X, M018.17.02.2X, M018.17.05.2X,
M018.17.06.2X, M018.15.01.2P, M018.15.02.2P, M018.15.03.2P, M018.15.04.2P,
M018.15.02.2X, M018.15.03.2X, M018.15.05.2X, M018.15.06.2X, M018.14.01.2P,
```

M018.14.02.2P, M018.14.03.2P, M018.14.04.2P, M018.14.01.2X, M018.14.02.2X, M018.14.04.2X, M018.14.05.2X)

write.csv(counts_raw, "M018_counts_raw.csv", row.names=FALSE)

Inspect read depth across all samples:

colSums(counts_raw[,-1])

read_depth <- colSums(counts_raw[,-1])

write.csv(read_depth, file="M018_edgeR_read_depth.csv", row.names=FALSE)

summary(colSums(counts_raw[,-1]))

Filter lowly expressed genes based on minimum read depth: There are between max(colSums) = 66940510 and min(colSums) = 10079404 counts per library. So we have a minimum read depth of 10079404. We set a cpm threshold for genes at the following: 1 count per million (cpm) reads must be present in at least 4 libraries. This means for a gene to be filtered it must have a minimum of 1*min(colSums)/10^6 = 1*10079404/10^6 = 10.079404 counts in at least four libraries. This cutoff technique is indicated in the edgeR user's guide (Chen *et al.* 2017) and Anders *et al.* (2013).

cpm <- cpm(counts_raw[,-1])

summary(cpm)

keep <- rowSums(cpm>1)>=4 # Only keep in the analysis those genes which have >1 read per million mapped reads in at least 4 libraries.

fcounts <- counts_raw[keep,]

write.csv(fcounts, file="M018_edgeR_counts_1cpm.csv", row.names=FALSE)

Create DGEList object: The DGEList object contains RNA-seq count data with the associated treatments. First, we create a group variable that tells edgeR which samples belong to which group and then supply that to DGEList in addition to the count matrix.

group <- c("HighN_Fungus_2P", "HighN_Fungus_2P", "HighN_Fungus_2P", "HighN_Fungus_2P", "HighN_Fungus_2X", "HighN_Fungus_2X", "HighN_Fungus_2X", "HighN_Fungus_2X", "HighN_Wound_2P",

"HighN_Wound_2P", "HighN_Wound_2P", "HighN_Wound_2P",
"HighN_Wound_2X", "HighN_Wound_2X", "HighN_Wound_2X",
"HighN_Wound_2X", "LowN_Fungus_2P", "LowN_Fungus_2P",
"LowN_Fungus_2P", "LowN_Fungus_2P", "LowN_Fungus_2X",
"LowN_Fungus_2X", "LowN_Fungus_2X", "LowN_Fungus_2X",
"LowN_Wound_2P", "LowN_Wound_2P", "LowN_Wound_2P",
"LowN_Wound_2P", "LowN_Wound_2X", "LowN_Wound_2X",
"LowN_Wound_2X", "LowN_Wound_2X")

d <- DGEList(counts=fcounts[,-1], group=group)

d$samples # contains a summary of your samples

~~~~~Normalize counts, calculate common dispersion and create glm~~~~~

Estimate normalization factors using the default trimmed mean of M-values (TMM): method (Robinson and Oshlak, 2010):

dnorm <- calcNormFactors(d)

dnorm$samples

Construct the experimental design matrix: As our data set has multiple possible group comparisons, we use a general linear model versus an exact test for differential expression testing. For this method, we must provide an experimental design matrix that incorporates all 8 groups separately. Note the "0" in the command below indicates there is no one single reference or control group for all comparisons.

design <- model.matrix(~0+group, data=dnorm$samples)

colnames(design) <- levels(dnorm$samples$group)

write.csv(design, file="M018_edgeR_glm_design.csv", row.names=FALSE) # The glm design.

Estimate common and trended dispersions over all genes, and the tagwise dispersion using the Cox-Reid (CR) adjusted likelihood (McCarthy *et al.* 2012):

