2020

Trees, Fungi, Insects: How Host Plant Genetics Builds a Community

Sandra Jeanne Simon

West Virginia University, ssimon4@mix.wvu.edu

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Part of the [Entomology Commons](#), [Evolution Commons](#), [Genetics Commons](#), [Plant Biology Commons](#), and the [Population Biology Commons](#)

Recommended Citation

Simon, Sandra Jeanne, "Trees, Fungi, Insects: How Host Plant Genetics Builds a Community" (2020). *Graduate Theses, Dissertations, and Problem Reports*. 7779.

<https://researchrepository.wvu.edu/etd/7779>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

2020

Trees, Fungi, Insects: How Host Plant Genetics Builds a Community

Sandra Jeanne Simon

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Part of the [Entomology Commons](#), [Evolution Commons](#), [Genetics Commons](#), [Plant Biology Commons](#), and the [Population Biology Commons](#)

Trees, Fungi, Insects: How Host Plant Genetics Builds a Community

Sandra Jeanne Simon

**Dissertation submitted
to the Eberly College of Arts and Sciences
at West Virginia University
in partial fulfillment of the requirements for
the degree of**

**Doctor of Philosophy
In
Biology**

**Stephen P. DiFazio, Ph.D., Chair
Yong-Lak Park, Ph.D.
Gina M. Wimp, Ph.D.
Jonathan R. Cumming, Ph.D.
Jennifer S. Hawkins, Ph.D.**

**Morgantown, West Virginia
2020**

Keywords: *Populus*, *Salix*, biotic interactions, comparative genomics, tandem duplication, community genetics, floral visitors, *Andrena*, phenolics, volatile organic compounds

Copyright © Sandra Jeanne Simon, 2020. All rights reserved.

Abstract

Trees, Fungi, Insects: How Host Plant Genetics Builds a Community

Sandra Jeanne Simon

Organisms, such as fungi and insects, can cause millions of acres of agricultural and forest damage, while others provide billions of dollars in ecological services such as education, aesthetic enjoyment, pollination, and gardening. Plant breeding and biotechnology can potentially help establish a balance between the proliferation of detrimental pests and attraction of beneficial insects. Variation in plant physiological and morphological characteristics are extremely important in the ability of host tissues to support many different types of organisms. When that variation is genetically heritable in a plant population, shifts in the underlying genes can have predictable consequences in structuring entire ecosystems. The field of community genetics seeks to study these interactions and identify the genes important in host plants, which will ultimately allow for the prediction of community level responses to changing conditions. The main goal of my dissertation was to identify the genetic underpinnings of host plant-biotic community organization in species belonging to the Salicaceae family, which contains many species of trees and shrubs of ecological and economic importance. To date, community genetic research has established the ability of hybrid plants to have wide-ranging heritable effects on communities and ecosystems. However, only a few publications have identified the genes underlying these relationships in pure species.

In chapter 2, I utilized a pseudo-backcross hybrid family of *Populus* and quantitative trait analysis (QTL) as well as genomic comparisons of the *P. trichocarpa* and *P. deltoides* parents to identify potential candidate genes mediating their relationship with several insect herbivores and fungal pathogens. I found that many gene candidates had undergone recent tandem duplication and this pattern was enriched relative to the rest of the genome in the native parent QTL intervals. Additionally, I found the hybrids were mediating interactions between pathogens leading to unique genetic associations that would not normally be observed in a single species population, which may contribute to the elevated community effects that have been previously observed in natural hybrid zones. In chapter 3, I used surveys that were conducted in multiple common gardens of a population of *P. trichocarpa*, genome wide association analysis (GWAS), and networks to identify genes and potential biological functions underlying arthropod community composition. I found that genes associated with individual arthropods appeared to be very functionally targeted with rare variants related to metabolite production and manipulation of tissue nutrition. Genes that associated with arthropod richness and community composition have biological functions that may allow them to more broadly target multiple groups of arthropods, such as terpenoid synthesis, RNA inhibition, and transmembrane protein activity. In chapter 4, I used visual observation and pan-traps to survey the tree species *Salix nigra* and explore the impact of dioecy on the assembly of floral insect communities. I found that male trees supported higher diversity of floral visitors on their catkins when compared to females due to visual cues of yellow pollen. I also identified the main cross-pollinators to be three species of *Andrena* bees, one of which (*A. nigrae*) showed a preference for female flowers and was correlated to specific VOC cues from catkins. Finally, I detected an asynchrony in catkin bloom and insect emergence in early spring that threatens not only the sexual reproduction of *S. nigra* trees, but also the

survival of local *A. nigrae* populations. Overall, I found that the dynamic plant-pathogen-herbivore-pollinator relationships are dependent on combinations of plant genetic effects with spatial and temporal environmental variability.

Acknowledgements

Thank you, Steve, for all your guidance on my project and patience in helping me write this dissertation. I would never have stayed at WVU if not for your support and confidence that I would be successful in science, and I wouldn't have been successful without you ensuring I was always well hydrated in the field and well caffeinated in the office. I also would like to thank my committee members for all your guidance and input. In particular I would like to thank Dr. Wimp for all the phone time you carved out of your schedule, whether it be normal working hours or spring break, to drop everything and talk about statistics, bugs, trees and how to find bugs on trees; Dr. Park for always being available to critique my methods and help me to improve them to capture the best version of an insect community possible for my research (and always reminding me that larvae and nymphs are not the same!); Dr. Cumming for showing me during my undergrad a look at how the experimental process develops (with an equal number of failures and successes) and sharing your knowledge of plant stress responses which helped me to structure my research; and finally Dr. Hawkins for supporting me through the Genomics Core assistantships in my final years, having one of my favorite labs in the department to collaborate with, and reminding me to fill out my defense declaration form which I am doing as I type out these acknowledgements.

I would like to thank my lab including Ran Zhou, Hari Chhetri, David Macaya-Sanz, Roshan Abeyratne, and Julianne Grady. I can't imagine a better group of people to work with in science than all of you. Also, thanks to Ryan Percifield, I learned so much about laboratory technique from you during my time in the Genomics Core it wouldn't be a stretch to say it is one of my most valuable skill sets and an easy one to have sacrificed a lot of chocolate to acquire. Additionally, for all the people that helped me with the field and lab surveys including Dr. Jared LeBoldus, Dr. Luke Evans, Dr. Ken Keefover-Ring, Dr. William MacDonald, Dr. Nesatalu Hiese, Jacob Miller, Margo Folwick, Tanita Cheevaphantusri, and Sunita Mahat.

I want to give a special shout-out to two people that I could talk to for hours about any kind of science. Ashley Henderson you have been a constant joy to work with and an even better friend through the hot mess express that is a doctorate in biology. Brandon Sinn you are easily one of the most unique and talented researchers I have had the opportunity to work closely with, I hope your career continues to inspire you the same way you did me in my last few semesters. Finally, I want to acknowledge Mark Burnham (and Maya) who has been a constant support through the end of my degree. You have helped me turn countless jumbled thoughts into many of the coherent ideas conveyed throughout this document. After my defense, I promise I will wait at least one hour after waking up before I start asking you science questions in the morning. Also, to all my co-authors for the countless hours they have spent helping collect data and edit my research chapters for final publication. Funding for this work came from the Bioenergy Science Corporation (BESC), the Center for Bioenergy Innovation (CBI), the National Science Foundation (NSF) Dimensions of Biodiversity, the Morrissey Ropp Scholarship, the Core Research in Taxonomy of Vascular Plants Scholarship, the Department of Biology WVU genomics Core graduate fellowship, and the WVU Department of Biology and Eberly College of Arts and Sciences.

Table of Contents

Abstract.....	ii
Acknowledgements	iv
Table of Contents	v
List of Tables and Figures	vii
Chapter 1: Introduction: Biotic interactions in the Salicaceae	1
1.1 Community genetics.....	2
1.2 The Salicaceae family as a genetic model.....	3
1.3 Physiological and morphological features	4
1.4 Dioecy and floral visitor communities	5
1.5 Objectives.....	7
1.6 Literature cited	9
Chapter 2: Host plant genetic control of associated fungal and insect species in a <i>Populus</i>	
hybrid	14
2.1 Abstract	15
2.2 Introduction	16
2.3 Methods.....	18
2.4 Statistical analysis	21
2.5 Results	25
2.6 Discussion	28
2.7 Literature cited	38
2.8 Tables and Figures	48
Chapter 3: Genetic underpinnings of arthropod community distributions in <i>Populus</i>	
trichocarpa	65
3.1 Abstract	66
3.2 Introduction	67
3.3 Methods.....	69
3.4 Statistical analysis	70
3.5 Results	73
3.6 Discussion	75
3.7 Literature cited	81
3.8 Tables and Figures	91

Chapter 4: Characterization of <i>Salix nigra</i> floral insect community and activity of three native <i>Andrena</i> bees	111
4.1 Abstract	112
4.2 Introduction	113
4.3 Methods	115
4.4 Statistical analysis	119
4.5 Results	123
4.6 Discussion	126
4.7 Literature cited	132
4.8 Tables and Figures	136
Chapter 5: Conclusions	160
5.1 Overview of research goals	161
5.2 Hybrid mediation of biotic associations.....	161
5.3 <i>P. trichocarpa</i> community genetics	163
5.4 <i>S. nigra</i> floral community	164
5.5 Overall conclusions	165
5.6 Literature cited	167

List of Tables and Figures

Table 2-1. TPS and linear-mixed model output for biotic surveys. Broad-sense heritability (H^2) denotes the contribution of all host plant genetic factors to total variance in the biotic phenotype. R package RLRsim exactRLRT function was used to test significance of effects in mixed model.....	48
Table 2-2. Summary of QTL permutation test output. Percent variance in surveys for insects and fungi explained by significant marker indicated for composite interval mapping models. Positive (+) allele specifies genotype at significant interval that results in an increase in susceptibility. D indicates progeny are homozygous for <i>P. deltooides</i> alleles and T indicates progeny are heterozygous for <i>P. deltooides</i> and <i>P. trichocarpa</i> alleles.	49
Table 2-3. One-way ANOVA, with Genotype as a covariate, analyzing the effect of infection severity of the genus of one fungus on the infection severity of the competing leaf fungi in 2008.....	50
Table 2-4. Number of genes in QTL intervals in parental genomes for biotic associations.....	51
Table 2-5. Candidate genes under positive selection ($K_a/K_s > 1$) in genetic intervals associated with fungi and insects.	52
Table 2-6. Tandem duplication profiles for genetic intervals. Number of copies next to species gene name indicates the size of tandem expansion for the gene.....	53
Table 3-1. Sum of arthropods observed on tree tissue within plantations.	91
Table 3-2. Linear-mixed model output for community surveys across all gardens and years. Broad-sense heritability (H^2) denotes the contribution of all host plant genetic factors to total variance in phenotype.	92
Table 3-3. PERMANOVA results for NMDS configuration within all 2012 gardens with replicated genotype observations and block to account for spatial position.....	93
Table 3-4. PERMANOVA results for NMDS configuration among gardens with replicated genotype observations and block to account for spatial position.	94
Table 3-5. Significant SNPs ($FDR < 0.1$) and nearest <i>P. trichocarpa</i> genes discovered through GWAS for <i>P. populiella</i> trait in Clatskanie garden.	95
Table 3-6. Significant SNPs ($FDR < 0.1$) and nearest <i>P. trichocarpa</i> genes discovered through GWAS for <i>Phyllonorycter</i> sp. trait.	96
Table 3-7. Significant SNPs ($FDR < 0.1$) and closest <i>P. trichocarpa</i> genes discovered through GWAS for Community metric traits.....	97
Table 3-8. Significant SNPs ($FDR < 0.1$) and closest <i>P. trichocarpa</i> genes discovered through GWAS for arthropod community composition multitrait.....	98

Table 3-9. Candidate genes for mediating biotic interactions selected for functional analysis from networks. Breadth indicates number of connected network layers, depth refers to the number of neighbor genes in network connected to anchor gene, and edges indicate the number of connections (co-methylation, co-expression, metabolite GWAS and pyMBMS GWAS) among genes in networks.....	99
Table 3-10. Genome function enrichment analysis of co-expressed neighbor genes from network analysis. Significance of enrichment was determined based on Fisher Exact p-value with multiple testing correction to meet a threshold of FDR < 0.1.	100
Table 4-1. NMDS and ANOSIM (Bray-Curtis dissimilarity) results for multivariate floral visitor composition tested against sex grouping (male vs female). p-values < 0.05 (bolded) indicate that floral visitor composition is more similar within replicate observations of sex group rather than among all observations.	136
Table 4-2. NMDS and ANOSIM (Bray-Curtis dissimilarity) results for multivariate chemistry composition tested against sex grouping (male vs female). p-values < 0.05 (bolded) indicate that chemistry composition is more similar within replicate observations of sex group rather than among all observations.	137
Table 4-3. Mantel test results comparing pairwise Bray-Curtis dissimilarity matrices among insect communities and chemistry composition of flowers and leaves. Bolded p-values and positive rm indicate a significant test, suggesting that similarity in chemistry composition relates to similarity in insect community assemblage.....	138
Table 4-4. Environmental fit results for 2019 floral community NMDS analysis. Bolded p-values (<0.05) indicate a significant correlation of the independent variable with the NMDS configuration.	139
Table 4-5. Nested ANCOVA model results for 2019 analysis of individual native bee abundances. Bolded p-values (<0.05) indicate significant model effects. $y \sim \text{Sex} + \text{Location} + \text{Tree}(\text{Sex}) \& \text{Random} + \text{Julian date}$	140
Table 4-6. Nested ANCOVA model results for 2019 analysis of calculated community metrics. Bolded p-values (<0.05) indicate significant model effects. $y \sim \text{Sex} + \text{Location} + \text{Tree}(\text{Sex}) \& \text{Random} + \text{Julian date}$	141
Table 4-7. Test of random effects p-values extracted from nested ANCOVA for most abundant floral visitors as well as calculated species richness and Shannon Weaver diversity for 2019 survey analysis. Bolded values indicate that the independent variable had a significant effect on the dependent variable (p-value < 0.05).	142
Table 4-8. Environmental fit results for floral community NMDS analysis across years (2017-2019). Bolded p-values (<0.05) indicate a significant correlation of the independent variable with the NMDS configuration.....	143
Table 4-9. Nested ANCOVA model results for year analysis. Bolded p-values (<0.05) indicate significant model effects. $y \sim \text{Sex} + \text{Year} + \text{Tree}(\text{Sex}) \& \text{Random} + \text{Julian date}$	144

Table 4-10. Test of random effects p-values extracted from nested ANCOVA for most abundant floral visitors as well as calculated species richness and Shannon Weaver diversity. Bolded values indicate that the independent variable had a significant effect on the dependent variable (p-value < 0.05).....	145
Figure 2-1. Biotic phenotype symptoms observed on trees including (a) leaf symptoms of <i>Melampsora</i> sp. fungal leaf rust, (b) leaf spot symptoms of the <i>Sphaerulina</i> sp. fungus, (c) <i>S. musiva</i> canker symptoms, (d) branch gall created by the <i>M. vagabunda</i> aphid (e) petiole gall created by the <i>P. populitransversus</i> aphid, and (f) <i>Phyllocolpa</i> sp. leaf folding gall.	58
Figure 2-2. Collinear genes in <i>P. deltoides</i> (Pd) and <i>P. trichocarpa</i> (Pt) chromosomes, based on MCSanX alignments. (a) Chr02; (b) Chr05.	59
Figure 2-3. Change in number of tandemly duplicated genes discovered with increasing window size.	60
Figure 2-4. QTL interval plots showing peaks across the genome that associate with 2008 biotic surveys. Lines on the plots indicate p-value thresholds as determined by running 1000 permutations of mapping model for (a) <i>Melampsora</i> sp. model with all individuals and clones subset to exclude individuals with <i>S. musiva</i> infection and (b) <i>S. musiva</i> leaf spot model with all individuals and clones subset to exclude individuals with <i>Melampsora</i> sp. infection.	61
Figure 2-5. QTL interval plots showing peaks across the genome that associate with biotic surveys. Lines on the plots indicate p-value thresholds as determined by running 1000 permutations of mapping model for (a) binary fungal survey of <i>S. musiva</i> canker, (b) binary insect surveys of <i>M. vagabunda</i> and <i>P. populitransversus</i> , (c) and insect surveys of <i>Phyllocolpa</i> sp. leaf gall counts and overlapping peaks for gentisyl alcohol 5-O-glucoside compound QTL.	62
Figure 2-6. Mean and standard error of fungal infection for individuals with varying category levels of competing fungus infection. Letters indicate significantly different means as determined by Tukey's honest significance test for each one-way ANOVA test.	63
Figure 2-7. Comparison of gene content in <i>P. trichocarpa</i> grandparent 93-968 (left line) and <i>P. deltoides</i> grandparent Ill-101 (right line) for significant genetic intervals for (a) <i>Melampsora</i> sp. chromosome 4, (b) <i>S. musiva</i> chromosome 16, (c) <i>M. vagabunda</i> chromosome 5, (d) <i>Phyllocolpa</i> sp. chromosome 10, and (e) <i>Phyllocolpa</i> sp. chromosome 13 associations. QTL intervals were defined as 1 Mb regions centered on the marker with the highest LOD score. Size of gene point is relative to the number of genes in the tandem duplication expansion.	64
Figure 3-1. Map displaying genotype origin populations and survey sites – circles represent collection locations in the wild, stars represent common gardens, the survey sites.	103
Figure 3-2. Arthropod community NMDS plot depicting 3 dimensions within all gardens surveyed in 2012. Color groupings indicate tree surveyed in specified garden.	104
Figure 3-3. Arthropod community NMDS plots depicting 3 dimensions for (a) 2012 Clatskanie garden survey and (b) 2012 Placerville garden survey. PERMANOVA analysis, run with only genotypes with replicate observations, output indicated for each garden in the upper right corner	

of each plot for each common garden. Size of species text indicates position relative to the dimensional space, for example, larger font indicates forward projection along the positive values of axes.....	105
Figure 3-4. Arthropod community NMDS plots depicting 3 dimensions for (a) 2012 Corvallis garden survey and (b) 2015 Corvallis garden survey. PERMANOVA analysis, run with only genotypes with replicate observations, output indicated for each garden in the upper right corner of each plot for each common garden. Size of species text indicates position relative to the dimensional space, for example, larger font indicates forward projection along the positive values of axes.....	106
Figure 3-5. Manhattan plot output from GWAS for (a) individual arthropods (single-trait; three tests), (b) calculated community metrics for each garden (single-trait; three tests), and (c) community composition (multi-trait; two tests). SNPs that passed the red dotted line were significantly associated with indicated trait. Clatskanie community (c) did not pass threshold but contained multiple suggestive markers.	107
Figure 3-6. Co-expression network for Clatskanie <i>Phyllonorycter</i> sp. single-trait gene candidates. Co-expression genes were grouped based on their biological function as determined by Gene-ontology terms.	108
Figure 3-7. Co-expression network for Corvallis arthropod richness single-trait gene candidates. Co-expression genes were grouped based on their biological function as determined by Gene-ontology terms.	109
Figure 3-8. Co-expression network for Corvallis community composition multi-trait gene candidates. Co-expression genes were grouped based on their biological function as determined by Gene-ontology terms.....	110
Figure 4-1. Map of all thirty <i>Salix nigra</i> tree locations in WVU Core Arboretum. Pin shapes indicate tree selected for floral insect community survey. Points indicate additional individuals in the Arboretum.	149
Figure 4-2. Tree canopy pan-trap design. Camouflaged buckets were added to hang below river traps to add weight, preventing winds from jostling traps.....	150
Figure 4-3. Non-metric multidimensional plot (dimensions = 3; stress = 0.12) of insect floral community with groupings indicated by color for sex of tree (ANOSIM $R = 0.3077$, p -value = 0.001) for 2019 analysis. Size of species text indicates position relative to the dimensional space, for example, larger font indicates forward projection along the positive values of all axes.	151
Figure 4-4. Non-metric multidimensional plot (dimensions = 4; stress = 0.12) of 2017-2019 insect floral community with groupings indicated by color for sex of tree (ANOSIM $R = 0.1232$, p -value = 0.011). Size of species text indicates position relative to the dimensional space, for example, larger font indicates forward projection along the positive values of all axes.	152

Figure 4-5. Total concentration averages of catkin metabolites, leaf metabolites and catkin VOCs for male and female trees. Letters to the left of boxes indicate significantly different means (p-value < 0.05) determined by one-way ANOVA test.	153
Figure 4-6. Total concentration averages of monoterpenes and sesquiterpenes VOCs for male and female trees. Letters to the left of boxes indicate significantly different means (p-value < 0.05) determined by one-way ANOVA test.	154
Figure 4-7. (a) Non-metric multidimensional plot (dimensions = 3; stress = 0.12) of insect floral community with groupings indicated by color for survey method (ANOSIM R = 0.6656, p-value = 0.001). (b) Breakdown of insect capture/observation for two canopy survey methods. Additional pie charts around pan-trap pie chart indicate the percentage of that order captured in different colors of pan-traps.	155
Figure 4-8. Average activity of most common floral visitors and calculated community metrics from 2019 <i>S. nigra</i> surveys. Letters to the left of boxes indicate significantly different means as determined by a Tukey's HSD (p-value < 0.05).	156
Figure 4-9. Correlation plots of <i>A. nigrae</i> with VOC compounds.	157
Figure 4-10. Average values of species richness and Shannon-Weaver diversity for female and male trees from year analysis model. Letters to the left of boxes indicate significantly different means as determined by a Tukey's HSD (p-value < 0.05) for each separate nested ANCOVA model.	158
Figure 4-11. (a) Amount of time and date in each year (2017, 2018, and 2019) in which catkins were actively attractive to insects. End date for male individuals indicates trees have dropped all catkins while female trees no longer have receptive stigmas (brown and shriveled) or active insect in canopies. (b) Average count of <i>A. nigrae</i> visits to <i>S. nigra</i> catkins in survey years 2017, 2018 and 2019 from year analysis model. Letters to the left of boxes indicate significantly different means as determined by a Tukey's HSD (p-value < 0.05).	159

Chapter 1: Introduction: Biotic interactions in the Salicaceae

1.1 Community genetics

The average person often overlooks insects and fungi, but it is difficult to ignore the impact of these organisms on our planet. In agricultural systems it is estimated that 16-18% of all crop losses globally are due to insect and fungal pathogens and, in an attempt to mitigate attack, over \$13 billion was spent on pesticide application in 2004 in the United States alone (Oerke, 2006). Additionally, in forests in 2015, more than 6 million acres of trees were destroyed by herbivores (Karel & Man, 2017), a problem that has the potential to grow as climate change progresses (Garrett *et al.*, 2006). Despite the cost associated with detrimental organisms, insects also provide over \$57 billion in ecological services associated with human recreation such as education, aesthetic enjoyment, and gardening in the United States (Zhang *et al.*, 2007; Mace *et al.*, 2012). Furthermore, plants receive a number of useful services from insects including pest control, nutrient cycling, pollination and seed dispersal (Losey & Vaughan, 2006; Zhang *et al.*, 2007; Potts *et al.*, 2010). It is estimated that over 80% of wild plant productivity is reliant upon insects for seed and fruit set (Kremen *et al.*, 2007; Potts *et al.*, 2010). Research focused on the biotic-host plant relationship is imperative to understanding and managing ecosystems to balance the harmful interactions and beneficial services that insects provide to both humans and plants.

To identify a host, insects undergo a series of sequential behaviors including locating habitat, locating a host, assessing and accepting an individual host, and finally feeding. Host location is of particular importance in insect-plant interactions and can be driven by any combination of physiological and morphological cues (Panda & Khush, 1995). Although the host selection process seems simple, natural ecosystems are complex as they are composed of interacting assemblages of plant, fungal, insect, and animal species. One method for disentangling the organization of these relationships begins at the plant genetic level. In natural plant populations there is often genetic

variation in characteristics that attract or deter biotic association (Crutsinger *et al.*, 2006). If the variation in a trait is strongly heritable, plants are able to extend their phenotypes and act as a mediator of multi-biotic interactions which can in turn structure whole communities of organisms (Whitham *et al.*, 2006). Determining how shifts in genes contribute to the biotic-plant relationship effectively links ecology and evolution, and allows for prediction of community level responses to changing conditions (Shuster *et al.*, 2006; Allan *et al.*, 2013).

1.2 The Salicaceae family as a genetic model

The Salicaceae family, which includes aspens/cottonwoods (*Populus*) and willows (*Salix*), contains trees and shrubs that are of both agricultural and ecological importance. With rapid growth and capability for clonal reproduction, many species of *Populus* and *Salix* have become a focus for research into biofuel production making them important commercial crops (Rubin, 2008; Hinchey *et al.*, 2011; Langholtz *et al.*, 2016). Additionally, *Populus* and *Salix* are considered foundational and dominant species in riparian environments elevating their impact on natural ecosystems (Whitham *et al.*, 2006; Crutsinger, 2016). Finally, the completion of full genome sequencing of both *Populus trichocarpa* and *Populus deltoides* has made the family an important genetic model for ecological research of forest trees (Tuskan *et al.*, 2006). Over the last decade, *Populus* has been utilized to explore the genetic relationship between plants and communities of insects and animals. Many species of *Populus* have overlapping distributions where they form naturally occurring hybrid zones that have been a target for community assemblage research (Stettler *et al.*, 1996; Floate *et al.*, 2016). Studies of these zones have shown that tree genetic composition can both directly and indirectly influence community interactions. For example, hybrid zones have been found to directly mediate cryptic speciation of arthropods such as the mite *Aceria parapopuli* (Evans *et al.*, 2008). Similarly, beaver browsing of particular *Populus* genotypes modifies

palatability of plant tissue which indirectly influences *Chrysomela confluenta* beetle feeding (Martinsen *et al.*, 1998; Bailey *et al.*, 2004).

Despite the number of papers demonstrating the community assembling potential of plant genotypes in hybrid *Populus* systems, only a few publications have emerged that identify the underlying genes important in these interactions in pure species and even fewer determine if the same community effects are detectable for insects that interact with floral tissue rather than leaf, wood, or soil (Crutsinger *et al.*, 2014; Crutsinger, 2016; Barker *et al.*, 2019). Additionally, a frequent observation of the multi-species *Populus* hybrid zones is that hybrids are more susceptible to insect/fungal attack and show more significant responses to community interactions than their pure species counterparts (Whitham *et al.*, 1999). A possible explanation is that elevated community responses are the result of hybrid production of novel traits (or expression of traits) that are qualitatively or quantitatively different from the pure parental species (Rieseberg *et al.*, 1993; Cheng *et al.*, 2011). Hybrids therefore can be a valuable tool in initial discovery of underlying loci and genetic structure influencing biotic distributions prior to determining if mechanisms exist in a pure species (Whitham *et al.*, 1999).

1.3 Physiological and morphological features

There are several genetically controlled traits in Salicaceae that are candidates for mediating insect-plant and community level interactions. Secondary metabolites are one such physiological trait with large literature support for their importance in insect-plant relationships. Phenolics are the most plentiful secondary metabolites in the Salicaceae family (Freeman & Beattie, 2008), and their main function is to act as a feeding deterrent to generalist insects and animals (Lindroth & Peterson, 1988; Bailey *et al.*, 2006; Boeckler *et al.*, 2011). However, in certain cases specialist insects prefer

trees high in these defenses as they have either adapted to or utilize them for their own protection (Pasteels *et al.*, 1983; Kearsley & Whitham, 1989; Matsuki & MacLean, 1994). Secondary metabolites have also been determined to vary dramatically after hybridization. For example, across 30 studies and 1,112 secondary metabolites it was found that only 70.3% were present in both hybrids and their parents, whereas 24.3% were not present in hybrids and 5.5% of compounds were completely novel to the system (Cheng *et al.*, 2011). Phenolics are also highly variable among pure species with some, such as salicin, found in many but others restricted to only a few (Boeckler *et al.*, 2011). Similarly, simple genetically heritable morphological traits, such as leaf or petiole shape, can impact insect distributions and are highly variable among pure species and hybrids in the Salicaceae (Whitham, 1989; Robinson *et al.*, 2012).

The Salicaceae family also presents a unique opportunity to study the community impact of a relatively rare plant characteristic, that is only present in 6% of angiosperms: dioecy (Renner & Ricklefs, 1995; Barrett & Hough, 2013). All species of *Populus* and *Salix* are dioecious, having separate male and female individuals. There have been several studies that considered the role that dioecy and sex differences in resource allocation can play on community level interactions. For example, male individuals in both *S. lasiolepis* and *S. cinerea* supported higher densities of tenthredinid sawflies than their female counterparts (Alliende, 1989; Boecklen *et al.*, 1990). Additionally, *Phratora vulgatissima* prefers to feed on female hosts of *S. cinerea* despite higher egg mortality associated with higher predator presence on female plant (Kabir, 2012). However, these studies have focused solely on leaf feeding herbivores, thereby neglecting the floral visitor community, which is of great potential importance to the fitness of plant species in the family.

