




12-2020

Phylogenetics and association analyses illustrate substantial cryptic diversity of a newly isolated collection of *Cenococcum geophilum*

Jessica Velez

University of Tennessee, Knoxville, jvelez@vols.utk.edu

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss

 Part of the [Bioresource and Agricultural Engineering Commons](#), and the [Other Biomedical Engineering and Bioengineering Commons](#)

Recommended Citation

Velez, Jessica, "Phylogenetics and association analyses illustrate substantial cryptic diversity of a newly isolated collection of *Cenococcum geophilum*. " PhD diss., University of Tennessee, 2020.
https://trace.tennessee.edu/utk_graddiss/6095

This Dissertation is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Jessica Velez entitled "Phylogenetics and association analyses illustrate substantial cryptic diversity of a newly isolated collection of *Cenococcum geophilum*." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Energy Science and Engineering.

Christopher W. Schadt, Major Professor

We have read this dissertation and recommend its acceptance:

Mark Radosevich, Sarah Lebeis, Jessy Labbe, Wellington Muchero, Scott Emrich

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Phylogenetics and association analyses
illustrate substantial cryptic diversity of
a newly isolated collection of
Cenococcum geophilum

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Jessica Marie Velez

December 2020

Copyright © by Jessica Marie Velez, 2020
All rights reserved.

To my father.

Acknowledgments

To my family, whose support and pride inspired me on my journey.

To my father, whose love of science and excitement for new discoveries and technology informed my approach to research and life.

To my mother, who provided me with emotional support, love, and patience during this five year journey.

To my PI and mentor Dr. Chris Schadt, who when I asked if he would consider taking me on as a student, replied, "Well, I know that would be a huge benefit to *me*, but let's make sure it would benefit you too." It was a huge benefit to both of us.

This work was performed using soil samples gathered from lands where the traditional village sites of the Multnomah, Kathlamet, Clackamas, Tualatin, Kalapuya, Molalla, bands of the Chinook, and many other Tribes made their homes along the Columbia River. I acknowledge the systemic policies of genocide, relocation, and assimilation that still impact many Indigenous/Native American families today. As settlers and guests on these lands we must respect the work of Native Nations, leaders and families, and make ongoing efforts to center Indigenous knowledge, creativity, resilience, and resistance.

Abstract

The ectomycorrhizal fungus *Cenococcum geophilum* is distributed worldwide across multiple climates and soil types and is known to positively associate with a multitude of plant genera, possibly contributing to plant ability to tolerate inorganic contaminants in a soil environment. New *C. geophilum* isolates are easily cultured from soils in a laboratory setting, making this an ideal candidate for a model species with which to study multiple plant-fungal effects across a collection of novel isolates. However, *C. geophilum* is also genetically complex and, at 178Mbp, features one of the largest fungal genomes, necessitating the use of the novel restriction-associated DNA sequencing (RADseq) technique which produces robust *de novo* single nucleotide polymorphism (SNP) detection. A preliminary investigation into the phylogenetic relationship of >200 new *C. geophilum* isolates from the United States Pacific Northwest (PNW) region using glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) strongly resolved (>80%) 15 cryptic clades. An investigation of the worldwide *C. geophilum* collection using *GAPDH* resolved >30 cryptic clades. In both collections, at least two cryptic clades incorporated extreme spatial diversity in the form of cross-regional, cross-country, and international strains. Phylogenetic analyses of 171 PNW isolates conducted using RADseq strongly supported (>80%) the 15 PNW clades using the *de novo* dataset assembly at a per-site depth of at least 10%. Direct comparison of the PNW ITS and *GAPDH* gene regions indicated strong evidence of sexual recombination and additional analyses confirmed high levels of incongruency between the two genes. However, when these same analyses were conducted on the RADseq *de novo* dataset, no strong evidence of recombination was detected across the PNW collection, suggesting this collection represents a hybridized clonal population with rare localized sexual recombination. A association study linked heavy metal resistance of 56 *C. geophilum* isolates to significant SNP associations

detected from the *de novo* and RADseq assembly, finding that 20 significant ($p < 0.05$) SNPs were detected in the presence of cadmium. These SNPs are linked to a series of metabolic, transcription and translation, and ion-binding protein coding regions as well as two proteins which may be directly involved in resistance to cadmium within isolates from two PNW sites.

Table of Contents

1	Introduction	1
1.1	Agriculture in a global context	1
1.2	<i>Populus</i> spp. as a biofuel crop	2
1.3	Microbial applications to bioremediation and biofuel crop systems	2
1.4	<i>Cenococcum geophilum</i> as a model organism	3
1.5	Phylogenetic studies of the cryptic fungus <i>Cenococcum geophilum</i>	4
1.6	RADseq applied to <i>C. geophilum</i>	5
1.7	Expanding phylogenetic analyses in <i>C. geophilum</i> studies	7
2	Phylogenetic diversity of 200+ isolates of the ectomycorrhizal fungus <i>Cenococcum geophilum</i> associated with <i>Populus trichocarpa</i> soils in the Pacific Northwest, USA and comparison to globally distributed representatives.	8
2.1	Preface	8
2.2	Author Affiliations	8
2.3	Author Contributions	9
2.4	Abstract	9
2.5	Introduction	10
2.6	Materials and Methods	13
2.6.1	Site selection and soil sampling	13
2.6.2	Sclerotia separation	14
2.6.3	DNA extraction	14
2.6.4	PCR amplification	15

2.6.5	Determination of native soil properties	15
2.6.6	Generation of phylogenies	16
2.6.7	Comparison of <i>GAPDH</i> , ITS, and <i>GAPDH</i> + ITS phylogenies	17
2.6.8	Phylogeographic variation within PNW isolates	17
2.6.9	Global <i>GAPDH</i> phylogenetic analyses	18
2.6.10	Recombination patterns within <i>GAPDH</i> and ITS data	18
2.7	Results	19
2.7.1	A diverse culture collection of PNW <i>Cenococcum isolates</i> associated with <i>Populus</i>	19
2.7.2	Phylogeographic variation within PNW isolates	19
2.7.3	Recombination analyses in the PNW isolate collection	21
2.7.4	Phylogeographic variation within the global <i>C. geophilum</i> isolate collection	21
2.8	Discussion	22
2.8.1	<i>Cenococcum</i> phylogeographic diversity in the Pacific Northwest	22
2.8.2	Incongruent genes and indications of recombination within the PNW isolate collection	23
2.8.3	<i>Cenococcum</i> phylogeographic diversity within a global context	25
2.9	Conclusions and future directions	26
2.10	Acknowledgements	27
3	Phylogeographic patterns across regional isolates of the ectomycorrhizal fungus <i>Cenococcum geophilum</i> differ between genome wide restriction associated DNA sequencing and single gene approaches	28
3.1	Preface	28
3.2	Author Affiliations	28
3.3	Author Contributions	29
3.4	Abstract	29
3.5	Introduction	30
3.6	Methods	33

3.6.1	Isolate Growth and DNA Extraction	33
3.6.2	Homemade Serapure (SpeedBead) preparation for 3RAD enzymatic reaction clean-up	33
3.6.3	Optimization of the 3RAD protocol for <i>C. geophilum</i>	34
3.6.4	RADseq data demultiplexing and quality filtering	36
3.6.5	Reference-based and <i>de novo</i> downstream assembly of sequences	36
3.6.6	Clustering and phylogenetic analyses	37
3.6.7	Single gene phylogeny and comparison to <i>de novo</i> assembly	37
3.6.8	Linkage disequilibrium analyses	38
3.7	Results	38
3.7.1	Phylogeographic patterns and distribution in the Pacific Northwest	39
3.7.2	Single gene (GPD) and RADseq <i>de novo</i> (RDN) phylogenetic comparison	39
3.7.3	Linkage disequilibrium within the PNW isolate collection	41
3.8	Discussion	41
3.8.1	Viability of RADseq usage for phylogenetic analysis in fungal studies	41
3.8.2	Phylogenetic structure within the Pacific Northwest isolate collection	42
3.8.3	Single gene compared to RADseq <i>de novo</i> phylogenetic approaches	43
3.8.4	Evidence for clonality and recombination in the PNW isolate collection	44
3.9	Conclusions	46
4	Responses of <i>Cenococcum geophilum</i> to heavy metals and identification of associated genomic markers	48
4.1	Preface	48
4.2	Author Affiliations	48
4.3	Author Contributions	49
4.4	Abstract	49
4.5	Introduction	50
4.6	Methods	53
4.6.1	Growth assay development and strain selection	53
4.6.2	Radial growth rate assay using 4 quadrant plates	54

4.6.3	Statistical analyses of radial growth	54
4.6.4	Determination of significant phenotype-associated SNPs	54
4.7	Results	55
4.7.1	Heavy metal tolerance across the Pacific Northwest isolates	55
4.7.2	SNP detection and phenotype association	56
4.8	Discussion	57
4.9	Conclusion	60
5	Conclusions and Future Directions	61
5.1	Phylogenetic complexity of <i>Cenococcum geophilum</i>	61
5.2	Future directions	62
	Bibliography	64
	Appendix	100
	Vita	131

List of Tables

1	Model with best fit analysis, Bayesian Information Criterion, and Akaike Information Criterion (AICc) for each alignment per an analysis using MEGA X software. Lower BIC and AICc values indicate the best model fit for use in analyses.	106
2	Previously resolved clades per study, number of isolates, geographic region, and host plant. The Pacific Northwest (PNW) isolate collection resolved 15 clades which appear to be uniquely associated with <i>Populus trichocarpa</i> within its host range in the PNW.	112
3	Pacific Northwest <i>Cenococcum geophilum</i> isolates used in this study with the rivershed of origin, <i>de novo</i> assembly cluster, <i>de novo</i> assembly clade and bootstrap support value, and single gene (GPD) clade and bootstrap support value. Unresolved isolates are designated with a dash (-).	113
4	Pacific Northwest (PNW) <i>Cenococcum geophilum</i> isolates used in this study with the rivershed of origin and single gene (GPD) clade. Unresolved isolates are designated with a dash (-). In total, 55 isolates were used representing 10 PNW riversheds and Lake Whatcom, along with the reference 1.58 strain.	123
5	Twenty significant associations were detected in the presence of cadmium. The <i>Cenococcum geophilum</i> scaffold number and coding region base pair start and end are indicated. The majority of the associated gene coding regions include proteins involved in metabolic processes including protein transport, nucleic transcription and translation, and heme binding	126

List of Figures

1	Agricultural security is critical for economic security, which includes food security, sovereign independence, and increased general prosperity of the population. In the United States, agriculture accounted for \$992 billion of the gross domestic product in 2015. Source: USDA Economic Research Service	101
2	There are up to 1.73 million hectares of arable land in the United States, increasing the need for sustainable agricultural practices which maintain and increase soil health [7].	102
3	A cropping system involving planting biofuel crops allows for remediation of agricultural soils without extreme negative impact on the quality of the product produced [15].	103
4	Pyrolysis of harvested biomass captures heavy metal contaminants in the char byproduct which accounts for 10 of the total product [33].	104
5	The 3RAD protocol developed by Bayona-Vasquez et al. (2019) [116] includes digestion of genomic DNA with three restriction enzymes, two of which share a cut site sticky end (A). Internal adapters are designed to ligate to one of the two sticky ends, and one internal adapter is also designed to self-ligate at the cut site for the 3rd restriction enzyme, leading to lesser chances of self-ligation and more efficient ligation of internal adapters (B).	105

6	<p>Correlogram of PNW clade, rivershed of origin, site latitude, soil percent moisture content, C/N weight percent, temperature, elemental metal (lead, zinc, copper, cadmium, strontium) counts per minute, total sclerotia isolated, and total <i>C. geophilum</i> isolation success of 105 PNW soil samples. Positive correlations are highlighted in blue and negative correlations are highlighted in red, with color intensity proportional to the correlation coefficient. Only those correlation coefficients with $p < 0.05$ are shown in color. No measured soil conditions or qualities were determined to correlate with the total sclerotia obtained or <i>C. geophilum</i> successfully isolated.</p>	107
7	<p>The Pacific Northwest (PNW) isolate <i>GAPDH</i> RAxML phylogeny mapped by latitude to the PNW region. Clades with $>80\%$ bootstrap support values are indicated above the associated clade. A total of 15 clades encompassing 155 isolates were identified in the PNW isolate collection, with 74 isolates remaining unresolved. Isolates from smaller clades tended to group by region of origin with a few notable exceptions present over large north-south distances.</p>	108
8	<p>Principle components analysis (PCoA) of the PNW <i>GAPDH</i> RAxML phylogeny revealed three distinct isolate clusters, with clade 11 segregating as a separate cluster from the remaining PNW collection (A). A PCoA of the PNW <i>GAPDH</i> alignment also revealed three distinct clusters, with clades 10 and 11 segregating from the remaining PNW collection (B). A scatterplot of clade versus latitude did not reveal distinct patterns within the larger groups of the PNW collection (C), and a multiple components analysis showed some differentiation with clades 3, 12, 14, and 15 from the remaining PNW isolate collection, but revealed weak associations overall between latitude, isolate clade, and phylogenetic differentiation (D).</p>	109