ddisp <- estimateGLMCommonDisp(dnorm, design)

```
ddisp <- estimateGLMTrendedDisp(ddisp, design)

ddisp <- estimateGLMTagwiseDisp(ddisp, design)


# Fit a glm to each gene:

fit <- glmFit(ddisp, design)

colnames(fit) # This prints the order of the contrast coefficients for the glm. Specific contrasts for differential expression testing require that the user choose the correct coefficients.


# ~~~~~Assess differential expression~~~~~

# Count data with over dispersion can be modeled using gene-wise negative binomial generalized linear models. The “glmLRT” function conducts likelihood ratio tests for two or more coefficients (treatment types) in the linear model. The contrast argument is used to indicate which treatments to compare for differential expression testing. The end result is a table of differentially expressed genes (DEGs) with test statistics like log base two fold change (log2FC) and p-value. P-values are adjusted using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995).


# ~~~~~Pair-wise comparisons~~~~~


LowN_2P_FungusVsWound <- glmLRT(fit, contrast=c(0, 0, 0, 0, 1, 0, -1, 0)) # All coefficient values must add to zero in order to test the null hypothesis that no differential expression will occur. The experimental group is given a positive coefficient and the control group is given a negative coefficient.

LowN_2P_FungusVsWound_decideTestsDGE <- decideTestsDGE(LowN_2P_FungusVsWound, adjust.method="BH", p.value=0.01)

summary(LowN_2P_FungusVsWound_decideTestsDGE) # Sanity check for the chosen coefficients.

FDR1 <- p.adjust(LowN_2P_FungusVsWound$table$PValue, method="BH") # The false discovery rate is adjusted using the Benjamini-Hochberg procedure.
```

```
Result1 <- LowN_2P_FungusVsWound$table

Result1["PValue.adjust"] <- FDR1

Result1 <- cbind(fcounts[,1], Result1)

colnames(Result1)[1] <- "GeneID"

write.csv(Result1,
file="M018_edgeR_LowN_2P_FungusVsWound_DE_results.csv",
row.names=FALSE)

ResultSig1 <- Result1[which(Result1$PValue.adjust<0.01),]

Up1 <- ResultSig1[which(ResultSig1$logFC>=2),]

Up1["reg"] <- "up"

Down1 <- ResultSig1[which(ResultSig1$logFC<=-2),]

Down1["reg"] <- "down"

DEGs1 <- rbind(Up1, Down1)

colnames(DEGs1)[1] <- "GeneID"

write.csv(DEGs1[order(DEGs1$logFC),],
file="M018_edgeR_LowN_2P_FungusVsWound_sig_DEGs.csv",
row.names=FALSE) # Produce differential expression table of significant results,
ordered by logFC, for inspecting DEGs that are strongly up or down regulated.


LowN_2X_FungusVsWound <- glmLRT(fit, contrast=c(0, 0, 0, 0, 0, 1, 0, -1))

LowN_2X_FungusVsWound_decideTestsDGE <-
decideTestsDGE(LowN_2X_FungusVsWound, adjust.method="BH",
p.value=0.01)

summary(LowN_2X_FungusVsWound_decideTestsDGE)

FDR2 <- p.adjust(LowN_2X_FungusVsWound$table$PValue, method="BH")

Result2 <- LowN_2X_FungusVsWound$table

Result2["PValue.adjust"] <- FDR2

Result2 <- cbind(fcounts[,1], Result2)

colnames(Result2)[1] <- "GeneID"
```

```
write.csv(Result2,
file="M018_edgeR_LowN_2X_FungusVsWound_DE_results.csv",
row.names=FALSE)

ResultSig2 <- Result2[which(Result2$PValue.adjust<0.01),]

Up2 <- ResultSig2[which(ResultSig2$logFC>=2),]

Up2["reg"] <- "up"

Down2 <- ResultSig2[which(ResultSig2$logFC<=-2),]

Down2["reg"] <- "down"

DEGs2 <- rbind(Up2, Down2)

colnames(DEGs2)[1] <- "GeneID"

write.csv(DEGs2[order(DEGs2$logFC),],
file="M018_edgeR_LowN_2X_FungusVsWound_sig_DEGs.csv",
row.names=FALSE)


HighN_2P_FungusVsWound <- glmLRT(fit, contrast=c(1, 0, -1, 0, 0, 0, 0, 0))

HighN_2P_FungusVsWound_decideTestsDGE <-
decideTestsDGE(HighN_2P_FungusVsWound, adjust.method="BH",
p.value=0.01)

summary(HighN_2P_FungusVsWound_decideTestsDGE)

FDR3 <- p.adjust(HighN_2P_FungusVsWound$table$PValue, method="BH")

Result3 <- HighN_2P_FungusVsWound$table

Result3["PValue.adjust"] <- FDR3

Result3 <- cbind(fcounts[,1], Result3)

colnames(Result3)[1] <- "GeneID"

write.csv(Result3,
file="M018_edgeR_HighN_2P_FungusVsWound_DE_results.csv",
row.names=FALSE)

ResultSig3 <- Result3[which(Result3$PValue.adjust<0.01),]

Up3 <- ResultSig3[which(ResultSig3$logFC>=2),]

Up3["reg"] <- "up"
```

```
Down3 <- ResultSig3[which(ResultSig3$logFC<=-2),]

Down3["reg"] <- "down"

DEGs3 <- rbind(Up3, Down3)

colnames(DEGs3)[1] <- "GeneID"

write.csv(DEGs3[order(DEGs3$logFC),],
file="M018_edgeR_HighN_2P_FungusVsWound_sig_DEGs.csv",
row.names=FALSE)


HighN_2X_FungusVsWound <- glmLRT(fit, contrast=c(0, 1, 0, -1, 0, 0, 0, 0))

HighN_2X_FungusVsWound_decideTestsDGE <-
decideTestsDGE(HighN_2X_FungusVsWound, adjust.method="BH",
p.value=0.01)

summary(HighN_2X_FungusVsWound_decideTestsDGE)

FDR4 <- p.adjust(HighN_2X_FungusVsWound$table$PValue, method="BH")

Result4 <- HighN_2X_FungusVsWound$table

Result4["PValue.adjust"] <- FDR4

Result4 <- cbind(fcounts[,1], Result4)

colnames(Result4)[1] <- "GeneID"

write.csv(Result4,
file="M018_edgeR_HighN_2X_FungusVsWound_DE_results.csv",
row.names=FALSE)

ResultSig4 <- Result4[which(Result4$PValue.adjust<0.01),]

Up4 <- ResultSig4[which(ResultSig4$logFC>=2),]

Up4["reg"] <- "up"

Down4 <- ResultSig4[which(ResultSig4$logFC<=-2),]

Down4["reg"] <- "down"

DEGs4 <- rbind(Up4, Down4)

colnames(DEGs4)[1] <- "GeneID"
```

```
write.csv(DEGs4[order(DEGs4$logFC),],
file="M018_edgeR_HighN_2X_FungusVsWound_sig_DEGs.csv",
row.names=FALSE)


Wound_2P_HighNvsLowN <- glmLRT(fit, contrast=c(0, 0, 1, 0, 0, 0, -1, 0))

Wound_2P_HighNvsLowN_decideTestsDGE <-
decideTestsDGE(Wound_2P_HighNvsLowN, adjust.method="BH",
p.value=0.01)

summary(Wound_2P_HighNvsLowN_decideTestsDGE)

FDR5 <- p.adjust(Wound_2P_HighNvsLowN$table$PValue, method="BH")

Result5 <- Wound_2P_HighNvsLowN$table

Result5["PValue.adjust"] <- FDR5

Result5 <- cbind(fcounts[,1], Result5)

colnames(Result5)[1] <- "GeneID"

write.csv(Result5,
file="M018_edgeR_Wound_2P_HighNvsLowN_DE_results.csv",
row.names=FALSE)

ResultSig5 <- Result5[which(Result5$PValue.adjust<0.01),]

Up5 <- ResultSig5[which(ResultSig5$logFC>=2),]

Up5["reg"] <- "up"

Down5 <- ResultSig5[which(ResultSig5$logFC<=-2),]

Down5["reg"] <- "down"

DEGs5 <- rbind(Up5, Down5)

colnames(DEGs5)[1] <- "GeneID"

write.csv(DEGs5[order(DEGs5$logFC),],
file="M018_edgeR_Wound_2P_HighNvsLowN_sig_DEGs.csv",
row.names=FALSE)


Wound_2X_HighNvsLowN <- glmLRT(fit, contrast=c(0, 0, 0, 1, 0, 0, 0, -1))
```

```
Wound_2X_HighNvsLowN_decideTestsDGE <-
decideTestsDGE(Wound_2X_HighNvsLowN, adjust.method="BH",
p.value=0.01)

summary(Wound_2X_HighNvsLowN_decideTestsDGE)

FDR6 <- p.adjust(Wound_2X_HighNvsLowN$table$PValue, method="BH")

Result6 <- Wound_2X_HighNvsLowN$table

Result6["PValue.adjust"] <- FDR6

Result6 <- cbind(fcounts[,1], Result6)

colnames(Result6)[1] <- "GeneID"

write.csv(Result6,
file="M018_edgeR_Wound_2X_HighNvsLowN_DE_results.csv",
row.names=FALSE)

ResultSig6 <- Result6[which(Result6$PValue.adjust<0.01),]

Up6 <- ResultSig6[which(ResultSig6$logFC>=2),]

Up6["reg"] <- "up"

Down6 <- ResultSig6[which(ResultSig6$logFC<=-2),]

Down6["reg"] <- "down"

DEGs6 <- rbind(Up6, Down6)

colnames(DEGs6)[1] <- "GeneID"

write.csv(DEGs6[order(DEGs6$logFC),],
file="M018_edgeR_Wound_2X_HighNvsLowN_sig_DEGs.csv",
row.names=FALSE)


Fungus_2P_HighNvsLowN <- glmLRT(fit, contrast=c(1, 0, 0, 0, -1, 0, 0, 0))

Fungus_2P_HighNvsLowN_decideTestsDGE <-
decideTestsDGE(Fungus_2P_HighNvsLowN, adjust.method="BH",
p.value=0.01)

summary(Fungus_2P_HighNvsLowN_decideTestsDGE)

FDR7 <- p.adjust(Fungus_2P_HighNvsLowN$table$PValue, method="BH")

Result7 <- Fungus_2P_HighNvsLowN$table
```

```
Result7["PValue.adjust"] <- FDR7

Result7 <- cbind(fcounts[,1], Result7)

colnames(Result7)[1] <- "GeneID"

write.csv(Result7,
file="M018_edgeR_Fungus_2P_HighNvsLowN_DE_results.csv",
row.names=FALSE)

ResultSig7 <- Result7[which(Result7$PValue.adjust<0.01),]

Up7 <- ResultSig7[which(ResultSig7$logFC>=2),]

Up7["reg"] <- "up"

Down7 <- ResultSig7[which(ResultSig7$logFC<=-2),]

Down7["reg"] <- "down"

DEGs7 <- rbind(Up7, Down7)

colnames(DEGs7)[1] <- "GeneID"

write.csv(DEGs7[order(DEGs7$logFC),],
file="M018_edgeR_Fungus_2P_HighNvsLowN_sig_DEGs.csv",
row.names=FALSE)


Fungus_2X_HighNvsLowN <- glmLRT(fit, contrast=c(0, 1, 0, 0, 0, -1, 0, 0))

Fungus_2X_HighNvsLowN_decideTestsDGE <-
decideTestsDGE(Fungus_2X_HighNvsLowN, adjust.method="BH",
p.value=0.01)

summary(Fungus_2X_HighNvsLowN_decideTestsDGE)

FDR8 <- p.adjust(Fungus_2X_HighNvsLowN$table$PValue, method="BH")

Result8 <- Fungus_2X_HighNvsLowN$table

Result8["PValue.adjust"] <- FDR8

Result8 <- cbind(fcounts[,1], Result8)

colnames(Result8)[1] <- "GeneID"

write.csv(Result8,
file="M018_edgeR_Fungus_2X_HighNvsLowN_DE_results.csv",
row.names=FALSE)
```

```
ResultSig8 <- Result8[which(Result8$PValue.adjust<0.01),]

Up8 <- ResultSig8[which(ResultSig8$logFC>=2),]

Up8["reg"] <- "up"

Down8 <- ResultSig8[which(ResultSig8$logFC<=-2),]

Down8["reg"] <- "down"

DEGs8 <- rbind(Up8, Down8)

colnames(DEGs8)[1] <- "GeneID"

write.csv(DEGs8[order(DEGs8$logFC),],
file="M018_edgeR_Fungus_2X_HighNvsLowN_sig_DEGs.csv",
row.names=FALSE)


# ~~~~~Multiple treatment comparisons~~~~~


# We are using a more strenuous cpm cutoff of >2 when combining multiple treatment groups for differential expression analysis. This requires a repeat of data preparation and dispersion calculations.

keep2 <- rowSums(cpm>2)>=4 # Only keep in the analysis those genes that have >2 read per million mapped reads in at least 4 libraries.

fcounts2 <- counts_raw[keep2,]

write.csv(fcounts2, file="M018_edgeR_counts_2cpm.csv", row.names=FALSE)

d2 <- DGEList(counts=fcounts2[,-1], group=group)

dnorm2 <- calcNormFactors(d2)

design2 <- model.matrix(~0+group, data=dnorm2$samples)

colnames(design2) <- levels(dnorm2$samples$group)


ddisp2 <- estimateGLMCommonDisp(dnorm2, design2)

ddisp2 <- estimateGLMTrendedDisp(ddisp2, design2)

ddisp2 <- estimateGLMTagwiseDisp(ddisp2, design2)

fit2 <- glmFit(ddisp2, design2)
```

```
# All 2P (regardless of fungus) HighNvsLowN

All_2P_HighNvsLowN <- glmLRT(fit2, contrast=c(.5, 0, .5, 0, -.5, 0, -.5, 0)) # Fraction values are used to "combine" treatment groups for differential expression comparison. Here, treatment groups have been "combined" by caluculating the average of their expression values.

All_2P_HighNvsLowN_decideTestsDGE <- decideTestsDGE(All_2P_HighNvsLowN, adjust.method="BH", p.value=0.01)

summary(All_2P_HighNvsLowN_decideTestsDGE)

FDR9 <- p.adjust(All_2P_HighNvsLowN$table$PValue, method="BH")

Result9 <- All_2P_HighNvsLowN$table

Result9["PValue.adjust"] <- FDR9

Result9 <- cbind(fcounts2[,1], Result9)

colnames(Result9)[1] <- "GeneID"

write.csv(Result9, file="M018_edgeR_All_2P_HighNvsLowN_DE_results.csv", row.names=FALSE)

ResultSig9 <- Result9[which(Result9$PValue.adjust<0.01),]