1.4 Dioecy and floral visitor communities

Although not a trait common to *Populus*, there are many species of *Salix* that rely heavily upon communities of insects to carry pollen among flowers for successful sexual reproduction (Tollsten & Knudsen, 1992; Füssel *et al.*, 2007; Ashman, 2009). *Salix* species rely on a complex combination of volatile organic compounds as olfactory signals, and male flowers are often yellow in color which acts as additional visual stimuli that female flowers lack (Tollsten & Knudsen, 1992; Füssel *et al.*, 2007). *Salix* floral resources are also extremely important in supplying food for many native herbivores, as flowers both act as a nutritive resource sink and offer rewards such as nectar and pollen (Ostaff *et al.*, 2015).

Work that has been done on *Salix* floral communities has almost exclusively focused on bee interactions while neglecting the diverse groups of insects that are also known to visit flowers such as flies, beetles, and aphids (Dötterl *et al.*, 2005; Füssel *et al.*, 2007; Ostaff *et al.*, 2015). Many of these papers also focus closely on volatiles and have not considered additional characteristics such as floral chemical defenses in insect distributions. The potential interactions and competition between bees and other floral herbivores could have major implications for both the plant sexual reproduction and bee survival. With the current global declines in bee species populations (Potts *et al.*, 2010) understanding how species like *Salix* attract and preserve resources for their cross-pollinators, despite the potential diversity of insects interacting with floral resources, is an important area of research that remains relatively unexplored.

1.5 Objectives

Building on the community structuring potential of the host plant genome that has already been established in *Populus* and *Salix*, I studied the host genetic influence on biotic communities by addressing three primary objectives in this dissertation:

Objective 1 (Chapter 2): Identify and understand the genetic mechanisms of fungal and insect association in *Populus*. This objective was accomplished by using a hybrid family cross of *P. trichocarpa* x *P. deltoides* and quantitative trait loci analysis (QTL) as a tool to identify genetic regions of interest. I then compared the QTL interval composition among the parental genomes to look for structural and functional enrichments to identify candidate genes important to biotic-biotic interactions. This manuscript has now been published in the journal *Ecology and Evolution* (Simon et al. 2020).

Objective 2 (Chapter 3): Identifying genes underlying arthropod community composition in a pure species population of *Populus trichocarpa*. I accomplished this objective by surveying arthropods in three common gardens containing over 1,000 genotypes from across the species range and using both single-trait and multi-trait genome-wide-association analysis (GWAS) and functional networks to identify SNPs associated with abundant insects and community composition.

Objective 3 (Chapter 4): Determining the impact of dioecy in *Salix* on assembly of floral insect communities and activity of native bee species. This objective was accomplished by visually surveying insects that visited flowers of the tree form willow *Salix nigra* across three years. Additionally, I utilized pan-trapping techniques to compare to visual surveys and capture a fuller community estimate in the last year of surveys. This coupled with non-metric multidimensional

scaling (NMDS) analysis and Mantel tests allowed me to assess the effects of tree sex, floral scent, defensive chemistry and inter-annual variability of flower emergence on distributions of floral visitors and primary cross-pollinators important in sexual productivity. This manuscript has been submitted to *Oecologia*.

1.6 Literature cited

- Allan GJ, Shuster SM, Woolbright S, Walker FM, Meneses N, Keith A, Bailey JK, Whitham TG. 2013.** Interspecific indirect genetic effects (iiges): Linking genetics and genomics to community ecology and ecosystem processes. In: Ohgushi T, Schmitz O, Holt RD, eds. *Trait-Mediated Indirect Interactions Ecological and Evolutionary Perspectives*. Cambridge, UK: Cambridge University Press, 295–323.
- Alliende MC. 1989.** Demographic studies of a dioecious tree. II. The distribution of leaf predation within and between trees. *The Journal of Ecology* **77**: 1048–1058.
- Ashman TL. 2009.** Sniffing out patterns of sexual dimorphism in floral scent. *Functional Ecology* **23**: 852–862.
- Bailey JK, Schweitzer JA, Rehill BJ, Lindroth RL, Martinsen GD, Whitham TG. 2004.** Beavers as molecular geneticists: a genetic basis to the foraging of an ecosystem engineer. *Ecology* **85**: 603–608.
- Bailey JK, Wooley SC, Lindroth RL, Whitham TG. 2006.** Importance of species interactions to community heritability: a genetic basis to trophic-level interactions. *Ecology Letters* **9**: 78–85.
- Barker HL, Riehl JF, Bernhardsson C, Rubert-Nason KF, Holeski LM, Ingvarsson PK, Lindroth RL. 2019.** Linking plant genes to insect communities: Identifying the genetic bases of plant traits and community composition. *Molecular Ecology* **28**: 4404–4421.
- Barrett SC, Hough J. 2013.** Sexual dimorphism in flowering plants. *Journal of Experimental Botany* **64**: 67–82.
- Boecklen WJ, Price PW, Mopper S. 1990.** Sex and drug and herbivores: Sex-biased herbivory in arroyo willow (*Salix lasiolepis*). *Ecology* **71**: 581–588.
- Boeckler GA, Gershenzon J, Unsicker SB. 2011.** Phenolic glycosides of the Salicaceae and their

role as anti-herbivore defenses. *Phytochemistry* **72**: 1497–1509.

Cheng D, Vrieling K, Klinkhamer PGL. 2011. The effect of hybridization on secondary metabolites and herbivore resistance: implications for the evolution of chemical diversity in plants. *Phytochemistry Reviews* **10**: 107–117.

Crutsinger GM. 2016. A community genetics perspective: opportunities for the coming decade. *New Phytologist* **210**: 65–70.

Crutsinger GM, Collins MD, Fordyce JA, Gompert Z, Nice CC, Sanders NJ. 2006. Plant genotypic diversity predicts community structure and governs an ecosystem process. *Science* **313**: 966–968.

Crutsinger GM, Rodriguez-Cabal MA, Roddy AB, Peay KG, Bastow JL, Kidder AG, Dawson TE, Fine PV, Rudgers JA. 2014. Genetic variation within a dominant shrub structures green and brown community assemblages. *Ecology* **95**: 387–398.

Dötterl S, Füssel U, Jürgens A, Aas G. 2005. 1,4-Dimethoxybenzene, a floral scent compound in willows that attracts an oligolectic bee. *Journal of Chemical Ecology* **31**: 2993–2998.

Evans LM, Allan GJ, Shuster SM, Woolbright SA, Whitham TG. 2008. Tree hybridization and genotypic variation drive cryptic speciation of a specialist mite herbivore. *Evolution: International Journal of Organic Evolution* **62**: 3027–3040.

Floate KD, Godbout J, Lau MK, Isabel N, Whitham TG. 2016. Plant–herbivore interactions in a trispecific hybrid swarm of *Populus*: assessing support for hypotheses of hybrid bridges, evolutionary novelty and genetic similarity. *New Phytologist* **209**: 832–844.

Freeman BC, Beattie GA. 2008. An overview of plant defenses against pathogens and herbivores. *The Plant Health Instructor*.

Füssel U, Dötterl S, Jürgens A, Aas G. 2007. Inter- and intraspecific variation in floral scent in the genus *Salix* and its implication for pollination. *Journal of Chemical Ecology* **33**: 749–

- Garrett KA, Dendy SP, Frank EE, Rouse MN, Travers SE. 2006.** Climate change effects on plant disease: genomes to ecosystems. *Annual Review of Phytopathology* **44**: 489–509.
- Hinchee M, Rottmann W, Mullinax L, Zhang C, Chang S, Cunningham M, Pearson L, Nehra N. 2011.** Short-rotation woody crops for bioenergy and biofuels applications. In: Biofuels. New York, NY: Springer New York, 139–156.
- Kabir M. 2012.** The influence of plant sex on the performance of a detrimental herbivore and two biocontrol agents in the dioecious grey willow, *Salix cinerea*.
- Karel T, Man G. 2017.** *Major forest insect and disease conditions in the United States: 2015*. Washington, D.C.
- Kearsley MJ, Whitham TG. 1989.** Developmental changes in resistance to herbivory: implications for individuals and populations. *Ecology* **70**: 422–434.
- Kremen C, Williams NM, Aizen MA, Gemmill-Herren B, LeBuhn G, Minckley R, Packer L, Potts SG, Roulston T, Steffan-Dewenter I, *et al.* 2007.** Pollination and other ecosystem services produced by mobile organisms: a conceptual framework for the effects of land-use change. *Ecology Letters* **10**: 299–314.
- Langholtz MH, Stokes BJ, Eaton LM. 2016.** *2016 Billion-ton report: Advancing domestic resources for a thriving bioeconomy, Volume 1: Economic availability of feedstock* (Oak Ridge National Laboratory, Ed.). Oak Ridge, TN (United States): UT-Battelle, LLC for the US Department of Energy.
- Lindroth RL, Peterson SS. 1988.** Effects of plant phenols on performance of southern armyworm larvae. *Oecologia* **75**: 185–189.
- Losey JE, Vaughan M. 2006.** The economic value of ecological services provided by insects. *Bioscience* **56**: 311–323.

- Mace GM, Norris K, Fitter AH. 2012.** Biodiversity and ecosystem services: a multilayered relationship. *Trends in Ecology & Evolution* **27**: 19–26.
- Martinsen GD, Driebe EM, Whitham TG. 1998.** Indirect interactions mediated by changing plant chemistry: beaver browsing benefits beetles. *Ecology* **79**: 192–200.
- Matsuki M, MacLean SF. 1994.** Effects of different leaf traits on growth rates of insect herbivores on willows. *Oecologia* **100**: 141–152.
- Oerke EC. 2006.** Crop losses to pests. *The Journal of Agricultural Science* **144**: 31–43.
- Ostaf DP, Mosseler A, Johns RC, Javorek S, Klymko J, Ascher JS. 2015.** Willows (*Salix* spp.) as pollen and nectar sources for sustaining fruit and berry pollinating insects. *Canadian Journal of Plant Science* **95**: 505–516.
- Panda N, Khush GA. 1995.** *Host plant resistance to insects*. Wallingford, UK: CAB International.
- Pasteels JM, Rowell-Rahier M, Braekman JC, Dupont A. 1983.** Salicin from host plant as precursor of salicylaldehyde in defensive secretion of chrysomeline larvae. *Physiological Entomology* **8**: 307–314.
- Potts SG, Biesmeijer JC, Kremen C, Neumann P, Schweiger O, Kunin WE. 2010.** Global pollinator declines: trends, impacts and drivers. *Trends in Ecology & Evolution* **25**: 345–353.
- Renner SS, Ricklefs RE. 1995.** Dioecy and it correlates in the flowering plants. *American Journal of Botany* **82**: 596–606.
- Rieseberg LH, Ellstrand NC, Arnold M. 1993.** What can molecular and morphological markers tell us about plant hybridization? *Critical Reviews in Plant Sciences* **12**: 213–241.
- Robinson KM, Ingvarsson PK, Jansson S, Albrechtsen BR. 2012.** Genetic variation in functional traits influences arthropod community composition in aspen (*Populus tremula* L.). *PLoS ONE* **7**: e37679.

- Rubin EM. 2008.** Genomics of cellulosic biofuels. *Nature* **454**: 841–845.
- Shuster SM, Lonsdorf E V., Wimp GM, Bailey JK, Whitham TG. 2006.** Community heritability measures the evolutionary consequences of indirect genetic effects on community structure. *Evolution* **60**: 991–1003.
- Stettler R, Bradshaw T, Heilman P, Hinckley T. 1996.** *Biology of Populus and its implications for management and conservation*. Ottawa, ON, Canada: NRC Research Press.
- Tollsten L, Knudsen JT. 1992.** Floral scent in dioecious *Salix* (Salicaceae)- a cue determining pollination system? *Plant Systematics and Evolution* **182**: 229–237.
- Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, *et al.* 2006.** The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596–1604.
- Whitham TG. 1989.** Plant hybrid zones as sinks for pests. *Science* **244**: 1490–1493.
- Whitham TG, Bailey JK, Schweitzer JA, Shuster SM, Bangert RK, LeRoy CJ, Lonsdorf E V., Allan GJ, DiFazio SP, Potts BM, *et al.* 2006.** A framework for community and ecosystem genetics: from genes to ecosystems. *Nature Reviews Genetics* **7**: 510–523.
- Whitham TG, Martinsen GD, Keim P, Floate KD, Dungey HS, Potts BM. 1999.** Plant hybrid zones affect biodiversity: tools for a genetic-based understanding of community structure. *Ecology* **80**: 416–428.
- Zhang W, Ricketts TH, Kremen C, Carney K. 2007.** Ecosystem services and dis-services to agriculture. *Ecological Economics* **64**: 253–260.

Chapter 2: Host plant genetic control of associated fungal and insect species in a *Populus* hybrid

In press: Simon S, Tschaplinski T, Leboldus J, Keefover-Ring K, Azeem M, Chen J, Macaya-Sanz D., MacDonald W., Muchero W., & DiFazio S. 2020. Host plant genetic control of associated fungal and insect species in a *Populus* hybrid. *Ecology and Evolution*.

Authors' contributions: S.S., S.D., and D.M.-S. identified, collected, and analyzed insect surveys. W.M., J.L., J.C., and W.M. identified and collected fungal species datasets. K.K.-R., M.A., and T.T. collected/analyzed metabolite data. S.S. and S.D. analyzed the data. W.M., G.T. and S.D. designed and supervised the study. S.S., S.D., J.L., T.T., and K.K.-R. wrote the manuscript.

2.1 Abstract

Plants employ a diverse set of defense mechanisms to mediate interactions with insects and fungi. These relationships can leave lasting impacts on host plant genome structure, such as rapid expansion of gene families through tandem duplication. These genomic signatures provide important clues about the complexities of plant/biotic stress interactions and evolution. We used a pseudo-backcross hybrid family to identify Quantitative Trait Loci (QTL) controlling associations between *Populus* trees and several common *Populus* diseases and insects. Using whole genome sequences from each parent, we identified candidate genes that may mediate these interactions. Candidates were partially validated using mass spectrometry to identify corresponding QTL for defensive compounds. We detected significant QTL for two interacting fungal pathogens and three insects. The QTL intervals contained candidate genes potentially involved in physical and chemical mechanisms of host-plant resistance and susceptibility. In particular, we identified overlapping QTLs for a phenolic glycoside and *Phyllocolpa* sawfly abundance. There was also significant enrichment of recent tandem duplications in the genomic intervals of the parent that was native to the pest range, but not the exotic parent. Tandem gene duplication may be an important mechanism for rapid response to biotic stressors, enabling trees with long juvenile periods to reach maturity despite many coevolving biotic stressors.

2.2 Introduction

In natural ecosystems, the dynamics of plant interactions with other living organisms are complex. This is especially true for organisms that rely on plants for shelter, nutrition, and reproduction, such as fungi and insects (Panda & Khush, 1995; Chisholm *et al.*, 2006). Although fungi and insects can provide some of the same useful services in return, those that do not can be extremely harmful to plant productivity. To mitigate the effects of biotic stress, plants employ a diverse set of defense mechanisms, including chemical, protein-derived molecules, and physical barriers (Panda & Khush, 1995). Insects and fungi must develop strategies in parallel to overcome these obstacles to survive (Mello & Silva-Filho, 2002; Chisholm *et al.*, 2006). The theory of gene-for-gene coevolution has frequently been used to describe this host plant genetic relationship to its arthropod and fungal communities (Ehrlich & Raven, 1964; Thompson, 1988; Mello & Silva-Filho, 2002; Chisholm *et al.*, 2006).

The gene-for-gene theory suggests a very simple dynamic for the genetic interactions that occur between two species. A gene in the host plant that is important in biotic relationships has a corresponding, coevolving gene from a pathogen/insect, which can lead to resistance or susceptibility depending on the life history of the pathogen/insect (Flor, 1971; Friesen *et al.*, 2007). Much of the evidence for these interactions has been found in crop systems where plant species often have dominant, single-gene trait conferring resistance to feeding (Thompson, 1988). For example, there are over twenty different genes in wheat (*Triticum aestivum* L.) that each confer resistance to the Hessian fly, *Mayetiola destructor* (Thompson & Burdon, 1992). Exposure of Hessian fly populations to these resistant varieties of wheat created selection pressure that led to increased virulent gene combinations in the pest (Gallun, 1977; Panda & Khush, 1995). Similarly, in plant-fungal systems, breeding for dominant resistance in cereal crops resulted in new selective

forces that increased virulence gene frequencies in *Puccinia* spp. cereal rusts (Pretorius *et al.*, 2000; Chen, 2005). This in turn can lead to an evolutionary arms race between plants, insects, and fungi with the continual development of mechanisms to overcome both genetic defenses and virulent attacks (Thompson & Burdon, 1992; Bergelson *et al.*, 2001).

The relationships of host plant genetics and biotic associations can also be more complex than these crop breeding systems suggest, and they can leave a lasting impact on genome structure (Lefebvre & Chèvre, 1995). Host plant and biotic associations can lead to the expansion of gene families responsible for the host plant response to biotic stress. For example, the Kunitz protease inhibitors (KPIs) in *Populus* are important in defense responses against insects by inhibition of herbivore digestion (Haruta *et al.*, 2001; Major & Constabel, 2008). The KPI gene family has greatly expanded in response to insect attack through tandem duplication events (Philippe *et al.*, 2009). Similarly, plant resistance (*R*) genes, which are important in the defense response of plants to fungal pathogen attack, have also expanded through tandem and segmental duplication events due to biotic pressures (Hulbert *et al.*, 2001; Leister, 2004). Analyzing how the genome is structured in the host plant when it associates with fungi and insects is important for studying these relationships and understanding the complexities of their genetic interactions.

Given their rapid growth and vegetative reproduction, *Populus* species have become a focus for research into biofuel production, making them a valuable commercial crop (Meilan *et al.*, 2002; Taylor, 2002; Stanton *et al.*, 2010). *Populus* has also become an important genetic model for research into a wide variety of ecological and adaptive traits (McKown *et al.*, 2014), including interactions with the biotic community (Whitham *et al.*, 2006; Crutsinger *et al.*, 2014). In particular, interspecific hybrids of *P. trichocarpa* x *P. deltoides* segregate for a wide variety of

traits, including resistance to insect and fungal attack (Newcombe, 1998; Newcombe & Ostry, 2001). Such hybrid family crosses can be used to identify regions of the genome that are important in mediating biotic stress.

In this study, we investigated the genome composition of loci associated with insect and fungal species in an interspecific *Populus* family. We surveyed insects and fungal pathogens in a *P. deltoides* x *P. trichocarpa* pseudo-backcross family and used quantitative trait locus (QTL) analysis and comparative genomics to address three main questions: (1) Is there heritable, host genetic control of fungal and insect species? (2) What protein domains and gene ontology terms are enriched in the QTL intervals in the genomes of each *Populus* species? (3) What candidate genes in the intervals are unique to each species when comparing the *P. trichocarpa* and *P. deltoides* genomes?

2.3 Methods

QTL mapping pedigree

The 52-124 family was developed by crossing a male *Populus deltoides*, ILL-101 from southern Illinois, with a female *Populus trichocarpa* clone, 93-968 from western Washington State. The F₁ clone, 52-225, was then crossed with a male *Populus deltoides* clone, D124 from Minnesota, to generate the final 749 progeny in the pseudo-backcross population.

Field site descriptions

Clonal cuttings of the 52-124 progeny were obtained from the University of Minnesota and the University of Florida. The individuals were propagated in the West Virginia University (WVU) greenhouses in March 2006. A hay production field was tilled and disked at the WVU Agronomy

farm (39°39'32"N 79°54'19"W) in Morgantown, West Virginia for planting. In July 2007 the rooted cuttings were planted with two clonal replicates for each of 749 genotypes at 2 m x 2 m spacing in an interlocking block design. The plantation was thinned in December of 2008 by removing 50% of the trees in a diamond fashion resulting in 2.83 m x 4 m spacing of the remaining trees. A plantation of the same family cross was established at the Westport Research Station in the Columbia River floodplain (46°07'58.8"N 123°22'05.0"W) in April, 2016. The trial was planted with a total of 339 progeny replicated in a randomized complete block design with three blocks.

SNP genotyping and genetic map construction

SNP loci were selected from whole genome resequencing data generated for the parent trees of the pedigree, focusing on loci that were fixed for different alleles in *P. deltoides* and *P. trichocarpa*. Sequencing was performed on the Illumina GAII system with single end read lengths of 75 bp and a total depth of ~35X on average. Reads were aligned using MAQ and SNPs were called using Samtools mpileup with a minimum quality of 30 and a minimum depth of 5 reads per allele, and a subset of loci were confirmed by Sanger sequencing (Slavov *et al.*, 2012). Polymorphic loci were selected that were maternally informative for *P. trichocarpa*. These were incorporated into an Illumina Infinium Bead Array, which was used to genotype 3,568 SNP loci in 692 of the progeny. This data was then used to create the genetic map composed of 19 linkage groups corresponding to 19 *Populus* chromosomes. Genotyping and map construction are described in more detail elsewhere (Muchero *et al.*, 2015).

Family 52-124 parent and progeny phenotyping

In order to identify regions of the *Populus* genome associated with biotic stress, a variety of fungal pathogens and insect herbivore species were surveyed in the WVU Morgantown and Westport

Oregon plantation sites. *Melampsora* sp. rust was identified visually by local pathologist Dr. William MacDonald. *Sphaerulina musiva* was identified by sequencing of the ITS region (Verkley *et al.*, 2013). Insect identification was completed using insect morphological features, feeding symptoms, and known hosts/species distributions for *Phyllocolpa* spp. (Smith & Fritz, 1996; Kopelke, 2007), *Mordwilkoja vagabunda* (Ignoffo & Granovsky, 1961a,b), and *Pemphigus populitransversus* (Bird *et al.*, 1979; Faith, 1979).

Melampsora sp. leaf rust (Figure 1a) and *Sphaerulina musiva* fungal leaf spot symptoms (Figure 1b) were scored on a 0-3 scale of disease severity, with 0 indicating absence of symptoms and 3 indicating high degree of pathogen leaf damage, in the 2008 growing season for all 1353 tree canopies. In the fall of 2014, stem canker symptoms caused by the same fungus *Sphaerulina musiva* (Figure 1c) were scored on a 0-5 disease severity scale for a subset 498 unique genotypes and a total of 580 trees. Upon further examination of field conditions for *S. musiva* disease severity, it was determined that none of the progeny displayed complete resistance. The original 0-5 scale was binarized with scores from 2.5-5 scaled to 1 and 0-2 scaled to 0. The new scale indicated the progression of infection, with 1 specifying severe canker development and 0 indicating less aggressive canker symptoms.

Two abundant species of galling aphid were also surveyed during the 2016 growing season. To equalize the biomass surveyed for the aphid insect observations on each tree, branches of equal diameter were selected for insect counts and the remaining branches in the canopy were not surveyed. Tree canopy presence or absence for the petiole galling aphid

Pemphigus populitransversus (Figure 1d) and the branch galling aphid *Mordwilkoja vagabunda* (Figure 1e) was recorded for 201 unique genotypes and a total of 218 trees. For all of the damage

scores in every survey year, 0 indicated a complete absence of fungal or insect presence in tree canopy.

In early August 2017 at the Westport site, full canopies were scored by counting galls of the leaf folding sawfly *Phyllocolpa* sp. (Figure 1f), on 534 unique genotypes and a total of 1020 trees. To estimate productivity of individuals and confirm that availability of resources did not drive insect attraction or feeding, main stem diameter in millimeters was recorded for all trees to be used as a covariate in the analysis.

Leaf metabolites

The metabolomic profiles for 211 genotypes of Family 52-124 were determined for leaves (leaf plastochron index 7 ± 2) sampled on September 14, 2006 at a USDA-FS field site near Grand Rapids, MI from plants in their first growing season in the field. Leaves were quickly collected and flash-frozen on dry ice, prior to shipping back to Oak Ridge National Laboratory, where ~200 mg fresh weight per sample were extracted in 80% ethanol (aqueous), as described in Zhao *et al.*, (2010) and analyzed for the major aromatic metabolites following trimethylsilylation and analysis by gas chromatography-mass spectrometry (GC-MS) using electron impact ionization (70 eV). The analytical and data extraction protocols were similar to those reported previously in Weston *et al.*, (2015). The major phenolic metabolites quantified included salicortin, α -salicyloylsalicin, salicyl-salicylic acid-2-O-glucoside, salirepin (gentisyl alcohol 2-O-glucoside), gentisic acid 5-O-glucoside, 1,2,4-benzenetriol, and gentisyl alcohol 5-O-glucoside.

2.4 Statistical analysis

Within-family broad-sense heritability (H^2) calculation

To prevent arbitrary score bias, *Melampsora* sp. and *S. musiva* values were normalized to have a mean of 0 and standard deviation 1 (*score.transform* function of the *CTT* R package) by applying the inverse of the cumulative distribution function of the normal distribution to the sample percentile score (Gianola & Norton, 1981). Surveys of the *Melampsora* sp. leaf rust, *S. musiva* leaf spot, and counts of the *Phyllocolpa* sp. galls had within-garden microsite variation removed using thin-plate spline regression (*fields* R package; Table 1). Residuals of the spatial correction were added to the mean of each survey dataset for each tree observation to spatially correct and rescale the values. These corrected values were then used to calculate the proportion of variance in the fungal and insect distributions that was due to genotype using a linear-mixed model (*lmer* function of the *lme4* R package) with insect counts and fungal scores as the response, tree genotype as the predictor, and either competing fungus score or stem diameter biomass estimates as a covariate where applicable. Broad-sense heritability was calculated as $H^2 = \delta_g^2 / (\delta_g^2 + \delta_e^2)$, where δ_g^2 is the genetic variance due to genotype and δ_e^2 is the residual variance. Rapid, simulation-based exact likelihood ratio tests were used to evaluate the significance of variation due to genotype for each linear model (*exactRLRT* function of the *RLRsim* R package). SAS software version 9.4 (2013) was used to test for the normality of all data sets and for all subsequent statistical analyses, and transformations were conducted when necessary. Finally, best linear unbiased predictors (BLUPs) were extracted from these models to use in the QTL analysis.

Quantitative trait loci (QTL) analysis

The R software package *R/qtl* (Broman *et al.*, 2003) was used for all QTL analyses described below. Composite interval mapping (CIM) was used to associate *Melampsora* sp. leaf rust, *S. musiva* leaf spot, and *Phyllocolpa* sp. to QTL positions (*cim* function). An additional CIM QTL analysis was conducted for fungi surveyed in 2008 to further evaluate potential competitive bias of

co-occurring fungal pathogens. This was done by sub-setting individuals out of the full dataset to exclude trees with symptoms of both pathogens, leaving us with 90 individuals that only showed *Melampsora* sp. symptoms and 434 individuals only infected with *S. musiva* leaf spot. Single QTL mapping was used to associate the binary scores for the *S. musiva* canker, *M. vagabunda*, and *P. populitransversus* to QTL positions (*scanone* function). The method used for both mapping approaches was the expectation-maximization (EM) algorithm. Estimation of QTL interval significance was completed by performing 1000 permutations. Intervals with logarithm of odds (LOD) scores that were above the p-value threshold ($\alpha = 0.05$), as determined from the permutation tests, were selected for further analysis. The percent variance explained by significant markers for fungal and insect surveys, that were mapped using CIM, were calculated by extracting significant marker positions and creating a fit QTL model (*fitQTL* function). The positive allele contributing to an increase in susceptibility to fungi and insects was found by generating effect plots for each phenotype and its significant marker position (*effectplot* function).

Physical genome intervals

Physical genome intervals in the *P. trichocarpa* genome (v3.0) were examined for each significant QTL for biotic associations. The intervals were defined as 1 Mb regions centered on the marker with the highest LOD score. Fixed physical genome sizes were used rather than intervals defined based on LOD scores due to the large variation in magnitude of LOD observed for the significant QTL. For example, intervals of 1 LOD centered on the QTL ranged in size from 169 to 4620 kb. Much of this variation was likely due to variation in marker density and local recombination rates, in addition to phenotyping and genotyping error. We believe that a fixed 1 Mb interval is a more consistent and conservative approach given the size of the family and the variation in strength of

the QTL (Yin *et al.*, 2004b). On average, this represents approximately 6.34 cM, based on a total map size of 2617 cM and a total assembled genome length of 420 Mb.