9	Phylogenetic incongruence between the ITS (left) and GAPDH (right) RAxML phylogenies (A). Heatmap showing pairwise distances for ITS and GAPDH genes from <i>Cenococcum geophilum</i> . Genetic distances (HKY85) were calculated among all sequence pairs for ITS (above diagonal) and GAPDH (below diagonal). Darker color indicates higher sequence divergence. The first is sequence order based on UPGMA clustering of ITS distances, the second is sequence order based on UPGMA clustering of GAPDH distances (B). A scatterplot of pairwise HKY85 distances for ITS vs GAPDH datasets shows low correlation between the ITS and GAPDH gene regions (C). The parsimony-based inter-partition length difference (PTP-ILD) test indicated that the ITS and GAPDH data sets are incongruent due to no significant difference between the observed data set parsimony trees and the distribution of the randomized tree length calculations ($p = 0.001$) (D).	110
10	Global isolate collection GAPDH RAxML phylogenetic tree. Clades with strong bootstrap support ($>80\%$) are labeled on the outer ring. Strongly supported clades implicated in this study are highlighted in orange and designated with V. Clades V16-V19 represent newly designated clades within the global <i>C. geophilum</i> isolate collection. The PNW isolates are highlighted in orange. Isolates are highlighted per the most recent published study of origin as follows: D (Douhan et al., 2007b; Douhan and Rizzo, 2005) highlighted in gold; dFP (de Freitas Pereira et al., 2018) highlighted in pink; M (Matsuda et al., 2015) highlighted in purple; and O (Obase et al., 2016) highlighted in blue.	111
11	The RADseq <i>de novo</i> RAxML phylogeny directly compared to the reference-based RADseq phylogeny. Neither assembly retained the <i>Pseudocenococcum floridanum</i> isolate BA4b018 due to insufficient read depth. The <i>de novo</i> assembly strongly resolved ($>80\%$ support) 76% of the retained 171 isolates into 15 clades with strongly supported substructure. The reference-based assembly resolved 42% of the retained 112 isolates into 8 clades.	118

12	The single gene (GPD) RAxML phylogeny directly compared to the RADseq <i>de novo</i> assembly RAxML phylogeny. Of 171 retained isolates, a total of 34% were unresolved using GPD versus 24% unresolved using RADseq data. A greater lever of sub-structure was also observed within clades DN1 and DN9 of the <i>de novo</i> assembly phylogeny when compared to the largest GPD3 and GPD6 clades. Clade GPD3 is predominantly grouped into clade DN1, and clades GPD6-8 are predominantly grouped into clade DN9.	119
13	The single gene (GPD) RAxML phylogeny compared to the reference-based RADseq RAxML phylogeny. The GPD phylogeny strongly resolved (>80%) 15 clades and 66% of 171 isolates. The reference-based assembly only resolved 42% of the retained 112 isolates into 8 clades.	120
14	A discriminant analysis of principle components of the <i>de novo</i> assembly revealed four distinct phylogenetic clusters. When <i>Cenococcum geophilum</i> isolates were mapped to the original Pacific Northwest site of isolation, there were no patterns of segregation found within the clusters based on latitude. Many sites also contained more than one cluster, a pattern which occurred regardless of the size of the site-associated isolate collection. Sites of isolation are coded from blue to red running North to South, and the site-specific color is indicated in the circular outline of each site cluster pie chart.	121

- 15 The Index of Association (I_a) was calculated for each *de novo* assembly cluster, clade, and Pacific Northwest site of origin with at least three isolates. There was no evidence of recombination within the full *de novo* assembly ($\bar{r}d = 0.0376$, $p < 0.001$) (A). *De novo* clade 5 fell within the expected distribution with no allelic linkage ($\bar{r}d = 0.274$, $p < 0.01$), indicating evidence of recombination (B). The Snohomish rivershed isolate collection fell within the expected distribution, however the results were not significant ($\bar{r}d = 0.149$, $p = 0.07$) (C). In contrast, Snohomish river site six (SN6) fell within the expected distribution ($\bar{r}d = 0.0885$, $p < 0.05$) (D). The White River collection fell outside of the expected distribution, ($\bar{r}d = 0.161$, $p < 0.001$) (E), while White River site WH4 fell within the expected distribution ($\bar{r}d = 0.0624$, $p < 0.01$), implicating a clonal population with site-specific recombination (F). 122
- 16 Average radial growth in cm at 70 days of all *Cenococcum geophilum* isolates used in this study per heavy metal condition. The average per condition is indicated with a black dot and value. All heavy metal conditions were at 25 ppm save for cadmium (Cd), which was at 0.5 ppm. All conditions significantly impacted average growth when compared to the control condition of each isolate. Generally isolates grew less in the presence of a heavy metal with cadmium (Cd) having the lowest average at 0.4 cm followed by copper (Cu) at 0.5 cm and lead (Pb) and zinc (Zn) at 0.7 cm. While sodium (Na) did have a significant impact on overall growth, the average was closest to the control at 1.1 cm. Isolates tended to grow as well or better in the presence of strontium (Sr) at 1.4 cm average radial growth. 128

17 Average radial growth per *Cenococcum geophilum* isolate and heavy metal treatment at 70 days of growth. All heavy metal conditions were at 25 ppm save for cadmium (Cd), which was at 0.5 ppm. Generally isolates grew as much as the control treatment in the sodium (Na) treatment and as much or more in the strontium (Sr) treatment. Most isolates had decreased average radial growth in the cadmium (Cd), copper (Cu), lead (Pb), and zinc (Zn) treatments except for isolates which originated from either White River 5 isolation site (WH5) or Skagit River 5 isolation site (SK5). 129

18 Average radial growth in cm at 70 days of *Cenococcum geophilum* isolates originating from either White River 5 isolation site (WH5) or Skagit River 5 isolation site (SK5). The average per condition is indicated with a black dot and value. All heavy metal conditions were at 25 ppm save for cadmium (Cd), which was at 0.5 ppm. All conditions significantly impacted average growth when compared to the control condition of each isolate. Isolates grew less than the control in both copper (Cu) and zinc (Zn) conditions at 0.4 cm and 0.7 cm respectively, but nearly equal average growth in the 0.5 ppm Cd condition in contrast to the average of the entire isolate collection at 0.4 cm. Higher average radial growth was also seen in the sodium (Na), lead (Pb), and strontium (Sr) conditions at 2.2 cm, 1.5 cm, and 2.2 cm respectively. . . 130

List of Attachments

1. Supplemental Table 2.1 - PNW Isolation Sites. Pacific Northwest isolate sites and total *Cenococcum geophilum* isolated. The overall isolation success rate of *C. geophilum* from 105 soil samples was 46%. Total *C. geophilum* isolates >0 are bolded in the Total *Cenococcum* column.
2. Supplemental Table 2.2 - PNW Isolates Table. Pacific Northwest *Cenococcum geophilum* isolates, rivershed of origin, associated clade and bootstrap support value. Bootstrap values of >85% are bolded. Unresolved isolates or isolates not grouped into a nested clade are designated with a dash (-). The ITS and *GAPDH* GenBank accession numbers are included.
3. Supplemental Table 2.3 - Global Isolates Table. Global population *GAPDH* maximum likelihood analysis of 790 *Cenococcum geophilum* strains. Groupings implicated or identified in previous studies are indicate as follows: D (Douhan et al., 2007a; Douhan and Rizzo, 2005); dFP (de Freitas Pereira et al., 2018); M (Matsuda et al., 2015); and O (Obase et al., 2016). Strongly supported clades designated in this study are designated with V. Clades V16-V19 represent newly designated clades within the global *C. geophilum* population. Bootstrap support values of >85% are bolded. Isolates not grouped into a nested clade are designated with a dash (-).
4. Supplemental Table 3.1 - Restriction-associated DNA sequencing files by isolate. Raw Illumina forward sequence reads were deposited into the NCBI Short Read Archive (SRA, <http://www.ncbi.nlm.nih.gov/sra/>) under Bioproject accession number PRJNA668013. Included are each sample name, object ID, and URL.

5. Supplemental Table 4.1 - Reference-based and *de novo* contig sequences associated with phenotypes were analyzed against the *Cenococcum geophilum* 1.58 genome v.2 using the MycoCosm BLAST feature within the Joint Genome Institute portal. The results are separated by *de novo* and reference-based assembly in the following order: contig or unique sequence ID, scaffold ID and begin/end base pair, presence per phenotype denoted by “x”, protein or transcript ID, KOG protein description, additional annotations and notes, protein name, and Joint Genome Institute (JGI) BLAST result values per MycoCosm result.
6. Supplemental Table 4.2: Associated proteins per SNP called and their functions. The results are separated by *de novo* and reference-based assembly in the following order: contig or unique sequence ID, scaffold ID and begin/end base pair, presence per phenotype denoted by “x”, protein or transcript ID, KOG protein description, additional annotations and notes, protein name, and Joint Genome Institute (JGI) BLAST result values per MycoCosm result.
7. Supplemental Table 4.3: Significant GWAS associations which were detected in both the *de novo* and reference-based assemblies at the same site(s) with the same alternate nucleotide. Results are separated by reference-based or *de novo* assembly in the following order: protein ID, scaffold number, sequence begin, sequence end, associated treatment, scaffold base pair, SNP (original/alternate), KOG protein description, and protein function.
8. Supplemental Table 4.4: All *de novo* assembly significant GWAS associations per treatment and control condition, separated by treatment or significance. Data included are: *de novo* contig, SNP, base pair, original nucleotide, alternate nucleotide, and number of isolates in which the SNP was detected.
9. Supplemental Table 4.5: All reference-based assembly significant GWAS associations per treatment and control condition, separated by treatment or significance. Data included are: *de novo* contig, SNP, base pair, original nucleotide, alternate nucleotide, and number of isolates in which the SNP was detected.

Chapter 1

Introduction

1.1 Agriculture in a global context

Population growth in the 20th century has led to increased efficiencies in agricultural development [1] (Fig 1 in Appendix), which in turn increases the demand for arable lands for sustainable farming [2, 3]. Agricultural development also serves as a primary source of economic security for many countries, encouraging country independence from outside resources in times of crisis [4]. Significant tracts of land are closed off to agricultural development due to contamination within soils [5, 6]. In the United States alone, there are approximately 1.73 million hectares of closed landfills, abandoned mining lands and areas which are closed to future development due to pre-existing contamination [7]. A significant portion of this land contains severe heavy metal contamination within the soil [6, 8–10], which prevents use of this soil for food crop systems. As a result, there is a finite amount of land available for either biofuel or food crop systems, and land availability is further limited due to the presence of toxic pollutants [11, 12], leading to phytotoxicity and other concerns of bioaccumulation in the food chain [13]. Investigations of both potential remediation techniques as well as agricultural systems which can be used in contaminated soils is of great interest in agricultural development [5, 12, 14, 15] (Fig 2 in Appendix).

1.2 *Populus* spp. as a biofuel crop

The planting of non-consumable biofuel crops would eliminate the dangers inherent in using these lands for food resources and open these lands for economic development [15–17]. One such biofuel crop genus, *Populus* spp. (poplar), consists of 30 fast-growing deciduous softwood species with natural hybridization and distinct sexes [18–21]. poplar is of particular interest in reforestation and phytoremediation for the recovery of contaminated sites due to their fast growth [22] and overall hardiness in the presence of heavy metal contaminants [23–29] (Fig 3 in Appendix). Hybrid poplar are currently grown in agroforestry settings for both pulp and softwood lumber applications, and currently being studied and optimized for biofeedstock production [25, 26, 30]. Indeed, the increased acidity of the poplar biomass may lead to more efficient processing for later biofuel applications [31, 32]. Additionally, when biomass is processed using pyrolysis, heavy metal contaminants are captured within the char byproduct [16, 26, 33–36], mitigating potential risks in further downstream processing (Fig 4 in Appendix). This char byproduct represents only 10% of the total yield from the pyrolysis of biomass, allowing for the collection and processing of bio-oil with negligible heavy metal contaminant carry-over [16, 33, 35].

1.3 Microbial applications to bioremediation and biofuel crop systems

While poplar is generally tolerant of heavy metal contaminants, this tolerance can be further increased through microbial community modification and soil additives [37–39]. Manipulation of host plant symbionts has been shown to increase plant tolerance to soil conditions which would otherwise be toxic to the plant [38, 40–44], and particular fungal-plant symbioses may increase the ability of poplar to thrive in soils containing high heavy metal concentrations [39, 40, 42, 44–49]. Poplar serves as host to a diverse community of microbes including fungi [22, 50–52] which are capable of metabolizing and/or immobilizing soil compounds which the plant cannot [38, 50, 53–57]. Some mycorrhizal fungi directly interact with heavy metal contaminants within the soil [37, 39, 49, 55, 58], and fungal

isolates from poplar roots have previously displayed increased resistance to heavy metal contaminants [22, 55].

While promising, these plant-fungal relationships are often difficult to disentangle. A single plant species may be associated with hundreds of fungal species with varying influences on host plant survival depending on the environmental and physical conditions of the soil [52, 57, 59–62]. The plant may also influence the rhizosphere directly through the secretion of exudates which promote or inhibit the presence of certain species [47, 63–68].