Up9 <- ResultSig9[which(ResultSig9$logFC>=2),]

Up9["reg"] <- "up"

Down9 <- ResultSig9[which(ResultSig9$logFC<=-2),]

Down9["reg"] <- "down"

DEGs9 <- rbind(Up9, Down9)

colnames(DEGs9)[1] <- "GeneID"

write.csv(DEGs9, file="M018_edgeR_All_2P_HighNvsLowN_sig_DEGs.csv", row.names=FALSE)


# All 2X (regardless of fungus) HighNvsLowN

All_2X_HighNvsLowN <- glmLRT(fit2, contrast=c(0, .5, 0, .5, 0, -.5, 0, -.5))
```

```
All_2X_HighNvsLowN_decideTestsDGE <-
decideTestsDGE(All_2X_HighNvsLowN, adjust.method="BH", p.value=0.01)

summary(All_2X_HighNvsLowN_decideTestsDGE)

FDR10 <- p.adjust(All_2X_HighNvsLowN$table$PValue, method="BH")

Result10 <- All_2X_HighNvsLowN$table

Result10["PValue.adjust"] <- FDR10

Result10 <- cbind(fcounts2[,1], Result10)

colnames(Result10)[1] <- "GeneID"

write.csv(Result10,
file="M018_edgeR_All_2X_HighNvsLowN_DE_results.csv",
row.names=FALSE)

ResultSig10 <- Result10[which(Result10$PValue.adjust<0.01),]

Up10 <- ResultSig10[which(ResultSig10$logFC>=2),]

Up10["reg"] <- "up"

Down10 <- ResultSig10[which(ResultSig10$logFC<=-2),]

Down10["reg"] <- "down"

DEGs10 <- rbind(Up10, Down10)

colnames(DEGs10)[1] <- "GeneID"

write.csv(DEGs10, file="M018_edgeR_All_2X_HighNvsLowN_sig_DEGs.csv",
row.names=FALSE)


# All 2P (regardless of nitrogen level) FungusVsWound

All_2P_FungusVsWound <- glmLRT(fit2, contrast=c(.5, 0, -.5, 0, .5, 0, -.5, 0))

All_2P_FungusVsWound_decideTestsDGE <-
decideTestsDGE(All_2P_FungusVsWound, adjust.method="BH", p.value=0.01)

summary(All_2P_FungusVsWound_decideTestsDGE)

FDR11 <- p.adjust(All_2P_FungusVsWound$table$PValue, method="BH")

Result11 <- All_2P_FungusVsWound$table

Result11["PValue.adjust"] <- FDR11
```

```
Result11 <- cbind(fcounts2[,1], Result11)

colnames(Result11)[1] <- "GeneID"

write.csv(Result11,
file="M018_edgeR_All_2P_FungusVsWound_DE_results.csv",
row.names=FALSE)

ResultSig11 <- Result11[which(Result11$PValue.adjust<0.01),]

Up11 <- ResultSig11[which(ResultSig11$logFC>=2),]

Up11["reg"] <- "up"

Down11 <- ResultSig11[which(ResultSig11$logFC<=-2),]

Down11["reg"] <- "down"

DEGs11 <- rbind(Up11, Down11)

colnames(DEGs11)[1] <- "GeneID"

write.csv(DEGs11,
file="M018_edgeR_All_2P_FungusVsWound_sig_DEGs.csv",
row.names=FALSE)


# All 2X (regardless of nitrogen level) FungusVsWound

All_2X_FungusVsWound <- glmLRT(fit2, contrast=c(0, .5, 0, -.5, 0, .5, 0, -.5))

All_2X_FungusVsWound_decideTestsDGE <-
decideTestsDGE(All_2X_FungusVsWound, adjust.method="BH", p.value=0.01)

summary(All_2X_FungusVsWound_decideTestsDGE)

FDR12 <- p.adjust(All_2X_FungusVsWound$table$PValue, method="BH")

Result12 <- All_2X_FungusVsWound$table

Result12["PValue.adjust"] <- FDR12

Result12 <- cbind(fcounts2[,1], Result12)

colnames(Result12)[1] <- "GeneID"

write.csv(Result12,
file="M018_edgeR_All_2X_FungusVsWound_DE_results.csv",
row.names=FALSE)
```

```
ResultSig12 <- Result12[which(Result12$PValue.adjust<0.01),]

Up12 <- ResultSig12[which(ResultSig12$logFC>=2),]

Up12["reg"] <- "up"

Down12 <- ResultSig12[which(ResultSig12$logFC<=-2),]

Down12["reg"] <- "down"

DEGs12 <- rbind(Up12, Down12)

colnames(DEGs12)[1] <- "GeneID"

write.csv(DEGs12,
file="M018_edgeR_All_2X_FungusVsWound_sig_DEGs.csv",
row.names=FALSE)


# ~~~~~Create master result files~~~~~


M018_edgeR_raw_DE_results_1cpm <- data.frame(Result1[,1], "::",
"LowN_2P_FungusVsWound", Result1[,2:6], "::",
"LowN_2X_FungusVsWound", Result2[,2:6], "::",
"HighN_2P_FungusVsWound", Result3[,2:6], "::",
"HighN_2X_FungusVsWound", Result4[,2:6], "::",
"Wound_2P_HighNvsLowN", Result5[,2:6], "::", "Wound_2X_HighNvsLowN",
Result6[,2:6], "::", "Fungus_2P_HighNvsLowN", Result7[,2:6], "::",
"Fungus_2X_HighNvsLowN", Result8[,2:6], "::")

colnames(M018_edgeR_raw_DE_results_1cpm)[1:58] <- c("GeneID", "::", "DE
Comparison", "logFC", "logCPM", "LR", "PValue", "PValue.adjust", "::", "DE
Comparison", "logFC", "logCPM", "LR", "PValue", "PValue.adjust", "::", "DE
Comparison", "logFC", "logCPM", "LR", "PValue", "PValue.adjust", "::", "DE
Comparison", "logFC", "logCPM", "LR", "PValue", "PValue.adjust", "::", "DE
Comparison", "logFC", "logCPM", "LR", "PValue", "PValue.adjust", "::", "DE
Comparison", "logFC", "logCPM", "LR", "PValue", "PValue.adjust", "::", "DE
Comparison", "logFC", "logCPM", "LR", "PValue", "PValue.adjust", "::", "DE
Comparison", "logFC", "logCPM", "LR", "PValue", "PValue.adjust", "::")


M018_edgeR_raw_DE_results_2cpm <- data.frame(Result9[,1], "::",
"All_2P_HighNvsLowN", Result9[,2:6], "::", "All_2X_HighNvsLowN",
```

```
Result10[,2:6], "::", "All_2P_FungusVsWound", Result11[,2:6], "::",
"All_2X_FungusVsWound", Result12[,2:6], "::")

colnames(M018_edgeR_raw_DE_results_2cpm)[1:30] <- c("GeneID", "::", "DE
Comparison", "logFC", "logCPM", "LR", "PValue", "PValue.adjust", "::", "DE
Comparison", "logFC", "logCPM", "LR", "PValue", "PValue.adjust", "::", "DE
Comparison", "logFC", "logCPM", "LR", "PValue", "PValue.adjust", "::", "DE
Comparison", "logFC", "logCPM", "LR", "PValue", "PValue.adjust", "::")


# ~~~~~Annotate results~~~~~

# TAIR (Lamesch et al. 2011), Ref-Seq (Pruitt et al. 2004), ConGenie (Sundell et
al. 2015), MapMan (Lohse et al. 2014), GO (The Gene Ontology Consortium 2015)
and InterPro domain (Apweiler et al. 2001) annotations were combined to form an
annotation master list called "LodgeAntn_v2.txt", which was created by Dr.
Rhiannon Peery. These annotations were assigned to the sig DEGs lists and the
master results files to annotate our transcripts using the merge function.:


LodgeAntn <- read.csv(file="LodgeAntn_v2.csv", header=TRUE)

colnames(LodgeAntn)[1] <- "GeneID"


DEGs_LowN_2P_FungusVsWound_Antd <- merge(DEGs1, LodgeAntn,
by="GeneID", all.x=TRUE)

DEGs_LowN_2P_FungusVsWound_Antd <-
data.frame("LowN_2P_FungusVsWound",
DEGs_LowN_2P_FungusVsWound_Antd)

colnames(DEGs_LowN_2P_FungusVsWound_Antd)[1] <- "DE comparison"

write.csv(DEGs_LowN_2P_FungusVsWound_Antd,
file="M018_edgeR_LowN_2P_FungusVsWound_sig_DEGs_Antd.csv",
row.names=FALSE)


DEGs_LowN_2X_FungusVsWound_Antd <- merge(DEGs2, LodgeAntn,
by="GeneID", all.x=TRUE)
```

```
DEGs_LowN_2X_FungusVsWound_Antd <-
data.frame("LowN_2X_FungusVsWound",
DEGs_LowN_2X_FungusVsWound_Antd)

colnames(DEGs_LowN_2X_FungusVsWound_Antd)[1] <- "DE comparison"

write.csv(DEGs_LowN_2X_FungusVsWound_Antd,
file="M018_edgeR_LowN_2X_FungusVsWound_sig_DEGs_Antd.csv",
row.names=FALSE)


DEGs_HighN_2P_FungusVsWound_Antd <- merge(DEGs3, LodgeAntn,
by="GeneID", all.x=TRUE)

DEGs_HighN_2P_FungusVsWound_Antd <-
data.frame("HighN_2P_FungusVsWound",
DEGs_HighN_2P_FungusVsWound_Antd)

colnames(DEGs_HighN_2P_FungusVsWound_Antd)[1] <- "DE comparison"

write.csv(DEGs_HighN_2P_FungusVsWound_Antd,
file="M018_edgeR_HighN_2P_FungusVsWound_sig_DEGs_Antd.csv",
row.names=FALSE)


DEGs_HighN_2X_FungusVsWound_Antd <- merge(DEGs4, LodgeAntn,
by="GeneID", all.x=TRUE)

DEGs_HighN_2X_FungusVsWound_Antd <-
data.frame("HighN_2X_FungusVsWound",
DEGs_HighN_2X_FungusVsWound_Antd)

colnames(DEGs_HighN_2X_FungusVsWound_Antd)[1] <- "DE comparison"

write.csv(DEGs_HighN_2X_FungusVsWound_Antd,
file="M018_edgeR_HighN_2X_FungusVsWound_sig_DEGs_Antd.csv",
row.names=FALSE)


DEGs_Wound_2P_HighNvsLowN_Antd <- merge(DEGs5, LodgeAntn,
by="GeneID", all.x=TRUE)

DEGs_Wound_2P_HighNvsLowN_Antd <-
data.frame("Wound_2P_HighNvsLowN",
DEGs_Wound_2P_HighNvsLowN_Antd)
```

```
colnames(DEGs_Wound_2P_HighNvsLowN_Antd)[1] <- "DE comparison"

write.csv(DEGs_Wound_2P_HighNvsLowN_Antd,
file="M018_edgeR_Wound_2P_HighNvsLowN_sig_DEGs_Antd.csv",
row.names=FALSE)


DEGs_Wound_2X_HighNvsLowN_Antd <- merge(DEGs6, LodgeAntn,
by="GeneID", all.x=TRUE)

DEGs_Wound_2X_HighNvsLowN_Antd <-
data.frame("Wound_2X_HighNvsLowN",
DEGs_Wound_2X_HighNvsLowN_Antd)

colnames(DEGs_Wound_2X_HighNvsLowN_Antd)[1] <- "DE comparison"

write.csv(DEGs_Wound_2X_HighNvsLowN_Antd,
file="M018_edgeR_Wound_2X_HighNvsLowN_sig_DEGs_Antd.csv",
row.names=FALSE)


DEGs_Fungus_2P_HighNvsLowN_Antd <- merge(DEGs7, LodgeAntn,
by="GeneID", all.x=TRUE)

DEGs_Fungus_2P_HighNvsLowN_Antd <-
data.frame("Fungus_2P_HighNvsLowN",
DEGs_Fungus_2P_HighNvsLowN_Antd)

colnames(DEGs_Fungus_2P_HighNvsLowN_Antd)[1] <- "DE comparison"

write.csv(DEGs_Fungus_2P_HighNvsLowN_Antd,
file="M018_edgeR_Fungus_2P_HighNvsLowN_sig_DEGs_Antd.csv",
row.names=FALSE)


DEGs_Fungus_2X_HighNvsLowN_Antd <- merge(DEGs8, LodgeAntn,
by="GeneID", all.x=TRUE)

DEGs_Fungus_2X_HighNvsLowN_Antd <-
data.frame("Fungus_2X_HighNvsLowN",
DEGs_Fungus_2X_HighNvsLowN_Antd)

colnames(DEGs_Fungus_2X_HighNvsLowN_Antd)[1] <- "DE comparison"
```

```
write.csv(DEGs_Fungus_2X_HighNvsLowN_Antd,
file="M018_edgeR_Fungus_2X_HighNvsLowN_sig_DEGs_Antd.csv",
row.names=FALSE)


DEGs_All_2P_HighNvsLowN_Antd <- merge(DEGs9, LodgeAntn,
by="GeneID", all.x=TRUE)

DEGs_All_2P_HighNvsLowN_Antd <- data.frame("All_2P_HighNvsLowN",
DEGs_All_2P_HighNvsLowN_Antd)

colnames(DEGs_All_2P_HighNvsLowN_Antd)[1] <- "DE comparison"

write.csv(DEGs_All_2P_HighNvsLowN_Antd,
file="M018_edgeR_All_2P_HighNvsLowN_sig_DEGs_Antd.csv",
row.names=FALSE)


DEGs_All_2X_HighNvsLowN_Antd <- merge(DEGs10, LodgeAntn,
by="GeneID", all.x=TRUE)

DEGs_All_2X_HighNvsLowN_Antd <- data.frame("All_2X_HighNvsLowN",
DEGs_All_2X_HighNvsLowN_Antd)

colnames(DEGs_All_2X_HighNvsLowN_Antd)[1] <- "DE comparison"

write.csv(DEGs_All_2X_HighNvsLowN_Antd,
file="M018_edgeRAll_2X_HighNvsLowN_sig_DEGs__Antd.csv",
row.names=FALSE)


DEGs_All_2P_FungusVsWound_Antd <- merge(DEGs11, LodgeAntn,
by="GeneID", all.x=TRUE)

DEGs_All_2P_FungusVsWound_Antd <-
data.frame("All_2P_FungusVsWound", DEGs_All_2P_FungusVsWound_Antd)

colnames(DEGs_All_2P_FungusVsWound_Antd)[1] <- "DE comparison"

write.csv(DEGs_All_2P_FungusVsWound_Antd,
file="M018_edgeR_All_2P_FungusVsWound_sig_DEGs_Antd.csv",
row.names=FALSE)
```

```
DEGs_All_2X_FungusVsWound_Antd <- merge(DEGs12, LodgeAntn,
by="GeneID", all.x=TRUE)

DEGs_All_2X_FungusVsWound_Antd <-
data.frame("All_2X_FungusVsWound", DEGs_All_2X_FungusVsWound_Antd)

colnames(DEGs_All_2X_FungusVsWound_Antd)[1] <- "DE comparison"

write.csv(DEGs_All_2X_FungusVsWound_Antd,
file="M018_edgeR_All_2X_FungusVsWound_sig_DEGs_Antd.csv",
row.names=FALSE)


M018_edgeR_raw_DE_results_1cpm_Antd <-
merge(M018_edgeR_raw_DE_results_1cpm, LodgeAntn, by="GeneID",
all.x=TRUE)

write.csv(M018_edgeR_raw_DE_results_1cpm_Antd,
file="M018_edgeR_raw_DE_results_1cpm_Antd.csv", row.names=FALSE)


M018_edgeR_raw_DE_results_2cpm_Antd <-
merge(M018_edgeR_raw_DE_results_2cpm, LodgeAntn, by="GeneID",
all.x=TRUE)

write.csv(M018_edgeR_raw_DE_results_2cpm_Antd,
file="M018_edgeR_raw_DE_results_2cpm_Antd.csv", row.names=FALSE)


save.image(file="M018_edgeR.RDATA") # Save R environment.
```