Orthologous intervals were identified in the *P. deltoides* clone WV94 reference genome (v2.1) obtained from Phytozome (Goodstein *et al.*, 2012). Orthology was determined using a combination of protein sequence conservation and synteny using MCScanX (Wang *et al.*, 2012). Briefly, all proteins were compared in all-vs-all searches using blastp both within genomes and between genomes. These were then chained into collinear segments using the MCScanX algorithm.

Orthologous segments were identified based on the presence of large numbers of gene pairs in collinear order with high sequence identity (median blastp E score $<1e-180$) (Figure 2).

Synonymous (K_s) and nonsynonymous (K_a) nucleotide substitution rates were calculated using the Bioperl DNASTatistics module (Stajich, 2002). Domain composition and Gene Ontology (GO) terms were obtained for each genome from Phytozome (v12.1). Intervals were customized for the grandparents of the pseudo backcross progeny (clones 93-968 for *P. trichocarpa* and D124 for *P. deltoides*) by converting the respective reference genome based on alignment of short-read sequences derived from each species. Specifically, we generated 243 and 248 million 250 bp paired end Illumina HiSeq sequences for 93-968 and D124, respectively. This yielded an average coverage of ~150X per genome. These were aligned to the respective reference genome for each species (Nisqually v3.0 for *P. trichocarpa*, and WV94 v2.0 for *P. deltoides*) using bwa mem with default parameters. SNPs and small indels were identified using samtools *mpileup* and bcftools *call* with default multiallelic variant settings (Li *et al.*, 2009; Li, 2011), and sequence depth was extracted using vcftools (Danecek *et al.*, 2011). Sequences were converted using the vcftools utility vcf-consensus. Genes with no coverage in the alignments were excluded from the intervals for each species.

Tandem duplications

Tandemly duplicated genes were identified using all-vs-all blastp searches within each genome for biotic stress-associated intervals. Genes with blastp E scores $< 1e-180$ that were located within 500 kb of one another were considered to be recent tandem duplications. These were considered recent because *P. trichocarpa* and *P. deltoides* orthologs also have a median blastp E score in this range, suggesting that these tandem duplications mostly occurred after these species diverged from a common ancestor. The window size was determined by testing a range of values and choosing a window size at which the number of newly discovered tandem duplicates began to decline (Figure 3). The stringent E score cutoff was intended to focus the analysis on genes that are recently duplicated and, therefore, potentially differentially duplicated between the species. The QTL intervals were tested for significant enrichment of tandem duplicates by using a Monte Carlo simulation. Sets of contiguous genes equal in number to those contained in each QTL interval were randomly selected from the whole genome, and the number of sampled tandem duplications was counted for each iteration. This was repeated 10,000 times, and the observed number of tandem duplicates was compared to the simulated distribution to derive an empirical P-value.

2.5 Results

Heritability of fungal and insect associations

Clonal repeatability (or within-family broad-sense heritability) was estimated for each categorical survey trait (Table 1). There was a significant host-plant genetic contribution of moderate effect controlling the association of *S. musiva* leaf spot disease severity ($H^2 = 0.250$, p-value < 0.0001). There was a strong effect of host-plant genetics on associations of *Melampsora* sp. disease severity

($H^2 = 0.609$, p-value < 0.0001) and the continuous count of leaf-folding galls of *Phyllocolpa* sp. ($H^2 = 0.391$, p-value < 0.0001).

QTL mapping of fungal and insect surveys

Four QTL intervals containing a total of 38 markers were significantly associated with the abundance of fungal pathogens (Table 2). An overlapping QTL on Chr04 (Figure 4a,b) was associated with both *Melampsora* sp. (marker position = 3.50687, p-value = 0.001) and *S. musiva* leaf spot disease severity (marker position = 15.4912824, p-value = 0.026) for analysis that included all clones. The marker with the highest LOD score explained 54.1% of the variance in the disease severity of *Melampsora* sp. within the Chr04 QTL interval, whereas the top marker within the overlapping interval only explained 3.31% of the variance in the disease severity of *S. musiva* in the analysis with all individuals.

Upon exclusion of individuals that showed symptoms of the competing fungus, the strength of the association on Chr04 increased for *Melampsora* sp. (marker position marker position = 3.50687, p-value = 0.0001), whereas the association was lost for *S. musiva* leaf spot and a new association was revealed on Chr06 (marker position = 142.6761866, p-value = 0.017). The top Chr04 marker explained 57.9% of variation in *Melampsora* sp. infection severity while the new Chr06 association explained 4.77% of the variation in *S. musiva* leaf spot disease severity. The positive allele contributing to an increase in fungal infection on Chr04 was derived from *P. deltoides* whereas the positive allele contributing to severity of *S. musiva* leaf spot infection on Chr06 was derived from *P. trichocarpa*. Binary presence of the canker symptoms caused by *S. musiva* (Figure 5a) was found to be associated with a QTL located on Chr16 (marker position = 60.81429, p-value =

0.022). The positive allele contributing to increase in presence of *S. musiva* canker was from *P. trichocarpa*.

Three QTL intervals containing a total of 40 markers were significantly associated with insect abundance (Table 2, Figure 5b,c). A QTL on Chr05 (marker position = 21.83874, p-value = 0.001) was associated with binary presence of *M. vagabunda* branch galls (Figure 5b). The positive allele contributing to an increase in presence of *M. vagabunda* was from *P. trichocarpa*. Two QTLs, one located on Chr10 (marker position = 57.69914343, p-value = 0.015) and one located on Chr13 (marker position = 84.43857329, p-value = 0.045), were associated with the number of leaf-folding galls from *Phyllocolpa* sp. (Figure 5c). The marker with the highest LOD score explained 9.65% of the variance in oviposition gall count while 8.82% of the variance was explained by the top marker located on Chr13. For both markers, the positive allele contributing to an increase in the number of female oviposition galls was from *P. deltoides*. There were no QTL intervals that passed the permutation threshold for the binary presence of *P. populitransversus* petiole galls (Figure 5c).

Interaction between Melampsora sp. and S. musiva leaf infection

Melampsora sp. leaf rust infection severity was dependent upon the disease severity of *S. musiva* leaf spot symptoms ($F = 35.2$, p-value = 0.0001) (Table 3). The severity of *Melampsora* sp. infection for individuals with no *S. musiva* leaf spot was significantly higher than for individuals with *S. musiva* leaf spot symptoms (Figure 6). Similarly, *S. musiva* leaf spot disease severity was dependent upon the levels of *Melampsora* sp. leaf rust ($F = 31.5$, p-value = 0.0001). *S. musiva* leaf spot infection severity was significantly lower for individuals with a *Melampsora* sp. leaf spot score of 3 compared to individuals with less severe *Melampsora* sp. leaf spot symptoms (Figure 6).

PFAM, GO term and tandem duplication enrichment

The total number of genes present in the QTL intervals of each parental species was 174 on Chr04, 81 on Chr16, 161 on Chr10, 445 on Chr13, and 156 on Chr05 (Table 4). There were 40 PFAM domains enriched relative to the rest of the *P. trichocarpa* genome across all QTL intervals ($\alpha = 0.05$; p-value threshold < 0.00014), and 25 PFAM domains enriched in the *P. deltoides* QTL intervals ($\alpha = 0.05$, p-value < 0.000059). Additionally, there were 21 GO terms enriched in the *P. trichocarpa* intervals ($\alpha = 0.05$, p-value < 0.000226) and 46 GO terms enriched in the *P. deltoides* intervals ($\alpha = 0.05$, p-value < 0.000214). Finally, seven of the genes in these intervals showed evidence of positive selection based on the ratio of nonsynonymous to synonymous nucleotide substitutions between the *P. deltoides* and *P. trichocarpa* orthologs (Table 5).

In total, there were 107 recent tandem duplicates in the *P. trichocarpa* intervals and 195 in the *P. deltoides* intervals for all biotic QTLs (Figure 7, Table 6). This includes 23 and 41 recent tandem duplicates for the *Melampsora* sp. intervals; 6 and 10 for the *S. musiva* intervals; 32 and 96 for the *Phyllocolpa* sp. intervals; and 47 and 46 for the *M. vagabunda* intervals, for *P. trichocarpa* and *P. deltoides*, respectively. The total number of recent tandem duplicates was significantly enriched relative to tandem counts for random intervals of the same size as the QTL intervals for the *P. deltoides* grandparent (p-value = 0.0118), but not for the *P. trichocarpa* grandparent (p-value = 0.1191).

2.6 Discussion

The goal of our research was to utilize QTL analysis as a tool to identify regions of the *Populus* genome that were important in mediating biotic interactions. Upon identification of these regions, we were able to directly compare the parental genomes of the hybrid cross to look for similarity in

content and potential gene-for-gene interactions reflected in recent tandem duplication expansion. We found that the host plant genotype had a significant effect on fungi and insects in our study. Additionally, the progeny segregated for varying resistance to fungal and insect pathogens and pests that were inherited from the grandparents and parents of the 52-124 family cross.

Novel alleles (i.e. those from the grandparent that were not native to the region where the trials occurred) appear to be important for resistance to *Melampsora* sp. fungus, *M. vagabunda* galls, and *Phyllocolpa* sp. oviposition. This is not the first case of potential novel alleles being important in resistance to biotic stress. In a *P. trichocarpa* GWAS population, three genes have been found that confer novel resistance to *S. musiva* canker infection and an additional single gene that, when inherited, can suppress that resistance (Muchero *et al.*, 2018). Although cases such as this are still under investigation to understand mechanisms of novel resistance can also originate from a variety of simple traits inherent to a non-native species, such as delayed emergence due to phenology (Mercader *et al.*, 2009) or changes in secondary metabolites that are important cues in insect recognition of the host important in oviposition (Nahrstedt, 1989). In contrast, the interval associated with *S. musiva* canker symptoms was the only case in which susceptibility to the fungus was dominant and inherited from the non-coevolved host (*P. trichocarpa*), as has been previously observed (Newcombe & Ostry, 2001; Muchero *et al.*, 2018).

We detected a major QTL for *Melampsora* sp. resistance on Chr04. Although the specific strain that infected the trees is unknown, this genomic interval is known to contain the *MXC3* locus, which confers resistance to infection of multiple species of the *Melampsora* leaf rusts (Newcombe *et al.*, 2001; Yin *et al.*, 2004a). Based on mean parental infection scores and the allelic effects at the QTL, progeny in family 52-124 inherited this resistance from the *P. trichocarpa* grandmother 93-

968. Furthermore, the QTL interval contained two tandem repeats of stigma-specific proteins (Stig1) in both the *P. trichocarpa* genome and the orthologous interval in the *P. deltoides* genome. Several of these genes showed evidence of positive selection based on K_a/K_s ratios (Table 5). Stigma-specific proteins, specifically Stig1, have been found to be associated with female sterility in tobacco (*Nicotiana tabacum*) and petunia (*Petunia hybrida*) (Goldman *et al.*, 1994; Verhoeven *et al.*, 2005). Stig1 is known to mediate secretion of exudate lipids in the intercellular spaces and high expression of the protein inhibits pollen grains from penetrating style tissue preventing fertilization (Verhoeven *et al.*, 2005). Most lipid transfer proteins, such as Stig1, are important in plant cell-wall loosening and their expression can prevent penetration of plant tissues (Nieuwland *et al.*, 2005). Diversification of the protein family containing Stig1 may play an important role in lowering fungal infection in *Populus* by providing a physical barrier to resist the *Melampsora* sp. hyphae. Another protein domain that was found to be enriched in the *P. trichocarpa* Chr04 interval was the malonyl-CoA decarboxylase C-terminal domain. Similarly, the gene ontology function for malonyl-CoA decarboxylase activity was also enriched in *P. trichocarpa* in the same interval. Genes products that are capable of transforming malonyl-CoA are important precursors in the production of several pathogen defensive compounds, such as isoprenoids in the mevalonate pathway (Dixon, 2001; Chen *et al.*, 2011).

Interestingly, an overlapping interval on Chr04 was found to be associated with the activity of the leaf spot symptoms of the *S. musiva* fungus. Upon further investigation, we found that trees that were not infected by the *S. musiva* leaf spot had more severe symptoms of the *Melampsora* sp. leaf rust. This suggests that a competitive interaction occurred between the two pathogens in the field during the year of survey and was reflected in an inflated association on Chr04 with the *S. musiva* leaf spot score. Furthermore, when individuals with presence of *Melampsora* sp. infection were

removed from the *S. musiva* analysis the association with Chr04 is no longer significant, and a new QTL appeared on Chr06. Competition among fungal pathogens is not uncommon in field conditions, with most examples focused on different genotypes within the same fungal species (Abdullah *et al.*, 2017). The outcome of within host tissue colonization of multiple strains or species often relies upon the genetic similarity of the pathogens (Koskella *et al.*, 2006; Abdullah *et al.*, 2017). In our case, we had two very different pathogens that were utilizing the leaf tissue in vastly dissimilar ways. *S. musiva* is a necrotrophic fungus, which requires dead tissue to reproduce, and *Melampsora* sp. is biotrophic, requiring living tissue to reproduce. The presence of *S. musiva* on the trees had a much larger effect on the occurrence of *Melampsora* sp. infection. This may indicate that the *S. musiva* fungus had a competitive advantage over *Melampsora* sp. in utilization of the *Populus* leaf tissue at our site. Further supporting this finding was the complete loss of *Melampsora* sp. symptoms at the plantation site over the course of ten years and the continual presence of the *S. musiva* leaf spot. A similar interaction has been reported in wheat between the biotroph *Blumeria graminis* f.sp. *tritici* (Bgt), the powdery mildew pathogen, and necrotroph *Zymoseptoria tritici*, the cause of *Septoria tritici* blotch (Orton & Brown, 2016). Similarly, the outcome of the interaction of the two pathogens in wheat was competitive. However, the necrotroph was found to actually be capable of reducing the reproductive capability of the biotroph, which indicates pathogen-pathogen interactions can be more direct rather than relying solely on the host plant genetics (Orton & Brown, 2016).

The *S. musiva* leaf spot and stem canker symptoms were found to be associated with QTL intervals on Chr06 and Chr16, respectively. In this study, susceptibility to the necrotrophic fungi appeared to be originating from the presence of *P. trichocarpa* alleles in the progeny for both symptoms. Previous work on a similar *Populus* hybrid crosses support this, with susceptibility to *S. musiva*

necrotrophic fungi originating primarily from dominant alleles derived from *P. trichocarpa* (Newcombe, 1998; Newcombe & Ostry, 2001; Muchero *et al.*, 2018). The Chr06 leaf spot association contained numerous genes with methyltransferase activity, which were higher in the *P. deltoides* interval. Methyltransferase enzymes are important in plant secondary metabolism and have been found to be key in the production of a variety of anti-microbial compounds (Noel *et al.*, 2003). Interestingly, the loci conferring susceptibility in family 52-124 in the Chr16 canker interval did not overlap with the four loci that were uncovered in a previous genome-wide association study of *S. musiva* susceptibility in *P. trichocarpa* (Muchero *et al.*, 2018), suggesting that different mechanisms may be involved in hybrid interactions with this pathogen. However, the QTL did contain a tandem repeat of a G-type lectin receptor-like protein kinase that was expanded in *P. deltoides*. This protein could play a similar role to a receptor-like kinase from the same family (Yang *et al.*, 2016) that was associated with susceptibility to *S. musiva* in the *P. trichocarpa* study (Muchero *et al.*, 2018).

M. vagabunda has been recorded completing its life cycle on several species of *Populus*, including *P. deltoides* and *P. tremuloides* (Floate, 2010). Although the aphid's life history has been documented, little is known about the influence of host plant genetics on gall formation or resistance to feeding (Ignoffo & Granovsky, 1961b; Floate, 2010). We detected a QTL on Chr05 in which alleles inherited from *P. deltoides* were positively associated with gall occurrence. Genes conferring lipoxygenase and oxidoreductase activity were enriched in both the *P. deltoides* as well as the *P. trichocarpa* QTL intervals. Lipoxygenase genes are known to be associated with the *Populus* response to both abiotic and biotic stressors (Ralph *et al.*, 2006; Cheng *et al.*, 2006). They are often upregulated in the presence of mechanical damage, fungal pathogen invasion, and exposure to simulated insect feeding (Cheng *et al.*, 2006; Chen *et al.*, 2009). The lipoxygenases are

important in the formation of jasmonic acid, the signaling molecule that upregulates plant defenses against herbivore feeding (Chen *et al.*, 2009).

The *M. vagabunda* QTL interval on Chr05 also contained a tandem array of resistance genes (R-genes) that encoded disease resistance proteins (TIR-NBS-LRR class) that was greatly expanded in *P. trichocarpa* compared to *P. deltoides*, as well as repeats of the leucine-rich repeat protein kinase family proteins in both species. These protein families are well known for their roles in the recognition and upregulation of host plant defenses against bacterial and fungal infection (Bergelson *et al.*, 2001; Martin *et al.*, 2003). Neofunctionalization of R-genes through tandem duplication due to gene-for-gene coevolution has also been demonstrated in many plant-fungal pathosystems (Leister, 2004). Although R-genes have been more frequently related to plant-fungal interactions, they are also becoming commonly associated in mediating plant-insect interactions as well, especially in insects that utilize piercing-sucking feeding (Harris *et al.*, 2003; Kaloshian, 2004).

In addition to the R-genes, there were a series of genes encoding cytochrome P450 family proteins in both species as well as a unique tandem set that was only present in the *P. trichocarpa* genome. P450 enzymes are important in the production of many classes of secondary metabolites, such as furanocoumarins and terpenoids, which are highly toxic to insects (Keeling & Bohlmann, 2006; Schuler, 2011). Alternatively, cytochrome P450's may also be implicated in the susceptibility of the host plant to galling aphids. They are important in the synthesis of fatty acids and production of suberin in plant tissues (Höfer *et al.*, 2008; Pinot & Beisson, 2011). Typically, suberin is important in separation of different tissues as well as in the establishment of apoplastic barriers that restrict nutrient/water loss as well as pathogen invasion (Höfer *et al.*, 2008; Qin & LeBoldus, 2014).

Several insects are known to produce suberized spherical galls on leaves, including hymenopteran pests of Rosacea species and dipteran pests of Fabaceae (Krishnan & Franceschi, 1988; Oliveira *et al.*, 2016). If aphids induce the suberization mechanism of the plant genome, may lead to the increased toughening of aphid galls, much like the woody structures *M. vagabunda* leaves behind on branches once they have finished feeding.

We detected two QTL for *Phyllocolpa* sp., with *P. trichocarpa* alleles being positively associated with leaf fold occurrence in both cases. The *P. trichocarpa* genomic interval corresponding to the Chr10 QTL was enriched for several domains and GO terms that could be involved in gall development. For example, the interval was enriched for sugar transporters as well as GO terms for sucrose transporter activity. Often, in the case of galling insects, plant tissue is modified in such a way as to act as a sugar sink, thereby enhancing its nutritional value to larvae (Larson & Whitham, 1991; Nyman *et al.*, 2000; Wool, 2004) . The presence of these combinations of sugar transporter genes may be mediating a similar interaction between the *Phyllocolpa* sp. female sawflies and their chosen *Populus* hosts.

Phyllocolpa sp. galls are formed early in the season when a female sawfly selects a leaf and injects the longitudinal fold with small amounts of fluid on the underside of young leaves (Fritz & Price, 1988; Kopelke, 2007). The adult sawfly will proceed to oviposit near the base of the leaf, and after 1-2 days, the leaf fold forms and the newly hatched larvae feeds on the inside of the gall (Smith & Fritz, 1996). The *Phyllocolpa* sp. galls were a unique biotic phenotype to this study as they were an estimate of female sawfly ovipositional choice rather than feeding success. Host selection for oviposition is initially driven by visual cues and reinforced by females assessing the nutrition and chemical cues of foliage (Panda & Khush, 1995; Boeckler *et al.*, 2011). Previous research in *Salix*

(closely related to *Populus*) has shown that common phenolic glycosides in leaf tissue are important in host choice in both the free-feeding *Nematus oligospilus* and galling *Euura amerinae* specialist sawflies (Kolehmainen *et al.*, 1994; Fernández *et al.*, 2019). We therefore pursued QTL mapping of metabolites to determine if phenolic compounds might be important determinants of the potential ovipositional relationship.

We detected three significant QTL for gentisyl alcohol 5-O-glucoside levels. One of these overlapped with the *Phyllocolpa* sp. leaf fold QTL on Chr10. The QTL on Chr14 also overlapped with a suggestive QTL peak for *Phyllocolpa* sp. (LOD=3.96). This is the first case of an association of gentisyl alcohol 5-O-glucoside with a specialist arthropod in *Populus*, and is, in fact, the first report of this metabolite in *Populus*. Salirepin (gentisyl alcohol 2-O-glucoside), a closely related metabolite, is a well-known constituent in *Populus* spp., (Busov *et al.*, 2006; Veach *et al.*, 2018; Tschaplinski *et al.*, 2019) and leaves of *P. deltoides* and *P. trichocarpa x deltoides* have a lower abundance, later eluting metabolite with a nearly identical fragmentation pattern to salirepin that we tentatively identify as gentisyl alcohol 2-O-glucoside. The QTL contains three putative candidate genes, including an aldehyde dehydrogenase 5F1 (Podel.10G175800) that may be involved in the reduction of gentisyl aldehyde to the alcohol, and two UDP-glycosyltransferases (Podel.10G184800, Podel.10G185000) that may be involved in the gentisyl alcohol conjugation to glucose. Specific substrates have yet to be determined for these genes, but a previous report suggests aldehyde dehydrogenase 5F1 genes are likely involved in the basic metabolism of *Populus* (Tian *et al.*, 2015).

Phyllocolpa sp. sawflies are considered a keystone species as the abandoned or unused leaf folds are often used as a habitat for many other species, such as aphids and spiders (Bailey & Whitham,

2007). The presence of folds in aspen forests is associated with a two-fold increase in arthropod species richness and around a four-fold increase in arthropod abundance relative to forests where the insect is absent (Bailey & Whitham, 2003). This in turn makes the host plant and sawfly relationship important in examining how shifts in the genes of a population ultimately structure whole communities, effectively linking ecology and evolutionary biology. Further investigation of this potential relationship could be key to connecting *Populus* genetics to the assemblage of the surrounding communities of organisms.

A striking finding in this study was an elevated number of recent tandem duplications in the *P. deltoides* genome, but not the *P. trichocarpa* genome, for the biotic QTL intervals. Out of the six chromosomes that yielded significant QTL results, four were associated with phenotypes that were fungi and insects native to the distribution of *P. deltoides*, but not *P. trichocarpa*. Given that *P. deltoides* has been co-evolving with the majority of the surveyed fungi and insects, it was not unexpected that there were more recent tandem duplicates in biotic intervals in its genome as there is more selective pressure on the native species to overcome biotic stress (Constabel & Lindroth, 2010; Newcombe *et al.*, 2010). However, given the high amount of novel resistance occurring in the progeny, recent tandem duplication may also be important in naïve host resistance.

In our study, we demonstrated how host plant genetics directly affect associated fungi and insects in the field, as well as how *Populus* progeny indirectly structured interactions between pathogens (Whitham *et al.*, 2006). The competition between *Melampsora* sp. and *S. musiva* highlights the complexities of how hybrid genetics are capable of strongly mediating multiple species interactions, which can result in inflated genetic associations. Finally, we have shown that many recent tandem duplications, found across biotic stress QTL intervals, have functional annotations

that are involved in host plant physical/chemical resistance and tolerance as well as a few that may be implicated in host plant susceptibility. The enrichment of recent tandem duplications is a signature of gene-for-gene interactions, and a mechanism that is essential to protect long-lived plants such as trees, enabling them to reach maturity despite many coevolving biotic stressors.

2.7 Literature cited

- Abdullah AS, Moffat CS, Lopez-Ruiz FJ, Gibberd MR, Hamblin J, Zerihun A. 2017.** Host–multi-pathogen warfare: pathogen interactions in co-infected plants. *Frontiers in Plant Science* **8**: 1806.
- Bailey JK, Whitham TG. 2003.** Interactions among elk, aspen, galling sawflies and insectivorous birds. *Oikos* **101**: 127–134.
- Bailey JK, Whitham TG. 2007.** Biodiversity is related to indirect interactions among species of large effect. In: *Ecological Communities: Plant Mediation in Indirect Interaction Webs*. Cambridge: Cambridge University Press, 306–328.
- Bergelson J, Kreitman M, Stahl EA, Tian D. 2001.** Evolutionary dynamics of plant R-genes. *Science* **292**: 2281–2285.
- Bird J, Faith DP, Rhomberg L, Riska B, Sokal RR. 1979.** The morphs of *Pemphigus populitransversus*: allocation methods, morphometrics, and distribution patterns. *Annals of the Entomological Society of America* **72**: 767–774.
- Boeckler GA, Gershenzon J, Unsicker SB. 2011.** Phenolic glycosides of the Salicaceae and their role as anti-herbivore defenses. *Phytochemistry* **72**: 1497–509.
- Broman KW, Wu H, Sen S, Churchill GA. 2003.** R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**: 889–890.
- Busov V, Meilan R, Pearce DW, Rood SB, Ma C, Tschaplinski TJ, Strauss SH. 2006.** Transgenic modification of *gai* or *rgll* causes dwarfing and alters gibberellins, root growth, and metabolite profiles in *Populus*. *Planta* **224**: 288–299.
- Chen XM. 2005.** Epidemiology and control of stripe rust [*Puccinia striiformis* f. sp. *tritici*] on wheat. *Canadian Journal of Plant Pathology* **27**: 314–337.
- Chen H, Kim HU, Weng H, Browse J. 2011.** Malonyl-CoA synthetase, encoded by ACYL

- ACTIVATING ENZYME13, is essential for growth and development of *Arabidopsis*. *The Plant Cell* **23**: 2247–62.
- Chen F, Liu C-J, Tschaplinski TJ, Zhao N. 2009.** Genomics of secondary metabolism in *Populus*: interactions with biotic and abiotic environments. *Critical Reviews in Plant Sciences* **28**: 375–392.
- Cheng Q, Zhang B, Zhuge Q, Zeng Y, Wang M, Huang M. 2006.** Expression profiles of two novel lipoxygenase genes in *Populus deltoides*. *Plant Science* **170**: 1027–1035.
- Chisholm ST, Coaker G, Day B, Staskawicz BJ. 2006.** Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* **124**: 803–814.
- Constabel CP, Lindroth RL. 2010.** The impact of genomics on advances in herbivore defense and secondary metabolism in *Populus*. In: Jansson S, Bhalerao R, Groover A, eds. Genetics and Genomics of *Populus*. NY: Springer New York, 279–305.
- Crutsinger GM, Rudman SM, Rodriguez-Cabal MA, McKown AD, Sato T, MacDonald AM, Heavyside J, Geraldine A, Hart EM, LeRoy CJ, *et al.* 2014.** Testing a ‘genes-to-ecosystems’ approach to understanding aquatic-terrestrial linkages. *Molecular Ecology* **23**: 5888–5903.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, *et al.* 2011.** The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.
- Dixon RA. 2001.** Natural products and plant disease resistance. *Nature* **411**: 843–847.
- Ehrlich PR, Raven PH. 1964.** Butterflies and plants: a study in coevolution. *Evolution* **18**: 586–608.
- Faith DP. 1979.** Strategies of gall formation in *Pemphigus* aphids. *Journal of the New York Entomological Society* **87**: 21–37.

- Fernández PC, Braccini CL, Dávila C, Barrozo RB, Aráoz MVC, Cerrillo T, Gershenzon J, Reichelt M, Zavala JA. 2019.** The use of leaf surface contact cues during oviposition explains field preferences in the willow sawfly *Nematus oligospilus*. *Scientific Reports* **9**: 4946.
- Floate KD. 2010.** Gall-inducing aphids and mites associated with the hybrid complex of cottonwoods, *Populus* spp. (Salicaceae) on Canada's grasslands. *Arthropods of Canadian Grasslands* **1**: 281–300.
- Flor HH. 1971.** Current status of the gene-for-gene concept. *Annual Review of Phytopathology* **9**: 275–296.
- Friesen TL, Meinhardt SW, Faris JD. 2007.** The *Stagonospora nodorum*-wheat pathosystem involves multiple proteinaceous host-selective toxins and corresponding host sensitivity genes that interact in an inverse gene-for-gene manner. *Plant Journal* **51**: 681–692.
- Fritz RS, Price PW. 1988.** Genetic variation among plants and insect community structure: willows and sawflies. *Ecology* **69**: 845–856.
- Gallun RL. 1977.** Genetic basis of Hessian fly epidemics. *Annals of the New York Academy of Sciences* **287**: 223–229.
- Gianola D, Norton HW. 1981.** Scaling threshold characters. *Genetics* **99**: 357–364.
- Goldman MH, Goldberg RB, Mariani C. 1994.** Female sterile tobacco plants are produced by stigma-specific cell ablation. *The EMBO Journal* **13**: 2976–2984.
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, *et al.* 2012.** Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40**: D1178–D1186.
- Harris MO, Stuart JJ, Mohan M, Nair S, Lamb RJ, Rohfritsch O. 2003.** Grasses and gall midges: plant defense and insect adaptation. *Annual Review of Entomology* **48**: 549–577.