1.4 *Cenococcum geophilum* as a model organism

The complexity of these plant-fungal relationships may be disentangled through targeted understanding of the interactions of exemplar species [57, 69–72], which can function as representatives for exploring plant-fungal interactions and characteristics in various conditions. Such a model species should be globally distributed in multiple soil conditions, associated with multiple plant species, and be easily culturable in a laboratory setting. The fungal species *Cenococcum geophilum* presents such a species as an ubiquitously distributed ectomycorrhizal fungus which is positively associated with plant health, growth, and increased contaminant resistance [46, 49, 73–76]. This fungus is known to associate with over 200 species of both angio- and gymnosperms across 40 genera [75, 77], and is generally tolerant across a salinity gradient [78], water stress conditions [79], and extreme heavy metal contamination conditions [46, 49, 80]. This wide-ranging resistance to multiple soil environment stressors is often attributed to a high melanin content [78, 79, 81, 82], which has been directly implicated to play a role in increased resistance to heavy metal contaminants [39, 45, 81].

There are significant complications regarding phylogenetic studies of *C. geophilum* [75]. The genome of *C. geophilum* is among the largest in the fungal kingdom due to an abundance of repeated sequences in the form of transposable elements [83–85]. The published genome has a mapped size of 178 mpb, and is estimated to range up to 203 mbp [83–85], making whole-genome sequencing of newly-isolated large collections of *C. geophilum* cost-prohibitive. This ubiquitous species is considered asexual despite high levels of genetic and physiological

diversity in the absence of documented sexual or asexual spore production [75, 86, 87]. Due to these high levels of intraspecies diversity found among *C. geophilum* isolates, even within isolates from a single tree [88], *C. geophilum* is proposed to be a species complex [75, 84, 86, 89] with significant variations in cultured isolate characteristics and physiology [75, 86]. This type of cryptic species population implicates the incorporation of morphologically identical members which are separate species on the genetic level [75, 86, 90, 91]. As a result, while *C. geophilum* appears to have a myriad of potential benefits for agricultural applications, this fungus has been difficult to study phylogenetically, as this research is necessarily resource-intensive in order to delineate the phylogenetic relationships observed within the species complex.

1.5 Phylogenetic studies of the cryptic fungus *Cenococcum geophilum*

The extreme genomic and phylogenetic complexity of *C. geophilum* encourages comparative studies incorporating multiple genetic analyses and sequencing technologies. In the past, studies have relied on single genes or concatenated alignments of several genes for phylogenetic inferences [75, 76, 86, 89–93]. While the internal transcribed spacer (ITS) region is a universal standard used in both the identification and often phylogenetic analyses of many fungi including *C. geophilum* [94–98], issues arise when intragenomic variation occurs within a collection and even within the same isolate [99, 100]. Cryptic species such as *C. geophilum* are difficult to delineate based on ITS region phylogenetic analyses [101, 102], a limitation further confounded by the presence of multiple ITS region copies within nuclear genomes [99, 103]. As a result, the glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) gene is most commonly used in comparative studies and is considered the standard gene for isolation and phylogenetic analyses of *C. geophilum* isolates [89, 91, 93]. This protein catalyzes the first step of glycolysis, synthesizing D-glyceraldehyde 3-phosphate to pyruvate. The *GAPDH* gene is a prime target for phylogenetic analyses as this protein is well-studied

and understood across multiple species [104] and there is generally only a single copy of this gene found within a filamentous fungal genome [105–108].

Despite the potential limitations of ITS as a single phylogenetic site, a combination of *GAPDH* and ITS serves as a powerful tool to determine potential recombination patterns within an isolate collection. Direct comparisons between *GAPDH* and ITS may reveal incongruence between the sites, leading to further recombination analyses to determine gene linkage patterns. Gene distance matrices may be graphed against each other to view correlations between pairwise distances of two genes. Additionally, the inter-partition length difference (ILD) test, partitioned by gene region, assesses phylogenetic congruence between gene regions by indicating whether the summed lengths of the parsimony trees created from the dataset are significantly shorter than the distribution of combined tree lengths calculated for a series of randomizations [109]. Should this summed length fall within the randomized tree lengths distributions, the genes may be considered congruent and recombination is less likely to be occurring within the population.

In summation, phylogenetic analyses completed using either *GAPDH* or ITS are powerful tools which provide an informative basis for further studies. However, these studies do have a limited scope when used with a cryptic species due to the incomplete genetic information provided by a single gene, a limitation highlighted by the assignment of a new taxonomic species *Pseudocenococcum floridanum* gen. et sp. nov. based on a multigene concatenated alignment of previously characterized *C. geophilum* isolates [91]. This study highlights the need to increase genetic information to properly delineate cryptic species within any *C. geophilum* collection. Whole-genome sequencing is an effective technique for species with smaller genomes but is cost-prohibitive when investigating a *C. geophilum* collection of >10 isolates due to genome size.

1.6 RADseq applied to *C. geophilum*

Restriction-associated DNA sequencing (RADseq) is a novel method of genomic DNA processing for sequencing analysis [110–115] featuring multiple optimized protocols which have been successfully applied across species [112, 116–119] but rarely applied to fungi

[112, 120–122]. This low-cost alternative to whole genome sequencing allows for robust *de novo* single nucleotide polymorphism (SNP) detection [123, 124] and increases intra-species resolution in phylogenetic analyses [112, 120, 125, 126]. This increased resolution is particularly valuable in cryptic species, leading to higher phylogenetic resolution in otherwise difficult to study populations [125, 127–129]. Because this improved resolution will promote the determination of species-level differentiation within the species complex of *C. geophilum*, RADseq has been proposed as a viable phylogenetic analysis technique for application [75]. Proof of concept of RADseq applied to *C. geophilum* will open more avenues of phylogenetic analyses, strengthening the case for use of this phylogenetically fascinating species as a model organism in future studies.

The Adapterama 3RAD protocol developed by Bayona-Vásquez et al. (2019) [116] has not yet been applied to a fungal species, making optimization of this protocol a priority. The full technique involves the digestion of genomic DNA (gDNA) by three restriction enzymes [116] (Fig 5A-B in Appendix), two of which share cut site sticky ends sequences (Fig 5B in Appendix). Internal adapters are then ligated to the ends of the gDNA fragments with one set ligating to the shared sticky end, and the other ligating to the alternate sticky end [116] (Fig 5B in Appendix). Fragments are then PCR amplified using primers designed with unique external adapters for isolate-specific identification in downstream analyses [116]. Fragments are then size-selected for approx. 550 base pair length, pooled into a library, and sent for sequencing [116] (Fig 5A in Appendix). Downstream demultiplexing and analyses on the resulting sequencing files may be completed using several pipelines, including ipyrad [130] and Stacks 2 [131–133]. While *C. geophilum* has a published genome for Swedish isolate 1.58 [84], RADseq does not require a pre-existing genome for analyses [123, 130, 133, 134] which allows a comparison of both reference-based and *de novo* RADseq datasets in order to optimize protocols for an isolate collection and determine which dataset provides the more robust phylogenetic inferences.

1.7 Expanding phylogenetic analyses in *C. geophilum* studies

As a cryptic species, isolates classified as *Cenococcum geophilum* present a variety of phenotypes ranging from isolate color, growth rate, sclerotia formation, and tolerance across a variety of stressors. Such phenotypes may be mapped back to either a reference genome or a *de novo* assembly produced through RADseq to complete association analyses and detect significant associations across the genome [75, 122, 135]. These associations may elucidate the roles of *C. geophilum* in soil environments, implicate protein function(s) in particular measured phenotypes, and inform more complex phylogenetic studies of larger isolate collections within the global *C. geophilum* collection.

Chapter 2

Phylogenetic diversity of 200+ isolates of the ectomycorrhizal fungus *Cenococcum geophilum* associated with *Populus trichocarpa* soils in the Pacific Northwest, USA and comparison to globally distributed representatives.

2.1 Preface

A version of this chapter was submitted for publication by Jessica M. Vélez ^{1,2}, Reese M. Morris ¹, Jessy Labbé ^{1,4}, Rytas Vilgalys ³, and Christopher W. Schadt ^{1,4,*}.

2.2 Author Affiliations

¹ Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

² The Bredesen Center for Interdisciplinary Research and Graduate Education, University

of Tennessee, Knoxville, TN 37996-3394, USA

³ Biology Department, Duke University, Raleigh, NC

⁴ Dept of Microbiology, University of Tennessee, Knoxville TN 37996-0845, USA

* Corresponding Author: Email: schadtcw@ornl.gov; Phone: 865.576.3982

2.3 Author Contributions

Initial soil collections were completed by Rytas Vilgalys and Christopher W. Schadt in 2015, followed by an additional collection in 2016 by Chris and Jessica M. Vélez during which Christopher W. Schadt and Jessica M. Vélez recorded soil temperature. Isolations of *Cenococcum geophilum* sclerotia from soil was completed by Reese M. Morris and Jessica M. Vélez. Reese M. Morris also prepared soil samples for carbon/nitrogen analyses and extracted genomic DNA for PCR amplification. Rytas Vilgalys completed and provided documentation for recombination analyses within this study, including pairwise comparisons of the gene regions. Jessy Labbé provided fungal growth, genomic DNA extraction, and genomic DNA PCR amplification guidance. Jessica M. Vélez completed x-ray fluorescence analyses on soil samples, recorded water weight content, extracted genomic DNA for PCR amplification, PCR amplified, sent PCR products to Eurofins for sequencing, and performed phylogenetic analyses. The manuscript was primarily written by Jessica M. Vélez. revised by Rytas Vilgalys, Jesse Labbé, and primarily revised by Christopher W. Schadt.

2.4 Abstract

The ectomycorrhizal fungal symbiont *Cenococcum geophilum* is of high interest as it is globally distributed, associates with many plant species, and has resistance to multiple environmental stressors. *C. geophilum* is only known from asexual states but is often considered a cryptic species complex, since extreme phylogenetic divergence is often observed within nearly morphologically identical strains. Alternatively, *C. geophilum* may represent a highly diverse single species, which would suggest cryptic but frequent recombination. Here we describe a new isolate collection of 229 *C. geophilum* isolates from soils under

Populus trichocarpa at 123 collection sites spanning approx. 283 mile north-south transect in Western Washington and Oregon, USA (PNW). To further understanding of the phylogenetic relationships within *C. geophilum*, we performed maximum likelihood phylogenetic analyses to assess divergence within the PNW isolate collection, as well as a global phylogenetic analysis of 789 isolates with publicly available data from the United States, Europe, Japan, and other countries. Phylogenetic analyses of the PNW isolates revealed three distinct phylogenetic groups, with 15 clades that strongly resolved at >80% bootstrap support based on a *GAPDH* phylogeny and one clade segregating strongly in two principle component analyses. The abundance and representation of PNW isolate clades varied greatly across the North-South range, including a monophyletic group of isolates that spanned nearly the entire gradient at approx. 250 miles. A direct comparison between the *GAPDH* and ITS rRNA gene region phylogenies, combined with additional analyses revealed stark incongruence between the ITS and *GAPDH* gene regions, consistent with intra-species recombination between PNW isolates. In the global isolate collection phylogeny, 34 clades were strongly resolved at >80% bootstrap support and >0.90 posterior probability, with some clades having intra- and intercontinental distributions. Together these data are highly suggestive of divergence within multiple cryptic species, however additional analyses such as higher resolution genotype-by-sequencing approaches are needed to distinguish potential species boundaries and the mode and tempo of recombination patterns.

2.5 Introduction

Plant-fungal relationships are often difficult to disentangle. A single plant species may be associated with hundreds of fungal species and each of these associations can have varying influences on host plant survival and growth that co-vary with the environmental and physical conditions of soil ([136–139]). The complexity of these plant-fungal relationships may be at least partially illuminated through targeted understanding of the interactions of exemplar fungal species. Such model species can serve as representatives for exploring plant-fungal interactions and characteristics across various conditions. The genus *Populus* is an excellent plant model for such studies because *Populus trichocarpa* has a fully sequenced

genome [21, 140] and hosts a diverse community of microbes including bacteria, archaea and fungi [52, 136, 141] which are capable of accessing, metabolizing, producing, and/or immobilizing compounds which the plant cannot [53]. With over 30 species of deciduous, fast-growing softwoods with distinct sexes and natural hybridization [19], *Populus* also has great importance in agroforestry industries for pulp and paper products, as well as potential as a bioenergy feedstock [26, 30].

Similar to *Populus* spp., a correspondingly robust model ectomycorrhizal (ECM) fungal group for paired studies with *Populus* should be widespread, interact with many plant hosts, and be easily culturable in a laboratory setting. The ECM fungal species *Cenococcum geophilum* is a ubiquitously distributed fungus which is positively associated with plant health, growth, and increased soil contaminant resistance [49, 74], and as such has the potential to serve well as a model organism for genetic, physiological, and ecological studies. *Cenococcum geophilum* is known to associate with both angiosperm and gymnosperm species across 40 plant genera, representing over 200 host species [86], and is generally tolerant across salinity gradients [78], under water stress conditions [79], and in extreme soil contamination conditions [46, 78, 79]. This wide-ranging resistance to stress conditions in the soil environment is often associated with a high melanin content [78, 79, 81, 82] which may also contribute to hyphal and ECM longevity and resistance to decomposition in soils [79]. *Cenococcum geophilum* is readily culturable in a laboratory setting and capable of growing on multiple standard solid media, including both defined media such as Modified Melin-Norkrans (MMN) and complex media such as potato dextrose agar (PDA). Additionally, laboratory methods for targeted and relatively rapid isolation of this fungus from soils via its abundant and phenotypically characteristic sclerotia [142, 143] allow for efficient isolation of new strains directly from soil samples. All of these characteristics make *C. geophilum* an ideal model species and candidate for population-level genomic studies. However most existing regional-level isolate collections have focused primarily on coniferous species [89, 91, 92] rather than angiosperms such as *Populus*.