M018_DESeq2_2P.R

This script was also implemented for secondary xylem (2X) samples.

Install and load the DESeq2 package:

install.packages("BiocManager")

BiocManager::install("DESeq2")

library ("DESeq2")

Load count data: Counts were produced using the Rsubread featureCounts function and filtered using the edgeR cpm function (counts per million). We will use 1 cpm cutoff for comparing two treatments and 2 cpm for comparing groups of treatments.

countdata_2cpm <- read.csv("M018_edgeR_counts_2cpm.csv", header=TRUE)

names(countdata_2cpm) # View samples names

countmatrix_2cpm <- data.matrix(countdata_2cpm[,c(1,2:5,10:13,18:21,26:29)])
# Extract secondary phloem (2P) samples and convert to matrix

countdata_1cpm <- read.csv("M018_edgeR_counts_1cpm.csv", header=TRUE)

countmatrix_1cpm <- data.matrix(countdata_1cpm[,c(1,2:5,10:13,18:21,26:29)])
# Extract secondary phloem (2P) samples and convert to matrix

Load meta data: Tables assign treatments to sample codes.

sampleInfo_All_2P <- read.csv("meta_All_2P.csv", header=TRUE)

sampleInfo_Fungus_2P <- read.csv("meta_Fungus_2P.csv", header=TRUE)

sampleInfo_Wound_2P <- read.csv("meta_Wound_2P.csv", header=TRUE)

sampleInfo_HighN_2P <- read.csv("meta_HighN_2P.csv", header=TRUE)

sampleInfo_LowN_2P <- read.csv("meta_LowN_2P.csv", header=TRUE)

Generate the DESeqData set object from the count matrix, meta data and desired design: The ddsMat object contains all the input data, the information about the

procedures implemented on the data and will store the differential expression results and glm coefficients following implementation of the functions estimateSizeFactors and DESeq.

ddsMat_All_2P_HighNvsLowN <- DESeqData setFromMatrix(countData=countmatrix_2cpm[,-1], colData=sampleInfo_All_2P, design=~ fungus + nitrogen) # HighN vs LowN regardless of inoculation type. Note that ~fungus means inoculation type is being controlled for.

ddsMat_All_2P_HighNvsLowN$nitrogen <- relevel(ddsMat_All_2P_HighNvsLowN$nitrogen, ref="low") # Set LowN (low) as the control group being compared with the HighN (high) experimental group.

ddsMat_All_2P_FungusVsWound <- DESeqData setFromMatrix(countData=countmatrix_2cpm[,-1], colData=sampleInfo_All_2P, design=~ nitrogen + fungus) # Fungus vs Wound regardless of nitrogen level. Note that ~nitrogen means nitrogen treatment is being controlled for.

ddsMat_All_2P_FungusVsWound$fungus <- relevel(ddsMat_All_2P_FungusVsWound$fungus, ref="no") # Set Wound (no) as the control group being compared with the Fungus (yes) experimental group.

ddsMat_Fungus_2P_HighNvsLowN <- DESeqData setFromMatrix(countData=countmatrix_1cpm[,c(2:5,10:13)], colData=sampleInfo_Fungus_2P, design=~ 0 + nitrogen) # HighN vs LowN for fungal inoculated trees. Note that ~0 means there is no control group that all samples are being contrasted with.

ddsMat_Fungus_2P_HighNvsLowN$nitrogen <- relevel(ddsMat_Fungus_2P_HighNvsLowN$nitrogen, ref="low")

ddsMat_Wound_2P_HighNvsLowN <- DESeqData setFromMatrix(countData=countmatrix_1cpm[,c(6:9,14:17)], colData=sampleInfo_Wound_2P, design=~ 0 + nitrogen) # HighN vs LowN for wounded trees