- Haruta M, Major IT, Christopher ME, Patton JJ, Constabel CP. 2001.** A Kunitz trypsin inhibitor gene family from trembling aspen (*Populus tremuloides* Michx.): cloning, functional expression, and induction by wounding and herbivory. *Plant Molecular Biology* **46**: 347–359.
- Höfer R, Briesen I, Beck M, Pinot F, Schreiber L, Franke R. 2008.** The *Arabidopsis* cytochrome P450 *CYP86A1* encodes a fatty acid ω -hydroxylase involved in suberin monomer biosynthesis. *Journal of Experimental Botany* **59**: 2347–2360.
- Hulbert SH, Webb CA, Smith SM, Sun Q. 2001.** Resistance gene complexes: evolution and utilization. *Annual Review of Phytopathology* **39**: 285–312.
- Ignoffo CM, Granovsky AA. 1961a.** Life history and gall development of *Mordwilkoja vagabunda* (Homoptera: Aphidae) on *Populus deltoides*. Part II—gall development. *Annals of the Entomological Society of America* **54**: 635–641.
- Ignoffo CM, Granovsky AA. 1961b.** Life history and gall development of *Mordwilkoja vagabunda* (Homoptera: Aphidae) on *Populus deltoides*. *Annals of the Entomological Society of America* **54**: 486–499.
- Kaloshian I. 2004.** Gene-for-gene disease resistance: bridging insect pest and pathogen defense. *Journal of Chemical Ecology* **30**: 2419–2438.
- Keeling CI, 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.
- Kolehmainen J, Roininen H, Julkunen-Tiitto R, Tahvanainen J. 1994.** Importance of phenolic glucosides in host selection of shoot galling sawfly, *Euura amerinae*, on *Salix pentandra*. *Journal of Chemical Ecology* **20**: 2455–2466.
- Kopelke J-P. 2007.** The European species of the genus *Phyllocolpa*, Part I: The leucosticta-group

- (Insecta, Hymenoptera, Tenthredinidae, Nematinae). *Senckenbergiana Biologica* **87**: 75–109.
- Koskella B, Giraud T, Hood ME. 2006.** Pathogen relatedness affects the prevalence of within-host competition. *The American Naturalist* **168**: 121–126.
- Krishnan HB, Franceschi VR. 1988.** Anatomy of some leaf galls of *Rosa woodsii* (Rosaceae). *American Journal of Botany* **75**: 396–376.
- Larson KC, Whitham TG. 1991.** Manipulation of food resources by a gall-forming aphid: the physiology of sink-source interactions. *Oecologia* **88**: 15–21.
- Lefebvre V, Chèvre A. 1995.** Tools for marking plant disease and pest resistance genes: a review. *Agronomie* **15**: 3–19.
- Leister D. 2004.** Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends in Genetics* **20**: 116–122.
- Li H. 2011.** A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**: 2987–2993.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009.** The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
- Major IT, Constabel CP. 2008.** Functional analysis of the Kunitz trypsin inhibitor family in poplar reveals biochemical diversity and multiplicity in defense against herbivores. *Plant Physiology* **146**: 888–903.
- Martin GB, Bogdanove AJ, Sessa G. 2003.** Understanding the functions of plant disease resistance proteins. *Annual Review of Plant Biology* **54**: 23–61.
- McKown AD, Klapste J, Guy RD, Gerald A, Porth I, Hannemann J, Friedmann M,**

- Muchero W, Tuskan GA, Ehling J, et al. 2014.** Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*. *New Phytologist* **203**: 535–553.
- Meilan R, Han KH, Ma C, DiFazio SP, Eaton JA, Hoiem EA, Stanton BJ, Crockett RP, Taylor ML, James RR, et al. 2002.** The *CP4* transgene provides high levels of tolerance to Roundup® herbicide in field-grown hybrid poplars. *Canadian Journal of Forest Research* **32**: 967–976.
- Mello MO, Silva-Filho MC. 2002.** Plant-insect interactions: an evolutionary arms race between two distinct defense mechanisms. *Brazilian Journal of Plant Physiology* **14**: 71–81.
- Mercader RJ, Aardema ML, Scriber JM. 2009.** Hybridization leads to host-use divergence in a polyphagous butterfly sibling species pair. *Oecologia* **158**: 651–662.
- Muchero W, Guo J, DiFazio SP, Chen J-G, Ranjan P, Slavov GT, Gunter LE, Jawdy S, Bryan AC, Sykes R, et al. 2015.** High-resolution genetic mapping of allelic variants associated with cell wall chemistry in *Populus*. *BMC Genomics* **16**: 24.
- Muchero W, Sondreli KL, Chen JG, Urbanowicz BR, Zhang J, Singan V, Yang Y, Brueggeman RS, Franco-Coronado J, Abraham N, et al. 2018.** Association mapping, transcriptomics, and transient expression identify candidate genes mediating plant-pathogen interactions in a tree. *Proceedings of the National Academy of Sciences of the United States of America* **115**: 11573–11578.
- Nahrstedt A. 1989.** The significance of secondary metabolites for interactions between plants and insects. *Planta Medica* **55**: 333–338.
- Newcombe G. 1998.** A review of exapted resistance to diseases of *Populus*. *Forest Pathology* **28**: 209–216.
- Newcombe G, Martin F, Kohler A. 2010.** Defense and nutrient mutualisms in *Populus*. In:

- Jansson S, Bhalerao R, Groover A, eds. Genetics and Genomics of Populus. NY: Springer New York, 247–277.
- Newcombe G, Ostry M. 2001.** Recessive resistance to *Septoria* stem canker of hybrid poplar. *Phytopathology* **91**: 1081–1084.
- Newcombe G, Stirling B, Bradshaw Jr. HD. 2001.** Abundant pathogenic variation in the new hybrid rust *Melampsora* × *columbiana* on hybrid poplar. *Phytopathology* **91**: 981–985.
- Nieuwland J, Feron R, Huisman BAH, Fasolino A, Hilbers CW, Derksen J, Mariani C. 2005.** Lipid transfer proteins enhance cell wall extension in tobacco. *The Plant Cell Online* **17**: 2009–2019.
- Noel JP, Dixon RA, Pichersky E, Zubieta C, Ferrer JL. 2003.** Structural, functional, and evolutionary basis for methylation of plant small molecules. In: Romeo J, ed. Recent Advances in Phytochemistry. Oxford: Elsevier Science, 37–58.
- Nyman T, Widmer A, Roininen H. 2000.** Evolution of gall morphology and host-plant relationships in willow-feeding sawflies (Hymenoptera: Tenthredinidae). *Evolution* **54**: 526–533.
- Oliveira DC, Isaias RM, Fernandes GW, Ferreira BG, Carneiro RG, Fuzaro L. 2016.** Manipulation of host plant cells and tissues by gall-inducing insects and adaptive strategies used by different feeding guilds. *Journal of Insect Physiology* **1**: 103–113.
- Orton ES, Brown JKM. 2016.** Reduction of growth and reproduction of the biotrophic fungus *Blumeria graminis* in the presence of a necrotrophic pathogen. *Frontiers in Plant Science* **7**: 742.
- Panda N, Khush GS. 1995.** *Host plant resistance to insects*. Wallingford, UK: CAB International.
- Philippe RN, Ralph SG, Külheim C, Jancsik SI, Bohlmann J. 2009.** Poplar defense against insects: genome analysis, full-length cDNA cloning, and transcriptome and protein analysis

- of the poplar Kunitz-type protease inhibitor family. *New Phytologist* **184**: 865–884.
- Pinot F, Beisson F. 2011.** Cytochrome P450 metabolizing fatty acids in plants: characterization and physiological roles. *The FEBS journal* **278**: 195–205.
- Pretorius ZA, Singh RP, Wagoire WW, Payne TS. 2000.** Detection of virulence to wheat stem rust resistance gene *Sr31* in *Puccinia graminis* f. sp. *tritici* in Uganda. *Plant Disease* **84**: 203–203.
- Qin R, LeBoldus JM. 2014.** The infection biology of *Sphaerulina musiva*: clues to understanding a forest pathogen. *PLoS ONE* **9**: e103477.
- Ralph S, Oddy C, Cooper D, Yueh H, Jancsik S, Kolosova N, Philippe RN, Aeschliman D, White R, Huber D, et al. 2006.** Genomics of hybrid poplar (*Populus trichocarpa* × *deltoides*) interacting with forest tent caterpillars (*Malacosoma disstria*): normalized and full-length cDNA libraries, expressed sequence tags, and a cDNA microarray for the study of in. *Molecular Ecology* **15**: 1275–1297.
- Schuler MA. 2011.** P450s in plant–insect interactions. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1814**: 36–45.
- Slavov G, DiFazio S, Martin J, Schackwitz W, Muchero W, Rodgers-Melnick E, Lipphardt M, Pennacchio C, Hellsten U, Pennacchio L, et al. 2012.** Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*. *New Phytologist* **196**: 713–725.
- Smith DR, Fritz RS. 1996.** Review of the eastern United States species of the leaf-folding sawflies of the genus *Phyllocolpa* Benson (Hymenoptera: Tenthredinidae). *Proceedings of the Entomological Society of Washington* **98**: 695–707.
- Stajich JE. 2002.** The Bioperl Toolkit: perl modules for the life sciences. *Genome Research* **12**: 1611–1618.

- Stanton BJ, B. Neale D, Li S. 2010.** *Populus* breeding: from the classical to the genomic approach. In: Jansson S, Bhalerao R, Groover A, eds. *Genetics and Genomics of Populus*. Springer New York, 309–348.
- Taylor G. 2002.** *Populus: Arabidopsis* for forestry. Do we need a model tree? *Annals of Botany* **90**: 681–689.
- Thompson JN. 1988.** Coevolution and alternative hypotheses on insect/plant interactions. *Ecology* **69**: 893–895.
- Thompson JN, Burdon JJ. 1992.** Gene-for-gene coevolution between plants and parasites. *Nature* **360**: 121–125.
- Tian F-X, Zang J-L, Wang T, Xie Y-L, Zhang J, Hu J-J. 2015.** Aldehyde dehydrogenase gene superfamily in *Populus*: organization and expression divergence between paralogous gene pairs (A Palsson, Ed.). *PLoS ONE* **10**: e0124669.
- Tschaplinski TJ, Abraham PE, Jawdy SS, Gunter LE, Martin MZ, Engle NL, Yang X, Tuskan GA. 2019.** The nature of the progression of drought stress drives differential metabolomic responses in *Populus deltoides*. *Annals of Botany* **124**: 617–626.
- Veach AM, Yip D, Engle NL, Yang ZK, Bible A, Morrell-Falvey J, Tschaplinski TJ, Kalluri UC, Schadt CW. 2018.** Modification of plant cell wall chemistry impacts metabolome and microbiome composition in *Populus PdKOR1* RNAi plants. *Plant and Soil* **429**: 349–361.
- Verhoeven T, Feron R, Wolters-Arts M, Edqvist J, Gerats T, Derksen J, Mariani C. 2005.** STIG1 controls exudate secretion in the pistil of petunia and tobacco. *Plant Physiology* **138**: 153–160.
- Verkley GJM, Quaedvlieg W, Shin H-D, Crous PW. 2013.** A new approach to species delimitation in *Septoria*. *Studies in Mycology* **75**: 213–305.
- Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, et al.**

- 2012.** MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* **40**: 1–14.
- Weston DJ, Rogers A, Tschaplinski TJ, Gunter LE, Jawdy SA, Engle NL, Heady LE, Tuskan GA, Wullschlegel SD. 2015.** Scaling nitrogen and carbon interactions: what are the consequences of biological buffering? *Ecology and Evolution* **5**: 2839–2850.
- Whitham TG, Bailey JK, Schweitzer JA, Shuster SM, Bangert RK, LeRoy CJ, Lonsdorf E V., Allan GJ, DiFazio SP, Potts BM, *et al.* 2006.** A framework for community and ecosystem genetics: from genes to ecosystems. *Nature Reviews Genetics* **7**: 510–523.
- Wool D. 2004.** Gallling aphids: specialization, biological complexity, and variation. *Annual Review of Entomology* **49**: 175–192.
- Yang Y, Labbé J, Muchero W, Yang X, Jawdy SS, Kennedy M, Johnson J, Sreedasyam A, Schmutz J, Tuskan GA, *et al.* 2016.** Genome-wide analysis of lectin receptor-like kinases in *Populus*. *BMC Genomics* **17**: 699.
- Yin T-M, DiFazio SP, Gunter LE, Jawdy SS, Boerjan W, Tuskan GA. 2004a.** Genetic and physical mapping of *Melampsora* rust resistance genes in *Populus* and characterization of linkage disequilibrium and flanking genomic sequence. *New Phytologist* **164**: 95–105.
- Yin T-M, DiFazio SP, Gunter LE, Riemenschneider D, Tuskan GA. 2004b.** Large-scale heterospecific segregation distortion in *Populus* revealed by a dense genetic map. *Theoretical and Applied Genetics* **109**: 451–463.
- Zhao N, Guan J, Ferrer JL, Engle N, Chern M, Ronald P, Tschaplinski TJ, Chen F. 2010.** Biosynthesis and emission of insect-induced methyl salicylate and methyl benzoate from rice. *Plant Physiology and Biochemistry* **48**: 279–287.

2.8 Tables and Figures

Table 1. TPS and linear-mixed model output for biotic surveys. Broad-sense heritability (H^2)

denotes the contribution of all host plant genetic factors to total variance in the biotic phenotype.

R package RLRsim exactRLRT function was used to test significance of effects in mixed-model.

Fungus/Insect	TPS R^2	TPS RMSE	H^2	Genetic variance	Error variance	RLRatio	p-value
<i>Melampsora</i> sp. leaf rust	18.96%	1.037	0.609	0.271	0.174	294	< 0.0001
<i>S. musiva</i> leaf spot	15.33%	0.6757	0.250	0.153	0.459	32.0	< 0.0001
<i>Phyllocolpa</i> sp.	4.537%	5.469	0.391	11.9	18.5	129	< 0.0001

Table 2. Summary of QTL permutation test output. Percent variance in surveys for insects and fungi explained by significant marker indicated for composite interval mapping models. Positive (+) allele specifies genotype at significant interval that results in an increase in susceptibility. D indicates progeny are homozygous for *P. deltooides* alleles and T indicates progeny are heterozygous for *P. deltooides* and *P. trichocarpa* alleles.

Model	Chrom.	LOD score	p-value	% variance	+ allele	Garden
<i>Melampsora</i> sp. leaf rust all individuals	Chr04	19.7	0.001	52.5	D	Morgantown
<i>Melampsora</i> sp. leaf rust subsetted individuals	Chr04	30.1	<0.0001	57.9	D	Morgantown
<i>S. musiva</i> leaf spot all individuals	Chr04	7.80	0.001	5.29	D	Morgantown
<i>S. musiva</i> leaf spot subsetted individuals	Chr06	4.60	<0.0001	4.77	T	Morgantown
<i>S. musiva</i> canker	Chr16	4.34	0.022	NA	T	Morgantown
<i>M. vagabunda</i>	Chr05	5.39	0.001	NA	D	Morgantown
<i>P. populitransversus</i>	Chr03	2.70	>0.05	NA	NA	Morgantown
<i>Phyllocolpa</i> sp.	Chr10	6.60	0.015	9.65	T	Westport
<i>Phyllocolpa</i> sp.	Chr13	4.99	0.045	8.82	T	Westport
Gentisyl alcohol 5-O-glucoside	Chr17	12.2	<0.0001	25.760	T	Grand Rapids
Gentisyl alcohol 5-O-glucoside	Chr10	10.3	<0.0001	23.104	T	Grand Rapids
Gentisyl alcohol 5-O-glucoside	Chr14	5.96	0.0100	7.296	T	Grand Rapids

Table 3. One-way ANOVA, with Genotype as a covariate, analyzing the effect of infection severity of the genus of one fungus on the infection severity of the competing leaf fungi in 2008.

Factor	DF	Sum of squares	Mean square	F-ratio	p-value
Dependent Variable- <i>Melampsora</i> sp. leaf rust infection					
Genotype	732	1120	1.53	4.69	<0.0001
<i>S. musiva</i> score	3	34.4	11.5	35.2	<0.0001
Residuals	602	197	0.326		
Dependent Variable- <i>S. musiva</i> leaf spot infection					
Genotype	732	379	0.518	1.77	<0.0001
<i>Melampsora</i> sp. score	3	27.6	9.19	31.5	<0.0001
Residuals	602	176	0.292		

Table 4. Number of genes in QTL intervals in parental genomes for biotic associations.

Fungus/Insect	QTL Interval	# Genes in <i>P. trichocarpa</i>	# Genes in <i>P. deltoides</i>
<i>Melampsora</i> sp.	Chr04	159	147
<i>S. musiva</i> canker	Chr16	71	68
<i>S. musiva</i> leaf spot	Chr06	129	132
<i>Phyllocolpa</i> sp.	Chr10	146	154
<i>Phyllocolpa</i> sp.	Chr13	252	425
<i>M. vagabunda</i>	Chr05	129	147

Table 5. Candidate genes under positive selection ($K_a/K_s > 1$) in genetic intervals associated with fungi and insects.

QTL	Arabidopsis function	<i>P. trichocarpa</i> gene	<i>P. deltoides</i> gene	K_a/K_s
Chr04	Chitin elicitor receptor kinase 1	Potri.004G005800	Podel.04G004900.1.p	1.35
Chr04	Cysteine-rich RLK (RECEPTOR-like protein kinase)	Potri.004G012600	Podel.04G010900.1.p	2.23
Chr04	Stigma-specific Stig1 family protein	Potri.004G006800	Podel.04G005700.1.p	1.31
Chr04	Stigma-specific Stig1 family protein	Potri.004G007200.1	Podel.04G006100.1.p	1.00
Chr04	Stigma-specific Stig1 family protein	Potri.004G007200	Podel.04G006100.1.p	1.00
Chr05	Pentatricopeptide repeat (PPR) superfamily protein	Potri.005G038500.1	Podel.05G042200.1.p	1.52
Chr10	Copper amine oxidase family protein	Potri.010G088700	Podel.10G084500.1.p	1.19

Table 6. Tandem duplication profiles for genetic intervals. Number of copies next to species gene name indicates the size of tandem expansion for the gene.

Arabidopsis function	P. trichocarpa gene	Copy #	P. deltoides gene	Copy #	QTL Chr
ERD (early- responsive to dehydration stress) family protein	Potri.004G005900	3	Podel.04G005000.1.p	3	Chr04
Pentatricopeptide repeat (PPR) superfamily protein	NA	0	Podel.04G010100.1.p	2	Chr04
Peroxidase superfamily protein	Potri.004G015300	1	Podel.04G013500.1.p		Chr04
Disease resistance protein (TIR-NBS- LRR class)	Potri.005G031900	7	Podel.05G035500.1.p	12	Chr05
Shikimate O- Hydroxycinnamoyl-transferase	Potri.005G028000	3	Podel.05G029500.1.p	2	Chr05
Protein tyrosine kinase (Pkinase_Tyr) Leucine rich repeat (LRR_8)	Potri.005G030600	4	Podel.05G030600.1.p	2	Chr05
Lipoxygenase	Potri.005G032400	4	Podel.05G037200.1.p	3	Chr05

Cytochrome P450, family 721, subfamily A, polypeptide 1	Potri.005G034500	2	Podel.05G039300.1.p	1	Chr05
Cytochrome P450, family 76, subfamily G, polypeptide 1	Potri.005G029200	5	Podel.05G034500.1.p	0	Chr05
Cytochrome P450, family 93, subfamily D, polypeptide 1	Potri.005G037100	2	NA	0	Chr05
Cytochrome P450, family 76, subfamily G, polypeptide 1	NA	0	Podel.05G034500.1.p	4	Chr05
Disease resistance protein (TIR-NBS class), putative	Potri.005G032000	1	Podel.05G035800.1.p	12	Chr05
Receptor-like kinase in flowers	Potri.005G040200	1	Podel.05G043500.1.p	2	Chr05
Pentatricopeptide repeat (PPR) superfamily protein	Potri.010G083800	3	Podel.10G079300.1.p	3	Chr10

Copper amine oxidase family protein	Potri.010G088800	3	NA	0	Chr10
Ankyrin repeat family protein	Potri.013G133400	5	Podel.13G142400.1.p	10	Chr13
O-methyltransferase family protein	Potri.013G143700	4	Podel.13G181100.1.p	1	Chr13
O-methyltransferase family protein	NA	0	Podel.13G144000.1.p	4	Chr13
G-type lectin receptor-like protein kinase	Potri.016G102500	4	Podel.16G106900.1.p	4	Chr16

Figure 1. Biotic phenotype symptoms observed on trees including (a) leaf symptoms of *Melampsora* sp. fungal leaf rust, (b) leaf spot symptoms of the *S. musiva* fungus, (c) *S. musiva* canker symptoms, (d) branch gall created by the *M. vagabunda* aphid (e) petiole gall created by the *P. populitransversus* aphid, and (f) *Phyllocolpa* sp. leaf folding gall.

Figure 2. Collinear genes in *P. deltoides* (Pd) and *P. trichocarpa* (Pt) chromosomes, based on MCScanX alignments. (a) Chr02; (b) Chr05.

Figure 3. Change in number of tandemly duplicated genes discovered with increasing window size.

Figure 4. QTL interval plots showing peaks across the genome that associate with 2008 biotic surveys. Lines on the plots indicate p-value thresholds as determined by running 1000 permutations of mapping model for (a) *Melampsora* sp. model with all individuals and clones subset to exclude individuals with *S. musiva* infection and (b) *S. musiva* leaf spot model with all individuals and clones subset to exclude individuals with *Melampsora* sp. infection.

Figure 5. QTL interval plots showing peaks across the genome that associate with biotic surveys. Lines on the plots indicate p-value thresholds as determined by running 1000 permutations of mapping model for (a) binary fungal survey of *S. musiva* canker, (b) binary insect surveys of *M. vagabunda* and *P. populitransversus*, (c) and insect surveys of *Phyllocolpa* sp. leaf gall counts and overlapping peaks for gentisyl alcohol 5-O-glucoside compound QTL.

Figure 6. Mean and standard error of fungal infection for individuals with varying category levels of competing fungus infection. Letters indicate significantly different means as determined by Tukey's honest significance test for each one-way ANOVA test.

Figure 7. Comparison of gene content in *P. trichocarpa* grandparent 93-968 (left line) and *P. deltoides* grandparent Ill-101 (right line) for significant genetic intervals for (a) *Melampsora* sp. chromosome 4, (b) *S. musiva* chromosome 16, (c) *M. vagabunda* chromosome 5, (d) *Phyllocolpa* sp. chromosome 10, and (e) *Phyllocolpa* sp. chromosome 13 associations. QTL intervals were defined as 1 Mb regions centered on the marker with the highest LOD score. Size of gene point is relative to the number of genes in the tandem duplication expansion.

Figure 1.

Figure 2.

Figure 3.

Figure 4.

Figure 5.

Figure 6.

Figure 7.

**Chapter 3: Genetic underpinnings of arthropod community
distributions in *Populus trichocarpa***

3.1 Abstract

Community genetics seeks to understand the mechanisms by which natural genetic variation in plant populations can extend phenotypes to predictably establish entire assemblages of organisms. Until now, much of the research in this field has focused on plant genotypes in naturally occurring hybrid zones of *Populus*, which show elevated influence on arthropod communities, with less work completed on single species populations and even fewer having identified candidate genes controlling hierarchical ecosystem organization. We surveyed arthropod herbivores and predators on genotypes of an association population of *P. trichocarpa* planted in three common gardens. Plant genotype was influential in structuring arthropod community composition among all garden sites and years. Additionally, using GWAS and functional networks, we were able to identify several candidate genes mediating plant-arthropod community relationships. Genes associated with occurrence of *Phyllonorycter* sp. suggest insect manipulation of host tissue nutrition as well as a host resistance response to attack. Candidate genes important for higher-level organization of communities had more broadly applicable functions, such as terpenoid biosynthesis and production of dsRNA binding proteins and protein kinases, that may be capable of targeting multiple arthropod species. Through our research, we have shown that natural genetic variation in a single species, *P. trichocarpa*, is just as capable of strongly organizing the assemblage of herbivores and predators as hybrid populations.

3.2 Introduction

Genetic variation in a plant species is capable of extending beyond traditional phenotypes, ultimately structuring entire assemblages of organisms (Whitham *et al.*, 2003). The field of community genetics links ecology into this evolutionary framework to predict community structure and function as underlying genes shift within plant populations (Whitham *et al.*, 2006; Hughes *et al.*, 2008). Prior research of these relationships in agricultural systems has shown that plant genotype has a strong effect on herbivore distributions (Barbour *et al.*, 2016; Crutsinger, 2016). However, the impacts of genotype are expected to increase when a species of plant is either dominant or foundational in the environment (Ellison *et al.*, 2005; Whitham *et al.*, 2006). Given the depth of genetic information on *Populus* and their dominant role in riparian ecosystems, *Populus* trees have been a common model to explore the heritability of community composition (Wimp *et al.*, 2005; Shuster *et al.*, 2006; DeWoody *et al.*, 2013).

Arthropod communities of *Populus* are known to be structured by a number of heritable plant, traits including tissue chemistry, morphology, and phenology (Whitham, 1978; Floate *et al.*, 1993; Boeckler *et al.*, 2011). Surveys of *Populus* populations showed that plant genotype is capable of structuring arthropod communities (Wimp *et al.*, 2005), which in turn can cascade out into additional trophic levels such as the distribution of avian predators (Bailey *et al.*, 2006) and fungi (Dickson & Whitham, 1996). Similarly, *Populus* genotypes influence beaver foraging (Bailey *et al.*, 2004), which can indirectly benefit leaf feeding beetles (Martinsen *et al.*, 1998). In addition to these multi-trophic interactions, genetic variation in tree populations has also been able to foster cryptic speciation in specialist mites (Evans *et al.*, 2008).

A noticeable pattern in early studies is the use of trees from natural *Populus* hybrid zones for community genetics research. Hybrids tend to display stronger community effects due to higher susceptibility/resistance to fungal and insect associations (Whitham *et al.*, 1999). The hybrid relationship with the tree insect community is so elevated it has been suggested as an alternative mechanism to distinguish closely-related plant taxa in *Populus* instead of costly genetic markers (Floate & Whitham, 1995). More recent work has established similar community interactions are also detectable in single species populations of *Populus*, including structuring arthropod-avian predator distributions (Smith *et al.*, 2011), microbes and lichens (Lamit *et al.*, 2015), as well as aquatic-terrestrial linkages (Crutsinger *et al.*, 2014). However, for many of these studies, the reliance on estimates of genotypic richness, and assumption of equal relatedness among genets, limited the ability to identify the genes underlying arthropod community interactions (Barker *et al.*, 2019; Wimp *et al.*, 2019).

The abundant genetic resources of *Populus trichocarpa* make it ideal for discovery of genes that contribute to community composition of arthropods. With a full reference genome sequence (Tuskan *et al.*, 2006) and the sequencing of over 1,100 individual trees collected throughout the species range, a genetic map containing millions of segregating single nucleotide polymorphisms (SNPs) is available to answer the “genes-to-ecosystem” questions (Evans *et al.*, 2014). Coupling these genetic resources with traditional single-trait genome-wide association analysis (GWAS), new multi-trait GWAS methods (Chhetri *et al.*, 2019), and networks built from a diverse set of plant phenotypes will allow us to discover previously unknown genes contributing to arthropod community composition.

In this study, we surveyed arthropod communities across three common gardens and two years to address three main questions: (1) Is there heritable genetic control of individual arthropods, calculated community metrics, and arthropod community composition? (2) Which genes underlie these associations? (3) What possible biological functions could candidate genes confer based on co-expression networks?