The first sequenced genome of *C. geophilum* strain 1.58 (isolated from Switzerland) was published in 2016 [84] with support from the Joint Genome Institute (JGI). This genome is among the largest in the fungal kingdom, with a mapped size of 178 Mbp and a total

estimated size of up to 203 Mbp [84, 85]. *Cenococcum geophilum* has no documented means of sexual spore production [86, 87], and is considered asexual as a species despite high levels of genetic and physiological diversity. Due to high levels of diversity that are reported among *C. geophilum* isolates, even within isolates from beneath a single tree [88], *C. geophilum* has been suggested to represent a complex of indistinguishable cryptic species [75, 84, 86, 89], with many studies finding significant variation in cultured isolate characteristics and physiology [86]. However, patterns supporting recombination have also been observed by previous studies, suggesting a cryptic sexual state or other mechanisms for intrapopulation recombination [83, 84, 86, 87, 144]. Together these studies have led many to suggest that *Cenococcum* represents an unknown number of cryptically separate species on the genetic level [86, 90]. Further supporting this suggestion, a 2016 multigene phylogeny of a previously characterized *C. geophilum* isolate collection [142] revealed a divergent clade described as *Pseudocenococcum floridanum* [91]. This discovery highlights the need to further explore the phylogenetic diversity among diverse regional isolates of *C. geophilum* in order to better characterize this species and determine several factors, including 1) if *C. geophilum* is a single highly outcrossed species, or a variety of cryptic species; 2) if a variety of species, how diverse the populations are and what are the patterns of speciation; and 3) if there exists intra- or interspecies patterns of recombination.

Cenococcum geophilum thus appears to have a myriad of potential benefits as a model ECM and rhizosphere-associated species. However, this fungus has been difficult to study phylogenetically, and further work is needed to delineate its phylogenetic and functional relationships within what may be a highly heterogeneous species complex. Our laboratory set a long-term goal to build a genetically diverse collection of *C. geophilum* isolates from across the *Populus trichocarpa* range, mirroring the host GWAS population studies which have proven valuable for understanding the biology of these host trees [29, 145–151]. Towards this goal we have isolated 229 new *C. geophilum* strains from beneath *P. trichocarpa* stands over a range of approximately 283 miles of the host range in the United States Pacific Northwest (PNW) states of Oregon and Washington and confirmed their identity using sequencing of the internal transcribed spacer (ITS) region and the glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) gene. Maximum likelihood phylogenies of the *GAPDH* gene and ITS region

of our PNW collection were compared directly in order to identify potential patterns of intra-species recombination. Furthermore, the *GAPDH* genes of our PNW collection were additionally compared to over 500 *C. geophilum* isolates with comparable available data from published isolates and studies primarily in the United States, Europe, Japan, but also other sites where data were publicly available [86, 89–92, 142], as well as 16 additional European isolates recently sequenced by JGI and provided by Drs. Francis Martin and Martina Peter [92].

2.6 Materials and Methods

2.6.1 Site selection and soil sampling

Primary sampling was carried out over a six-day period in late July of 2016. At each site, three approximately one-gallon soil samples were collected directly under *P. trichocarpa* with at least 50 yards distance in between each collection. Typically, at least three sampling sites were selected along watersheds of the Willamette River in Central Oregon ranging to the Nooksack river near the Canadian border, along a north-south gradient (i.e. the Interstate 5 corridor). Several watersheds and sites corresponded to those sampled in Evans et. al. [146] where possible to allow for testing of associations with *P. trichocarpa* GWAS populations in future studies. Soil samples were collected to approximately a 20 cm depth. A trowel was used to fill one-gallon Ziploc freezer bags which were kept on ice and refrigerated at 4°C until analysis. Site soil temperatures were recorded using an electronic thermometer (OMEGA model RDXL4SD). Upon return to the lab, two 15 mL tubes of the 105 samples in 2016 were subsampled for soil moisture, carbon (C), nitrogen (N), and soil elemental characterization. Additionally, several isolates derived from smaller exploratory soil samples from within the southern end of this geographic range (sampled by Drs. R. Vilgalys and C. Schadt) for methods development during the prior year were also included in this study, for a total of 123 soil samples used for isolation attempts.

2.6.2 Sclerotia separation

Soil samples were prepared using a procedure described by Obase et al. [142] with modifications. Samples were manually sieved and rinsed with distilled water to retain particles between 2mm and 500 μM , and the resulting slurry allowed to soak in distilled water for 10-30 seconds. Floating debris could be decanted off the top. Portions of the slurry were placed into gridded square Petri dishes (VWR 60872-310) which were partially filled with water and a dissection scope was used to remove sclerotia using tweezers. Sclerotia were submerged in undiluted Clorox bleach for 40 minutes using a VWR 40 μm nylon cell strainer (Sigma-Aldrich Z742102), then rinsed 3 times with sterile distilled water. Sclerotia which did not turn white after bleach treatment were then plated onto Modified Melin-Norkrans (MMN) media [152] composed of 3 g l⁻¹ malt extract, 1.25 g l⁻¹ glucose, 0.25 g l⁻¹ (NH₄)₂HPO₄, 0.5 g l⁻¹ KH₂PO₄, 0.15 g l⁻¹ MgSO₄·7H₂O, 0.05 g l⁻¹ CaCl₂, 1 mL⁻¹ FeCl₃ of 1% aqueous solution, and 10 g l⁻¹ agar, adjusted to 7.0 pH using 1N NaOH. After autoclaving media were allowed to cool to 60°C, then 1 g l⁻¹ thiamine was added along with the antibiotics Ampicillin and Streptomycin at 100 ppm each. Plates with sclerotia were stored in the dark at 20°C.

2.6.3 DNA extraction

Isolates with dark black growth were considered viable and allowed to grow 20°C until approx. 5 mm diameter (approx. 6 weeks to 3 months). Colonies were transferred onto a cellulose grid filter (GN Metrical 28148-813) on MMN plates using the previous recipe except adding 7 g l⁻¹ dextrose and omitting antibiotics and allowed to grow for 1-3 months for DNA extraction. The Extract-N-Amp kit (Sigma-Aldrich XNAP2-1KT) was used as per manufacturer instructions to extract genomic DNA, with the modification to use 20 μL of the Extraction and Dilution solutions [153]. DNA samples were stored at -20°C until use in PCR and sequencing efforts below.

2.6.4 PCR amplification

Ribosomal DNA (rRNA) was amplified using fungal-specific ITS primers ITS1 and ITS4 [94], and the *GAPDH* gene was amplified using the *gpd1* and *gpd2* primers [154] using the Promega GoTaq © Master Mix kit to amplify DNA. The thermocycling conditions consisted of an initial hold of 94°C for 5 min, followed by 35 cycles of 94°C (30 s), 55°C (30 s), and 72°C (2 min), with a final elongation of 72°C for 10 min. Amplified PCR products were analyzed on a 1% agarose gel using TAE buffer to confirm band size prior to cleanup. PCR products were then cleaned using the Affymetrix USB ExoSAP-IT © kit and sequenced on an ABI3730 Genetic Analyzer at the University of Tennessee at Knoxville (UTK), or at Eurofins Genomics (Louisville, Kentucky, USA). Sequences generated were analyzed against the NCBI database using the BLAST feature in Geneious version 10.2.3 (<https://www.geneious.com>, [155]) to verify fungal identity as *C. geophilum*. Confirmed *C. geophilum* isolates (marked “CG” with number) were stored on MMN plates at 20°C and re-plated quarterly to ensure continued viability.

2.6.5 Determination of native soil properties

For the 105 soil samples collected in July of 2016, soil temperature, concentrations of carbon, nitrogen, elemental metals, and soil water content were analyzed. A C/N analysis was conducted on approximately 18g of each soil sample. The samples were oven-dried at 70°C and ground to a fine powder. Approximately 0.2 g of ground sample were analyzed for carbon and nitrogen on a LECO TruSpec elemental analyzer (LECO Corporation, St. Joseph, MI). Duplicate samples and a standard of known carbon and nitrogen concentration (Soil lot 1010, LECO Corporation, carbon = 2.77 % ± 0.06 % SD, nitrogen = .233 % ± 0.013 % SD) were used to ensure the accuracy and precision of the data.

Soil elemental metal concentrations were determined using the Bruker Tracer III-SD XRF device. Approximately 1g of dried, homogenized soil was placed into Chemplex 1500 series sample cups with Chemplex 1600 series vented caps and 6 µM Chemplex Mylar® Thin-Film. Cups were placed against the XRF examination window and scanned for 60 seconds at 40 kV with a vacuum and no filter. Elemental spectra were collected using

the Bruker S1PXRf S1 MODE v. 3.8.30 software and analyzed using the Bruker Spectra ARTAX v. 7.4.0.0 software. Correlation coefficients relating the total sclerotia isolated, total *Cenococcum* isolates per site, soil temperature, percent moisture content, C and N weight percent, and total counts per minute (cpm) of a range of elements were then calculated using the R v.3.4.1 statistical analysis software [156] and corrplot package v. 0.84 (<https://cran.r-project.org/web/packages/corrplot/index.html>) in order to determine whether relationships existed between soil quality and content and *C. geophilum* abundance and isolation success.

2.6.6 Generation of phylogenies

The ITS1-5.8s-ITS2 rRNA gene region and *GAPDH* gene sequences of 228 PNW *C. geophilum* isolates were successfully amplified and sequenced, trimmed and aligned to the published genome of the *C. geophilum* strain 1.58 using Geneious. Data sets were deposited into the NCBI GenBank database and made accessible (Supplemental Table 2.1). The ITS and *GAPDH* phylogenies were concatenated for downstream analyses using Geneious, and isolates lacking either ITS or *GAPDH* sequence data were excluded from the multigene concatenated analysis. For comparison with globally distributed isolates, ITS and *GAPDH* sequence data of 543 *C. geophilum* strains were obtained from GenBank including strains from Japan, Europe, the United States, and 16 additional European isolates recently sequenced by JGI were provided by Drs. Francis Martin and Martina Peter (personal communication, [157]). These isolates were aligned with the Pacific Northwest (PNW) isolate collection using ITS and *GAPDH* separately, as well as a multigene concatenation of ITS and *GAPDH*. All phylogenies were rooted using the outgroups *Glonium stellatum*, *Hysterium pulicare*, and *Pseudocenococcum floridanum* isolate BA4b018 [91], which were downloaded from either GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) or MycoCosm [158, 159] (<https://genome.jgi.doe.gov/programs/fungi/index.jsf>).

The best models for maximum likelihood analyses were determined for each individual gene alignment and concatenated gene region alignments using the Find Best DNA/Protein Models feature in MEGA X [160] to determine the Akaike Information Criterion (AICc) value [161] and Bayesian Information Criterion (BIC) [162], with the best model indicated by the lowest AICc and BIC values (1 in Appendix). A maximum likelihood (ML) phylogenetic

analysis was produced for the PNW isolate collection ITS, *GAPDH* and concatenated phylogenies using the MEGA X software with the determined best model settings using 1000 bootstrap replications. Bayesian probabilities of the PNW and global single gene alignments were also inferred using the MrBayes v. 3.2.7a software [163] with the determined best model settings. Isolates lacking either ITS or *GAPDH* sequence data were excluded from the concatenated ML analysis of the global isolate collection for a total of 499 isolates, and the concatenated alignment was partitioned by gene region for analyses. The alignments and phylogenetic tree files are deposited in Treebase and may be accessed at <http://purl.org/phylo/treebase/phyloids/study/TB2:S25906>.

2.6.7 Comparison of *GAPDH*, ITS, and *GAPDH* + ITS phylogenies

The produced ITS and *GAPDH* ML analyses of the PNW isolate collection were directly compared to both the concatenated phylogeny separately, and to each other, using the TreeGraph2 software [164] and the R v.3.4.1 statistical analysis software. These analyses indicated a large disagreement between the two gene datasets as well as weak phylogenetic signal from the ITS dataset, and thus only *GAPDH* was used for further phylogenetic relationship inferences as this dataset contained the most phylogenetically informative sites.

2.6.8 Phylogeographic variation within PNW isolates

A correlation plot was created using R statistical software packages Hmisc v. 4.2.0 and corrplot v. 0.84 to determine whether resolved *GAPDH* clades correlated with the latitude of the strain isolation site, and a multiple correspondence analysis (MCA) was conducted to determine clade-specific correlations to latitude. Additionally, a principle components analysis (PCoA) was conducted using a distance matrix generated from the PNW *GAPDH* phylogeny with and without species outgroups included in order to determine potential speciation patterns within the PNW isolate collection. The PNW ML analysis was mapped by isolate site latitude to the PNW region using the GenGIS 2.5.3 software [165] in order

to co-visualize the spatial and phylogenetic diversity within the overall PNW isolates and within resolved clades.