ddsMat_Wound_2P_HighNvsLowN$nitrogen <- relevel(ddsMat_Wound_2P_HighNvsLowN$nitrogen, ref="low")

```
ddsMat_HighN_2P_FungusVsWound <- DESeqData
setFromMatrix(countData=countmatrix_1cpm[,2:9],
colData=sampleInfo_HighN_2P, design=~ 0 + fungus) # Fungus vs Wound for
HighN trees

ddsMat_HighN_2P_FungusVsWound$fungus <-
relevel(ddsMat_HighN_2P_FungusVsWound$fungus, ref="no")


ddsMat_LowN_2P_FungusVsWound <- DESeqData
setFromMatrix(countData=countmatrix_1cpm[,10:17],
colData=sampleInfo_LowN_2P, design=~ 0 + fungus) # Fungus vs Wound for
LowN trees

ddsMat_LowN_2P_FungusVsWound$fungus <-
relevel(ddsMat_LowN_2P_FungusVsWound$fungus, ref="no")


# Estimate size factors required for differential expression analyses: These
normalization factors are calculated using the "median ratio method" described by
Equation 5 in Anders and Huber (2010).

ddsMat_All_2P_HighNvsLowN <-
estimateSizeFactors(ddsMat_All_2P_HighNvsLowN)

ddsMat_All_2P_FungusVsWound <-
estimateSizeFactors(ddsMat_All_2P_FungusVsWound)

ddsMat_Fungus_2P_HighNvsLowN <-
estimateSizeFactors(ddsMat_Fungus_2P_HighNvsLowN)

ddsMat_Wound_2P_HighNvsLowN <-
estimateSizeFactors(ddsMat_Wound_2P_HighNvsLowN)

ddsMat_HighN_2P_FungusVsWound <-
estimateSizeFactors(ddsMat_HighN_2P_FungusVsWound)

ddsMat_LowN_2P_FungusVsWound <-
estimateSizeFactors(ddsMat_LowN_2P_FungusVsWound)


# This single command performs differential expression analysis, which includes
estimating dispersions, estimating gene-wise dispersions, determining the mean-
dispersion relationship, fitting the glm and hypothesis testing for DE.
```

```
ddsMat_All_2P_HighNvsLowN <- DESeq(ddsMat_All_2P_HighNvsLowN)

ddsMat_All_2P_FungusVsWound <- DESeq(ddsMat_All_2P_FungusVsWound)

ddsMat_Fungus_2P_HighNvsLowN <-
DESeq(ddsMat_Fungus_2P_HighNvsLowN)

ddsMat_Wound_2P_HighNvsLowN <-
DESeq(ddsMat_Wound_2P_HighNvsLowN)

ddsMat_HighN_2P_FungusVsWound <-
DESeq(ddsMat_HighN_2P_FungusVsWound)

ddsMat_LowN_2P_FungusVsWound <-
DESeq(ddsMat_LowN_2P_FungusVsWound)


# Extract all results:

res_All_2P_HighNvsLowN <- results(ddsMat_All_2P_HighNvsLowN)

res_All_2P_HighNvsLowN <- cbind(countdata_2cpm[,1],
res_All_2P_HighNvsLowN)

colnames(res_All_2P_HighNvsLowN)[1] <- "GeneID"

write.csv(res_All_2P_HighNvsLowN,
"M018_DESeq2_All_2P_HighNvsLowN_DE_results.csv", row.names=FALSE)


res_All_2P_FungusVsWound <- results(ddsMat_All_2P_FungusVsWound)

res_All_2P_FungusVsWound <-
cbind(countdata_2cpm[,1],res_All_2P_FungusVsWound)

colnames(res_All_2P_FungusVsWound)[1] <- "GeneID"

write.csv(res_All_2P_FungusVsWound,
"M018_DESeq2_All_2P_FungusVsWound_DE_results.csv",
row.names=FALSE)


res_Fungus_2P_HighNvsLowN <- results(ddsMat_Fungus_2P_HighNvsLowN)

res_Fungus_2P_HighNvsLowN <-
cbind(countdata_1cpm[,1],res_Fungus_2P_HighNvsLowN)
```

```
colnames(res_Fungus_2P_HighNvsLowN)[1] <- "GeneID"

write.csv(res_Fungus_2P_HighNvsLowN,
"M018_DESeq2_Fungus_2P_HighNvsLowN_DE_results.csv",
row.names=FALSE)


res_Wound_2P_HighNvsLowN <- results(ddsMat_Wound_2P_HighNvsLowN)

res_Wound_2P_HighNvsLowN <-
cbind(countdata_1cpm[,1],res_Wound_2P_HighNvsLowN)

colnames(res_Wound_2P_HighNvsLowN)[1] <- "GeneID"

write.csv(res_Wound_2P_HighNvsLowN,
"M018_DESeq2_Wound_2P_HighNvsLowN_DE_results.csv",
row.names=FALSE)


res_HighN_2P_FungusVsWound <-
results(ddsMat_HighN_2P_FungusVsWound)

res_HighN_2P_FungusVsWound <-
cbind(countdata_1cpm[,1],res_HighN_2P_FungusVsWound)

colnames(res_HighN_2P_FungusVsWound)[1] <- "GeneID"

write.csv(res_HighN_2P_FungusVsWound,
"M018_DESeq2_HighN_2P_FungusVsWound_DE_results.csv",
row.names=FALSE)


res_LowN_2P_FungusVsWound <-
results(ddsMat_LowN_2P_FungusVsWound)

res_LowN_2P_FungusVsWound <-
cbind(countdata_1cpm[,1],res_LowN_2P_FungusVsWound)

colnames(res_LowN_2P_FungusVsWound)[1] <- "GeneID"

write.csv(res_LowN_2P_FungusVsWound,
"M018_DESeq2_LowN_2P_FungusVsWound_DE_results.csv",
row.names=FALSE)
```

```
# Export files with sig DEG results (adjusted p-value < 0.01, |log2FC| >= 2):

sigres_All_2P_HighNvsLowN <-
res_All_2P_HighNvsLowN[which(res_All_2P_HighNvsLowN$padj<0.01),]

Up_sigres_All_2P_HighNvsLowN <-
sigres_All_2P_HighNvsLowN[which(sigres_All_2P_HighNvsLowN$log2FoldC
hange>=2),]

Up_sigres_All_2P_HighNvsLowN["reg"] <- "up"

down_sigres_All_2P_HighNvsLowN <-
sigres_All_2P_HighNvsLowN[which(sigres_All_2P_HighNvsLowN$log2FoldC
hange<=-2),]

down_sigres_All_2P_HighNvsLowN["reg"] <- "down"

DEGs_All_2P_HighNvsLowN <- rbind(Up_sigres_All_2P_HighNvsLowN,
down_sigres_All_2P_HighNvsLowN)

write.csv(DEGs_All_2P_HighNvsLowN[order(DEGs_All_2P_HighNvsLowN$lo
g2FoldChange),], file="M018_DESeq2_All_2P_HighNvsLowN_sig_DEGs.csv",
row.names=FALSE)


sigres_All_2P_FungusVsWound <-
res_All_2P_FungusVsWound[which(res_All_2P_FungusVsWound$padj<0.01),
]

Up_sigres_All_2P_FungusVsWound <-
sigres_All_2P_FungusVsWound[which(sigres_All_2P_FungusVsWound$log2F
oldChange>=2),]

Up_sigres_All_2P_FungusVsWound["reg"] <- "up"

down_sigres_All_2P_FungusVsWound <-
sigres_All_2P_FungusVsWound[which(sigres_All_2P_FungusVsWound$log2F
oldChange<=-2),]

down_sigres_All_2P_FungusVsWound["reg"] <- "down"

DEGs_All_2P_FungusVsWound <- rbind(Up_sigres_All_2P_FungusVsWound,
down_sigres_All_2P_FungusVsWound)

write.csv(DEGs_All_2P_FungusVsWound[order(DEGs_All_2P_FungusVsWoun
d$log2FoldChange),],
```

```
file="M018_DESeq2_All_2P_FungusVsWound_sig_DEGs.csv",
row.names=FALSE)


sigres_Fungus_2P_HighNvsLowN <-
res_Fungus_2P_HighNvsLowN[which(res_Fungus_2P_HighNvsLowN$padj<0.
01),]

Up_sigres_Fungus_2P_HighNvsLowN <-
sigres_Fungus_2P_HighNvsLowN[which(sigres_Fungus_2P_HighNvsLowN$lo
g2FoldChange>=2),]

Up_sigres_Fungus_2P_HighNvsLowN["reg"] <- "up"

down_sigres_Fungus_2P_HighNvsLowN <-
sigres_Fungus_2P_HighNvsLowN[which(sigres_Fungus_2P_HighNvsLowN$lo
g2FoldChange<=-2),]

down_sigres_Fungus_2P_HighNvsLowN["reg"] <- "down"

DEGs_Fungus_2P_HighNvsLowN <-
rbind(Up_sigres_Fungus_2P_HighNvsLowN,
down_sigres_Fungus_2P_HighNvsLowN)

write.csv(DEGs_Fungus_2P_HighNvsLowN[order(DEGs_Fungus_2P_HighNvs
LowN$log2FoldChange),],
file="M018_DESeq2_Fungus_2P_HighNvsLowN_sig_DEGs.csv",
row.names=FALSE)


sigres_Wound_2P_HighNvsLowN <-
res_Wound_2P_HighNvsLowN[which(res_Wound_2P_HighNvsLowN$padj<0.
01),]

Up_sigres_Wound_2P_HighNvsLowN <-
sigres_Wound_2P_HighNvsLowN[which(sigres_Wound_2P_HighNvsLowN$lo
g2FoldChange>=2),]

Up_sigres_Wound_2P_HighNvsLowN["reg"] <- "up"

down_sigres_Wound_2P_HighNvsLowN <-
sigres_Wound_2P_HighNvsLowN[which(sigres_Wound_2P_HighNvsLowN$lo
g2FoldChange<=-2),]

down_sigres_Wound_2P_HighNvsLowN["reg"] <- "down"
```

```
DEGs_Wound_2P_HighNvsLowN <-
rbind(Up_sigres_Wound_2P_HighNvsLowN,
down_sigres_Wound_2P_HighNvsLowN)

write.csv(DEGs_Wound_2P_HighNvsLowN[order(DEGs_Wound_2P_HighNvs
LowN$log2FoldChange),],
file="M018_DESeq2_Wound_2P_HighNvsLowN_sig_DEGs.csv",
row.names=FALSE)


sigres_HighN_2P_FungusVsWound <-
res_HighN_2P_FungusVsWound[which(res_HighN_2P_FungusVsWound$padj
<0.01),]

Up_sigres_HighN_2P_FungusVsWound <-
sigres_HighN_2P_FungusVsWound[which(sigres_HighN_2P_FungusVsWound
$log2FoldChange>=2),]

Up_sigres_HighN_2P_FungusVsWound["reg"] <- "up"

down_sigres_HighN_2P_FungusVsWound <-
sigres_HighN_2P_FungusVsWound[which(sigres_HighN_2P_FungusVsWound
$log2FoldChange<=-2),]

down_sigres_HighN_2P_FungusVsWound["reg"] <- "down"

DEGs_HighN_2P_FungusVsWound <-
rbind(Up_sigres_HighN_2P_FungusVsWound,
down_sigres_HighN_2P_FungusVsWound)

write.csv(DEGs_HighN_2P_FungusVsWound[order(DEGs_HighN_2P_Fungus
VsWound$log2FoldChange),],
file="M018_DESeq2_HighN_2P_FungusVsWound_sig_DEGs.csv",
row.names=FALSE)


sigres_LowN_2P_FungusVsWound <-
res_LowN_2P_FungusVsWound[which(res_LowN_2P_FungusVsWound$padj
<0.01),]

Up_sigres_LowN_2P_FungusVsWound <-
sigres_LowN_2P_FungusVsWound[which(sigres_LowN_2P_FungusVsWound$
log2FoldChange>=2),]

Up_sigres_LowN_2P_FungusVsWound["reg"] <- "up"
```

```
down_sigres_LowN_2P_FungusVsWound <-
sigres_LowN_2P_FungusVsWound[which(sigres_LowN_2P_FungusVsWound$
log2FoldChange<=-2),]

down_sigres_LowN_2P_FungusVsWound["reg"] <- "down"

DEGs_LowN_2P_FungusVsWound <-
rbind(Up_sigres_LowN_2P_FungusVsWound,
down_sigres_LowN_2P_FungusVsWound)

write.csv(DEGs_LowN_2P_FungusVsWound[order(DEGs_LowN_2P_FungusV
sWound$log2FoldChange),],
file="M018_DESeq2_LowN_2P_FungusVsWound_sig_DEGs.csv",
row.names=FALSE)


save.image(file="M018_DESeq2_2P.RDATA") # Save R environment.
```