3.3 Methods

Association population plantation

A total of 1,100 *P. trichocarpa* genotypes were previously collected from across the natural range of *P. trichocarpa* from northern California to northern British Columbia (Slavov *et al.*, 2012; Evans *et al.*, 2014). These were planted in 2009 in a randomized, three block design with three replicates of each genotype in Clatskanie, OR, Corvallis, OR and Placerville, CA (Figure 1). The trees were planted at a 2 m x 3 m spacing in Corvallis, and 3 m x 3 m in the other two sites.

Arthropod surveys

Surveys were conducted in the Clatskanie, Placerville, and Corvallis plantations in July 2012 and again in the Corvallis plantation in July 2015. The 2012 plantation surveys were performed on five branches of approximately equal size for each tree while surveys conducted in 2015 were performed on ten branches. All insects and leaf modifier damage present on surveyed branches were counted and recorded (Bangert *et al.*, 2008) (Table 1). Surveys completed in Clatskanie in 2012 consisted of 712 trees and a total of 654 genotypes (36 with replicate observations), in Corvallis 102 trees and 40 genotypes (35 with replicate observations) were

observed, and in Placerville 100 trees and 40 genotypes (34 with replicate observations) were observed. Surveys completed in Corvallis in 2015 consisted of 557 trees and 496 genotypes (61 with replicate observations).

Genotypic data collection

Methods for obtaining genotypic data were described elsewhere previously (Chhetri *et al.*, 2019; Weighill *et al.*, 2019). Briefly the genotypes were sequenced to a minimum of 15X depth for a total of 1053 trees using Illumina genetic analyzers at the DOE Joint Genome Institute. We removed trees that were related more closely than first cousins and removed highly differentiated California trees. This left a total of 869 trees that were used for the GWAS analyses. We further removed SNPs with minor allele frequency ≤ 0.05 and markers with severe departures from Hardy-Weinberg expectations.

3.4 Statistical analysis

Within-garden microsite variation was removed from survey datasets using thin-plate spline regression (*fields* R package) for insect species that were abundant in plantations. Residuals of the spatial correction were added to the mean of each trait to spatially correct values. For genotypes with replicate observations, the genotypic component of variance in insect traits was estimated using a linear-mixed model (*lmer* function of the *lme4* R package) with the spatially corrected insect trait as the response, tree genotype as the predictor, and a stem diameter proxy for biomass (Ruark *et al.*, 1987) as a covariate, where applicable. Broad-sense heritability was calculated as:

$$H^2 = \delta_g^2 / (\delta_g^2 + \delta_e^2)$$

where δ^2_g is the genetic variance due to genotype and δ^2_e is the residual variance. Rapid, simulation-based exact likelihood ratio tests were used to evaluate the significance of variation due to genotype for each linear model (*exactRLRT* function of the *RLRsim* R package). The same models were also used to estimate heritability of calculated community metrics from the original dataset including abundance (sum of all insects), richness (total number of unique species), and Shannon-Weaver diversity.

Arthropod community composition was evaluated using non-metric multidimensional scaling (NMDS) with Bray-Curtis dissimilarity both across all gardens surveyed for the same 44 genotypes in 2012 and within each garden across sites and years. A PERMANOVA (*adonis* function of the *vegan* R package) was run for replicated genotypes to determine the influence of tree genotypes on arthropod community composition among all NMDS configurations. Coordinates for dependent variables were extracted from the NMDS configurations and depicted in plots using a rescaled font to represent the three-dimensional projection. This was accomplished by rescaling all axis scores to a zero origin and scaling the font size relative to the product of the three rescaled axis scores.

Association and network analysis

Genome-wide association tests were performed using the Genome-wide Efficient Mixed Model Association package (GEMMA) (Zhou & Stephens, 2012, 2014). The GWAS association test included phenotypic BLUPs, a genetic relationship matrix, and 6,741,160 SNPs. The genetic relationship matrix was also estimated using GEMMA. Single-trait GWAS was performed for insect and community traits from the Clatskanie 2012 surveys (654 genotypes) and Corvallis

2015 surveys (496 genotypes) (Table 2). Insect traits including *Chrysomela scripta*, *Phyllonorycter* sp. blotch miner, *Phyllocnistis populiella* serpentine miner, and *Corythucha* sp. lace wing counts were binarized for analysis given the non-normal distribution of their occurrence in the field. The tested model was:

$$y = W\alpha + x\beta + u + \epsilon$$

where y is an n -vector of phenotypic BLUP values, where n is the number of individuals tested; W is an $n \times c$ matrix of covariates; α is a c -vector of corresponding coefficients, where c is the number of principal coordinate axes used; x is an n -vector of marker genotypes, β is the effect size of the marker, u is an n -vector of random effects that includes a relatedness matrix and ϵ is an n -vector of errors.

Multitrait GWAS was performed for two multitrait sets, including the NMDS configuration from the Clatskanie 2012 (654 genotypes) and Corvallis 2015 (496 genotypes) sites. We combined 5 NMDS axes obtained from the community genetics analyses independently for each of the plantations in Corvallis and Clatskanie. Multitrait association was conducted with GEMMA using the same model as for single trait associations, except y is an $n \times d$ matrix of d phenotypes for n individuals. We used an FDR cutoff of 0.05 and a more liberal FDR cutoff of 0.1 to identify suggestive associations (Storey & Tibshirani, 2003).

To assess possible biological function of candidate genes identified for insect/community metric associations identified from GWAS, networks were constructed for genes from the same surveyed populations. Networks were based on genes that were co-expressed with candidate genes using the JGI Plant Gene Atlas for *P. trichocarpa* (<https://phytozome.jgi.doe.gov>),

methylation data for multiple *P. trichocarpa* tissue types (Vining *et al.*, 2012), as well as GWAS for pyMBMS (Weighill *et al.*, 2018) and metabolite profiles as determined by GC-MS (Tschaplinski *et al.*, 2012). Further detail of network construction can be found in Weighill *et al.* (2018). Candidate genes of functional interest to insect traits were selected using strength of SNP association (p-values) and SNP distance from gene determined from GWAS. Networks were merged based on shared nodes. Candidate genes with significant GWAS associations and with depth (number of connected layers) greater than 2 and breadth (number of neighbor genes) greater than 500 in the merged networks were selected for further analysis. Enrichment of GO functional categories in corresponding one-hop subnetworks was tested using Fisher's Exact Tests with a significance threshold of $FDR \leq 0.1$. Final networks were visualized using Cytoscape v.3.6.1 (Shannon *et al.*, 2003).

3.5 Results

Broad-sense heritability and NMDS configuration

Significant broad-sense heritability was detected for all insect and community traits in one or more of the surveys across gardens and years (Table 2). The abundance of the *Harmandia* sp. trait was heritable in two garden surveys in 2012 including Clatskanie ($H^2 = 0.511$, p-value < 0.001) and Corvallis ($H^2 = 0.394$, p-value < 0.001). In the 2015 Corvallis surveys, only two insect traits were genetically heritable, including *Phyllocolpa* sp. ($H^2 = 0.237$, p-value = 0.043) and *Corythucha* sp. ($H^2 = 0.145$, p-value = 0.05). Abundance of insects occurring on branches was significantly heritable in Placerville in 2012 ($H^2 = 0.181$, p-value = 0.042) and Corvallis in 2015 ($H^2 = 0.268$, p-value = 0.016). Insect richness was also heritable in Placerville in 2012 ($H^2 = 0.217$, p-value = 0.037). Shannon-Weaver diversity was heritable in Clatskanie in 2012 ($H^2 =$

0.231, p-value = 0.05), Placerville in 2012 ($H^2 = 0.244$, p-value = 0.026), and Corvallis in 2012 ($H^2 = 0.191$, p-value = 0.046).

The number of NMDS dimensions required to adequately explain dissimilarities in insect community was 4 across all gardens, and within each garden was 5 for Clatskanie 2012, 3 for Placerville 2012, 3 for Corvallis in 2012, and 5 for Corvallis in 2015. PERMANOVA revealed significant garden effects across 2012 surveys (Table 3; Figure 2) and significant genetic effects on arthropod community composition among all NMDS configurations (Table 4; Figure 3,4).

GWAS and network results

A total of 71 loci were significantly associated with insect and community traits. *P. populiella* yielded 21 significant SNPs across 19 loci and 13 chromosomes from the Clatskanie survey (Figure 5a; Table 5). *Phyllonorycter* sp. blotch miner yielded a combined total of 31 significant SNPs across 27 loci and 14 chromosomes; of these, 30 SNPs and 26 loci were discovered from the Clatskanie 2012 survey and 1 SNP and 1 locus was discovered from the Corvallis 2015 survey (Figure 5a; Table 6). Insect richness was associated with a combined total of 11 SNPs across 9 loci on 8 separate chromosomes; of these, 6 SNPs and 5 loci were discovered from the Clatskanie 2012 survey and 5 SNPs and 4 loci were discovered from the Corvallis 2015 survey. Insect abundance in the Clatskanie site also yielded 4 significant SNPs across 3 loci and 2 chromosomes (Figure 5b; Table 7). Both the Clatskanie and Corvallis sites yielded significant associations for the insect community composition (Figure 5c; Table 8). There were 11 significant SNPs across 4 loci and 4 chromosomes from the Clatskanie GWAS analysis, while the Corvallis site yielded 13 SNPs across 9 loci and 9 chromosomes. Of the GWAS SNPs which

were located on/near 66 *P. trichocarpa* genes, 6 genes were identified from networks as strong candidates for functional importance (Table 9) including Clatskanie survey of *Phyllonorycter* sp. blotch miner (Figure 6), Corvallis arthropod richness (Figure 7), and Corvallis community composition (Figure 8). Significant GO term functional enrichments of co-expressed genes from networks were detected for 4 of the 6 candidate genes (Table 10).

3.6 Discussion

Host plant genetics can have predictable effects on community assembly of living organisms at multiple hierarchical levels (Crutsinger *et al.*, 2008, 2009; Barbour *et al.*, 2009), especially in naturally occurring hybrid zones of *Populus* (Wimp *et al.*, 2005; Whitham *et al.*, 2006, 2012). In our study, we have demonstrated, across surveys of arthropods in multiple gardens and years, that host genotype has a significant effect on the community of associated arthropods, even with the unique assemblages of insects at each location. Additionally, we were able leverage the genomic resources of *P. trichocarpa* to identify the underlying genes and potential functions that are important in several levels of arthropod community assembly, including the distribution of the leaf mining specialist arthropod *Phyllonorycter* sp., the number of species (richness) capable of co-existing on a single genotype, and community composition.

Phyllonorycter sp. larvae develop in the leaf mesophyll and their first three early instar stages rely on sap-feeding while the last two feed on leaf tissue (Davis & Deschka, 2001). Two candidate genes that were associated with *Phyllonorycter* sp. in the Clatskanie site appear to both be important in cell proliferation processes that may impact feeding success of the blotch miner. A gene located on Chr01 was identified to be cyclin B1;4 and is known to be an important cell

cycle regulator in *Populus* (Dong *et al.*, 2011; Plavcova *et al.*, 2013), with further support for this function in the form of co-expression enrichment in our networks with processes involved in chromosome segregation/organization and multicellular development. Gall-forming and mining insects have been shown to be directly capable of manipulating molecular mechanisms important in cell proliferation via phytohormones (Erb *et al.*, 2012). Furthermore, some species of *Phyllonorycter* have already been implicated in indirectly manipulating cytokinin levels in host plant tissues, thereby preventing senescence and increasing the nutritive value of leaves for larval development (Kaiser *et al.*, 2010; Erb *et al.*, 2012).

Interestingly, the same candidate gene on Chr01 was directly associated with two metabolites including the dihydromyricetin. Although dihydromyricetin has been found to be detrimental to the survival of generalist herbivores (Peters & Constabel, 2002), flavonoids are also capable of being important ovipositional stimulants to specialist herbivores (Simmonds, 2001). An additional candidate gene located on Chr18 and identified to be an adenine phosphoribosyl transferase-like protein, may have implications in suppressing cell division. Adenine phosphoribosyl transferases are capable of utilizing cytokinins as substrates, converting them into inactive intermediate or storage forms (Allen *et al.*, 2002; Plavcova *et al.*, 2013). Moreover, the same gene has been associated with production of alanine, which, when used in the modification of cytokinin, has the ability to permanently deactivate the phytohormone (Srivastava, 2002), possibly preventing the tissue manipulating behavior of the leaf miners.

Although there were no definitive results for individual insect species with abundant distributions in the Corvallis garden site, we discovered multiple gene associations with two

community metrics. Arthropod richness was significantly associated with two potential candidate genes located on Chr07 and Chr18. Arthropod richness refers to the total number of unique species observed to be interacting with plant tissues, including a combination of insect predators as well as specialist and generalist herbivores and damage types. The gene located on Chr07 was identified to be an RNA-binding protein Musashi (MSI). Little is known regarding MSI protein function in plants, but the subnetwork of connected genes was enriched for RNA- and histone-binding functions. RNA-binding proteins are generally important in post-transcriptional gene regulation as they guide RNA stability/transport and affect chromatin modification and also have implications in abscisic acid hormone signaling (Thaler & Bostock, 2004; Lorković, 2009).

The gene on Chr18 that was associated with insect richness encodes a protein that is similar to a cytochrome P450 enzyme. Plant cytochrome P450's are known to have a wide diversity of biological catalyst functions. In *Populus* they have been frequently associated with recovery from abiotic stresses, such as drought and salt (Gu *et al.*, 2004; Chen *et al.*, 2013; Wang *et al.*, 2016), as well as dealing with biotic stress (Irmisch *et al.*, 2013, 2014a). The same cytochrome P450 gene was associated with an unidentified metabolite. The associated subnetwork was enriched for terpenoid (isoprenoid) biosynthetic processes as well as molecular mevalonate kinase activity. This may provide valuable clues about the possible identity of the metabolite as well as the function of this uncharacterized gene.

Cytochrome P450 genes are known to be involved in terpenoid biosynthesis in plants (Keeling & Bohlmann, 2006). The relationship of terpenoids and the number of species observed on plant tissues is highly dependent on the evolutionary history of the insect with its host. Generalist

insects that are not adapted to *Populus* chemistry will potentially be deterred by constitutive/induced terpenoid compounds, as consumption of these can have detrimental effects on their physiology and development (Wada & Munakata, 1971; Huber & Bohlmann, 2004; Keeling & Bohlmann, 2006). Conversely, specialist insects that have coevolved to utilize *Populus* as a host are capable of utilizing terpenoid compounds as oviposition/location cues (Brilli *et al.*, 2009). Finally, terpenoids released as volatiles from crushed plant tissues have been found to be important indirect plant defenses as they can aid in the attraction of predators and parasites of insect herbivores (Huber & Bohlmann, 2004; Irmisch *et al.*, 2014b). This is particularly interesting in the case of our surveys given that most predators present in the garden site were made up of lady bugs (Coccinellidae) and spiders belonging to Salticidae and Araneidae families. There are many examples of the importance of plant volatiles to lady bug foraging of aphids (Ninkovic *et al.*, 2001; de Vos & Jander, 2010), whereas the relationship between spiders and volatiles has not been as thoroughly explored. However, more literature is being published supporting the importance of volatile compounds in the ability of spiders to locate and hunt prey (Zhao *et al.*, 2002; Nelson *et al.*, 2012).

We also identified two promising candidate genes that were mediating the arthropod community composition. One gene, located on Chr17, is a double-stranded RNA (dsRNA) binding protein. dsRNA-binding proteins have been implicated in RNA interference (RNAi), which is capable of silencing genes in not only herbivore insects but also disease vectors, including fungal pathogens, viruses, and bacteria (Niu *et al.*, 2010; Yoon *et al.*, 2018). The wide diversity of potential targets of dsRNA-binding proteins could enable them to act as a potential first response to biotic stress. Moreover, the subnetwork contained sucrose transporter genes that were

negatively co-expressed with the dsRNA binding protein. This could cause a reduction in nutritive value of the tissue, thereby providing a generic response that would reduce attractiveness of the tissue to a broad spectrum of herbivores (Niu *et al.*, 2010).

Another gene associated with community composition was a wall-associated receptor kinase C-terminal gene that acts as a transmembrane protein. This gene family is greatly expanded in *Populus* and has been implicated in pathogen responses in several plant systems (Li *et al.*, 2009; Tocquard *et al.*, 2014; Hurni *et al.*, 2015). Although pathogens were not a part of these surveys, herbivores are capable of predisposing trees to secondary pathogen infection and pathogen infection can weaken plant immune systems making them more susceptible to herbivore feeding (Busby *et al.*, 2019). As a result, the function of wall-associated receptor kinases may be important as a general response to biotic stress preventing secondary infections due to weakened immune systems and mass attack.

Overall, we have shown across three common gardens, each of which had unique assemblages of arthropod species, that plant genotype within a single species has strong influence over its community of herbivores and predators. We also discovered several candidate genes in *P. trichocarpa* important in underlying individual herbivore distributions, total number of species on plant tissues, and community composition. Candidates mediating the specialist herbivore *Phyllonorycter* sp. blotch miner reflected very targeted responses important in the insect's ability to manipulate host plant growth and the potential host response to resist attack. Conversely, genes that were associated with species richness and community composition have potential functions, such as terpenoid and dsRNA-binding protein production, that may be capable of

more broadly targeting groups of insects. Future work should aim to confirm the functions of these candidates as well as determine how genes of similar function may contribute to community genetics effects in additional pure species of *Populus*.

3.7 Literature cited

- Allen M, Qin W, Moreau F, Moffatt B. 2002.** Adenine phosphoribosyltransferase isoforms of *Arabidopsis* and their potential contributions to adenine and cytokinin metabolism. *Physiologia Plantarum* **115**: 56–68.
- Bailey JK, Schweitzer JA, Rehill BJ, Lindroth RL, Martinsen GD, Whitham TG. 2004.** Beavers as molecular geneticists: a genetic basis to the foraging of an ecosystem engineer. *Ecology* **85**: 603–608.
- Bailey JK, Wooley SC, Lindroth RL, Whitham TG. 2006.** Importance of species interactions to community heritability: a genetic basis to trophic-level interactions. *Ecology Letters* **9**: 78–85.
- Bangert, RK, Lonsdorf, EV, Wimp, GM, Shuster, SM, Fischer, D, Schweitzer, JA, Whitham, TG. 2008.** Genetic structure of a foundation species: scaling community phenotypes from the individual to the region. *Heredity* **100**: 121-131.
- Barbour MA, Fortuna MA, Bascompte J, Nicholson JR, Julkunen-Tiitto R, Jules ES, Crutsinger GM. 2016.** Genetic specificity of a plant–insect food web: implications for linking genetic variation to network complexity. *Proceedings of the National Academy of Sciences* **113**: 2128–2133.
- Barbour RC, Storer MJ, Potts BM. 2009.** Relative importance of tree genetics and microhabitat on macrofungal biodiversity on coarse woody debris. *Oecologia* **160**: 335–342.
- Barker HL, Riehl JF, Bernhardsson C, Rubert-Nason KF, Holeski LM, Ingvarsson PK, Lindroth RL. 2019.** Linking plant genes to insect communities: identifying the genetic bases of plant traits and community composition. *Molecular Ecology* **28**: 4404–4421.

- Boeckler GA, Gershenzon J, Unsicker SB. 2011.** Phenolic glycosides of the Salicaceae and their role as anti-herbivore defenses. *Phytochemistry* **72**: 1497–1509.
- Brilli F, Ciccioli P, Frattoni M, Prestininzi M, Spanedda AF, Loreto F. 2009.** Constitutive and herbivore-induced monoterpenes emitted by *Populus × euroamericana* leaves are key volatiles that orient *Chrysomela populi* beetles. *Plant, Cell & Environment* **32**: 542–552.
- Busby PE, Crutsinger G, Barbour M, Newcombe G. 2019.** Contingency rules for pathogen competition and antagonism in a genetically based, plant defense hierarchy. *Ecology and Evolution* **9**: 6860–6868.
- Chen J, Song Y, Zhang H, Zhang D. 2013.** Genome-wide analysis of gene expression in response to drought stress in *Populus simonii*. *Plant Molecular Biology* **31**: 946–962.
- Chhetri HB, Macaya-Sanz D, Kainer D, Biswal AK, Evans LM, Chen J, Collins C, Hunt K, Mohanty SS, Rosenstiel T, et al. 2019.** Multitrait genome-wide association analysis of *Populus trichocarpa* identifies key polymorphisms controlling morphological and physiological traits. *New Phytologist* **223**: 293–309.
- Crutsinger GM. 2016.** A community genetics perspective: opportunities for the coming decade. *New Phytologist* **210**: 65–70.
- Crutsinger GM, Cadotte MW, Sanders NJ. 2009.** Plant genetics shapes inquiline community structure across spatial scales. *Ecology Letters* **12**: 285–292.
- Crutsinger GM, Reynolds WN, Classen AT, Sanders NJ. 2008.** Disparate effects of plant genotypic diversity on foliage and litter arthropod communities. *Oecologia* **158**: 65–75.
- Crutsinger GM, Rudman SM, Rodriguez-Cabal MA, McKown AD, Sato T, MacDonald AM, Heavyside J, Geraldine A, Hart EM, LeRoy CJ, et al. 2014.** Testing a ‘genes-to-ecosystems’ approach to understanding aquatic–terrestrial linkages. *Molecular Ecology*

23: 5888–5903.

- Davis DR, Deschka G. 2001.** Biology and systematics of the North American *Phyllonorycter* leafminers on Salicaceae, with a synoptic catalog of the Palearctic species (Lepidoptera: Gracillariidae). *Smithsonian Contributions to Zoology* **614**: 1–89.
- DeWoody J, Viger M, Lakatos F, Tuba K, Taylor G, Smulders MJ. 2013.** Insight into the genetic components of community genetics: QTL mapping of insect association in a fast-growing forest tree. *PLoS ONE* **8**: e79925.
- Dickson LL, Whitham TG. 1996.** Genetically-based plant resistance traits affect arthropods, fungi, and birds. *Oecologia* **106**: 400–406.
- Dong Q, Zhao Y, Jiang H, He H, Zhu S, Cheng B, Xiang Y. 2011.** Genome-wide identification and characterization of the cyclin gene family in *Populus trichocarpa*. *Plant Cell, Tissue and Organ Culture (PCTOC)* **107**: 55–67.
- Ellison AM, Bank MS, Clinton BD, Colburn EA, Elliott K, Ford CR, Foster DR, Kloeppel BD, Knoepp JD, Lovett GM, et al. 2005.** Loss of foundation species: consequences for the structure and dynamics of forested ecosystems. *Frontiers in Ecology and the Environment* **3**: 479–486.
- Erb M, Meldau S, Howe GA. 2012.** Role of phytohormones in insect-specific plant reactions. *Trends in Plant Science* **17**: 250–259.
- Evans LM, Allan GJ, Shuster SM, Woolbright SA, Whitham TG. 2008.** Tree hybridization and genotypic variation drive cryptic speciation of a specialist mite herbivore. *Evolution: International Journal of Organic Evolution* **62**: 3027–3040.
- Evans LM, Slavov GT, Rodgers-Melnick E, Martin J, Ranjan P, Muchero W, Brunner AM, Schackwitz W, Gunter L, Chen JG, et al. 2014.** Population genomics of *Populus*

- trichocarpa* identifies signatures of selection and adaptive trait associations. *Nature Genetics* **46**: 1089.
- Floate KD, Kearsley MJ, Whitham TG. 1993.** Elevated herbivory in plant hybrid zones: *Chrysomela confluenta*, *Populus* and phenological sinks. *Ecology* **74**: 2056–2065.
- Floate KD, Whitham TG. 1995.** Insects as traits in plant systematics: their use in discriminating between hybrid cottonwoods. *Canadian Journal of Botany* **73**: 1–13.
- Gu R, Fonseca S, Puskás LG, Hackler Jr L, Zvara Á, Dudits D, Pais MS. 2004.** Transcript identification and profiling during salt stress and recovery of *Populus euphratica*. *Tree Physiology* **24**: 265–276.
- Huber DP, Bohlmann J. 2004.** Terpene synthases and the mediation of plant-insect ecological interactions by terpenoids: a mini-review. In: Cronk Q, Whitton J, Ree R, Taylor I, eds. Plant adaptation: molecular genetics and ecology. Vancouver, Canada: National Research Council Canada, 70–81.
- Hughes AR, Inouye BD, Johnson MT, Underwood N, Vellend M. 2008.** Ecological consequences of genetic diversity. *Ecology Letters* **11**: 609–623.
- Hurni S, Scheuermann D, Krattinger SG, Kessel B, Wicker T, Herren G, Fitze MN, Breen J, Presterl T, Ouzunova M, et al. 2015.** The maize disease resistance gene Htn1 against northern corn leaf blight encodes a wall-associated receptor-like kinase. *Proceedings of the National Academy of Sciences* **112**: 8780–8785.
- Irmisch S, Clavijo McCormick A, Günther, J., Schmidt A, Boeckler GA, Gershenzon J, Unsicker SB, Köllner TG. 2014a.** Herbivore-induced poplar cytochrome P450 enzymes of the CYP 71 family convert aldoximes to nitriles which repel a generalist caterpillar. *The Plant Journal* **80**: 1095–1107.