2.6.9 Global *GAPDH* phylogenetic analyses

In order to directly compare the new PNW and with prior global isolate collections, a global *GAPDH* phylogeny was constructed using a maximum likelihood (ML) analysis in the MEGA X software [160] with 1000 bootstrap replications and the previously determined best fit settings (Table 1 in Appendix). Bayesian probabilities of the global *GAPDH* alignment was also inferred using the MrBayes v. 3.2.7a software [163] with the determined best model settings (Table 1 in Appendix). Phylogenetic trees were visualized indicating either bootstrap support values or branch lengths using FigTree v. 1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>). Clades were designated as strongly grouped isolates with bootstrap support values of >80%.

In addition to avoiding conflicts between *GAPDH* and ITS identified in the PNW isolates, usage of the *GAPDH* rather than multigene concatenated phylogeny allowed for a more comprehensive global isolate collection analysis, as the concatenated global isolate phylogeny excluded >300 isolates which lacked both sequences while the *GAPDH* global isolate phylogeny included a total of 789 isolates. As a result, only the *GAPDH* dataset was used for global collection analyses.

2.6.10 Recombination patterns within *GAPDH* and ITS data

The ITS and *GAPDH* RAxML phylogenies were visually compared and analyzed using R package phytools v. 0.6.99. A PCoA was completed for the ITS, *GAPDH*, and concatenated phylogenies and alignments using distance matrices generated using the ape v. 5.3 and seqinr v. 3.6-1 packages in the R statistical software in order to compare patterns of recombination within the PNW isolate collection based on either or both gene regions. Additionally, HKY85 distance matrices were generated using Phylogenetic Analysis Using Parsimony (PAUP) v. 4 [109] and graphed against each other in the R statistical software using the ggplot2 package [166]. The inter-partition length difference (ILD) test was performed on the

PNW concatenated alignment, partitioned by gene region, to assess phylogenetic congruence between ITS and *GAPDH* data sets using the PTP-ILD option in PAUP* Version 4 [109] with 1000 permutations and default settings. We additionally used the HKY85 distance matrices to generate cluster distance matrices based on either ITS or *GAPDH* only and then plotted them against each other to determine any pattern incongruence between the two genes using the software Mathematica v. 12.0.0.0.

2.7 Results

2.7.1 A diverse culture collection of PNW *Cenococcum isolates* associated with *Populus*

A total of 229 PNW isolates were obtained from 56 out of 123 soil samples (105 primary soil samples from 2016 + 18 preliminary samples from 2015 - Supplemental Table 2.1), accounting for a 46% overall *C. geophilum* isolation success rate, as some soil samples had few or no sclerotia. The 105 PNW primary soil samples for which associated soils data were also generated, the total *C. geophilum* isolation success rate did not positively correlate to any measured soil condition or quality (Fig 6 in Appendix). Isolation success also did not strongly correlate to the total number of sclerotia recovered ($r = 0.38$, $p > 0.05$). Between the soil values however there were expected relationships. For example, the strongest correlation was between the soil C and N weight percentages ($r = 0.99$, $p = < 0.05$). Strong correlations also existed between the percent moisture content and C content ($r = 0.88$, $p = < 0.05$) or N content ($r = 0.86$, $p = < 0.05$), C content and zinc counts per minute (cpm) ($r = 0.80$, $p = < 0.05$), and N content and zinc cpm ($r = 0.81$, $p = < 0.05$) suggesting that despite the lack of correlation with sclerotia and isolate numbers our measurement approaches were robust (Fig 6 in Appendix).

2.7.2 Phylogeographic variation within PNW isolates

A total of 438 GAPDH positions were represented in the alignment for the GAPDH ML analyses of the PNW and global isolate collections. In the PNW isolates, the phylogeny

backbone strongly resolved the PNW collection at 97.1%. A total of 155 isolates grouped into 15 clades where two or more strains were resolved at >80% bootstrap values and 1.0 posterior probability apart from clade 10 (0.8 posterior probability) (Fig 7 in Appendix, Supplemental Table 2.2). Of the 15 resolved clades, two contain nested clades of two or more isolates which resolved at >80% bootstrap support and 1.0 posterior probability (Supplemental Table 2.2). Nonetheless, 74 of our 229 PNW *Populus* isolates were not strongly resolved by these analyses. Within these 15 *C. geophilum* clades resolved, across the 283 mile transect of our sites selected along rivers in the PNW, smaller clades tended to group latitudinally by site and watershed of origin, although notable exceptions exist, with two groups containing numerous nearly identical isolates despite distances of >100 miles between their sites of origin. In the PNW isolate collection, 74 isolates were outside of any other strongly supported clades (Fig 7 in Appendix), Supplemental Table 2.2), but the majority grouped strongly (1.0 posterior probability and >80% bootstrap).

Phylogenetic distance matrices analyzed via PCoA revealed that the PNW isolates group separately as three distinct groups. When using a distance matrix based on the RAxML phylogeny, the PNW 11 clade segregates as a distinct phylogenetic group (Fig 8A in Appendix), and when using an alignment-generated distance matrix, clades 10 and 11 both segregate as distinct phylogenetic groups (Fig 8B in Appendix).

The *GAPDH* phylogenetic analysis mapped to the PNW soil sampling sites by latitude revealed that isolates did not appear to strongly correlate to original isolation site latitudes (Fig 7 in Appendix, Supplemental Table 2.2). High geographic latitude variation is seen within clade 11, which includes 38 strains isolated >60 miles apart (Fig 7 in Appendix, Supplemental Table 2.2). Clades three, four, five, seven, and nine include strains isolated >100 miles apart (Fig 7 in Appendix, Supplemental Table 2.2). The nested clade 8.1 has the overall largest spatial range, with a single Willamette river isolate WI7_83.9 grouping with several isolates from sites >200 miles to the north (Fig 7 in Appendix, Supplemental Table 2.2) on the Nooksack River near the Canadian border. Clade ten is the largest clade with a total of 45 isolates (Fig 7 in Appendix, Supplemental Table 2.2), representing the greatest spatial diversity between individual sites within a single clade. The direct mapping to the origin of isolation latitude revealed no clear patterns of segregation, particularly in the

largest clades. Further supporting this lack of association with latitude, an MCA found that the combination of clade and latitude were weakly associated with phylogenetic variation within the PNW isolate collection, with the combination of these two factors accounting for only 7.6% of the phylogenetic variation found (Fig 8D in Appendix). Analyses further revealed that clades three, 12, 14, and 15 are not strongly grouped with the majority of the PNW isolate collection (Fig 8D in Appendix).

2.7.3 Recombination analyses in the PNW isolate collection

While both the ITS and *GAPDH* phylogenies of the PNW isolates had a strongly resolved backbone (>80% bootstrap support), the PNW *GAPDH* phylogeny strongly resolved 15 clades within the PNW isolate collection while the PNW ITS phylogeny only strongly resolved only 3 clades and showed numerous apparent conflicts (Fig 9A in Appendix). Individually the HKY85 pairwise genetic distances of each gene indicate strong hierarchical clustering, but when plotted against each other that pattern is lost (Fig 9B in Appendix) and a very low correlation was observed between the ITS and *GAPDH* gene regions using a linear regression analysis ($R^2=0.185$, $p<.001$, Fig 9C in Appendix). To further explore this apparent incongruence between the two genes, a total of 102 informative sites of the concatenated PNW alignment were analyzed using the ILD test protocol in PAUP. The ILD test confirmed that the ITS and *GAPDH* data sets are incongruent (Fig 9D in Appendix), as the summed lengths of the two parsimony trees made from the observed data set were significantly shorter than the distribution of combined tree lengths calculated for 1000 randomizations of the data set ($p = 0.001$).

2.7.4 Phylogeographic variation within the global *C. geophilum* isolate collection

Phylogenetic analyses of the PNW isolate collection together with the larger global isolate collection revealed 34 clades of two or more strains resolved at >80% bootstrap support (Fig 10 in Appendix, Supplemental Table 2.3). The major PNW clades 1, 2, 5, 6, 7, 11, 12, 13, 14, and 15 grouped identically in the global *GAPDH* analysis (clades V1, V2, V5, V6,

V7, V11, V12, V13, V14, and V15 respectively (Fig 10 in Appendix, Supplemental Table 2.3). The PNW clade 8 persisted as clade V8 and additionally included isolates CAA022, CAA006, CLW001 and CLW033 from the Florida/Georgia region [91] (Fig 10 in Appendix, Supplemental Table 2.3).

2.8 Discussion

2.8.1 *Cenococcum* phylogeographic diversity in the Pacific Northwest

This study is the first comprehensive look at the genetic relationships within a regional population of *Cenococcum geophilum* isolates associated with a single host tree *Populus trichocarpa*. Previous *C. geophilum* genetic studies have primarily focused on gymnosperm species such as pines and Douglas fir [89, 91], with two studies incorporating isolates collected under an angiosperm (oak) as well as other local gymnosperm host species [77, 90] and a second incorporating isolates collected under *Fagus sylvatica* ([92] (Table 2 in Appendix). Interestingly, the genetic diversity encountered within this isolate collection exceeds the diversity observed in gymnosperm-associated populations by over twofold (Table 2 in Appendix) [89, 91] implying that *P. trichocarpa* may exert fundamentally different selective pressures on the ECM *C. geophilum* than gymnosperm hosts.

While geographic patterns are clearly evident within the phylogenetic analysis of our PNW isolates, interestingly the soil variables examined seemed to have no correlation with either the number of *C. geophilum* sclerotia recovered or the number of isolates obtained from our samples or their phylogenetic relationships (Fig 6 in Appendix). Other researchers have previously found aluminum concentrations to be related to the formation and abundance of sclerotia, primarily in pine forests [167–169], however none of the soil chemistry data was relatable to the isolate diversity within or between sites. We also observed a much lower abundance of *Cenococcum* sclerotia under *Populus trichocarpa* hosts than reported for other systems. Our methods were only semi-quantitative as we were focused on diversity rather than biomass, however we were typically only able to recover less than 20 sclerotia, and in

approximately half our samples zero sclerotia, per gallon (approx. 3.5 kg) of soil sampled. While much of the literature uses similar semi-quantitative methods, one previous study of Douglas Fir forests, also conducted in western Oregon, averaged 0.91 sclerotia per gram of soil [170] translating to 2785 kg ha⁻¹ of *Cenococcum* sclerotial biomass. This level of abundance associated with a different host in the same region would seem to be approximately two orders of magnitude greater than those recovered in our study under *Populus*.

2.8.2 Incongruent genes and indications of recombination within the PNW isolate collection

A side by side visual comparison of the ITS and *GAPDH* phylogenies showed in some cases up to three ITS types associated with a *GAPDH* gene group, and numerous groupings in the respective genes that were inconsistent with each other, a pattern which is strongly suggestive of ongoing or historic intra-species recombination events among the PNW isolates (Fig 9A in Appendix). While the HKY85 pairwise genetic distances of each gene individually reveal clear hierarchical clustering, when plotted against each other that pattern is lost (Fig 9B in Appendix), further supporting a high level of incongruence between the two genes. A linear analysis indicated no linear relationships between pairwise distances within the *GAPDH* and ITS datasets (Fig 9C in Appendix), and an inter-partition length difference test [109, 171] confirmed that the *GAPDH* and ITS data are incongruent (Fig 9D in Appendix), indicating that the two genes have conflicting phylogenetic history or else are otherwise drawn from different distributions. These data reflect previous evidence of cryptic recombination observed in *C. geophilum* in some of the first studies of this species that used phylogenetic approaches over 20 years ago [144] and are largely consistent with studies using a variety of methods both in *Cenococcum* [75, 77, 83, 86] and other ascomycetes [172, 173]. Similarly, in our study recombination appears to have been active in the PNW isolate collection and is evident in the incongruent histories of inheritance between the ITS and *GAPDH* gene regions. However, while the history of these two genes show patterns consistent with recombination, only two genes were investigated. Gaining evidence across many more loci

within the population will be important to solidify this conclusion and fully quantify the rate and extent of gene flow.

Incomplete lineage sorting is unlikely within this population as the side by side phylogeny suggests many recombination events. Additionally, the ILD test represents a minimum estimate and suggests that there were multiple events, many of which are found deep within the phylogeny (Fig 9A-D in Appendix). If lineage sorting were likely, we would expect to see the more apparent recombination events in the tips of the tree rather than in the older branches.

Of course, a high level of incongruence between these two genes is not conclusive. Additional genomic sequencing could further elucidate the likelihood of recombination versus incomplete lineage sorting. Genes such as ITS are known to undergo conserved evolution after hybridization, so we would not expect ITS to accurately reflect the history of any rare recombination events. While the ITS region is used as the universal standard for fungal species identification [97], intragenomic variation presents serious concerns even when compared from the same isolate [99, 100]. Previous studies have observed that both closely related and cryptic species are difficult to delineate based on ITS region phylogenetic analyses [101, 102], and the presence of multiple copies of the ITS region within nuclear genomes increases the risk of inaccurate delineation [99, 103]. As a result, the *GAPDH* gene is considered more phylogenetically informative in analyses [89].

This discrepancy in apparent rates of recombination is consistent with the increased grouping shown by *GAPDH* (Fig 9A in Appendix), which is not known to undergo conserved evolution in the same method as ITS. Ultimately our comparison of the ITS and *GAPDH* gene phylogenies showed that those ITS resolves with less phylogenetic support and higher AICc [161] and BIC values [162] (Table 1 in Appendix), indicating incompatible modes and tempo of evolutionary change between the ITS region and the protein-coding *GAPDH* gene region.