M018_WGCNA.R

```
# Install and load the DESeq2 and WGCNA packages:

install.packages("BiocManager")

BiocManager::install(c("DESeq2", “WGCNA”))

library("WGCNA")

library("DESeq2")


options(stringsAsFactors=FALSE) # This is required for WGCNA.


# Allow multi-threading within WGCNA: This helps speed up certain
calculations.

allowWGCNAThreads()


# ~~~~~Input, inspect and prepare expression data~~~~~


# Load and inspect data: Expression data produced by featureCounts (Subread R
package v1.34.7). Counts were filtered by edgeR v3.20.9 using a 1 count per million
(cpm) in at least 4 libraries cutoff.

Exp_data <- read.csv("M018_edgeR_counts_1cpm.csv", header=TRUE) #  Note
that the data were alphabetized by GeneID prior to being imported. This order is
maintained throughout.

dim(Exp_data) # returns number of genes and number of samples

names(Exp_data) # returns name of samples


# Transforms and normalize the 1 cpm filtered expression data using a variance
stabilizing transformation, a DESq2 function: This is suggested by the creators of
WGCNA                                                                  here:
https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGC
NA/faq.html.
```

```
Exp_data_matrix <- data.matrix(Exp_data[,-1]) # Create matrix of integers for transformation.

Exp_data_matrix_1 <- Exp_data_matrix + 1 # Transformed data will be on a log base two scale where 0 values are not permitted.

Exp_data_norm <- varianceStabilizingTransformation(Exp_data_matrix_1)

Exp_data_norm <- as.data.frame(Exp_data_norm) # Turn matrix back into data frame.

Exp_data_norm <- cbind(Exp_data[,1], Exp_data_norm) # Reintroduce GeneIDs to transformed (normalized) data set.

colnames(Exp_data_norm)[1] <- "GeneID"


# Transpose normalized data:

Exp_data_0 <- as.data.frame(t(Exp_data_norm[,-1])) # Note that duplicate GeneIDs are given a tag (.1, .2, .3, etc.) when the data frame is transposed. We do not want this, so We are excluding the GeneID column at this time.

names(Exp_data_0) <- Exp_data_norm[,1] # Make GeneIDs column names.

rownames(Exp_data_0) <- names(Exp_data_norm)[-1]


# The function "goodSampleGenes" checks for missing entries, entries with weights below a threshold, and zero-variance genes, and returns a list of samples and genes that pass criteria on maximum number of missing or low weight values:

gsg <- goodSamplesGenes(Exp_data_0, verbose=3)

gsg$allOK # This returns TRUE, so all genes pass inspection.


# Hierarchical clustering using the average linkage method is used to detect outlier samples:

Exp_data_Tree <- Exp_data_0

rownames(Exp_data_Tree) <- c("HighN.Fungus.2P.1", "HighN.Fungus.2P.2", "HighN.Fungus.2P.3", "HighN.Fungus.2P.4", "HighN.Fungus.2X.1", "HighN.Fungus.2X.2", "HighN.Fungus.2X.3", "HighN.Fungus.2X.4", "HighN.Wound.2P.1", "HighN.Wound.2P.2", "HighN.Wound.2P.3",
```

```
"HighN.Wound.2P.4", "HighN.Wound.2X.1", "HighN.Wound.2X.2",
"HighN.Wound.2X.3", "HighN.Wound.2X.4", "LowN.Fungus.2P.1",
"LowN.Fungus.2P.2", "LowN.Fungus.2P.3", "LowN.Fungus.2P.4",
"LowN.Fungus.2X.1", "LowN.Fungus.2X.2", "LowN.Fungus.2X.3",
"LowN.Fungus.2X.4", "LowN.Wound.2P.1", "LowN.Wound.2P.2",
"LowN.Wound.2P.3", "LowN.Wound.2P.4", "LowN.Wound.2X.1",
"LowN.Wound.2X.2", "LowN.Wound.2X.3", "LowN.Wound.2X.4")

Tree <- hclust(dist(Exp_data_Tree), method="average")

tiff(filename="M018 WGCNA clustering for the detection of outliers.tiff",
res=1200, width=9, height=5.75, units="in")

par(cex=0.6)

par(mar=c(0, 4, 2, 0))

plot(Tree, main="Hierarchical Clustering of Samples", sub="", xlab="",
cex.lab=1.5, cex.axis=1.5, cex.main=2)

abline(h=7e05, col="red")

dev.off() # Sample M018.17.06.2X is an outlier. It will not be included in the
network analysis and must be removed.


Exp_data_raw_no_outlier <- Exp_data_norm[,-17] # M018.17.06.2X is the 17th
column in Exp_data_norm.

names(Exp_data_raw_no_outlier) # Verify that the correct sample was
removed.


# Transpose data without outliers:

Exp_data_1 <- as.data.frame(t(Exp_data_raw_no_outlier[,-1]))

names(Exp_data_1) <- Exp_data_raw_no_outlier[,1]

rownames(Exp_data_1) <- names(Exp_data_raw_no_outlier)[-1]


# Check genes again using the function "goodSampleGenes":

gsg_no_outlier <- goodSamplesGenes(Exp_data_1, verbose=3)
```

gsg_no_outlier$allOK # Returns TRUE, and we are ready to continue with network construction.

~~~~~Network Construction~~~~~

Pick a soft-threshold power: The soft-threshold is a value used to power the gene-to-gene correlation values. The assumption is that raising the correlation values to a soft-threshold power will reduce statistical noise. We will use two different techniques to choose the soft-threshold from the set, the Scale-free topology fit and the Mean connectivity. We will choose the soft-threshold as a trade-off between these two techniques.

powers <- c(1:30)

soft_thresh <- pickSoftThreshold(Exp_data_1, dataIsExpr=TRUE, powerVector=powers, verbose=5, corFnc="bicor", corOptions="use='p'")

Scale-free topology fit index as a function of the soft-thresholding power: We want our network to follow an approximate scale-free topology. Scale-free networks of genes are robust in that they correlate highly connected genes in a manner that greatly exceeds the average correlation.

tiff(filename = "M018 WGCNA scale-free topology as a function of the soft-thresholding power.tiff", res=1200, width=9, height=5.75, units="in")

plot(soft_thresh$fitIndices[,1], -sign(soft_thresh$fitIndices[,3])*soft_thresh$fitIndices[,2], xlab="Soft Threshold", ylab="Scale Free Topology Model Fit (Signed R^2)", type="n")

text(soft_thresh$fitIndices[,1], -sign(soft_thresh$fitIndices[,3])*soft_thresh$fitIndices[,2], labels=powers, cex=0.9, col="red")

abline(h=0.90, col="red") # This line corresponds to using an R^2 cutoff of 0.90. R^2 values provide a goodness of fit statistic. We are fitting our soft-thresholding powers to a hypothetical scale-free network.

dev.off() # 15 was the first soft-thresholding power to exceed the R^2 cutoff.

Mean connectivity as a function of the soft-thresholding power: We want highly correlated genes to be connected and fall into the same modules. We will choose a soft-thresholding power that will ensure that highly correlated genes are connected in a manner that greatly exceeds the mean connectivity.

tiff(filename = "M018 WGCNA mean connectivity as a function of the soft-thresholding power.tiff", res=1200, width=9, height=5.75, units="in")

plot(soft_thresh$fitIndices[,1], soft_thresh$fitIndices[,5], xlab="Soft Threshold", ylab="Mean Connectivity", type="n")

text(soft_thresh$fitIndices[,1], soft_thresh$fitIndices[,5], labels=powers, cex=0.9, col="red")

dev.off() # An "elbow" exists at 7, which is indicative of a good soft-thresholding power choice.

We will choose 15 as our soft-threshold power for network construction as we are assuming a scale-free topology.

Generate topological overlap map (TOM) and identify expression modules: Pair-wise correlation (connectivity) values between genes are calculated using the biweight midcorrelation function with a 10% maximum outlier detection parameter. The Pearson correlation value is employed for any genes that have a median absolute deviation of zero, for which the biweight midcorrelation function fails. Correlation values are then weighted by raising them to the soft-threshold power of 15 in order to achieve a scale-free topology. The TOM is then constructed to calculate the interconnectedness between correlated genes proportional to the number of neighbors shared by a pair of genes. A signed network is employed to differentiate between positively correlated genes that are strongly connected, and negatively correlated genes that are not strongly connected. Module membership is calculated by correlating the gene and its associated module eigengene using the biweight midcorrelation function with average sensitivity (deepsplit=2) and p-value ratio of 1E-8 for reassigning genes to a closer module.

TOM <- blockwiseModules(Exp_data_1, maxBlockSize=70000, minBlockSize=0, corType="bicor", maxPOutliers=0.10, pearsonFallback="individual", power=15, networkType="signed", TOMType="signed", reassignThreshold=1E-8, minModuleSize=50, mergeCutHeight=0, deepsplit=2, numericLabels=TRUE, verbose=4, saveTOMs=TRUE, saveTOMFileBase="TOM") # TOM will be saved as "TOM-

block.1.RData". Note only one "block" was formed because the "maxBlockSize" exceeded the number of genes used for network construction. Multiple blocks mean multiple networks, and we just want one.

Label modules using colors:

moduleColors <- labels2colors(TOM$colors, zeroIsGrey=TRUE, colorSeq=c("black", "yellow", "darkred", "violet", "orange", "blue", "saddlebrown", "tan", "magenta", "pink", "darkgoldenrod", "deepskyblue", "darkcyan", "red", "darkmagenta", "steelblue", "purple", "cyan", "slateblue", "darkslateblue", "deeppink", "yellowgreen"))

table(moduleColors) # See the number of modules and their respective sizes. The grey color is reserved for unassigned genes.

Calculate module eigengenes: Module eigengenes represent the expression profile of their respective module. They are the first principal components of the modules.

MEs <- TOM$MEs

MEs0 <- moduleEigengenes(Exp_data_1, moduleColors)$eigengenes

MEs <- orderMEs(MEs0)

~~~~~Visualize results~~~~~

Plot the interconnectedness dendrogram and the module colors underneath: Densely interconnected genes will be clustered into modules and displayed in an interconnectedness dendrogram, where genes with the highest intramodular connectivity were located at the tip of each module branch. Note that modules are subnetworks in and of themselves. Intramodular connectivity is calculated by taking the sum of the pairwise biweight midcorrelation values (edge weights) of the gene to its module counterparts.