- Irmisch S, Jiang Y, Chen F, Gershenzon J, Köllner T. 2014b.** Terpene synthases and their contribution to herbivore-induced volatile emission in western balsam poplar (*Populus trichocarpa*). *BMC Plant Biology* **14**: 270.
- Irmisch S, McCormick AC, Boeckler GA, Schmidt A, Reichelt M, Schneider B, Block K, Schnitzler JP, Gershenzon J, Unsicker SB, *et al.* 2013.** Two herbivore-induced cytochrome P450 enzymes CYP79D6 and CYP79D7 catalyze the formation of volatile aldoximes involved in poplar defense. *The Plant Cell* **25**: 4737–4754.
- Kaiser W, Huguet E, Casas J, Commin C, Giron D. 2010.** Plant green-island phenotype induced by leaf-miners is mediated by bacterial symbionts. *Proceedings of the Royal Society B: Biological Sciences* **277**: 2311–2319.
- Keeling CI, 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.
- Lamit LJ, Busby PE, Lau MK, Compson ZG, Wojtowicz T, Keith AR, Zinkgraf MS, Schweitzer JA, Shuster SM, Gehring CA, *et al.* 2015.** Tree genotype mediates covariance among communities from microbes to lichens and arthropods. *Journal of Ecology* **103**: 840–850.
- Li H, Zhou SY, Zhao WS, Su SC, Peng YL. 2009.** A novel wall-associated receptor-like protein kinase gene, *OsWAK1*, plays important roles in rice blast disease resistance. *Plant Molecular Biology* **69**: 337–346.
- Lorković ZJ. 2009.** Role of plant RNA-binding proteins in development, stress response and genome organization. *Trends in Plant Science* **14**: 229–236.
- Martinsen GD, Driebe EM, Whitham TG. 1998.** Indirect interactions mediated by changing

- plant chemistry: beaver browsing benefits beetles. *Ecology* **79**: 192–200.
- Nelson XJ, Pratt AJ, Cheseto X, Torto B, Jackson RR. 2012.** Mediation of a plant-spider association by specific volatile compounds. *Journal of Chemical Ecology* **38**: 1081–1092.
- Ninkovic V, Al Abassi S, Pettersson J. 2001.** The influence of aphid-induced plant volatiles on ladybird beetle searching behavior. *Biological Control* **21**: 191–195.
- Niu JH, Jian H, Xu JM, Guo YD, Liu Q. 2010.** RNAi technology extends its reach: engineering plant resistance against harmful eukaryotes. *African Journal of Biotechnology* **9**: 7573–7582.
- Peters DJ, Constabel CP. 2002.** Molecular analysis of herbivore-induced condensed tannin synthesis: cloning and expression of dihydroflavonol reductase from trembling aspen (*Populus tremuloides*). *The Plant Journal* **32**: 701–712.
- Plavcova L, Hacke UG, Almeida-Rodriguez AM, Li E, Douglas CJ. 2013.** Gene expression patterns underlying changes in xylem structure and function in response to increased nitrogen availability in hybrid poplar. *Plant, Cell & Environment* **36**: 186–199.
- Ruark, GA, Martin, GL, Bockheim, JG. 1987.** Comparison of constant and variable allometric ratios for estimating *Populus tremuloides* biomass. *Forest Science* **33**: 294–300.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, 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.
- Shuster SM, Lonsdorf EV, Wimp GM, Bailey JK, Whitham TG. 2006.** Community heritability measures the evolutionary consequences of indirect genetic effects on community structure. *Evolution* **60**: 991–1003.
- Simmonds MS. 2001.** Importance of flavonoids in insect–plant interactions: feeding and

- oviposition. *Phytochemistry* **56**: 245–252.
- Slavov G, DiFazio S, Martin J, Schackwitz W, Muchero W, Rodgers-Melnick E, Lipphardt M, Pennacchio C, Hellsten U, Pennacchio L, *et al.* 2012.** Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*. *New Phytologist* **196**: 713–725.
- Smith DS, Bailey JK, Shuster SM, Whitham TG. 2011.** A geographic mosaic of trophic interactions and selection: trees, aphids and birds. *Evolutionary Biology* **24**: 422–429.
- Srivastava LM. 2002.** Cytokinins. In: Plant growth and development: hormones and environment. San Diego: Academic Press, 191–202.
- Storey JD, Tibshirani R. 2003.** Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* **100**: 9440–9445.
- Thaler JS, Bostock RM. 2004.** Interactions between abscisic-acid-mediated responses and plant resistance to pathogens and insects. *Ecology* **85**: 48–58.
- Tocquard K, Lafon-Placette C, Auguin D, Muries B, Bronner G, Lopez D, Fumanal B, Franchel J, Bourgerie S, Maury S, *et al.* 2014.** In silico study of wall-associated kinase family reveals large-scale genomic expansion potentially connected with functional diversification in *Populus*. *Tree Genetics & Genomes* **10**: 1135–1147.
- Tschaplinski TJ, Standaert RF, Engle NL, Martin MZ, Sangha AK, Parks JM, Smith JC, Samuel R, Jiang N, Pu Y, *et al.* 2012.** Down-regulation of the caffeic acid O-methyltransferase gene in switchgrass reveals a novel monolignol analog. *Biotechnology for Biofuels* **5**: 71.
- Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, *et al.* 2006.** The genome of black cottonwood, *Populus*

- trichocarpa* (Torr. & Gray). *Science* **313**: 1596–1604.
- Vining KJ, Pomraning KR, Wilhelm LJ, Priest HD, Pellegrini M, Mockler TC, Freitag M, Strauss SH. 2012.** Dynamic DNA cytosine methylation in the *Populus trichocarpa* genome: tissue-level variation and relationship to gene expression. *BMC Genomics* **13**: 27.
- de Vos M, Jander G. 2010.** Volatile communication in plant-aphid interactions. *Current Opinion in Plant Biology* **13**: 366–371.
- Wada K, Munakata K. 1971.** Feeding inhibitory activity of terpenoids in plants. *Agricultural and Biological Chemistry* **35**: 115–118.
- Wang C, Yang Y, Wang H, Ran X, Li B, Zhang J, Zhang H. 2016.** Ectopic expression of a cytochrome P450 monooxygenase gene PtCYP714A3 from *Populus trichocarpa* reduces shoot growth and improves tolerance to salt stress in transgenic rice. *Plant Biotechnology Journal* **14**: 1838–1851.
- Weighill D, Jones P, Shah M, Ranjan P, Muchero W, Schmutz J, Sreedasyam A, Macaya-Sanz D, Sykes R, Zhao N, et al. 2018.** Pleiotropic and epistatic network-based discovery: integrated networks for target gene discovery. *Frontiers in Energy Research* **6**: 30.
- Weighill D, Macaya-Sanz D, DiFazio SP, Joubert W, Shah M, Schmutz J, Sreedasyam A, Tuskan G, Jacobson D. 2019.** Wavelet-based genomic signal processing for centromere identification and hypothesis generation. *Frontiers in Genetics* **10**: 487.
- Whitham TG. 1978.** Habitat selection by *Pemphigus* aphids in response to response limitation and competition. *Ecology* **59**: 1164–1176.
- Whitham TG, Bailey JK, Schweitzer JA, Shuster SM, Bangert RK, LeRoy CJ, Lonsdorf**

- EV, Allan GJ, DiFazio SP, Potts BM, *et al.* 2006.** A framework for community and ecosystem genetics: from genes to ecosystems. *Nature Reviews Genetics* **7**: 510–523.
- Whitham TG, Gehring CA, Lamit LJ, Wojtowicz T, Evans LM, Keith AR, Smith DS. 2012.** Community specificity: life and afterlife effects of genes. *Trends in Plant Science* **17**: 271–281.
- Whitham TG, Martinsen GD, Keim P, Floate KD, Dungey HS, Potts BM. 1999.** Plant hybrid zones affect biodiversity: tools for a genetic-based understanding of community structure. *Ecology* **80**: 416–428.
- Whitham TG, Young WP, Martinsen GD, Gehring CA, Schweitzer JA, Shuster SM, Wimp GM, Fischer DG, Bailey JK, Lindroth RL, *et al.* 2003.** Community and ecosystem genetics: a consequence of the extended phenotype. *Ecology* **84**: 559–573.
- Wimp, GM, Tomasula, J, Hamilton, MB. 2019.** Putting the genes into community genetics. *Molecular Ecology* **28**: 4351–4353.
- Wimp GM, Martinsen GD, Floate KD, Bangert RK, Whitham TG. 2005.** Plant genetic determinants of arthropod community structure and diversity. *Evolution* **59**: 61–69.
- Yoon JS, Mogilicherla K, Gurusamy D, Chen X, Chereddy SC, Palli SR. 2018.** Double-stranded RNA binding protein, Staufen, is required for the initiation of RNAi in coleopteran insects. *Proceedings of the National Academy of Sciences* **115**: 8834–8339.
- Zhao D, Zong-mao C, Cheng J. 2002.** Isolation and activity identification of volatiles among tea plant-green leafhopper-*Evarcha* spider. *Journal of Tea Science* **22**: 109–114.
- Zhou X, Stephens M. 2012.** Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44**: 821–824.
- Zhou X, Stephens M. 2014.** Efficient algorithms for multivariate linear mixed models in

genome-wide association studies. *Nature Methods* **11**: 407–409.

3.8 Tables and Figures

Table 1. Sum of arthropods observed on tree tissue within plantations.

Order	Arthropod trait	Clatskanie 2012	Corvallis 2012	Placerville 2012	Corvallis 2015
Araneae	Araneae	0	0	0	59
	Salticidae	0	0	0	13
Coleoptera	<i>Chrysomela scripta</i>	642	0	0	0
	Coccinellidae	0	0	0	55
	<i>Polydrusus impressifrons</i>	0	0	0	34
	Curculionidae	0	0	0	18
Diptera	Agromyzid fly	5	0	0	0
	<i>Harmandia</i> sp.	5272	1492	0	0
	<i>Dasineura</i> sp.	0	0	0	167
Hemiptera	<i>Corythucha</i> sp.	0	0	0	417
	Cicadellidae	0	0	0	51
	Chloretic damage	0	0	0	108
	Mid-vein leaf damage	0	0	0	192
Hymenoptera	<i>Phyllocolpa</i> sp. leaf fold	1043	297	17	6138
	<i>Phyllocolpa</i> sp. larvae	0	0	0	14
	Sawfly larvae	0	0	0	8
Leaf damage	Chewing damage	943	143	512	0
	Egg mass	628	0	0	16
	Holes in leaves	1333	118	538	0
	Bottom of leaf damage	1802	21	68	183
	Top of leaf damage	259	22	6	0
Leaf modifier	Leaf folder	3	0	0	0
	Leaf roller	2	0	0	0
	Tip folder	63	6	5	0
Lepidoptera	<i>Phyllonorycter</i> sp.	96	216	0	496
	<i>Gypsonoma</i> sp.	38	0	0	10
	Leaf miner	191	0	0	0
	<i>Phyllocnistis populiella</i>	310	8	37	7
	<i>Sarothrips frigidana</i>	0	0	0	19

Table 2. Linear-mixed model output for community surveys across all gardens and years.

Broad-sense heritability (H^2) denotes the contribution of all host plant genetic factors to total variance in phenotype.

Garden	# genotypes	# genotypes with replicates	Phenotype	H^2	Gvar	Evar	p-value
Clatskanie 2012	654	36	Diversity	0.231	0.034	0.113	0.050
			<i>Harmandia</i> sp.	0.511	37.6	36.1	<0.001
Placerville 2012	40	34	Abundance	0.181	10.4	47.1	0.042
			Richness	0.217	0.126	0.453	0.037
			Diversity	0.244	0.024	0.075	0.026
Corvallis 2012	40	35	Abundance	0.363	95.2	166	<0.001
			Diversity	0.191	0.033	0.141	0.046
			<i>Harmandia</i> sp.	0.394	84.0	129.3	<0.001
Corvallis 2015	496	61	Abundance	0.268	28.8	78.4	0.016
			<i>Corythucha</i> sp.	0.140	0.493	2.90	0.050
			<i>Phyllocolpa</i> sp.	0.237	13.9	44.7	0.043

Table 3. PERMANOVA results for NMDS configuration within all 2012 gardens with replicated genotype observations and block to account for spatial position.

Factor	Df	Sum of squares	Mean square	F-model	R ²	p-value
Garden	2	24.6	12.3	65.1	0.287	0.001
Block	7	3.39	0.484	2.56	0.039	0.001
Genotype	44	13.5	0.307	1.62	0.157	0.001
Residuals	235	44.4	0.189	--	0.517	--

Table 4. PERMANOVA results for NMDS configuration among gardens with replicated genotype observations and block to account for spatial position.

Garden	Factor	Df	Sum of squares	Mean squares	F-model	R2	p-value
Clatskanie 2012	Block	2	0.790	0.395	1.70	0.035	0.044
	Genotype	34	9.77	0.287	1.23	0.435	0.029
	Residuals	51	11.9	0.233	--	0.529	--
Corvallis 2012	Block	2	0.360	0.180	1.14	0.016	0.298
	Genotype	39	12.0	0.307	1.95	0.549	0.001
	Residuals	60	9.48	0.158	--	0.434	--
Placerville 2012	Block	3	2.24	0.745	5.43	0.132	0.001
	Genotype	39	7.07	0.181	1.32	0.416	0.036
	Residuals	56	7.68	0.137	--	0.452	--
Corvallis 2015	Block	1	0.185	0.185	1.37	0.010	0.189
	Genotype	60	10.1	0.168	1.24	0.549	0.032
	Residuals	60	8.11	0.135	--	0.441	--

Table 5. Significant SNPs (FDR < 0.1) and nearest *P. trichocarpa* genes discovered through GWAS for *P. populiella* trait in Clatskanie garden.

Gene	Chr	Position	p-value	Description
Potri.001G166800	Chr01	14,020,499	3.29*10 ⁻⁷	Von Willebrand factor, Type A domain containing
Potri.001G166900	Chr01	14,023,753	7.05*10 ⁻⁸	Similar to hypothetical protein
Potri.001G385600	Chr01	40,092,552	4.23*10 ⁻⁸	Protein tyrosine kinase (Pkinase_Tyr) // Di-glucose binding within endoplasmic reticulum (Malectin)
Potri.002G244000	Chr02	23,611,478	1.56*10 ⁻⁷	Protein XRI1
Potri.003G062600	Chr03	9,092,517	6.24*10 ⁻⁷	O-fucosyltransferase family protein
Potri.003G062700	Chr03	9,096,040	1.15*10 ⁻⁷	Similar to pentatricopeptide (PPR) repeat-containing protein
Potri.004G007700	Chr04	497,265	1.22*10 ⁻⁷	Mitogen-activated protein kinase kinase kinase 17-related
Potri.005G033500	Chr05	2,494,360	3.95*10 ⁻⁷	Glycogenin glucosyltransferase
Potri.005G039300	Chr05	2,864,214	2.85*10 ⁻⁷	Tropinone reductase II / Tropinone reductase
Potri.005G170000	Chr05	18,393,885	4.94*10 ⁻⁷	Kelch repeat-containing F-box family protein
Potri.005G184700	Chr05	20,171,994	8.47*10 ⁻⁷	Protein stay-green like, Chloroplastic
Potri.006G187400	Chr06	20,199,567	6.10*10 ⁻⁷	Polyadenylation factor I complex, subunit FIP1
Potri.006G204100	Chr06	21,938,274	3.40*10 ⁻⁷	No domains
Potri.006G235700	Chr06	24,505,808	6.06*10 ⁻⁷	Nuclear poly(A) polymerase
Potri.008G020900	Chr08	1,094,017	4.16*10 ⁻⁷	Similar to expressed protein in Arabidopsis thaliana
Potri.011G027100	Chr11	2,218,904	4.24*10 ⁻⁷	NADPH-dependent diflavin oxidoreductase 1
Potri.012G046000	Chr12	4,224,452	3.00*10 ⁻⁷	Similar to expressed protein in Arabidopsis thaliana
Potri.014G118900	Chr14	9,233,521	7.22*10 ⁻⁷	No domains
Potri.015G103600	Chr15	12,174,850	4.16*10 ⁻⁷	Nitrate, fromate, iron dehydrogenase
Potri.019G020000	Chr19	2,300,230	4.13*10 ⁻⁷	Similar to Putative peroxisome assembly protein 12 (Peroxin-12)
Potri.T148400	scaffold_495	17,502	5.37*10 ⁻⁷	No domains

Table 6. Significant SNPs (FDR < 0.1) and nearest *P. trichocarpa* genes discovered through GWAS for *Phyllonorycter* sp. trait.

Trait	Gene	Chr	Position	p-value	Description
<i>Phyllonorycter</i> sp. (Clatskanie)	Potri.001G103400	Chr01	8,238,390	3.37*10 ⁻⁷	Phosphatidylinositol N-acetylglucosaminyltransferase subunit P-like protein
	Potri.001G179200	Chr01	15,465,196	3.16*10 ⁻⁷	Coatomer, subunit beta
	Potri.001G179300	Chr01	15,491,815	2.37*10 ⁻⁸	DNA-directed RNA polymerase V subunit 5C
	Potri.001G190500	Chr01	1,715,073	1.08*10 ⁻⁷	Calmodulin-binding family protein
	Potri.001G190600	Chr01	17,160,294	4.93*10 ⁻⁸	No domains
	Potri.001G190700	Chr01	17,171,087	3.23*10 ⁻⁷	Similar to expressed protein in Arabidopsis thaliana
	Potri.001G272000	Chr01	27,936,516	4.24*10 ⁻⁷	Cyclin-B1-4
	Potri.001G470700	Chr01	50,239,765	7.62*10 ⁻⁷	Similar to Ferredoxin 2; similar to chloroplast precursor.
	Potri.002G202600	Chr02	16,493,267	6.11*10 ⁻⁷	Pectinesterase family protein
	Potri.003G116800	Chr03	14,025,351	6.37*10 ⁻⁷	Calcium-binding protein CML24-related
	Potri.003G204900	Chr03	20,594,059	2.99*10 ⁻⁷	Outer envelope protein 80, Chloroplastic
	Potri.005G088400	Chr05	6,614,910	4.01*10 ⁻⁷	Gem-like protein 4-related
	Potri.005G088500	Chr05	6,622,054	1.02*10 ⁻⁷	No domains
	Potri.005G098600	Chr05	7,507,025	8.55*10 ⁻⁷	Leucine-rich repeat family protein
	Potri.006G053400	Chr06	3,867,149	3.36*10 ⁻⁷	Similar to guanylate kinase 1
	Potri.006G053500	Chr06	3,888,118	2.72*10 ⁻⁷	Transcription factor NF-Y alpha-related
	Potri.006G053600	Chr06	3,897,089	4.23*10 ⁻⁷	No domains
	Potri.006G147800	Chr06	12,715,987	3.87*10 ⁻⁷	Similar to casein kinase
	Potri.006G223400	Chr06	23,563,498	5.91*10 ⁻⁷	No domains
	Potri.007G093200	Chr07	11,997,824	9.12*10 ⁻⁷	No domains
	Potri.011G158300	Chr11	17,552,928	1.09*10 ⁻⁷	FAD-binding berberine family protein-related
	Potri.012G126300	Chr12	14,449,335	8.04*10 ⁻⁷	No domains
	Potri.013G108900	Chr13	12,228,211	9.49*10 ⁻⁷	ZF-HD protein dimerisation region (ZF-HD_dimer)
	Potri.013G126100	Chr13	13,887,304	9.52*10 ⁻⁷	Cysteine protease family C1-related
	Potri.014G130600	Chr14	9,962,648	2.33*10 ⁻⁷	Transducin family protein
	Potri.014G182800	Chr14	15,527,094	5.27*10 ⁻⁸	Glucan endo-1,3-beta-glucosidase 7-related
	Potri.015G120200	Chr15	13,506,582	1.90*10 ⁻⁷	Similar to lectin
	Potri.016G134100	Chr16	13,748,464	5.44*10 ⁻⁸	Co-expressed with genes in roots specific co-expression subnetwork
	Potri.018G026200	Chr18	2,090,844	2.10*10 ⁻⁸	Similar to Adenine phosphoribosyltransferase 2
	Potri.018G076400	Chr18	10,190,180	8.34*10 ⁻⁷	Copper-binding family protein; similar to copper homeostasis factor
<i>Phyllonorycter</i> sp. (Corvallis)	Potri.019G076500	Chr19	11,061,377	2.61*10 ⁻⁸	Domain of unknown function (DUF313) (DUF313)

Table 7. Significant SNPs (FDR < 0.1) and closest *P. trichocarpa* genes discovered through GWAS for community metric traits.

Trait	Gene	Chr	Position	p-value	Description
Richness (Clatskanie)	Potri.001G068000	Chr01	5,421,300	1.32*10 ⁻⁷	Retinaldehyde binding protein-related
	Potri.001G115500	Chr01	9,243,933	7.23*10 ⁻⁸	ATP-dependent CLP protease
	Potri.006G038400	Chr06	2,690,405	1.17*10 ⁻⁷	similar to ferredoxin-dependent glutamate synthase
	Potri.008G079700	Chr08	4,968,425	1.59*10 ⁻⁷	Conserved oligomeric golgi complex subunit 1
	Potri.008G079900	Chr08	4,989,054	2.19*10 ⁻⁷	Do-like 15 protein-related
	Potri.014G116100	Chr14	9,044,620	3.12*10 ⁻⁷	Zinc finger FYVE domain containing protein
Richness (Corvallis)	Potri.007G044400	Chr07	3,884,609	7.77*10 ⁻⁸	RNA-binding protein Musashi (MSI)
	Potri.007G044500	Chr07	3,890,366	2.00*10 ⁻⁸	CGI-141-related/lipase containing protein
	Potri.010G095000	Chr10	11,746,693	3.34*10 ⁻⁷	No domains
	Potri.013G119400	Chr13	13,255,200	8.10*10 ⁻⁷	TCP family transcription factor (TCP)
	Potri.018G149300	Chr18	16,746,330	7.48*10 ⁻⁸	Similar to cytochrome P450 family
Abundance (Clatskanie)	Potri.001G243600	Chr01	25,479,678	2.92*10 ⁻⁷	DNA-binding domain-containing protein-related
	Potri.001G243700	Chr01	25,480,842	3.40*10 ⁻⁸	No domains
	Potri.009G027700	Chr09	3,866,118	2.83*10 ⁻⁷	Similar to responsive to high light 41
	Potri.009G028800	Chr09	3,949,481	1.48*10 ⁻⁷	Similar to osmotin-like protein

Table 8. Significant SNPs (FDR < 0.1) and closest *P. trichocarpa* genes discovered through GWAS for arthropod community composition multitrait.

Trait	Gene	Chr	Position	p-value	Description
Community composition (Clatskanie)	Potri.003G028500	Chr03	3,553,415	2.88*10 ⁻⁷	Adhesion regulating molecule 1 110 KDA cell membrane glycoprotein
	Potri.003G028500	Chr03	3,555,411	5.56*10 ⁻⁷	
	Potri.003G028500	Chr03	3,555,593	4.27*10 ⁻⁷	
	Potri.003G028500	Chr03	3,555,643	1.96*10 ⁻⁷	
	Potri.003G028500	Chr03	3,555,663	7.72*10 ⁻⁷	
	Potri.003G028500	Chr03	3,555,755	5.90*10 ⁻⁷	
	Potri.003G028500	Chr03	3,555,770	1.70*10 ⁻⁷	
	Potri.003G028500	Chr03	3,556,231	8.56*10 ⁻⁷	
	Potri.014G080100	Chr14	6,406,993	1.24*10 ⁻⁷	K15095 - (+)-neomenthol dehydrogenase
	Potri.015G029200	Chr15	2,364,978	4.13*10 ⁻⁷	Nucleotide-diphospho-sugar transferase domain-containing protein-related
Community composition (Corvallis)	Potri.018G131600	Chr18	15,288,689	7.27*10 ⁻⁷	Peroxidase
	Potri.001G199700	Chr01	19,388,433	5.10*10 ⁻⁸	Similar to histidyl-tRNA synthetase
	Potri.001G199700	Chr01	19,388,444	1.34*10 ⁻⁷	
	Potri.001G199700	Chr01	19,388,477	5.92*10 ⁻⁸	
	Potri.002G202800	Chr02	16,515,234	9.06*10 ⁻⁷	Similar to 60S ribosomal protein L18
	Potri.008G067900	Chr08	4,132,807	4.48*10 ⁻⁹	ABC1 family protein
	Potri.010G040500	Chr10	6,904,193	1.20*10 ⁻⁷	Wall-associated receptor kinase galacturonan-binding (GUB_WAK_bind)
	Potri.010G103900	Chr10	12,466,417	9.33*10 ⁻⁷	Similar to expressed protein in Arabidopsis thaliana
	Potri.012G005900	Chr12	480,690	4.17*10 ⁻⁷	Two-component response regulator-like APRR5-related
	Potri.017G063700	Chr17	6,029,280	1.76*10 ⁻⁸	Eukaryote specific dsRNA binding protein
	Potri.017G063700	Chr17	6,033,420	2.15*10 ⁻⁸	
	Potri.017G063700	Chr17	6,036,353	7.81*10 ⁻⁷	
	Potri.017G124700	Chr17	13,673,803	4.29*10 ⁻⁷	Similar to GGPP synthase 2
	Potri.T112800	scaffold 178	21,170	2.10*10 ⁻⁹	Protein kinase domain (Pkinase)

Table 9. Candidate genes for mediating biotic interactions selected for functional analysis from networks. Breadth indicates number of connected network layers, depth refers to the number of neighbor genes in network connected to anchor gene, and edges indicate the number of connections (co-methylation, co-expression, metabolite GWAS and pyMBMS GWAS) among genes in networks.

Trait	Gene	Chr	Pos	Breadth	Depth	Edges
Clatskanie <i>Phyllonorycter</i> sp.	Potri.001G272000	Chr01	27,936,516	4	1196	1750
Clatskanie <i>Phyllonorycter</i> sp.	Potri.018G026200	Chr18	2,090,844	4	615	757
Corvallis richness	Potri.007G044400	Chr07	3,884,609	3	662	2133
Corvallis richness	Potri.018G149300	Chr18	16,746,330	4	1610	6074
Corvallis community	Potri.017G063700	Chr17	6,029,280	3	1725	5758
Corvallis community	Potri.T112800	scaffold_178	21,170	3	710	1418

Table 10. Genome function enrichment analysis of co-expressed neighbor genes from network

analysis. Significance of enrichment was determined based on Fisher Exact p-value with

multiple testing correction to meet a threshold of FDR < 0.1.

Gene	Phenotype	Expression	GO namespace	FDR p-value	Function
Potri.001G272000	<i>Phyllonorycter</i> sp. (Clatskanie)	Coex-pos	Biological process	6.38*10 ⁻²	Meiotic chromosome segregation
				3.97*10 ⁻²¹	Microtubule-based movement
				9.29*10 ⁻⁵	Protein polymerization
				5.22*10 ⁻³	Malate metabolic process
				3.86*10 ⁻⁴	Microtubule-based process
				2.40*10 ⁻²	Proteolysis involved in cellular protein catabolic process
			Molecular function	1.91*10 ⁻²	L-malate dehydrogenase activity
				1.98*10 ⁻²¹	Microtubule motor activity
				2.22*10 ⁻²	Threonine-type endopeptidase activity
				9.95*10 ⁻⁵	GTPase activity
				1.34*10 ⁻³	GTP binding
				2.65*10 ⁻²	ATP binding
Potri.007G044400	Richness (Corvallis)	Coex-pos	Biological process	1.35*10 ⁻²	Histone lysine methylation
			Molecular function	8.58*10 ⁻²	mRNA guanylyltransferase activity
				1.36*10 ⁻¹	Polynucleotide
					adenylyltransferase activity
				1.69*10 ⁻¹	Histone-lysine N-methyltransferase activity
				1.04*10 ⁻¹	Histone binding
				1.26*10 ⁻⁶	Helicase activity
				6.92*10 ⁻⁴	Nucleic acid binding
				1.43*10 ⁻¹	RNA binding
				0.0905	Zinc ion binding
				1.30*10 ⁻⁵	Protein binding
				1.42*10 ⁻¹	DNA binding
Potri.018G149300	Richness (Corvallis)	Coex-pos	Biological process	3.87*10 ⁻²	Terpenoid (Isoprenoid) biosynthetic process
				2.80*10 ⁻²	Phosphorylation
			Molecular function	2.82*10 ⁻³	Mevalonate kinase activity
Potri.T112800	Community composition (Corvallis)	Coex-pos	Biological process	1.80*10 ⁻²	Oxidation-reduction process
			Molecular function	6.21*10 ⁻²	Ferredoxin-NAD(P) reductase activity
				7.78*10 ⁻²	Sigma factor activity
				9.63*10 ⁻²	Antioxidant activity
				3.04*10 ⁻²	Oxidoreductase activity
Potri.017G063700	Community composition (Corvallis)	Coex-neg	Biological process	1.55*10 ⁻²	Sucrose metabolic process

Figure 1. Map displaying genotype origin populations and survey sites – circles represent collection locations in the wild, stars represent common gardens, the survey sites.

Figure 2. Arthropod community NMDS plot depicting 3 dimensions within all gardens surveyed in 2012. Color groupings indicate tree surveyed in specified garden.

Figure 3. Arthropod community NMDS plots depicting 3 dimensions for (a) 2012 Clatskanie garden survey and (b) 2012 Placerville garden survey. PERMANOVA analysis, run with only genotypes with replicate observations, output indicated for each garden in the upper right corner of each plot for each common garden. Size of species text indicates position relative to the dimensional space, for example, larger font indicates forward projection along the positive values of axes.

Figure 4. Arthropod community NMDS plots depicting 3 dimensions for (a) 2012 Corvallis garden survey and (b) 2015 Corvallis garden survey. PERMANOVA analysis, run with only genotypes with replicate observations, output indicated for each garden in the upper right corner of each plot for each common garden. Size of species text indicates position relative to the dimensional space, for example, larger font indicates forward projection along the positive values of axes.

Figure 5. Manhattan plot output from GWAS for (a) individual arthropods (single-trait; three tests), (b) calculated community metrics for each garden (single-trait; three tests), and (c) community composition (multi-trait; two tests). SNPs that passed the red dotted line were

significantly associated with indicated trait. Clatskanie community (c) did not pass threshold but contained multiple suggestive markers.

Figure 6. Co-expression network for Clatskanie *Phyllonorycter* sp. single-trait gene candidates. Co-expression genes were grouped based on their biological function as determined by Gene-ontology terms.

Figure 7. Co-expression network for Corvallis arthropod richness single-trait gene candidates. Co-expression genes were grouped based on their biological function as determined by Gene-ontology terms.

Figure 8. Co-expression network for Corvallis community composition multi-trait gene candidates. Co-expression genes were grouped based on their biological function as determined by Gene-ontology terms.

Figure 1.

Figure 2.

Figure 3.

Figure 4.

Figure 5.

Figure 6.

Figure 7.

Figure 8.

Chapter 4: Characterization of *Salix nigra* floral insect community and activity of three native *Andrena* bees

Submitted to *bioRxiv*: Simon S, Keefover-Ring K, Park YL, Wimp G, Grady J, & DiFazio S. 2020. Characterization of *Salix nigra* floral insect community and activity of three native *Andrena* bees.

Authors' contributions: S.S., S.D., P.Y.-L., and J.G. designed and supervised insect surveys/identifications. K.K.-R and J.G. collected/analyzed metabolite and VOC data. S.S., G.M., and S.D. analyzed the data. S.S., S.D., K.K.-R, P.Y.-L., and G.M. wrote the manuscript.