2.8.3 *Cenococcum* phylogeographic diversity within a global context

Our study increases the known diversity of *C. geophilum* within the global isolate collection, with many of the PNW isolate collection clades appearing to be unique to those shown previously (Fig 10 in Appendix). Distinct groups that were present in previous analyses are indicated using the last initial of the first author (Fig 10 in Appendix, Supplemental Table 2.3). Four newly designated clades, V16-V19, represent never-before-seen relationships determined through a ML analysis using best fit parameters (Fig 10 in Appendix, Supplemental Table 2.3). One of these clades, V16, includes isolates from both the North American and European continents, and both clade V8 and V17 include isolates from both the West and East coasts of the United States (Fig 10 in Appendix, Supplemental Table 2.3). The genetic diversity of our *C. geophilum* isolates from the Pacific Northwest also appears greater in comparison to that found in other regions, with previous studies resolving a maximum of 9 clades within any particular collection [77, 89–92] (Table 2 in Appendix).

Four of the identified clades in the PNW isolates grouped similarly but remained unaffiliated with any of those in the global study, implying that clades three, four, and nine may be specifically associated with *Populus* in the Pacific Northwest or part of a hypothesized “core” of *C. geophilum* isolate collections associated with diverse hosts but just not represented in the sparse global samplings to date. Additionally, our analyses suggest PNW clades 10 and particularly 11 may represent particularly divergent clonal groups or incipient species within the PNW isolate collection, as both clades segregated strongly from the rest of the PNW isolate collection.

A 2005 study completed on a collection of *C. geophilum* isolated from Browns Valley, CA in the United States based on GAPDH, showed a total of three clades identified at greater than 90% bootstrap support [90]. While isolates from this study collection tended to group together in our analyses as well, the associated clades did not persist save for one nested grouping in clade III, designated here as D2 (Fig 10 in Appendix, Supplemental Table 2.3), which includes 23 isolates at a bootstrap support value of 98.3% and posterior probability of 1.0.

The identified Obase et al. clade 7 (O7) was described as the new species designated as *Pseudocenococcum floridanum* gen. et sp. nov. We used this taxon as an outgroup for both the PNW and global studies, and none of our isolates appear to be phylogenetically affiliated with *P. floridanum* (Fig 10 in Appendix). The isolate sequences provided by Drs. Francis Martin and Martina Peter generally remained unresolved in the global isolate collection analysis with few exceptions. One such exception, clade dFP1, includes four isolates from Switzerland and mirrors the phylogenetic relationship observed in de Freitas Pereira et al. [92] (Fig 10 in Appendix, Supplemental Table 2.3).

Two of the newly designated clades, V16 and V17, encompass the greatest geographic diversity in the global isolate collection analysis. The clade V16 includes 22 intercontinental total isolates grouped into 4 strongly supported nested clades (>90% bootstrap value, 1.0 posterior probability) (Fig 10 in Appendix, Supplemental Table 2.3), and Clade V17 (>90% bootstrap value, 1.0 posterior probability) includes cross-continental isolates from Browns Valley, CA, US, and Florida/Georgia (Fig 10 in Appendix, Supplemental Table 2.3).

These large spatial distances across which several well-resolved clades were observed within both the PNW and global isolate analyses highlight the need for higher resolution genetic studies. There is a possibility that greater genetic resolution will subdivide such groups despite spatially distinct origins to strongly group together based on origin. However, the cryptic nature of *C. geophilum* also presents the possibility that these wide distribution patterns between isolates will become more extreme as well. Either case will provide further understanding of the patterns of speciation and genetic exchange within the *C. geophilum* species complex.

2.9 Conclusions and future directions

The genetic diversity present within both local and global isolate collections of *C. geophilum* isolates is striking. Our study reveals the existence of multiple cryptic clades of *C. geophilum* as well as distinct phylogenetic groups from the PNW which may be uniquely associated with *P. trichocarpa*, and confirms the common view of this species as a hyper-diverse group of ectomycorrhizal fungi on both regional and global scales. Our study additionally strongly

indicates patterns consistent with recombination within the PNW isolate collection based on analyses performed on the *GAPDH* and ITS gene regions. However, to provide further evidence as to whether this represents one global, hyper-diverse species, or a myriad of cryptic species, a more robust analysis with greater population-level resolution across many loci to more accurately quantify gene flow will be required. Future research could further elucidate these regional and global relationships through the use of genotype-by-sequencing (GBS) or restriction-associated DNA sequencing (RAD-seq) approaches for rigorous *de novo* single nucleotide polymorphism (SNP) analysis [116]. Such approaches could help shed a light on this ubiquitous fungal taxon which has proven historically difficult to classify both physiologically and phylogenetically, and allow us the opportunity to delineate the individual species which may currently be included in this greater *C. geophilum* species complex, or the other mechanisms by which it may maintain such diversity.

2.10 Acknowledgements

The authors would like to thank Keisuke Obase, Yosuke Matsuda, Matthew Smith, Francis Martin and Martina Peter for providing access to cultures and sequences for comparison with our isolates, as well as Allison Veach and Stephanie Kivlin for their assistance with statistical analyses. This research was sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research, as part of the Plant Microbe Interfaces Scientific Focus Area at ORNL (<http://pmi.ornl.gov>). Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DEAC05-00OR22725.

Chapter 3

Phylogeographic patterns across regional isolates of the ectomycorrhizal fungus *Cenococcum geophilum* differ between genome wide restriction associated DNA sequencing and single gene approaches

3.1 Preface

A version of this chapter will be submitted for publication by Jessica M. Vélez^{1,2}, Todd W. Pierson³, Reese M. Morris¹, Luz M. Serrato-Diaz⁴, Steven J. LeBreux¹, Rytas Vilgalys⁵, Scott J. Emrich⁶, Jessy Labbé^{1,8}, and Christopher W. Schadt^{1,2,7,*}.

3.2 Author Affiliations

¹ Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

² The Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN 37996-3394, USA

³ Dept. of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA

⁴ College of Natural Sciences, University of Puerto Rico - Río Piedras Campus, San Juan Puerto Rico 00931, USA

⁵ Biology Department, Duke University, Raleigh, NC

⁶ Min H. Kao Dept of Electrical Engineering and Computer Science, University of Tennessee, Knoxville TN 37996-0845, USA

⁷ Dept of Microbiology, University of Tennessee, Knoxville TN 37996-0845, USA

⁸ Genome Science and Technology, University of Tennessee, Knoxville TN 37996-0845, USA

* Corresponding Author: Email: schadtcw@ornl.gov; Phone: 865.576.3982

3.3 Author Contributions

In addition to roles detailed in chapter 1, Todd W. Pierson of the Fitzpatrick Lab at the University of Tennessee Knoxville provided restriction-associated DNA sequencing (RADseq) reagents and training to Jessica M. Vélez. Luz M. Serrato-Díaz assisted with RADseq protocol optimization. Steven J. LeBreux completed genomic DNA extraction for RADseq. Scott J. Emrich provided RADseq data training and initial analyses which were expanded upon by Jessica M. Vélez. Jessica M. Vélez completed the full RADseq dataset analyses and primarily wrote this chapter. Revisions were primarily completed by Christopher W. Schadt.

3.4 Abstract

The genome of the ectomycorrhizal fungus *Cenococcum geophilum* has a mapped size of 178 Mbp, making this one of the largest sequenced genomes in the fungal kingdom. This species is genetically interesting due to its broad host range and extreme phylogenetic variation, despite its apparent asexual nature. Previous single and multi-gene phylogenies indicate that *C. geophilum* may in fact represent cryptic but genetically distinct species when comparing up to five concatenated genes but have remained inconclusive. While

broad scale whole-genome sequencing of isolates could address this uncertainty, lower-cost alternative methods are needed given the large genome size, global distribution, and diversity within the species. Here we adapt and apply methods for assessment of genome-level variation using restriction-associated DNA sequencing (RADseq) to assess species diversity and phylogeographic patterns across 187 *C. geophilum* strains isolated from the Pacific Northwest (PNW) of the United States. A total of 15 clades were strongly resolved (>80%) in the *de novo* RADseq data assembly with 24% of 171 retained isolates unresolved, compared to previous efforts which strongly resolved (>80%) 15 clades in the single gene phylogeny with 34% of 171 isolates unresolved. A discriminant analysis of principle components (DAPC) indicated four phylogenetic clusters, with one cluster potentially representing a cryptic species due to both distinct grouping across multiple datasets and the phylogenetic patterns observed within the *de novo* dataset phylogeny. In contrast to previous single gene analyses, a linkage disequilibrium analysis found no strong evidence of sexual recombination across the entire PNW isolate collection when considered as a whole but did indicate site- and clade-specific evidence of rare recombination or hybridization events, implicating that these regional isolates likely represent multiple diverse but cryptic species. However, even with our rather large sampling and genome-wide RADseq dataset, it appears more isolates and data will be required to sort out the complexity of relationships in this important group of fungi.

3.5 Introduction

The ectomycorrhizal fungus *Cenococcum geophilum* has potential as a model species for understanding plant-fungal interactions [75, 84, 174] due to its overall stress tolerance [49, 78, 79, 81, 175], simple isolation methods using multiple standard media types for new strains [142, 143], and worldwide distribution and association with multiple genera of plants [75, 84, 176]. However, understanding of the ecology and phylogenetics of *C. geophilum* is difficult due its asexual nature, lack of morphologically distinct features, and the large amount of genetic divergence found within the species [75, 77, 84]. The above complexities are coupled with one of the largest known genomes in the fungal kingdom with an estimated

size of 178 Mbp due to large numbers of transposable elements [84]. In combination these features have led many to believe that *C. geophilum* is likely a species complex, representing a multitude of phenotypically similar species which are genetically distinct [75, 77, 83, 91].

The increase in quality and efficiency of DNA sequencing technology has led to a push to increase the total number of fungal genomes sequenced [158, 177, 178] and is significantly expanding our understanding of fungal biology and evolution. While whole genome sequencing can provide the highest resolution of phylogenetic relationships and allow for robust genotype and phenotype predictions [179–182], these methods can be difficult to implement for fungi [183, 184]. Fungi are phylogenetically complex due to several factors. Fungi may be asexual, sexual, or a mixture of both, and a single species may have both haplo- and diplo-type chromosomes present [126, 181, 183, 184]. Additionally, species which were traditionally classified as asexual or clonal have often been found to contain evidence of recombination as well as hybridization with genetic approaches [86, 90, 144, 185–188].

While the cryptic nature of *C. geophilum* necessitates DNA sequencing in order to properly characterize this species, whole-genome sequencing at the scales required are prohibitive due to the large genome combined with its global distribution and cryptic diversity. Previous studies have sequenced single genes, including most commonly the internal transcribed spacer (ITS) gene and the protein-coding glyceraldehyde 3-phosphate dehydrogenase (GPD) gene, and performed phylogenetic analyses based on single genes or multigene concatenations [77, 89–92]. In summation, these studies characterized isolate collections from both the western and eastern United States, Europe, and Japan, and resolved between three to nine phylogenetic clades based on strong support (>80%) using either GPD, ITS, or a concatenation of both genes. In Chapter 2, we determined that a newly-isolated Pacific Northwest (PNW) collection of >200 isolates contained 15 well-supported (>80%) phylogenetic clades based on the glyceraldehyde 3-phosphate dehydrogenase (GPD) gene and showed evidence consistent with recombination between GPD and ITS phylogenies [174]. However, analyses of these two commonly used phylogenetic markers were complicated by insufficient phylogenetic resolution.

Collectively such studies have highlighted the broad diversity present within this traditional taxon and within the global isolate collection and in one case the description

of a new taxon, *Pseudocenococcum floridanum* [91], highlighting the need for increased sequencing that increases the phylogenetic resolution within the *C. geophilum* complex. Phylogenetic groupings identified in these previous studies might be further supported or identified as cryptic species within the *C. geophilum* species complex using more robust approaches. Furthermore, as the potential for applications of *C. geophilum* as a model species increases, cost-effective and robust methods for population-level polymorphism detection should increase our understanding of the genetic boundaries and genomic potential within *C. geophilum* isolates.

Restriction-associated DNA sequencing (RADseq) methods have been used in various animal and plant species for single nucleotide polymorphism (SNP) detection and increased species-level resolution of phylogenetic patterns as well as understanding of population level gene flow [112, 117–120, 189–191], but rarely applied to fungi [112, 120, 192]. This method has been proposed as a potential alternative to whole-genome sequencing [75, 110] at a fraction of the cost [116, 119, 193, 194] while providing a snapshot of genetic variation across genomes. The RADseq method also does not require prior genomic information in order to complete downstream analyses [110], making this a powerful tool which can be used in either *de novo* or reference-based sequencing data assemblies.

In order to test the viability of this process with a collection of *C. geophilum* isolates, the 3RAD RADseq method [116] was applied to a total of 187 of the Vélez et al. (2020) [174] PNW strains in order alleviate previous limitations from low phylogenetic resolution within the collection based on gene-based approaches. Because this method has rarely been applied to fungi, we evaluated various approaches using both reference-based and *de novo* methods to determine the best bioinformatic approach for analyzing these sequencing data [124]. We additionally compare our results to the most common single gene marker (GPD) to directly assess the improvements in the phylogenetic signal based on this new approach.