tiff(filename="M018 WGCNA module dendrogram new.tiff", res=1200, width=9, height=5.75, units="in")

```
plotDendroAndColors(TOM$dendrograms[[1]], main=paste("M018 cluster dendrogram with modules"), colors=moduleColors, groupLabels="Dynamic Tree Cut", hang=0.03, dendroLabels=FALSE, addGuide=TRUE, guideHang=0.05)

dev.off()


# Plot the eigengene heatmap matrix: Modules are related pair-wise. Similar modules (positive/red side of correlation spectrum) and dissimilar modules (negative/blue side of correlation spectrum) are showcased.

tiff(filename = "M018 WGCNA eigengene correlation heatmap new.tiff", res=1200, width=6.75, height=5.75, units="in")

plotEigengeneNetworks(MEs, "Module Eigengene Correlation Heatmap", plotHeatmaps=TRUE, plotDendrograms=FALSE, plotAdjacency=FALSE, excludeGrey=TRUE, marHeatmap=c(6.5, 8, 3, 2), xSymbols=c("slate blue", "orange", "magenta", "violet", "dark magenta", "pink", "blue", "yellow", "dark cyan", "deep sky blue", "red", "dark goldenrod", "black", "dark red", "deep pink", "yellow-green", "purple", "cyan", "dark slate blue", "steel blue", "saddle brown", "tan"), ySymbols=c("slate blue", "orange", "magenta", "violet",  "dark magenta", "pink", "blue", "yellow", "dark cyan", "deep sky blue", "red", "dark goldenrod", "black", "dark red", "deep pink", "yellow-green", "purple", "cyan", "dark slate blue", "steel blue", "saddle brown", "tan"))

dev.off()


# Plot the eigengene dendrogram: Modules are clustered by eigengene correlation using average linkage. Highly correlated eigengenes represent modules with similar gene co-expression patterns.

tiff(filename = "M018 WGCNA eigengene correlation dendrogram new.tiff", res=1200, width=9, height=5.75, units="in")

plotEigengeneNetworks(MEs, "M018 eigengene correlation dendrogram", plotDendrograms=TRUE, plotHeatmaps=FALSE, plotAdjacency=FALSE, excludeGrey=TRUE, marDendro=c(1, 4, 4, 1))

abline(h=0.65, col="red") # This line corresponds to a discorrelation cutoff of 0.65. All MEs with a pairwise correlation <0.35 will be preserved.

dev.off()
```

Given the "messiness" of the original module dendrogram figure, the modules will be merged until a more clarified module figure is produced (i.e. until larger unbroken modules are obtained). The "cutHeight" parameter was optimized by checking values from 0.3 to 0.65 and looking at module dendrogram, ME heatmap and ME dendrogram figures. We wanted good separation between modules without over merging.

merge <- mergeCloseModules(Exp_data_1, moduleColors, cutHeight=0.65, verbose=4)

mergedColors <- merge$colors

table(mergedColors)

mergedMEs <- merge$newMEs # Extract new MEs

newColors <- data.frame(moduleColors, mergedColors)[TOM$blockGenes[[1]],] # Create two module color panels for plotting, one for the original unmerged colors and one for the merged colors.

Add differential expression logFC panels below the interconnectedness dendrogram module colors panel for all FungusVsWound comparisons: logFC values were generated by edgeR v3.20.9. Note that the differential expression data were alphabetized by GeneID prior to being imported to ensure that all GeneIDs properly aligned in the resulting figure.

DEGs_LowN_2P_WoundVsFungus <- read.csv(file="M018_edgeR_LowN_2P_FungusVsWound_DE_results.csv", header=TRUE)

LowN_2P_WoundVsFungus <- DEGs_LowN_2P_WoundVsFungus[,c(1, 2)]

LowN_2P_WoundVsFungus_colors <- numbers2colors(LowN_2P_WoundVsFungus$logFC, colors=blueWhiteRed(200))