4.1 Abstract

Salix nigra (black willow) is a widespread tree that hosts many species of polylectic hymenopterans and oligolectic bees of the genus *Andrena*. The early flowering time of *S. nigra* makes it an important nutritive resource for arthropods emerging from hibernation. However, since *S. nigra* is dioecious, not all insect visits will lead to successful pollination. Using both visual observation and pan-trapping, we characterized the community of arthropods that visited *S. nigra* flowers and assessed the differences among male and female trees as well as the chemical and visual drivers that influenced community composition across three years. We found that male trees consistently supported higher diversity of insects than female trees and only three insect species, all *Andrena* spp., consistently visited both sexes. Additionally, visits by *A. nigrae* correlated strongly to volatile cues suggesting that cross-pollinators cue into specific aspects of floral scent, but diversity of floral visitors is driven strongly by visual cues. Through time, the floral activity of two *Andrena* species remained stable, but *A. nigrae* visited less in 2017 when flowers bloomed earlier than other years. When native bee emergence does not synchronize with bloom time, activity appears to be greatly diminished, which could threaten species that only subsist on a single host. Despite the community diversity of *S. nigra* flowers, productivity depends on a small fraction of insect species that are not threatened by competition, but rather rapidly changing climate conditions that lead to host-insect asynchrony.

4.2 Introduction

Early spring emergence of flowers is extremely important in supplying nutritive rewards, such as pollen and nectar, for many native arthropods, while the host plant benefits with an increased chance of successful sexual reproduction. The most common cross-pollinators in agricultural systems are often made up of arthropods that collect pollen from many unrelated host plants, such as flies belonging to the family Syrphidae, eusocial bee species such as honey bees (*Apis mellifera* L.), and polylectic solitary bees (Ostaff *et al.*, 2015). Lack of discrimination among hosts allows these arthropod groups to more flexibly collect resources for survival and population growth.

Conversely, oligolectic solitary bees, which only collect pollen from either related plant species or a single species, rely heavily on predictable timing of available native floral resources (Danforth, 2007; Straka *et al.*, 2014). Upon emergence from nests, oligolectic bees must locate flowers, feed, breed, build new nests, lay eggs, and collect resources to provide their larvae with food provisioned for development throughout the remainder of the year, all within the bloom time of their specific host (Stevens, 1949; Linsley, 1958; Danforth, 2007). In addition to being a valuable resource for early emerging generalist floral insects, willow species belonging to the genus *Salix* are the primary hosts of many oligolectic bees, especially those belonging to one of the largest bee genera, *Andrena* (Stevens, 1949; Ostaff *et al.*, 2015).

Salix encompasses between 300-400 species of shrubs and trees with a dual pollination system, which occurs via wind (anemophilous), insects (entomophilous), or both (Tamura & Kudo, 2000; Karrenberg *et al.*, 2002; Argus, 2011). *Salix* biology creates a unique environment for insect

reward collection due to its dioecious nature. In order for sexual reproduction to occur, insects must locate male plants to collect pollen and then carry it to a separate female plant in the population (Dötterl *et al.*, 2014). Host location typically occurs through a combination of visual, olfactory, and reward cues. *Salix* species have a non-showy inflorescence arranged in a catkin form where male flowers are often yellow and female flowers tend to be green (Karrenberg *et al.*, 2002; Füssel *et al.*, 2007). Additionally, *Salix* species emit a complex mixture of volatile organic compounds that are important as olfactory signals to insects, and both male and female plants offer nectar rewards (Tollsten & Knudsen, 1992; Füssel *et al.*, 2007). However, upon locating a host plant, some insects may rob flowers of their resources and not carry pollen between male and female individuals (Galen & Butchart, 2003).

Salix nigra, known by its common name black willow, is a tree-form, entomophilous willow that grows throughout the Eastern United States north to Maine, west to North Dakota and south to Georgia (Burns & Honkala, 1990). The extensive range and productivity of *S. nigra* as well as its early bloom, typically February in its southern range through late June in more northern states, makes it an ideal resource for early emerging insects (Burns & Honkala, 1990; Ostaff *et al.*, 2015). Studying the mechanisms that *S. nigra* employs to attract insects as well as the influence of sex of tree on floral insect community through time is important in determining the competition for and potential stability of catkin resources, native oligolectic bee activity, and *S. nigra* reproductive success.

The goal of this study was to characterize the community of insects that visit *S. nigra* catkins and examine how the total floral community responded to tree sex, geographic position, volatile

organic compound (VOC) profiles, and secondary metabolites in catkins and leaves. For comparison, we also evaluated the community of floral insects captured using visual survey techniques and pan-traps placed in tree canopies. Finally, we examined the effect of tree sex, VOCs, and survey year on the activity of three native *Andrena* bee species, including the willow oligolectic bees *A. macoupinense* and *A. nigrae*.

4.3 Methods

Population and site description

The target population of *S. nigra* was located in the West Virginia University Core Arboretum in Morgantown, West Virginia (39.6462° N, 79.9811° W). The Core Arboretum is an old growth forest that contains 91 acres of native shrubs, trees, and herbaceous plants. It is located on a hillside that stretches between Monongahela Boulevard and the Monongahela River and contains riparian and floodplain sites with a small grove of *S. nigra*. The population contained thirty-two trees of which twelve were identified as female and twenty were male (Figure 1).

Visual survey technique

Visual surveys were performed in April through early May in years 2017, 2018, and 2019 on sunny days with minimal wind to help increase observation of small floral visitors. Visual observations were chosen as the survey method for this species due to the brittleness of the base of short shoots, which prevents the use of sweep nets (Beismann *et al.*, 2000). To equalize observations among trees, survey branches were flagged containing approximately 300 catkin flowers for each individual tree. Trees were visually observed for 16 minutes per survey throughout their bloom (~ 2 weeks) and order of surveyed trees was randomized to account for

time of day for two female and four male trees. Insect specimens were carefully hand collected from trees throughout the survey time for family/species level identification. Additionally, survey month and day were recorded for each observation to use as covariates in models analyzing individual insect activities as well as flower phenology. Finally, early season herbivore activity, which was made up of observations of insects that feed on trees during and after catkins were no longer present, was collected in 2019 at the end date of tree flowering by counting number of herbivore occurrences on visual survey branches.

Canopy pan-trap technique

In the year 2019, pan-traps were constructed by painting three-ounce plastic cups with either fluorescent blue, fluorescent yellow, or white paint (Guerra Paint and Pigment, New York). Fluorescent paints were a mixture of 16 ounces of the fluorescent dispersion to 1 gallon of silica flat paint. Three cups, one of each color, were affixed with velcro to a bucket lid as a platform to be raised into tree canopies below flowering branches (Figure 2) on five female and six male trees. A soapy trap solution was prepared by addition of approximately 5mL of organic unscented dish soap into one gallon of water. Cups were filled $\frac{3}{4}$ of the way full with soap solution. Traps were raised to the bottom of each tree canopies (2- 12 meters; average 4 meters) at 9:00 am in the morning and emptied daily at 5:00 pm. Visual surveys were performed on the same days. Captured insects from each cup were transferred to separate vials containing 70% ethanol for later identification.

Insect identification

Abundant insects were identified to species level while rare insects were identified to family. Native bee species identifications were validated by an expert (Sam Droege, Patuxent Wildlife Research Center, personal communication). Vouchers of collected insect specimens were submitted to the West Virginia University Entomology Collection.

Flower volatile and tissue collection

Dormant branches were collected from the field in early spring for 6 female and 9 male *Salix nigra* trees along the river. Branches were allowed to root and flower in the Department of Biology greenhouse in buckets of water. Volatile organic compounds (VOCs) were collected using the dynamic headspace method (Keefover-Ring, 2013). An oven bag was placed over the flowering branches and bags were secured with thin gauge wire around cotton pads that had been wrapped near the base of the stems. At the top of each bag, polytetrafluoroethylene (PTFE) ports were fixed and connected to chemical traps consisting of 65 mm long and 3 mm internal diameter glass tubes packed with 20 mg of Super Q adsorbent (80/100 mesh size, DVB/ethylvinylbenzene polymer) (Alltech Associates Inc., Deerfield, IL, USA). The traps were connected to calibrated flow meters (Aalborg Instruments & Controls, Inc., Orangeburg, New York, USA) and air was pulled through at a flow rate of 200 ml min⁻¹ with an AirLite pump (SKC Inc., Eighty Four, PA) modified to run with a 6 V battery. Controls consisted of oven bags with an identical setup, but without an enclosed branch. Volatiles were collected for a three-hour time period with flow meter maintaining a flow rate of 200 ml min⁻¹. At the conclusion of each sampling period, chemical traps were rinsed with 150 µl n-hexane (GC2, Honeywell Burdick & Jackson, Morristown, NJ, USA) into GC vials with PTFE lined screw caps. All catkins were counted in each bag, and maturation stage was noted. All catkins were subsequently lyophilized,

and the dry weight obtained of mature and immature (pre-anthesis) catkins separately. Prior to GC-MS analysis, 40 μl of each sample was combined with 2 μl of an internal standard solution (m-xylene in n-hexane). Collected floral volatile organic compounds (VOCs) were analyzed by gas chromatography (GC) with mass spectrometry (MS) detection.

VOC and metabolite characterization

Floral VOC samples were analyzed with a Thermo Trace 1310 GC coupled to a Thermo ISQ MS with electron ionization (EI) at 70.0 eV at 250 °C, using helium as the carrier gas at 1.0 ml min^{-1} with the injector temperature set at 250 °C. Oven conditions included an initial temperature of 40 °C followed by an immediate ramp of 3 °C min^{-1} to 200 °C. Available standards, 1 μl of samples, and a continuous series of n-alkanes (C8–C20; Sigma-Aldrich) were injected in the split mode onto a TR-5MS capillary column (30 m $\times$ 0.25 mm I.D., film thickness 0.25 μm ; Thermo Fisher Scientific). Compounds were identified with retention time matches to pure standards, mass spectra, and/or linear retention indexes calculated with the alkane series (Adams, 2001; NIST, 2008; El-Sayed, 2013). Standard curves of available compounds were used to calculate final VOC results, which were expressed as $\text{ng compound g}^{-1} \text{DW hr}^{-1}$.

Sample collection for chemical characterization

Catkins and leaves collected from 24 trees (14 males and 10 females) in the field and in the greenhouse were characterized for five different metabolites; salicin, isosalicin, salicortin, tremuloidin, and termulacin as well as total metabolites. Leaves were flash frozen in the field and later shipped on dry ice to the University of Wisconsin-Madison, WI. The flash frozen catkins and leaves were lyophilized, counted (catkins) and weighed (catkins and leaves), and then

ground with steel balls in plastic scintillation vials in a ball mill. Accurately weighed portions (~15 mg) of powdered leaf tissue were extracted with cold (4° C) methanol (1.00 ml) containing salicylic acid-d6 (Sigma-Aldrich, St. Louis, MO, USA) as an internal standard with sonication in an ice bath (15 min) and then centrifugation to obtain a clear supernatant for analysis.

Chemical analyses

2 µL of all standards and samples solutions were injected onto the UHPLC and separated peaks with a Waters Acquity CSH C-18 column (2.1 x 100 mm, 1.7 µm) at 40 °C with a flow rate of 0.5 mL min⁻¹, using a gradient of water (solvent A) and acetonitrile (solvent B), both containing 0.1% formic acid. The PDA was configured to scan from 210-400 nm, with 1.2-nm resolution and a sampling rate of 20 points·s⁻¹. The MS operating parameters for were as follows: cone potential, 30 V; capillary potential, 2500 V; extractor potential, 3 V; RF lens potential, 0.1 V; source temperature, 120 °C; desolvation temperature, 250 °C; desolvation gas flow, 500 L h⁻¹; cone gas flow, 10 L h⁻¹; infusion rate, 5 µL min⁻¹; dwell time, 0.025 s.

Standard curves of methanol solutions, also containing the salicylic acid-d6 internal standard, of various purified compounds were used to calculate the concentrations in the extracted leaves, which were then normalized by sample dry weight and expressed as mg compound g⁻¹.

Commercially available standards of salicin (Sigma-Aldrich), and salicortin, HCH-salicortin, tremuloidin, and tremulacin were used that had been previously isolated and purified from aspen foliage (Lindroth *et al.*, 1986).

4.4 Statistical analysis

Floral insect community and sex effect on composition

To determine whether there were any sex differences in multivariate floral visitor community, the R package *vegan* (Oksanen *et al.*, 2019) was used to ordinate the data using non-metric multidimensional scaling (NMDS) using Bray-Curtis dissimilarities. Differences in floral visitor communities among male and female flowers were determined using Analysis of Similarity (ANOSIM), and significance was determined using 999 permutations to determine whether group assignments were significantly different from those generated by chance. Four survey types were tested to determine the stability of sex effects on floral community composition, including: 1) 2019 visual observations, 2) 2019 pan-traps, 3) 2019 total insect community across survey types and 4) visually observed communities through time from 2017-2019. Coordinates for dependent variables were extracted from the NMDS configuration and depicted in plots using a rescaled font to represent the three-dimensional projection. This was accomplished by rescaling all axis scores to a zero origin and scaling the font size relative to the product of the three rescaled axis scores.

Sex differences in tree chemistry

NMDS and ANOSIM were also utilized to determine if there were any sex differences in multivariate VOC and metabolite compositions by testing point grouping by sex of tree. Total monoterpenes, sesquiterpenes, VOC emissions, and metabolites were tested using a one-way analysis of variance (ANOVA) in SAS software version 9.4 to determine if there were any differences among sexes.

Tree chemistry and insect community relationship

Pairwise geographic distances were calculated among trees from GPS coordinates. VOC production per catkin was scaled to branch-level production using average catkin counts per branch for each individual tree to correlate to insect activity. Bray-Curtis dissimilarity matrices were generated for all datasets, including floral community, early season herbivore community, pairwise geographic distances, VOCs, and catkin/leaf metabolites. A Mantel test was utilized to determine whether differences in floral insect community were a function of pairwise geographic distances, or differences in catkin VOCs, catkin metabolites, or leaf metabolites. This analysis was also repeated for the tree early season herbivore community and catkin/leaf metabolite dissimilarities.

2019 survey method comparison and floral community

Twelve trees (5 female and 7 male) were surveyed in 2019 with visual and pan-trap techniques in the riparian and floodplain sites in the WVU Core Arboretum for a total of one-hundred and forty observation dataset. NMDS and ANOSIM were utilized to determine the effect of survey type on floral community composition. Visual observations were then merged with pan-trap capture counts for overlapping trees and dates for a final fifty-four observations. Independent numeric variables associated with survey day, including Julian date, military time, and temperature, were correlated with the NMDS configuration using the environmental fit vector analysis in vegan (*envfit* function). Variables that were found to significantly correlate to the community dataset were added as covariates in the nested analysis of covariance (ANCOVA) models.

Insects that were abundant in surveys, including *A. macoupinense*, *A. morrisonella*, *Andrena nigrae*, *Lasioglossum coeruleum*, and parasitic wasps belonging to the *Braconidae* family, were extracted from datasets. Additionally, species richness was calculated from the dataset as total number of species to visit each tree, and Shannon-Weaver diversity was calculated using the R package *vegan*. A nested ANCOVA was used to analyze differences in transformed insect counts, richness, and Shannon-Weaver diversity with the following model:

$$y \sim \text{Sex} + \text{Location} + \text{Tree}(\text{Sex}) + \text{Random} + \text{Covariates}$$

where () indicates the independent variable is nested in another variable.

Additionally, an environmental fit was conducted using the R package *vegan* with floral volatile compounds that were overlaid on the floral visitor community. The resulting patterns were then evaluated to select specific volatile compounds to look for linear relationships to insect activity using Pearson correlations with Bonferroni p-value corrections to account for multiple testing. Insects chosen to test included *Bombus* sp., *Andrena macoupinense* & *Andrena morrisonella*, *Andrena nigrae*, *Miridae*, *Chalcosyrphus nemorum*, and *Sarcophagidae*. VOCs of interest included acetophenone, cis-b-terpineol-2, ethyl-1-hexanol, germacrene-D, hexenyl-AC, octanal, octen-2-ol, and trans-3-pinane.

Floral community through time

Among all three years, after accounting for mortality and branch loss, a total of six trees along the river overlapped among surveys, including two female and four male trees for 77

observation. NMDS and ANOSIM were utilized to determine the effect of survey year on floral insect community composition. Independent numeric variables associated with survey day including year, Julian date, military time, and temperature, and these variables were correlated with the NMDS configuration using the environmental fit vector analysis in vegan (*envfit* function). Variables that were found to significantly correlate to the community dataset were added as additional covariates in all nested ANCOVA models.

Insects that were abundant in surveys, including combined counts of *A. macoupinense*, *A. morrisonella*, and *A. nigrae* were extracted from the community dataset and a nested ANCOVA to analyze differences in transformed insect counts, richness, and Shannon-Weaver diversity with the following model:

$$y \sim \text{Year} + \text{Sex} + \text{Tree}(\text{Sex}) \& \text{Random} + \text{Covariates}$$

where () indicates the independent variable is nested in another variable.

4.5 Results

Floral insect community and sex effect on composition

For visual surveys, across all three years of data, 3,160 insects were observed to visit flowers. Of those observations, 88.9% were hymenopteran, 9.6% were dipteran, 1.3% were hemipteran, and 0.2% were coleopteran. Additionally, of all insects observed, bees belonging to the genus *Andrena* made up 69.4% of the visually surveyed community. NMDS indicated that that the appropriate number of dimensions for 2019 visual surveys, 2019 pan-traps, 2019 total

community (Figure 3), and 2017-2019 community (Figure 4) was four (stress = 0.11), three (stress = 0.10), three (stress = 0.12), and four (stress = 0.12), respectively. An ANOSIM indicated that the floral community composition was dependent upon the sex of the tree for all survey types, 2019 total community, and across years ($R > 0.1000$, $p\text{-value} < 0.05$; Table 1).

Sex differences in tree chemistry

The NMDS of catkin VOCs indicated the appropriate number of dimensions for analysis was three (stress = 0.10) while the number of dimensions for leaf metabolites and catkin metabolites was two (stress = 0.10). An ANOSIM (Table 2) indicated that the catkin VOC composition and leaf metabolite composition were not significantly different between male and female trees ($R = -0.1191$, $p\text{-value} = 0.914$; $R = -0.00471$, $p\text{-value} = 0.449$). However, the catkin metabolite composition was different between the sexes ($R = 0.1787$, $p\text{-value} = 0.027$; Figure 5). The total amount of catkin and leaf metabolites did not differ between the sexes (catkins, one-way ANOVA $F_{1,28} = 2.741$, $p\text{-value} = 0.109$; leaves, $F_{1,30} = 1.549$, $p\text{-value} = 0.2229$). Furthermore, total monoterpenes ($F_{1,13} = 0.02755$, $p\text{-value} = 0.8707$), sesquiterpenes ($F_{1,13} = 0.2623$, $p\text{-value} = 0.6171$), and VOC emissions ($F_{1,13} = 0.004865$, $p\text{-value} = 0.9455$) were not significantly different between the sexes (Figure 6).

Tree chemistry and insect community relationship

Mantel tests (Table 3) indicated that there was no relationship between pairwise geographic distances and floral visitor community ($r_m = 0.026$, $p\text{-value} = 0.450$). Similarly, distances between catkin volatile composition, leaf metabolite composition, and catkin metabolite composition were not related to differences in floral visitor community ($r_m = -0.200$, $p\text{-value} =$

0.100; $r_m = 0.002$, p-value = 0.480; $r_m = 0.062$, p-value = 0.380, respectively). There was a significant positive relationship between differences in catkin metabolite composition and early season herbivore community ($r_m = 0.416$, p-value = 0.010).

2019 survey method comparison and floral community

An NMDS with results from both survey methods indicated the appropriate number of dimensions for analysis was three (stress = 0.12; Figure 7a). An ANOSIM indicated that the floral community composition was dependent upon survey method ($R = 0.6656$, p-value = 0.001). Pan-traps captured 284 total insects with the majority of insects captured belonging to the orders Diptera (33%) and Coleoptera (41%). There were 1,531 insect observations made during visual surveys with the majority (87%) belonging to the order Hymenoptera (Figure 7b).

The NMDS vector analysis indicated that only Julian date was significantly correlated with the NMDS configuration (Vector Max $R = 0.3036$; p-value = 0.001; Table 4). Julian date was then selected for use as a covariate in all nested ANCOVA models. A nested ANCOVA revealed that the occurrence of *A. nigrae* ($F_{13,40} = 5.6233$; p-value = 0.0310), *A. morrisonella* ($F_{13,40} = 7.2722$, p-value = 0.0140), and *L. coeruleum* ($F_{13,40} = 4.2768$, p-value = 0.0500) was dependent on the sex of the tree (Table 5,7; Figure 7). Additionally, species richness ($F_{13,40} = 17.749$; p-value = 0.0004) and Shannon-Weaver diversity ($F_{13,40} = 28.936$; p-value < 0.0001) also differed between the sexes, with males demonstrating higher values (Table 6,7; Figure 8). Finally, the abundance of *A. macoupinense* differed significantly among trees ($F_{13,40} = 3.3835$, p-value = 0.0030).

A significant correlation was found between the abundance of *A. nigrae* and two volatile compounds: acetophenone ($\rho = 0.9122$, p-value = 0.05) and octen-2-ol ($\rho = 0.9393$, p-value = 0.01) (Figure 9). Additional VOCs and insect activities were not significantly correlated (p-value > 0.05).

Floral community through time

An ANOSIM indicated that the floral community composition differed among years ($R = 0.1669$, p-value = 0.001). The vector analysis indicated that only Julian date was significantly correlated with the NMDS dimensions (Vector Max $R = 0.0802$; p-value = 0.047; Table 8). A nested ANCOVA (Table 9,10; Figure 10) showed that average species richness ($F_{8,68} = 32.2841$, p-value = 0.0046) and Shannon-Weaver diversity ($F_{8,68} = 34.6569$; p-value = 0.0041) were dependent on the sex of the tree. Additionally, flowering time occurred substantially earlier in 2017 than in the other survey years (Figure 11a), which may be related to the dependence of *A. nigrae* activity on the survey year ($F_{8,68} = 5.1049$; p-value = 0.0086; Figure 11b).

4.6 Discussion

The total floral community composition was strongly influenced by the sex of tree across survey techniques and in all years. Male trees consistently attracted a more diverse and unique insect assemblage on their flowers when compared to female trees. However, there was no relationship detected between the sex of the tree and the floral VOC composition, indicating that the scent of male and female flowers among trees was not different. Total emission of monoterpenes, sesquiterpenes, and all VOCs were also found to be similar among sexes. Differences in floral VOCs and plant tissue defenses among trees were also not strong drivers of the community of

insects that visited the flowers. Thus, we found no evidence that the influence that sex of tree on floral insect assemblages was due to the effects of VOCs or metabolite composition, although more subtle relationships might be revealed with more intensive sampling.

Our results are unusual given that dioecious plants, in which both sexes offer rewards, are often dimorphic in both scent composition and emission levels (Füssel, 2007; Ashman, 2009; Okamoto *et al.*, 2013). Nevertheless, this pattern does not always hold true in the *Salix* genus. For example, *S. repens*, *S. bicolor*, *S. caprea*, and *S. cinerea* all have similar overall volatile composition among male and female flowers (Tollsten & Knudsen, 1992; Füssel, 2007). This may reflect differences in the balance between visual and olfactory cues among *Salix* species, although this hypothesis remains to be robustly tested.

Although male and female trees did not differ in VOC and leaf metabolite composition, there was a sex effect on the metabolite composition of catkins. We also found that catkin metabolites exerted a strong influence over herbivore composition, suggesting that floral defenses are utilized differently among male and female trees, in turn attracting unique assemblages of herbivores. In the early season, trees may be at risk of herbivores feeding directly on flowers which act as a large resource sink in plant tissues (Wäckers *et al.*, 2007). Floral larceny is a threat to female reproduction since trees must maintain flowers through seed production, and this relationship may be antagonized in a dioecious system where chance of accidental pollination is rare (Maloof & Inouye, 2000; Richardson, 2004). Thus, it is important that females invest more resources toward defense of catkins, as supported by our finding that female individuals emitted more (3*E*)-hexenyl acetate and (2*E*)-hexenyl acetate. Hexenyl acetates have been frequently

characterized as common green leaf volatiles emitted upon the crushing of plant tissue (Whitman & Eller, 1990; Heil *et al.*, 2008; Wei & Kang, 2011) and they appear to also be a common component in floral scent (Tollsten & Knudsen, 1992; Kaiser, 1994; Messinger, 2006). Hexenyl acetates are associated with the attraction of insect parasites, which may provide female flowers in *S. nigra* added protection benefits against floral resource theft (Whitman & Eller, 1990; Ngumbi *et al.*, 2009).

In our 2019 floral community surveys, pan-traps were employed to test differences in survey techniques and characterize total insect community shifts among male and female trees. We determined that pan-trapping and visual survey techniques captured vastly different samples of the floral visitor community. Pan-traps were very effective at capturing small insects which belonged to Diptera and Coleoptera, explaining much of the disparity between techniques. However, hymenopterans made up 87% of insects counted during visual surveys with 67% of bee observations coming from native *Andrena* spp. Conversely, hymenopterans only made up 22% of insects captured in pan-traps with only 8% of the bees captured belonging to *Andrena* spp. Unscented pan-traps filter the insect community, which was presumably attracted to the flowers on the tree, to those relying heavily on visual cues (Laubertie *et al.*, 2006; Tuell & Isaacs, 2009; Vrdoljak & Samways, 2012). The lack of native *Andrena* spp. attraction to pan-traps may indicate that both visual and volatile cues are important in their association rather than only visual cues. Additionally, given that the pan-trap data increased the diversity disparity observed between male and female flowers in 2019, it appears that the initial visual component of male pollen may be influencing the differences in the larger floral community shift observed

among male and female trees. This is further supported by the differential rates of capture of the hemipterans and hymenopterans in yellow cups and coleopterans in blue cups.

Surveys completed in 2019 revealed the sex of tree was important for many of the Hymenopteran distributions throughout our site. *A. morisonella*, *L. coeruleum*, and a parasitoid wasp species in the family Braconidae all showed higher activity on male trees indicating a prioritization of male rewards. *A. macoupinense* was equally observed among male and female trees. Interestingly, *A. nigrae* was more actively visiting female flowers on trees in the study site. Given the non-showy nature of the *Salix* female catkins, this suggests that the species may be more tightly coupled to the volatile rather than visual cues in order to locate an appropriate host. In support of this, *A. nigrae* was the only insect whose activity was significantly correlated to level of specific floral volatiles. The number of visits of *A. nigrae* increased with increasing levels of both acetophenone and octen-2-ol.

Despite detection of sex differences in activity of native bees in 2019, *Andrena* spp. were stable among male and female flowers through time, indicating that all three species are important in contributing to sexual reproduction of *S. nigra*. Additionally, the total activity of *A. macoupinense* and *A. morisonella* were stable across years, but *A. nigrae* had far fewer catkin visits in 2017 when compared to later years. Adult emergence of *A. macoupinense* has been recorded mid-March through May, *A. nigrae* is often found April through May, and *A. morisonella* is frequent from May to early June (Stevens, 1949; Ribble, 1968, 1974). Of the three species, *A. nigrae* emergence is more closely coupled to the bloom of local *S. nigra* trees. Additionally *A. nigrae* has been recorded as a primary pollinator of *S. nigra* in two other states,

Illinois and North Dakota, indicating that it may be more tightly coupled to the biology of *S. nigra* than other observed native bees in our site (Stevens, 1949; Ribble, 1968).

When examining the phenology of flowers through time, the bloom time of trees in 2017 occurred earlier than subsequent years due to uncharacteristically warm weather in early March followed by much cooler temperatures through April and early May. The abnormality of temperature may have decoupled the availability of floral resources from local *A. nigrae* population emergence. Additionally, cooler temperatures may have had a negative effect on volatilization of catkin scent, leading to the inability of native bees to effectively locate their host. The lack of activity difference detected for *A. macoupinense* and *A. morisonella* may indicate that local populations effectively supplement their diets with resources from additional plant species, such as other local *Salix*, including *S. discolor* and *S. bebbiana*, or *Prunus* and *Amelanchier* spp., both of which are native to West Virginia and listed as occasional hosts (Ribble, 1974).

While male trees attract a more diverse community of floral visitors, males also overproduce resources (average pollen:ovule ratio ~1500:1, J. Grady, unpublished data), thus minimizing competition between native pollinators and pollen/nectar thieves. Conversely, female individuals lack strong visual cues and, despite having similar floral volatile composition, appear to have different protective chemistry composition in their catkins allowing them to preserve their rewards directly for the cross-pollinating *Andrena* species. Competition for resources does not seem to be a direct threat to *Andrena* activity or successful sexual reproduction of *S. nigra*. However, we show that one native bee species may be vulnerable when local tree bloom time is

decoupled from their seasonal emergence. For most native pollinators and plants, partial asynchrony with bloom time is not a threat if additional hosts are present and insect phenology has time to adjust to permanently altered conditions such as those caused by climate change (Willmer, 2012; Bartomeus *et al.*, 2013). For oligolectic species like *A. nigrae* that rarely collect from hosts outside of a single plant group or species, even mild decoupling could have drastic implications for locally adapted bee populations. As a result, plant sexual reproduction will be negatively impacted, especially when the cross-pollinating community is made up of specialists, and accidental pollination becomes rare.