3.6 Methods

3.6.1 Isolate Growth and DNA Extraction

C. geophilum isolates were grown on a sterile cellulose grid filter (GN Metrical 28148-813) placed onto Modified Melin-Norkrans (MMN) media [152] composed of 3 g l⁻¹ malt extract, 1.25 g l⁻¹ glucose, 0.25 g l⁻¹ (NH₄)₂HPO₄, 0.5 g l⁻¹ KH₂PO₄, 0.15 g l⁻¹ MgSO₄·7H₂O, 0.05 g l⁻¹ CaCl₂, 1 mL⁻¹ FeCl₃ of 1% aqueous solution, and 10 g l⁻¹ agar, adjusted to 7.0 pH using 1N NaOH. Genomic DNA was successfully extracted from 187 PNW isolates representing 59 isolation sites [174] as well as the European 1.58 strain with known genome sequence for mapping [84] and *Pseudocococcus floridanum* isolate BA4b018 [91] with the Qiagen DNeasy PowerPlant Pro Kit per manufacturer instructions. DNA was quantified using the Invitrogen Qubit 2.0 fluorometer to confirm each extraction to ensure a minimum of concentration and quantity of 5 ng/μL and 50 ng total for 3RAD library preparations. While these isolates represent the majority of the original 229 member collection from our previous study [174], unfortunately 40 isolates (17%) were lost in the intervening time between the studies due failed growth during transfers, irrecoverable contamination by faster growing species, or other issues.

3.6.2 Homemade Serapure (SpeedBead) preparation for 3RAD enzymatic reaction clean-up

In order to clean up each enzymatic reaction product during the 3RAD process and efficiently remove DNA from the mixture, we used a homemade mix known as Serapure (http://baddna.uga.edu/Protocols/Speedbead_Protocol_June2016.docx) [195] produced from a mixture of Sera-mag SpeedBeads (Fisher 09-981-123) within a PEG/NaCl buffer. The stock TE solution was mixed in a 50 mL conical tube using sterilized stocks of 500 μM 1M Tris pH 8, 100 μM 0.5M EDTA, and deionized water to fill up to the 50 mL mark. The Sera-mag bead solution was mixed using a vortex and 1 mL transferred to a 1.5 mL tube. The 1.5 mL tube was placed in a magnetic stand and the supernatant removed with a 1000 mL pipette. To rinse the SpeedBeads, 1 mL TE was added and the solution

vortexed for 1 min, then the tube placed back in a magnetic stand and the supernatant removed. This process was repeated once more, then 1 mL TE was added and the solution vortexed for 1 min to fully resuspend the SpeedBeads. An additional solution was mixed in a sterile 50 mL conical tube using 9 g PEG-8000, 10 mL 5M NaCl, 500 μ L 1M Tris-HCl, 100 μ L 0.5M EDTA, and deionized water up to the 49 mL mark. This solution was mixed until the PEG went into the solution (5 min), then 27.5 μ L Tween 20 was added and the solution gently mixed. The 1 mL SpeedBead mix was added and this solution gently mixed until brown. The 50 mL conical tube was wrapped in aluminum foil in order to prevent exposure to light, and aliquots were removed and placed in 1.5 mL tubes as needed for reactions.

In order to test the Serapure mix, 2 μ L of the Fermentas 50 bp GeneRuler (Fisher FERSM0371) was mixed with 18 μ L of deionized water and mixed with 1 volume of the Serapure solution, incubated at 20°C for 5 min, then placed on a magnet stand. The supernatant was removed using a 1000 μ L pipette, then the SpeedBeads rinsed with 500 μ L freshly made 70% EtOH twice. A sterilized toothpick was used to absorb remaining 70% EtOH from the tube, then 20 μ L deionized water was added to rehydrate the SpeedBeads. The solution was once again placed on a magnet stand and the supernatant removed with a 200 μ L pipette and transferred to a new 1.5 mL tube. This was mixed with 1 μ L of loading dye and electrophoresed in 1.5% agarose for 60 min at 100 V. The resulting gel image was manually analyzed to ensure that bp fragments in the desired range of 550 bp were captured by the Serapure mixture.

3.6.3 Optimization of the 3RAD protocol for *C. geophilum*

The 3RAD technique described by Bayona-Vásquez et al. (2019) [116] was optimized for these fungi using the *C. geophilum* 1.58 reference isolate. We first tested both 5 and 10 μ L undigested genomic DNA with the three restriction enzymes to be used for 3RAD: 1.0 μ L ClaI (NEB R0197S), 0.5 μ L BamHI-HF (NEB R3136S), and 0.5 μ L MspI (NEB R0106S). Once a smear of DNA was confirmed, indicating that the restriction enzymes were able to cut *C. geophilum* gDNA, we normalized the quantity of gDNA present in the initial 3RAD digest reaction to either 50 or 100 ng. All three restriction enzymes were combined with 50 or 100 ng 1.58 gDNA and ClaI and BamHI-HF i5 and i7 internal adapters, which were

provided by the Fitzpatrick Laboratory (<http://web.utk.edu/~bfitzpa1/>) of the University of Tennessee Knoxville (UTK). The samples were incubated at 37° for 1 hour, then immediately combined with a ligation mix consisting of 2 μL deionized water, 1.5 μL ATP, 0.5 μL 10x ligase buffer (provided with T4 DNA ligase), and 1.0 μL T4 DNA ligase (NEB M0202S).

Ligation was completed using a thermocycler and the following conditions: 2 rounds of 22° for 20 min and 37° for 10 min followed by 80° for 10 min. Two volumes of Serapure were added to the reaction and incubated at 20° for 10 minutes. The reaction was set into a magnet tray and the supernatant removed with a 200 μL pipette, then rinsed with freshly made 180% EtOH which was immediately removed. A sterilized toothpick was used to remove remaining 180% EtOH from the bottom of the tube while carefully avoiding the Serapure beads, which were then resuspended in 20 μL TE.

Next, 10 μL of the ligation product was added to the Kapa HiFi PCR (Fisher KK2502) mix including: 5 μL Kapa HiFi Buffer, 0.75 μL dNTPs, 8.75 μL deionized water, 0.5 μL Kapa HiFi DNA Polymerase, and 2.5 μL i5 and i7 primer from iTru5 96-4 and iTru7 96-1 plates purchased from the Environmental Health Science (EHS) DNA laboratory at the University of Georgia (<https://baddna.uga.edu/services.html>). Full primer and adapter sequences may be found on the EHS DNA laboratory Adapterama Protocols site (<https://baddna.uga.edu/protocols.html>). Adapter-ligated DNA was amplified in a thermocycler with the following thermal profile: 95° for 2 min; 98° for 20 sec, 60° for 15 sec, 72° for 30 sec for 16-25 cycles; 72° for 5 min. These cycles were repeated for all strains and final PCR products were combined to ensure a stock solution total of at least 5 ng DNA/ μL . Successful 3RAD amplification was indicated when a smear was observed after the final 3RAD PCR amplification step using gel electrophoresis concentrated in the 300 to 600 bp range. Approximately 50 ng was found to be the optimal starting amount of fungal gDNA for 3RAD.

The thus optimized Adapterama 3RAD protocol was applied across the entire PNW isolated collection. Each isolate PCR amplification product was cleaned using Serapure and resuspended in 25 μL TE and a library pool was created by normalizing each isolate DNA quantity to 55 ng DNA in a single 1.5 mL tube, which was then cleaned using 2 volumes Serapure, 3 volumes of freshly made 80% EtOH, and resuspended in 60 μL for size selection.

Automated gel extraction for approx. 550 bp fragment sizes was completed using a Pippin Prep at the UT Knoxville Genomic Core facility. The final pool was cleaned using 1 volume Serapure, quantified using the Qubit dsDNA HS reagent and buffer, and sent for sequencing at the Georgia Genomics and Bioinformatics Core <https://dna.uga.edu/>. The PNW pooled library was combined and balanced with libraries from other studies to achieve a desired theoretical sequence depth of 500X and sequenced using an Illumina NovaSeq S4 at Hudson Alpha using the HiSeq X PE 300 cycles kit.

3.6.4 RADseq data demultiplexing and quality filtering

Raw sequence data were demultiplexed according to their Adapterama barcodes within the ipyrad package v .0.7.21 [130] on the UTK Advanced Computing Facility cluster <https://www.jics.utk.edu/acf>. All forward sequence reads were deposited to the NCBI Short Read Archive (SRA, <http://www.ncbi.nlm.nih.gov/sra/>) under Bioproject accession number PRJNA668013 (Supplemental Table 1). Data were then transferred to the UTK Helix cluster (Intel Xeon Silver 4116 X4) and trimmed and quality filtered using Trimmomatic v. 0.30 [196]. Sequences were trimmed with minimum thresholds of 60 bp in length and a minimum quality score per base of 15 with phred33 scoring.

3.6.5 Reference-based and *de novo* downstream assembly of sequences

For the reference-based assembly forward sequences were aligned using bwa (alignment via Burrows-Wheeler transformation) v. 0.5.9-r16 [197] and converted to SAM files, which were sorted and aligned to the *Cenococcum geophilum* 1.58 reference genome v. 2 <https://mycocosm.jgi.doe.gov/Cenge3/Cenge3.home.html> [84] using SAMtools v. 0.1.18 [198]. The generated binary variant call format (BCF) were then converted to variant call format (VCF) using BCFtools v. 0.1.17-dev [199] and further quality filtering was completed using VCFtools v. 0.1.16 [200] for a minimum read count of 3, a maximum missing data count of 50%, and a minimum quality score of 30. A cutoff of 5% minimum depth per site

coverage was applied to the sequenced isolates, and those with less were filtered out of the analysis.

For the *de novo* assembly, forward read RADseq data was assembled using the default settings of the Stacks 2 `denovo_map.pl` pipeline [131–133, 201] in the Galaxy GenOuest Bioinformatics Platform (<https://galaxy.genouest.org/>) [202] with a population map indicating the region of origin of the isolate as Switzerland (isolate 1.58), Florida/Georgia (*Pseudocococcus floridanum*), or PNW (187 isolates from Oregon and Washington in the USA). The generated VCF was processed using VCFtools with the same parameters as the reference-based assembly except that a minimum depth per site cutoff of 10% was used rather than 5%.

3.6.6 Clustering and phylogenetic analyses

A discriminant analysis of principle components (DAPC), which first transforms the data using a principle components analysis (PCA) and then uses retained PCA components to identify phylogenetic clusters [203], was completed on both the reference-based and *de novo* VCF files using the `vcfR` package [204] in the R statistical analysis software v. 3.6.2 [156]. A RAxML phylogeny using 1000 bootstrap replicates was generated for both VCF files using the `adegenet` v. 2.1.3 R package [205]. Phylogenies were directly compared between both datasets and against a single gene GPD analysis, and phylogenetic clades were assigned based on bootstrapping support of >80%. The RAxML phylogenies were deposited into the TreeBASE database (<http://purl.org/phylo/treebase/phylovs/study/TB2:S27050?x-access-code=e387164e9f75d539489421c346b5274d&format=html>).

The isolates were mapped to their original source of isolation and the clusters represented per isolation site were indicated using GenGIS v. 2.5.3 for both the *de novo* and reference-based RADseq assembly data sets.

3.6.7 Single gene phylogeny and comparison to *de novo* assembly

A RAxML phylogeny was generated using only the isolates retained in the *de novo* assembly in order to directly compare phylogenetic resolution between a single gene (GPD) and the *de*

novo assembly. A GPD PNW alignment was created using Geneious v. 10.2.3 [155]. A best fit analysis was run and a RAxML tree was constructed per the indicated best fit parameters with 1000 bootstraps using the MEGA X software [160].

3.6.8 Linkage disequilibrium analyses

In order to determine whether the PNW isolate collection represents a predominantly clonal or sexually reproducing population of *C. geophilum*, the *de novo* assembly was converted to a genind object using R package vcfR v.1.10.0 [204]. In addition, all clusters, well supported clades, and original sites of collections isolations with at least three isolates were extracted from the full dataset and analyzed separately. The Index of Association (IA) was calculated using 1000 permutations in the poppr v. 2.8.5 package [206, 207] and visualized in the R statistical software on the Oak Ridge National Laboratory CADES computational cluster.

3.7 Results

The reference-based assembly of the PNW isolate collection represents a less conservative data set with >18,000 SNPs but fewer isolates (112), while the *de novo* assembly represents a more conservative data set with 1,670 SNPs but more isolates (171). Additionally, the reference-based assembly had lower read depth (minimum of 250 – 5%) compared to the *de novo* assembly (minimum of 1500 – 10%). Likely, the difference of 18,000 SNPs determined for the 5% cutoff VCF are due to genomic repeat regions [208] as well as high genetic divergence between the PNW isolates and the 1.58 European reference strain.