DEGs_HighN_2P_WoundVsFungus <- read.csv(file="M018_edgeR_HighN_2P_FungusVsWound_DE_results.csv", header=TRUE)

```
HighN_2P_WoundVsFungus <- DEGs_HighN_2P_WoundVsFungus[,c(1, 2)]

HighN_2P_WoundVsFungus_colors <-
numbers2colors(HighN_2P_WoundVsFungus$logFC,
colors=blueWhiteRed(200))


DEGs_LowN_2X_WoundVsFungus <-
read.csv(file="M018_edgeR_LowN_2X_FungusVsWound_DE_results.csv",
header=TRUE)

LowN_2X_WoundVsFungus <- DEGs_LowN_2X_WoundVsFungus[,c(1, 2)]

LowN_2X_WoundVsFungus_colors <-
numbers2colors(LowN_2X_WoundVsFungus$logFC,
colors=blueWhiteRed(200))


DEGs_HighN_2X_WoundVsFungus <-
read.csv(file="M018_edgeR_HighN_2X_FungusVsWound_DE_results.csv",
header=TRUE)

HighN_2X_WoundVsFungus <- DEGs_HighN_2X_WoundVsFungus[,c(1, 2)]

HighN_2X_WoundVsFungus_colors <-
numbers2colors(HighN_2X_WoundVsFungus$logFC,
colors=blueWhiteRed(200))


dataColors <- data.frame(newColors, LowN_2P_WoundVsFungus_colors,
HighN_2P_WoundVsFungus_colors, LowN_2X_WoundVsFungus_colors,
HighN_2X_WoundVsFungus_colors)[TOM$blockGenes[[1]],]


tiff(filename="M018 WGCNA module dendrogram merged colors with
differential expression data.tiff", res=1200, width=9, height=5.75, units="in")

plotDendroAndColors(TOM$dendrograms[[1]], main=paste("Gene Expression
Correlation Dendrogram \nwith Module Colors and Differential Expression
Data"), colors=dataColors, groupLabels=c("Dynamic Tree Cut Modules",
"Merged Modules", "Low N 2P Fungus vs Wound", "High N 2P Fungus vs
Wound", "Low N 2X Fungus vs Wound", "High N 2X Fungus vs Wound"),
hang=0.03, dendroLabels=FALSE, addGuide=TRUE, guideHang=0.05,
marAll=c(1, 9, 4, 1))
```

```
dev.off()


# ~~~~~~Export module gene lists~~~~~~

# Create "for" loop to generate gene lists for each color in mergedColors (ie for each module), and export in txt format:

for (color in mergedColors) {

  inMod <- is.finite(match(mergedColors, color))

  gene.list <- network_genes[inMod,]

  write.table(gene.list, file=paste("M018_WGCNA_module_", color, ".txt", sep=""),   row.names=FALSE, col.names="GeneID", quote=FALSE)

}


# ~~~~~Hub gene analysis~~~~~

# Hub genes are highly connected to other genes in their module and are assumed to be involved in gene regulation. Therefore, we are looking for genes with high intramodular connectivity (top 10% most connected genes) plus significant module membership (>0.80).


# We will annotate the hub genes for data mining purposes using the master anotation file "lodgeAntn_v2.csv".

lodgeAntn <- read.csv(file="lodgeAntn_v2.csv", header=TRUE)

colnames(lodgeAntn)[1] <- "GeneID"

M018_all_FvW_DEG_results_antd <- merge(M018_all_FvW_DEG_results, lodgeAntn, by="GeneID", all.x=TRUE)


# Calculate connectivity for each gene: The function "intramodularConnectivity.fromExpr" calculates the connectivity of nodes to other nodes within the same module. The function uses the same soft-threshold parameter and correlation function utilized for network construction, along with the module colors ("mergedColors") as input.
```

```
intramodular_connectivity <- intramodularConnectivity.fromExpr(Exp_data_1, mergedColors, corFnc="bicor", corOptions="use='p'", power=15, networkType="signed")

intramodular_connectivity <- cbind(Exp_data_raw_no_outlier[,1], intramodular_connectivity)

colnames(intramodular_connectivity)[1] <- "GeneID"


# Extract connectivity results for the 4 largest modules, and choose the top 10% most connected genes:

orange_connectivity <- merge(orange_module_GeneID, intramodular_connectivity, by="GeneID", all.x=TRUE, all.y=FALSE)

orange_connectivity <- orange_connectivity[order(-orange_connectivity$kWithin),] # Order genes in desecnding order by intramodular connectivity (kwithin)

orange_top_connected <- as.data.frame(orange_connectivity[1:147,]) # Choose the 147 most connected genes (10% of orange module genes).

colnames(orange_top_connected)[1] <- "GeneID"


darkmagenta_connectivity <- merge(darkmagenta_module_GeneID, intramodular_connectivity, by="GeneID", all.x=TRUE, all.y=FALSE)

darkmagenta_connectivity <- darkmagenta_connectivity[order(-darkmagenta_connectivity$kWithin),]

darkmagenta_top_connected <- as.data.frame(darkmagenta_connectivity[1:168,]) # Top 168 most connected genes (10% of darkmagenta module genes)

colnames(darkmagenta_top_connected)[1] <- "GeneID"


darkslateblue_connectivity <- merge(darkslateblue_module_GeneID, intramodular_connectivity, by="GeneID", all.x=TRUE, all.y=FALSE)

darkslateblue_connectivity <- darkslateblue_connectivity[order(-darkslateblue_connectivity$kWithin),]
```

```
darkslateblue_top_connected <-
as.data.frame(darkslateblue_connectivity[1:67,]) # Top 67 most connected genes
(10% of darkslateblue module genes)

colnames(darkslateblue_top_connected)[1] <- "GeneID"


darkcyan_connectivity <- merge(darkcyan_module_GeneID,
intramodular_connectivity, by="GeneID", all.x=TRUE, all.y=FALSE)

darkcyan_connectivity <- darkcyan_connectivity[order(-
darkcyan_connectivity$kWithin),]

darkcyan_top_connected <- as.data.frame(darkcyan_connectivity[1:209,]) #
Top 209 most connected genes (10% of darkcyan module genes)

colnames(darkcyan_top_connected)[1] <- "GeneID"


# Calculate module membership for all genes: Calculation of (signed) eigengene-
based connectivity, also known as module membership (MM), is performed with
the WGCNA function "signedKME". This function produces a data frame in which
rows correspond to input genes and columns to module eigengenes, giving the
signed eigengene-based connectivity of each gene with respect to each eigengene.

module_membership <- signedKME(Exp_data_1, mergedMEs,
outputColumnName="MM.")

module_membership <- cbind(Exp_data_raw_no_outlier[,1],
module_membership)

colnames(module_membership)[1] <- "GeneID"


# Extract MM results, and filter genes using MM>0.80 to obtain hub gene set:

orange_MM <- module_membership[,c(1,4)]

orange_top_connected_MM <- merge(orange_top_connected, orange_MM,
by="GeneID", all.x=TRUE, all.y=FALSE)

orange_hub_genes <-
orange_top_connected_MM[which(orange_top_connected_MM$MM.orange>
0.8),c(1,3,6)]
```

```
orange_hub_genes <- merge(orange_hub_genes,
M018_all_FvW_DEG_results_antd, by="GeneID", all.x=TRUE, all.y=FALSE)

colnames(orange_hub_genes)[1:27] <- c("GeneID",
"orange_intramodular_connectivity", "orange_module_membership",
"DE_comparison", "logFC", "logCPM", "LR", "PValue", "PValue.adjust",
"DE_comparison", "logFC", "logCPM", "LR", "PValue", "PValue.adjust",
"DE_comparison", "logFC", "logCPM",  "LR", "PValue", "PValue.adjust",
"DE_comparison", "logFC", "logCPM", "LR", "PValue", "PValue.adjust")

orange_hub_genes <- merge(orange_hub_genes, WGCNA_categories,
by="GeneID", all.x=TRUE, all.y=FALSE)

write.csv(orange_hub_genes, file="M018_WGCNA_orange_hub_genes.csv",
row.names=FALSE)


darkmagenta_MM <- module_membership[,c(1,3)]

darkmagenta_top_connected_MM <- merge(darkmagenta_top_connected,
darkmagenta_MM, by="GeneID", all.x=TRUE, all.y=FALSE)

darkmagenta_hub_genes <-
darkmagenta_top_connected_MM[which(darkmagenta_top_connected_MM$
MM.darkmagenta>0.8),c(1,3,6)]

darkmagenta_hub_genes <- merge(darkmagenta_hub_genes,
M018_all_FvW_DEG_results_antd, by="GeneID", all.x=TRUE, all.y=FALSE)

colnames(darkmagenta_hub_genes)[1:27] <- c("GeneID",
"darkmagenta_intramodular_connectivity",
"darkmagenta_module_membership", "DE_comparison", "logFC", "logCPM",
"LR", "PValue", "PValue.adjust", "DE_comparison", "logFC", "logCPM", "LR",
"PValue", "PValue.adjust", "DE_comparison", "logFC", "logCPM", "LR",
"PValue", "PValue.adjust", "DE_comparison", "logFC", "logCPM", "LR",
"PValue", "PValue.adjust")

darkmagenta_hub_genes <- merge(darkmagenta_hub_genes,
WGCNA_categories, by="GeneID", all.x=TRUE, all.y=FALSE)

write.csv(darkmagenta_hub_genes,
file="M018_WGCNA_darkmagenta_hub_genes.csv", row.names=FALSE)


darkslateblue_MM <- module_membership[,c(1,10)]
```

```
darkslateblue_top_connected_MM <- merge(darkslateblue_top_connected,
darkslateblue_MM, by="GeneID", all.x=TRUE, all.y=FALSE)

darkslateblue_hub_genes <-
darkslateblue_top_connected_MM[which(darkslateblue_top_connected_MM$
MM.darkslateblue>0.8),c(1,3,6)]

darkslateblue_hub_genes <- merge(darkslateblue_hub_genes,
M018_all_FvW_DEG_results_antd, by="GeneID", all.x=TRUE, all.y=FALSE)

colnames(darkslateblue_hub_genes)[1:27] <- c("GeneID",
"darkslateblue_intramodular_connectivity",
"darkslateblue_module_membership", "DE_comparison", "logFC", "logCPM",
"LR", "PValue", "PValue.adjust", "DE_comparison", "logFC", "logCPM", "LR",
"PValue", "PValue.adjust", "DE_comparison", "logFC", "logCPM", "LR",
"PValue", "PValue.adjust", "DE_comparison", "logFC", "logCPM", "LR", PValue",
"PValue.adjust")

darkslateblue_hub_genes <- merge(darkslateblue_hub_genes,
WGCNA_categories, by="GeneID", all.x=TRUE, all.y=FALSE)

write.csv(darkslateblue_hub_genes,
file="M018_WGCNA_darkslateblue_hub_genes.csv", row.names=FALSE)


darkcyan_MM <- module_membership[,c(1,9)]

darkcyan_top_connected_MM <- merge(darkcyan_top_connected,
darkcyan_MM, by="GeneID", all.x=TRUE, all.y=FALSE)

darkcyan_hub_genes <-
darkcyan_top_connected_MM[which(darkcyan_top_connected_MM$MM.dark
cyan>0.8),c(1,3,6)]

darkcyan_hub_genes <- merge(darkcyan_hub_genes,
M018_all_FvW_DEG_results_antd, by="GeneID", all.x=TRUE, all.y=FALSE)

colnames(darkcyan_hub_genes)[1:27] <- c("GeneID",
"darkcyan_intramodular_connectivity", "darkcyan_module_membership",
"DE_comparison", "logFC", "logCPM", "LR",  "PValue", "PValue.adjust",
"DE_comparison", "logFC", "logCPM", "LR", "PValue", "PValue.adjust",
"DE_comparison", "logFC", "logCPM", "LR", "PValue", "PValue.adjust",
"DE_comparison", "logFC", "logCPM", "LR", "PValue", "PValue.adjust")

darkcyan_hub_genes <- merge(darkcyan_hub_genes, WGCNA_categories,
by="GeneID", all.x=TRUE, all.y=FALSE)
```

```
write.csv(darkcyan_hub_genes,
file="M018_WGCNA_darkcyan_hub_genes.csv", row.names=FALSE)


# ~~~~~Export to Cytoscape~~~~~

# The WGCNA function "exportNetworkToCytoscape" is used to export the edge file for the most connected genes in the largest color-coded module, orange, using a network threshold of 0.98 (out of 1).


# Load and prepare data:

TOM <- load(TOM-block.1.RData) # Note that WGCNA function “blockwiseModules” produces a dissimilarity TOM.

TOM_matrix <- as.matrix(TOM) # Convert TOM to matrix for Cytoscape export

TOM_matrix <- 1 - TOM_matrix # Since the TOM generated was a dissimilarity TOM, we subtract TOM_matrix from 1 to obtain the similarity TOM. We will overwrite TOM_matrix to save space.

probes <- as.data.frame(names(Exp_data_1)) # All genes that went into network construction in the original, correct order.

orange_module_GeneID <-
read.table(file="M018_WGCNA_module_orange.txt", sep="\t", header=TRUE)
# Load orange module genes produced with WGCNA (not necessarily in the correct order).

M018_all_sig_DEG_categories <-
read.csv(file="M018_All_DEG_categories_WGCNA.csv", header=TRUE) # Load DEG annotation data, including categories.

mergedColors <- mergedColors[,1] # Change mergedColors from dataframe into character vector.


# Select module probes and extract alternative node names (gene categories):

orange_categories <- merge(orange_module_GeneID,
M018_all_sig_DEG_categories, by="GeneID", all.x=TRUE) # Orange gene categories will be used as alternative node names.

inOrange <- is.finite(match(mergedColors, "orange"))
```

```
orangeProbes <- probes[inOrange,] # Orange module genes in the correct order.

orangeCategories <- orange_categories[,11][match(orangeProbes,
orange_categories[,1])] # This makes sure that categories are in same order as
the genes.


# Select the corresponding similarity topological overlap matrix:

orangeTOM <- TOM_matrix[inOrange, inOrange]

dimnames(orangeTOM) <- list(orangeProbes, orangeProbes)


# Export the orange network as an edge file that Cytoscape can read:

orangeCytoscape <- exportNetworkToCytoscape(orangeTOM, edgeFile =
"CytoscapeInput-edges-orange_0.98.txt", weighted = TRUE, threshold = 0.98,
nodeNames = orangeProbes, altNodeNames = orangeCategories)


save.image(file='M018_WGCNA.RData') # Save R environment.
```

References

Anders, S., Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, 11, 1-12.

Anders, S., McCarthy, D. J., Chen, Y. S., Okoniewski, M., Smyth, G. K., Huber, W., Robinson, M. D. (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature Protocols*, 8, 1765-1786.

Apweiler, R. Attwood, T. K., Bairoch A., Bateman A., Birney E., Biswas M., Bucher P., Cerutti L., Corpet F., Croning M. D., Durbin R., Falquet L., Fleischmann W., Gouzy J., Hermjakob H., Hulo N., Jonassen I., Kahn D., Kanapin A., Karavidopoulou Y., Lopez R., Marx B., Mulder N. J., Oinn T. M., Pagni M., Servant F., Sigrist C. J., Zdobnov E. M. (2001). The InterPro database, An integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Research,* 29, 37-40.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *Series B* (*Methodological*), 57, 289-300.

Chen, Y., McCarthy, D., Ritchie, M., Robinson, M., Smyth, G. K. (2017, Dec 26). EdgeR: Differential expression analysis of digital gene expression data: User's Guide. Retrieved from *https://www.bioconductor.org/packages/devel/bioc /vignettes/edgeR/inst/doc/edgeRUsersGuide.pdf*

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D. L., Garcia-Hernadez, M., Karthikeyan, A. S., Lee, C. H., Nelson, W. D., Ploetz, L., Singh, S., Wensel, A., Huala, E. (2011). The *Arabidopsis* Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Research*, 40, D1202-D1210.

Lohse, M., Nagel, A., Herter, T., May, P., Schroda, M., Zrenner, R., Tohge, T., Fernie, A. R., Stitt, M., Usadel, B. (2014). Mercator: A fast and simple web server

for genome scale functional annotation of plant sequence data. *Plant, Cell and Environment*, 37, 1250–1258.

McCarthy, D. J., Chen, Y., Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, 40, 4288-4297.

Pruitt, K. D., Tatusova, T., Maglott, D. R. (2004). NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 33, D501–D504.

Robinson, M. D., Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11, PAGES.

Sundell, D., Mannapperuma, C., Netotea, S., Delhomme, N., Lin, Y., Sjödin, A., de Peer, Y. V., Jansson, S., Hvidsten, T. R., Street, N. R. (2015). The plant genome integrative explorer resource: PlantGenIE.org. *New Phytologist*, 208, 1149-1156.

The Gene Ontology Consortium (2015). Gene ontology consortium: Going forward. *Nucleic Acids Research*, 43, D1049-D1056.