Overall, differences in floral community composition and diversity among male and female trees of *S. nigra* appears to be strongly driven by visual cues. However, the main cross pollinators, *Andrena* spp., rely on both visual and specific volatile cues to locate male individuals and carry pollen to the less showy female trees. Additionally, the interannual variability in flowering time highlights the potential effects of climate on pollinator activity and tree sexual productivity. Inter-annual variability in the *S. nigra* – *Andrena* interaction illustrates that shifting seasonal transitions could detrimentally affect plants that depend on early-emerging arthropods for sexual reproduction as well as the arthropods that depend on resources provided by early-flowering plants.

4.7 Literature cited

- Adams RP. 2001.** *Identification of essential oil components by gas chromatography/mass spectrometry*. Carol Stream, IL: Allured Publishing Corporation.
- Argus GW. 2011.** An experimental study of hybridization and pollination in *Salix* (willow). *Canadian Journal of Botany* **52**: 1613–1619.
- Ashman TL. 2009.** Sniffing out patterns of sexual dimorphism in floral scent. *Functional Ecology* **23**: 852–862.
- Bartomeus I, Park MG, Gibbs J, Danforth BN, Lakso AN, Winfree R. 2013.** Biodiversity ensures plant-pollinator phenological synchrony against climate change (M Eubanks, Ed.). *Ecology Letters* **16**: 1331–1338.
- Beismann H, Wilhelmi H, Baillères H, Spatz H, Bogenrieder A, Speck T. 2000.** Brittleness of twig bases in the genus *Salix*: fracture mechanics and ecological relevance. *Journal of Experimental Botany* **51**: 617–633.
- Burns RM, Honkala BH. 1990.** *Silvics of North America: Volume II: hardwoods*. Washington, D.C., USA: United States Department of Agriculture (USDA), Forest Service.
- Danforth B. 2007.** Bees. *Current Biology* **17**: 156–161.
- Dötterl S, Glück U, Jürgens A, Woodring J, Aas G. 2014.** Floral reward, advertisement and attractiveness to honey bees in dioecious *Salix caprea* (AG Dyer, Ed.). *PLoS ONE* **9**: e93421.
- El-Sayed AM. 2013.** The pherobase: database of insect pheromones and semiochemicals.
- Füssel U. 2007.** Floral scent in *Salix* L. and the role of olfactory and visual cues for pollinator attraction of *Salix caprea* L. [PhD-Thesis].
- Füssel U, Dötterl S, Jürgens A, Aas G. 2007.** Inter- and intraspecific variation in floral scent in

- the genus *Salix* and its implication for pollination. *Journal of Chemical Ecology* **33**: 749–765.
- Galen C, Butchart B. 2003.** Ants in your plants: effects of nectar-thieves on pollen fertility and seed-siring capacity in the alpine wildflower, *Polemonium viscosum*. *Oikos* **101**: 521–528.
- Heil M, Lion U, Boland W. 2008.** Defense-inducing volatiles: in search of the active motif. *Journal of Chemical Ecology* **34**: 601–604.
- Kaiser R. 1994.** *Trapping, investigation and reconstitution of flower scents*. Springer Netherlands.
- Karrenberg S, Edwards P, Kollmann J. 2002.** The life history of Salicaceae living in the active zone of floodplains. *Freshwater Biology* **47**: 733–748.
- Keefover-Ring K. 2013.** Making scents of defense: do fecal shields and herbivore-caused volatiles match host plant chemical profiles? *Chemoecology* **23**: 1–11.
- Laubertie EA, Wratten SD, Sedcole JR. 2006.** The role of odour and visual cues in the pan-trap catching of hoverflies (Diptera: Syrphidae). *Annals of Applied Biology* **148**: 173–178.
- Lindroth RL, Scriber JM, Hsia MTS. 1986.** Differential responses of tiger swallowtail subspecies to secondary metabolites from tulip tree and quaking aspen. *Oecologia* **70**: 13–19.
- Linsley EG. 1958.** The ecology of solitary bees. *Hilgardia* **27**: 543–599.
- Maloof JE, Inouye DW. 2000.** Are nectar robbers cheaters or mutualists? *Ecology* **81**: 2651–2661.
- Messinger OJ. 2006.** The role of visual and olfactory cues in host recognition for the specialist

- bee genus *Diadasia*, and implications for the evolution of host choice [PhD Thesis].
- Ngumbi E, Chen L, Fadamiro HY. 2009.** Comparative GC-ead responses of a specialist (*Microplitis croceipes*) and a generalist (*Cotesia marginiventris*) parasitoid to cotton volatiles induced by two caterpillar species. *Journal of Chemical Ecology* **35**: 1009–1020.
- NIST. 2008.** *National institute of standards and technology mass spectral library*. Washington, D.C.: National Institute of Standards and Technology, US Department of Commerce.
- Okamoto T, Kawakita A, Goto R, Svensson GP, Kato M. 2013.** Active pollination favours sexual dimorphism in floral scent. *Proceedings of the Royal Society B: Biological Sciences* **280**.
- Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O’Hara RB, Simpson GL, Solymos P, *et al.* 2019.** vegan: community ecology package. R package version 2.5-4. <https://cran.r-project.org/package=vegan>.
- Ostaff DP, Mosseler A, Johns RC, Javorek S, Klymko J, Ascher JS, Johns A, Javorek RC, And Ascher J. 2015.** Willows (*Salix* spp.) as pollen and nectar sources for sustaining fruit and berry pollinating insects. *Canadian Journal of Plant Science* **95**: 505–516.
- Ribble DW. 1968.** Revisions of two subgenera of *Andrena*: *Micrandrena* (Ashmead) and *Derandrena*. New subgenus (Hymenoptera: Apoidea). *Bulletin University of Nebraska Museum* **8**: 237–394.
- Ribble DW. 1974.** A revision of the bees of the genus *Andrena* of the western hemisphere subgenus *Scaphandrena*. *American Entomological Society* **100**: 101–189.
- Richardson SC. 2004.** Are nectar-robbers mutualists or antagonists? *Oecologia* **139**: 246–254.
- Stevens OA. 1949.** Native bees. *Experimental Station Bimonthly Bulletin* **12**: 90–98.

- Straka J, Rina Cern K, Mach L, Zemenov M, Keil P. 2014.** Life span in the wild: the role of activity and climate in natural populations of bees. *Functional Ecology* **28**: 1235–1244.
- Tamura S, Kudo G. 2000.** Wind pollination and insect pollination of two temperate willow species, *Salix miyabeana* and *Salix sachalinensis*. *Plant Ecology* **147**: 185–192.
- Tollsten L, Knudsen JT. 1992.** Floral scent in dioecious *Salix* (Salicaceae)- a cue determining pollination system? *Plant Systematics and Evolution* **182**: 229–237.
- Tuell JK, Isaacs R. 2009.** Elevated pan traps to monitor bees in flowering crop canopies. *Entomologia Experimentalis et Applicata* **131**: 93–98.
- Vrdoljak SM, Samways MJ. 2012.** Optimising coloured pan traps to survey flower visiting insects. *Journal of Insect Conservation* **16**: 345–354.
- Wäckers FL, Romeis J, van Rijn P. 2007.** Nectar and pollen feeding by insect herbivores and implications for multitrophic interactions. *Annual Review of Entomology* **52**: 301–323.
- Wei J, Kang L. 2011.** Roles of (Z)-3-hexenol in plant-insect interactions. *Plant Signaling and Behavior* **6**: 369–371.
- Whitman DW, Eller FJ. 1990.** Parasitic wasps orient to green leaf volatiles. *Chemoecology* **1**: 69–76.
- Willmer P. 2012.** Ecology: pollinator–plant synchrony tested by climate change. *Current Biology* **22**: R131–R132.

4.8 Tables and Figures

Table 1. NMDS and ANOSIM (Bray-Curtis dissimilarity) results for multivariate floral visitor composition tested against sex grouping (male vs female). p-values < 0.05 (bolded) indicate that floral visitor composition is more similar within replicate observations of sex group rather than among all observations.

Survey	# dimensions	Stress	ANOSIM R	p-value
2019 visual	4	0.11	0.3110	0.001
2019 pan-traps	3	0.10	0.1532	0.017
2019 total community	3	0.12	0.3077	0.001
2017-2019 community	4	0.12	0.1018	0.032

Table 2. NMDS and ANOSIM (Bray-Curtis dissimilarity) results for multivariate chemistry composition tested against sex grouping (male vs female). p-values < 0.05 (bolded) indicate that chemistry composition is more similar within replicate observations of sex group rather than among all observations.

Multivariate response	# dimensions	Stress	ANOSIM R	p-value
VOCs	3	0.10	-0.1191	0.914
Catkin metabolites	2	0.10	0.1787	0.027
Leaf metabolites	2	0.10	-0.00471	0.449

Table 3. Mantel test results comparing pairwise Bray-Curtis dissimilarity matrices among insect communities and chemistry composition of flowers and leaves. Bolded p-values and positive r_m indicate a significant test, suggesting that similarity in chemistry composition relates to similarity in insect community assemblage.

Insect matrix	Chemistry matrix	r_m	p-value
Floral community	Catkin VOCs	-0.300	0.07
Floral community	Catkin metabolites	0.062	0.380
Floral community	Leaf metabolites	0.002	0.480
Herbivore community	Catkin metabolites	0.416	0.010
Herbivore community	Leaf metabolites	0.275	0.07

Table 4. Environmental fit results for 2019 floral community NMDS analysis. Bolded p-values (<0.05) indicate a significant correlation of the independent variable with the NMDS configuration.

Independent variable	Vector Max R	p-value
Julian date	0.3036	0.001
Temperature	0.0582	0.241
Military time	0.0447	0.308

Table 5. Nested ANCOVA model results for 2019 analysis of individual native bee abundances.

Bolded p-values (<0.05) indicate significant model effects.

$y \sim \text{Sex} + \text{Location} + \text{Tree}(\text{Sex}) + \text{Random} + \text{Julian date}$

Dependent variable: <i>Andrena nigrae</i>					
Source	Degrees of freedom	Sum of squares	Mean square	F-Ratio	p-value
Model	13	43.48185	3.34476	2.4947	0.0134
Error	40	53.62887	1.34072		
C. Total	53	97.11072			
Dependent variable: <i>Andrena macoupinense</i>					
Source	Degrees of freedom	Sum of squares	Mean square	F-Ratio	p-value
Model	13	5.604838	0.431141	3.9674	0.0004
Error	40	4.34681	0.10867		
C. Total	53	9.951648			
Dependent variable: <i>Andrena morrisonella</i>					
Source	Degrees of freedom	Sum of squares	Mean square	F-Ratio	p-value
Model	13	4.347317	0.334409	2.9616	0.0042
Error	40	4.516573	0.112914		
C. Total	53	8.86389			
Dependent variable: <i>Lasioglossum coeruleum</i>					
Source	Degrees of freedom	Sum of squares	Mean square	F-Ratio	p-value
Model	13	3.101719	0.238594	2.2816	0.023
Error	40	4.182974	0.104574		
C. Total	53	7.284693			

Table 6. Nested ANCOVA model results for 2019 analysis of calculated community metrics.

Bolded p-values (<0.05) indicate significant model effects.

$y \sim \text{Sex} + \text{Location} + \text{Tree}(\text{Sex}) + \text{Random} + \text{Julian date}$

Dependent variable: Species richness					
Source	Degrees of freedom	Sum of squares	Mean square	F-Ratio	p-value
Model	13	307.5722	23.6594	3.3871	0.0015
Error	40	279.4093	6.9852		
C. Total	53	586.9815			

Dependent variable: Shannon-Weaver diversity					
Source	Degrees of freedom	Sum of squares	Mean square	F-Ratio	p-value
Model	13	9.700905	0.746223	5.2588	<0.0001
Error	40	5.675971	0.141899		
C. Total	53	15.37688			

Table 7. Test of random effects p-values extracted from nested ANCOVA for most abundant floral visitors as well as calculated species richness and Shannon Weaver diversity for 2019 survey analysis. Bolded values indicate that the independent variable had a significant effect on the dependent variable (p-value < 0.05).

	<i>Andrena nigrae</i>	<i>Andrena macoupinense</i>	<i>Andrena morrisonella</i>	<i>Lasioglossum coeruleum</i>	Species richness	Shannon-Weaver diversity
Sex	0.031	0.5447	0.014	0.0500	0.0004	<0.0001
Location	0.5645	0.3587	0.8326	0.9577	0.9536	0.7348
Tree	0.2046	0.003	0.4729	0.1635	0.5944	0.6917
Julian date	0.5645	0.6261	0.0600	0.6141	0.5800	0.2839

Table 8. Environmental fit results for floral community NMDS analysis across years (2017-2019). Bolded p-values (<0.05) indicate a significant correlation of the independent variable with the NMDS configuration.

Independent variable	Vector Max R	p-value
Year	0.0802	0.047
Julian date	0.0815	0.046
Temperature	0.0491	0.18

Table 9. Nested ANCOVA model results for year analysis. Bolded p-values (<0.05) indicate significant model effects.

$y \sim \text{Sex} + \text{Year} + \text{Tree}(\text{Sex}) + \text{Random} + \text{Julian date}$

Dependent variable: <i>Andrena macoupinense</i> & <i>Andrena morrisonella</i>					
Source	Degrees of freedom	Sum of squares	Mean square	F-Ratio	p-value
Model	8	27.17312	3.39664	1.6702	0.1218
Error	68	138.2903	2.03368		
C. Total	76	165.4635			
Dependent variable: <i>Andrena nigrae</i>					
Source	Degrees of freedom	Sum of squares	Mean square	F-Ratio	p-value
Model	8	46.10628	5.76328	4.9103	<0.0001
Error	68	79.81222	1.17371		
C. Total	76	125.9185			
Dependent variable: Species richness					
Source	Degrees of freedom	Sum of squares	Mean square	F-Ratio	p-value
Model	8	82.47375	10.3092	4.8422	<0.0001
Error	68	144.773	2.129		
C. Total	76	227.2468			
Dependent variable: Shannon-Weaver diversity					
Source	Degrees of freedom	Sum of squares	Mean square	F-Ratio	p-value
Model	8	5.76738	0.720923	7.3942	<0.0001
Error	68	6.62987	0.097498		
C. Total	76	12.39725			

Table 10. Test of random effects p-values extracted from nested ANCOVA for most abundant floral visitors as well as calculated species richness and Shannon Weaver diversity. Bolded values indicate that the independent variable had a significant effect on the dependent variable (p-value < 0.05).

	<i>Andrena nigrae</i>	Species richness	Shannon-Weaver diversity
Sex	0.1475	0.0046	0.0041
Year	0.0086	0.7303	0.9019
Tree	0.1134	0.4356	0.2488
Julian date	0.2171	0.2467	0.2617

Figure 1. Map of all thirty *Salix nigra* tree locations in WVU Core Arboretum. Pin shapes indicate tree selected for floral insect community survey. Points indicate additional individuals in the Arboretum.

Figure 2. Tree canopy pan-trap design. Camouflaged buckets were added to hang below river traps to add weight, preventing winds from jostling traps.

Figure 3. Non-metric multidimensional plot (dimensions = 3; stress = 0.12) of insect floral community with groupings indicated by color for sex of tree (ANOSIM $R = 0.3077$, p -value = 0.001) for 2019 analysis. Size of species text indicates position relative to the dimensional space, for example, larger font indicates forward projection along the positive values of all axes.

Figure 4. Non-metric multidimensional plot (dimensions = 4; stress = 0.12) of 2017-2019 insect floral community with groupings indicated by color for sex of tree (ANOSIM $R = 0.1232$, p -value = 0.011). Size of species text indicates position relative to the dimensional space, for example, larger font indicates forward projection along the positive values of all axes.

Figure 5. Total concentration averages of catkin metabolites, leaf metabolites and catkin VOCs for male and female trees. Letters to the left of boxes indicate significantly different means (p -value < 0.05) determined by one-way ANOVA test.

Figure 6. Total concentration averages of monoterpenes and sesquiterpenes VOCs for male and female trees. Letters to the left of boxes indicate significantly different means (p-value < 0.05) determined by one-way ANOVA test.

Figure 7. (a) Non-metric multidimensional plot (dimensions = 3; stress = 0.12) of insect floral community with groupings indicated by color for survey method (ANOSIM R = 0.6656, p-value = 0.001). (b) Breakdown of insect capture/observation for two canopy survey methods. Additional pie charts around pan-trap pie chart indicate the percentage of that order captured in different colors of pan-traps.

Figure 8. Average activity of most common floral visitors and calculated community metrics from 2019 *S. nigra* surveys. Letters to the left of boxes indicate significantly different means as determined by a Tukey's HSD (p-value < 0.05).

Figure 9. Correlation plots of *A. nigrae* with VOC compounds.

Figure 10. Average values of species richness and Shannon-Weaver diversity for female and male trees from year analysis model. Letters to the left of boxes indicate significantly different means as determined by a Tukey's HSD (p-value < 0.05) for each separate nested ANCOVA model.

Figure 11. (a) Amount of time and date in each year (2017, 2018, and 2019) in which catkins were actively attractive to insects. End date for male individuals indicates trees have dropped all catkins while female trees no longer have receptive stigmas (brown and shriveled) or active insect in canopies. (b) Average count of *A. nigrae* visits to *S. nigra* catkins in survey years 2017, 2018 and 2019 from year analysis model. Letters to the left of boxes indicate significantly different means as determined by a Tukey's HSD (p-value < 0.05).

Figure 1.

Figure 2.

Figure 3.

Figure 4.

Figure 5.

Figure 6.

Figure 7.

Figure 8.

Figure 9.

Figure 10.

Figure 11.

Chapter 5: Conclusions

5.1 Overview of research goals

For my dissertation, I used *Populus* and *Salix* as models in order to study the host genetic influence on biotic communities by: 1) utilizing a hybrid backcross family to identify loci underlying fungal and insect associations in *Populus*, and genomic comparisons of *P. trichocarpa* and *P. deltoides* to understand genetic mechanisms underlying host plant-biotic relationships; 2) using an association population of *P. trichocarpa* to identify and uncover potential biological function of candidate genes underlying individual arthropod associations and entire community composition; and 3) studying the impacts of dioecy on floral insect communities using insect pollinated *S. nigra*.

5.2 Hybrid mediation of biotic associations

In Chapter 2, I identified several fungal pathogens and insect species that were segregating among our *P. deltoides* $\times$ *P. trichocarpa* pseudo-backcross family progeny. QTL analysis revealed a combined 7 loci controlling the distributions of 5 surveyed species, including *Melampsora* sp. fungal leaf rust, *S. musiva* fungal leaf spot, *S. musiva* fungal stem canker, *M. vagabunda* branch galling aphid, and *Phyllocolpa* sp. leaf folding sawfly. An overlapping genetic interval on Chr04 was discovered to be associated with the biotrophic *Melampsora* sp. and necrotrophic *S. musiva* fungus. Upon further investigation I found that the hybrids were mediating the interaction between the two pathogens, resulting in competitive exclusion (Orton & Brown, 2016), which led to an inflation of the relationship between the Chr04 interval and occurrence of *S. musiva* leaf spot symptoms (Koskella *et al.*, 2006; Abdullah *et al.*, 2017). An additional QTL interval was also discovered to co-occur with the *Phyllocolpa* sp. Chr10 association, which was important in the host-plant production of the metabolite gentisyl alcohol

5-O-glucoside. Given that the positive allele contributing to an increase in both phenotypes came from *P. trichocarpa*, the metabolite may be acting as an ovipositional cue for egg-laying female sawflies.

Genomic comparisons of the parents of the backcross family revealed several interesting observations. The Chr16 interval that was important in the severity of *S. musiva* canker symptoms did not overlap with the four loci that have previously been implicated in *P. trichocarpa* susceptibility (Muchero *et al.*, 2018), which suggests that the hybrids are mediating the pathogen interactions using a potentially analogous, but distinct, mechanism. The QTL interval contained a tandem repeat of G-type lectin receptor-like kinases that were expanded in the *P. deltoides* parent, which may play a similar role to the receptor-like kinase family known to be important in controlling the infection severity of the pathogen in the *P. trichocarpa* study (Muchero *et al.*, 2018). Our other QTL intervals revealed functional enrichments and multiple recent tandem duplications of genes that are important in resistance to biotic stress such as cell-wall loosening, cell death signaling, secondary metabolite production, and R-genes as well as ones that may be important in susceptibility such as sugar transportation that can make source tissues into nutritive sinks. Finally, I found that recent tandem gene expansion was enriched in the *P. delotides* parent genome, but not in the *P. trichocarpa* parent. Most of the herbivores and pathogens surveyed in our study have been co-evolving with *P. deltoides*, leading to a longer history of gene-for-gene interactions that potentially left a lasting impact on the genome structure in the form of recent tandem gene duplications.

5.3 *P. trichocarpa* community genetics

In Chapter 3, I used an association population of *P. trichocarpa* and surveys of arthropod activity across three common gardens and two years to better understand the genetic control of abundant arthropods and community composition. I discovered that, despite the arthropod species distributions being highly variable among each garden, there was a strong host-genetic influence mediating the assembly of herbivores and predators in all gardens and years. Furthermore, the occurrence of abundant arthropods and community metrics, such as arthropod richness (total number of unique species on plant tissue), abundance (total number of arthropods on plant tissue), and Shannon Weaver diversity, were found to be strongly heritable across all sites. Two gardens, Clatskanie 2012 and Corvallis 2015, had an adequate number of genotypes surveyed for genome wide association analysis (GWAS). Single variable traits including abundance of arthropods such as *Harmandia* sp., *Phyllocolpa* sp., and binarized occurrences of *Chrysomela scripta*, *Phyllonorycter* sp. blotch miner, *Phyllocnistis populiella* serpentine miner, and *Corythucha* sp. lace bug. Additionally, community metrics calculated from each garden, such as richness, abundance, and diversity, were included in single-trait analyses. Finally, a multi-trait GWAS was also implemented for the axis scores extracted from the NMDS configurations for the Clatskanie and Corvallis sites to look for genes important in underlying arthropod community composition.

Significantly associated SNPs were identified for 10 of the 14 traits on or near 88 potential candidate genes of interest and, of these, only 32 yielded significant network results based on genes that were co-expressed with candidate genes, methylation data for multiple *P. trichocarpa* tissue types (Vining *et al.*, 2012), and GWAS for pyMBMS and metabolite profiles

(Tschaplinski *et al.*, 2012). Candidate genes with biologically meaningful functions were selected for further investigation from networks based on a depth (number of connected layers) cutoff of 3 and a breadth (number of neighbor genes) cutoff of 500, leaving us with 6 genes of interest associated with the distribution of *Phyllonorycter* sp. blotch miner in the Clatskanie garden and arthropod richness and composition from the Corvallis garden. Candidate genes important in the association of the specialist herbivore *Phyllonorycter* sp. were found to be important in tissue senescence and metabolite production, whereas genes important in arthropod richness and community composition had functions that were capable of more broadly targeting multiple species of arthropods, such as terpenoid production, RNA interference, and transmembrane signaling.

5.4 *S. nigra* floral community

In Chapter 4, I was interested in exploring the effect of plant dioecy on communities of insects in a natural population of *S. nigra*. To achieve this, I surveyed flowers for insect visitors using both visual and pan trapping techniques across three years, insect herbivores that feed on trees during flowering and after catkins are no longer present. Various traits collected from the *S. nigra* trees, including sex of the individual, catkin volatile organic compounds (VOCs), catkin metabolites, and leaf metabolites, were explored as potential candidates mediating the interactions of the trees with their arthropod communities. Total composition of floral visitors did not appear to be related to defensive or volatile chemistry composition of flowers or leaves but were strongly driven by sex of tree, indicating that visual cues, such as the yellow color of male pollen are important in driving the diversity disparity observed among male and female trees. Composition

of herbivores however, did appear to be related to tree defensive metabolites, especially in the case of catkin phenolics, which were also strongly driven by sex of tree.

I identified three insects that frequently visited both male and female trees that are the most likely cross-pollinators of *S. nigra* in my site. All three were identified as species of the solitary ground nesting bees *Andrena*, including *A. morisonella*, *A. macoupinense*, and *A. nigrae*. Two of the species, *A. morisonella* and *A. macoupinense*, showed higher activity on male catkins, while *A. nigrae* showed stronger preference for female trees. Female catkins lack attractive visual cues, suggesting that insects that can easily locate their resources may be relying on specific volatile cues. This was supported by my finding that *A. nigrae* activity was strongly correlated to the levels of two volatile compounds, including acetophenone and octen-2-ol. My final finding in this chapter was that, through time, the activity of *A. morisonella* and *A. macoupinense* stayed stable on *S. nigra* flowers, but the activity of *A. nigrae* was lower in 2017 when flowers bloomed the earliest. This created an asynchrony between the emergence of the native bees and availability of floral resources which is not only a threat to the sexual reproduction of *S. nigra* trees, but to the survival of local *A. nigrae* populations that rely on this single host species. Host-insect asynchrony in tightly-linked relationships, such as the *Salix-Andrena* system, threaten to become more frequent and widespread given global impacts of climate change altering their ecosystems (Willmer, 2012; Bartomeus *et al.*, 2013).

5.5 Overall conclusions

In conclusion, I found that hybrids can be a valuable tool for uncovering loci and genomic patterns associated with biotic distributions. However, hybrid mediations of biotic interactions can also lead to unique genetic associations that would not be found in a pure species population, which may contribute to their elevated community effects. I also showed, through common garden surveys in multiple years, that genetic variation in a single species is capable of strongly influencing the organization of arthropod community assembly. From this same population I discovered multiple underlying genes and their potential biological functions contributing to several different levels of community organization. Determining if functionally similar genes have the same outcome in other pure species of *Populus* will enhance the predictive power of host plant genetics on ecosystem assembly. Finally, I found that community assembly of insect floral visitors and herbivores is detectable in a natural population of *S. nigra* and is mediated by a multitude of genetically controlled characteristics, including dioecy, volatile organic compounds, and catkin/leaf metabolites. Additionally, I found abnormal seasonal temperature created an asynchrony in catkin bloom and insect emergence, which has the potential to alter tree-pollinator relationships in early spring. Overall: plant genotypes dictate plant physiology and morphology in ways that structure the greater biotic community. Host plant genetic effects combine with spatial and temporal environmental variability to drive dynamic plant-pathogen-herbivore-pollinator interactions.

5.6 Literature cited

- Abdullah AS, Moffat CS, Lopez-Ruiz FJ, Gibberd MR, Hamblin J, Zerihun A. 2017.** Host–multi-pathogen warfare: pathogen interactions in co-infected plants. *Frontiers in Plant Science* **8**: 1806.
- Bartomeus I, Park MG, Gibbs J, Danforth BN, Lakso AN, Winfree R. 2013.** Biodiversity ensures plant–pollinator phenological synchrony against climate change. *Ecology Letters* **16**: 1331–1338.
- Koskella B, Giraud T, Hood ME. 2006.** Pathogen relatedness affects the prevalence of within-host competition. *The American Naturalist* **168**: 121–126.
- Muchero W, Sondreli KL, Chen JG, Urbanowicz BR, Zhang J, Singan V, Yang Y, Brueggeman RS, Franco-Coronado J, Abraham N, et al. 2018.** Association mapping, transcriptomics, and transient expression identify candidate genes mediating plant-pathogen interactions in a tree. *Proceedings of the National Academy of Sciences of the United States of America* **115**: 11573–11578.
- Orton ES, Brown JKM. 2016.** Reduction of growth and reproduction of the biotrophic fungus *Blumeria graminis* in the presence of a necrotrophic pathogen. *Frontiers in Plant Science* **7**: 742.
- Tschaplinski TJ, Standaert RF, Engle NL, Martin MZ, Sangha AK, Parks JM, Smith JC, Samuel R, Jiang N, Pu Y, et al. 2012.** Down-regulation of the caffeic acid O-methyltransferase gene in switchgrass reveals a novel monolignol analog. *Biotechnology for biofuels* **5**: 71.

- Vining KJ, Pomraning KR, Wilhelm LJ, Priest HD, Pellegrini M, Mockler TC, Freitag M, Strauss SH. 2012. Dynamic DNA cytosine methylation in the *Populus trichocarpa* genome: tissue-level variation and relationship to gene expression. *BMC Genomics* **13**: 27.
- Willmer P. 2012. Ecology: pollinator–plant synchrony tested by climate change. *Current Biology* **22**: R131–R132.