The reference-based assembly retained isolates from 45 out of 59 sites, while the *de novo* assembly retained isolates from 52 out of 59 sites. Neither assembly retained the *Pseudocococcum floridanum* isolate BA4b018 due to insufficient read depth. The *de novo* and reference-based assemblies also differed in strengths in phylogenetic signal within the PNW population (Fig 11 in Appendix). The *de novo* assembly strongly resolved (>80% support) 76% of the retained 171 isolates into 15 clades, while the reference-based assembly only resolved 42% of the retained 112 isolates into 8 clades (Fig). There was also additional strongly supported sub-structure within the *de novo* clades, particularly clades RDN6 and

RDN9, which represent the majority of the resolved isolates-12 in Appendix). Comparatively, the single gene phylogeny strongly resolved (>80%) 15 clades and 66% of 171 isolates (Fig 13 in Appendix). The higher number of retained isolates and sites, combined with stronger phylogenetic signal, led us to only consider the *de novo* assembly in downstream analyses.

3.7.1 Phylogeographic patterns and distribution in the Pacific Northwest

Four phylogenetic clusters were identified based on a DAPC analysis of the *de novo* VCF file, with 2 clusters containing 34 isolates, one containing 47 isolates, and one containing 55 isolates. The defined clusters spanned the entire geographic region sampled in both data sets with no noticeable patterns of segregation relating to latitude between or within the 4 clusters (Fig 14 in Appendix). While no population overlap was seen within individual isolates (data not shown), many individual isolation sites contained more than one cluster within the isolates derived from same soil samples from a particular site, a pattern which occurred in sites with both small and large numbers of derived isolates (Fig 14 in Appendix).

3.7.2 Single gene (GPD) and RADseq *de novo* (RDN) phylogenetic comparison

Ten clades were strongly resolved at >80% bootstrap support based on the GPD gene (Fig 12 in Appendix). A total of 34% of the PNW isolates included were unresolved. The two largest clades representing the greatest spatial and riverbed diversity are GPD3 and GPD6 (Fig 12 in Appendix, Table 3 in Appendix). The GPD3 clade includes 29 PNW isolates from sites along the Sandy, Skykomish, White, and Willamette Rivers as well as Lake Whatcom (Fig 12 in Appendix, Table 3 in Appendix). The GPD6 clade includes 31 PNW isolates from sites along the Lewis, Puyallup, Sandy, Skykomish, Nisqually, White, and Toutle Rivers, in addition to isolate POR1_10.3 from a site along the Santiam river (Fig 12 in Appendix, Table 3 in Appendix). The 1.58 reference strain was unresolved in the GPD phylogeny (Fig 12 in Appendix).

The *de novo* assembly produced fifteen strongly resolved (>80%) clades with significantly increased resolution within the larger clades (Fig 12 in Appendix). The largest clade, RDN9, includes 44 PNW isolates from sites along the Columbia, Lewis, Puyallup, Sandy, Skykomish, Nisqually, White, and Toutle rivers, in addition to isolate POR1_10.3 from a site along the Santiam river (Fig 12 in Appendix, Table 3 in Appendix). The second largest clade, RDN1, includes 34 PNW isolates ranging sites of isolation along the Willamette, White, and Sandy rivers as well as Lake Whatcom (Fig 12 in Appendix, Table 3 in Appendix). Clade RDN1 also includes the 1.58 reference strain, indicating that these isolates may be more closely related to the Swedish reference strain based on additional genetic information (Fig 12 in Appendix).

While *de novo* assembly clusters 3 and 4 are composed of PNW isolates from multiple clades as well as unresolved isolates, clusters 1 and 2 are predominantly composed of a single *de novo* clade (Fig 14 in Appendix, Table 3 in Appendix). Cluster one is entirely composed of 35 isolates from clade RDN9 ranging sites of isolation along the Lewis, Nisqually, Santiam, Puyallup, Skykomish, Toutle, and White rivers (Fig 14 in Appendix, Table 3 in Appendix). Similarly, cluster 2 includes 33 of the RDN1 isolates with the sole exception of WI_102.10, which is included in phylogenetic cluster 3 (Fig 14 in Appendix, Table 3 in Appendix).

Overall, there is a negative correlation ($r = -0.45$, $p < 0.05$) between the *de novo* clade (RDN) and the clade assigned based on the GPD gene. However, when clades GPD3 and GPD6 are examined separately from the remaining GPD phylogeny there is a positive correlation ($r = 0.85$, $p < 0.05$) between the *de novo* assigned clade and the clade assigned based on the GPD gene. There was a slightly positive correlation between the *de novo* clade and phylogenetic cluster ($r = 0.19$, $p < 0.05$), as well as a slightly positive correlation between the latitude of the original isolation site and the *de novo* phylogenetic cluster ($r = 0.30$, $p < 0.05$) (Fig 14 in Appendix, Table 3 in Appendix). The strongest correlation was observed between *de novo* strongly supported (>80%) clades as a whole and the GPD strongly supported (>80%) clades as a whole ($r = 0.66$, $p < 0.05$).

3.7.3 Linkage disequilibrium within the PNW isolate collection

We found significant support for allele linkage across loci in the overall *de novo* assembly with the observed $\bar{r}d$ falling outside of the expected distribution with no allelic linkage ($\bar{r}d = 0.0376$, $p < 0.001$), strongly indicating predominantly clonal rather than sexual reproduction across the entire collection of PNW isolates when considered as a whole (Fig 15A in Appendix). However, when individual *de novo* assembly clades, individual rivershed sites of origin, and individual isolation sites with at least three isolates were analyzed separately, only the observed $\bar{r}d$ of clade RDN5 (N=4 members) fell within the expected distribution associated with no allelic linkage ($\bar{r}d = 0.274$, $p < 0.01$) (Fig 15B in Appendix). Additionally, when analyses were further partitioned by geographic origin, the Snohomish rivershed isolate collection (N=4) fell within the expected distribution, however the results were not significant ($\bar{r}d = 0.149$, $p = 0.07$) (Fig 15C in Appendix). The $\bar{r}d$ of the isolates obtained from Snohomish river site six (SN6) (N=3) also fell within the expected distribution ($\bar{r}d = 0.0885$, $p < 0.05$) (Fig 15D in Appendix) to indicate no allelic linkage. Similarly, the observed $\bar{r}d$ of the total White River population (N = 36) fell outside of the expected distribution, ($\bar{r}d = 0.161$, $p < 0.001$), while isolates obtained from White River site four (WH4, N = 3) fell within the expected distribution ($\bar{r}d = 0.0624$, $p < 0.01$) (Fig 15E-F in Appendix).

3.8 Discussion

3.8.1 Viability of RADseq usage for phylogenetic analysis in fungal studies

The RADseq technique has rarely been applied to fungal species to date [112, 120, 192]. We found that, with optimization, this technique provides a low-cost alternative to whole-genome sequencing, particularly in phylogenetically complex species such as *Cenococcum geophilum*. The modified 3RAD technique used here provided a semi-random selection of genetic variation across the genome of *C. geophilum*, accounting for approx. 1% total genomic coverage. Despite this seemingly low coverage, stronger phylogenetic signal was

observed in the *de novo* assembly as compared to the single gene GPD phylogeny. Strikingly, the reference-based assembly produced fewer strongly supported (>80%) clades and fewer phylogenetic signals within the phylogeny, which may reflect the high repeated sequence content due to the presence of transposable elements within the *C. geophilum* genome [84]. While these repeats increase difficulty compared to more traditional single-gene or concatenated multi-gene sequencing applications [75], the ability to complete both a reference-based assembly using a published genome as well as a *de novo* assembly without relying upon a previously published genome mitigates the risk of phylogenetic signal loss making this an exciting potential technique for use with a myriad of unsequenced fungal species.

3.8.2 Phylogenetic structure within the Pacific Northwest isolate collection

Overall, we found that the *de novo* RADseq assembly methodology provided more robust phylogenetic signal and further supported the presence of cryptic species boundaries within our isolate collection. A total of 15 clades were strongly resolved (>80%) in the *de novo* RADseq data assembly with 24% of 171 retained isolates unresolved, compared to 10 strongly resolved (>80%) clades in the single gene phylogeny with 34% of 171 isolates unresolved, and 8 strongly resolved (>80%) clades in the reference-based RADseq data assembly with 58% of 112 retained isolates unresolved (Fig 11-13 in Appendix, Table 3 in Appendix). Additionally, we found four phylogenetic clusters, with at least one cluster potentially representing a cryptic species due to both consistent grouping across multiple datasets, consistent phylogenetic clustering in previous studies [174] and the current study, and phylogenetic patterns observed within the *de novo* dataset phylogeny (Fig 14 in Appendix, Table 3 in Appendix). The *de novo* phylogenetic clusters had no noticeable patterns of segregation relating to latitude and ranged across the entire geographic region sampled (Fig 14 in Appendix). While no phylogenetic cluster overlap was seen within individual isolates (data not shown), many PNW isolation sites contained more than one cluster within the

isolates derived from same soil samples from a particular site, a pattern which occurred in sites with both small and large numbers of derived isolates (Fig 14 in Appendix).

When we calculated the Index of Association (Ia) for both the reference-based and *de novo* assemblies, we found no strong evidence of sexual recombination within the PNW isolate collection as a whole (Fig 15 in Appendix). Deeper Ia analyses of site-specific isolate collections as well as strongly supported (>80%) *de novo* clades revealed evidence of recombination, indicating a clonal population with hybridization patterns and some local-level sexual recombination (Fig 15 in Appendix). Additional genome-wide association studies (GWAS) studies can now be performed with these data in order to map measurable phenotypic traits to genomic information and further determine whether phenotypic traits are correlated to specific groupings within the PNW population.

We observed further evidence of extreme phylogenetic diversity within the PNW *Cenococcum geophilum* isolate collection based on the *de novo* assembly of the genome-wide RADseq analyses. The *de novo* RADseq data assembly (RDN) resolved 15 well-supported clades (>80%) compared to 10 well-supported clades in a single gene (GPD) phylogeny and featured greater sub-structure and support (Fig 12 in Appendix, Table 3 in Appendix). Upon further investigation, there was some evidence of recombination at specific local-level sites and within one *de novo* clade, possibly implicating a collection of clonal groups with possible recombination occurring between isolates of sub-groups within the collection originating in close geographic proximity (Fig 15A-F in Appendix).

3.8.3 Single gene compared to RADseq *de novo* phylogenetic approaches

Our previous efforts had applied phylogenetic analyses of the ITS region GPD gene across this largely same population of isolates. Incongruence patterns and comparison of the ITS and GPD phylogenies had also suggested recombination may be occurring within the population. However, there is much congruence to be noted between the *de novo* RADseq analyses and the GPD single gene phylogeny, especially in the larger, more common and well supported groups. In general however, the RADseq based approach provided more support

for understanding the identity of previously unresolved isolates, the relationships between groups, and the structure of their fine scale differentiation within groups. When a direct comparison between the approaches is made, isolates with data from both approaches may show similar patterns. For example, apart from isolate WH2_54.10, all the PNW isolates that grouped into clade GPD6 are also grouped into clade RDN9 with strong support (81.1%) (Fig 12 in Appendix, Table 3 in Appendix). Similarly, 24 of the 29 total GPD3 isolates are grouped into the strongly supported clade RDN1 (99.4%) in the *de novo* analyses, and an additional six isolates which were unresolved based on the single gene phylogeny are also grouped into the new clade RDN1 (Fig 12 in Appendix, Table 3 in Appendix). Additionally, clade GPD6 is predominantly assigned to DAPC cluster 1 with few exceptions (Fig 14 in Appendix, Table 3 in Appendix).

Interestingly, we also observed that reference strain 1.58 resolved into clade RDN1, although on a very long branch, and grouped into DAPC cluster 2. This may imply that this strain is closely related to the PNW isolates of clade RDN1 despite an origin located in Switzerland (Fig 12 in Appendix, Table 3 in Appendix). However due to the large number of differences with other isolates in this group and the lack of additional European isolates for comparison, this may be artifactual due to long branch attraction or similar phenomena.

3.8.4 Evidence for clonality and recombination in the PNW isolate collection

The widespread dispersal of many fungal species has led to the development of asexual clonal populations, which are thought to account for approximately 20% of total known fungi [144, 188, 209]. While *C. geophilum* is often referred to as an asexual species complex, evidence of sexual recombination has been detected in multiple studies based on single genes or multigene alignments [76, 83, 84, 86, 174]. A pattern of sexual recombination is hypothesized to indicate the origin of a eukaryotic population, whereas a clonal eukaryotic population may display patterns of hybridization and diversification [20, 144, 185, 210–212]. One such example determined that the global origin of the plant pathogen *Phytophthora infestans* was located

within central Mexico due to the inference of sexual reproduction within the Mexican isolates [210].

Overall, our data suggests a predominantly clonal rather than sexually recombining population of *C. geophilum* (Fig 15A in Appendix). However, within such a diverse collection of potentially cryptic species, it is possible that such analyses may be blurred. For example a prior study by Douhan et al. (2007a) [86] determined that when sub-clades were considered separately in analyses from the full population set, patterns of recombination could be detected within the clades as might be expected when cryptic species boundaries may exist. When we analyzed each collection of isolates based on either site of origin, rivershed of origin, or *de novo* clade membership, we indeed found potential evidence for recombination (Fig 15A-F in Appendix). For example, the small group of clade RDN5 isolates (n=4) have origins at a site on the Nooksack River as well as a site relatively nearby (15km) at the Whatcom Creek outflow of Lake Whatcom (Fig 15B in Appendix). Additionally, when considered on a site level there was evidence for recombination within isolates originating from site SN6 on the Snohomish and within site WH4 on the White River (Fig 15B, D, F in Appendix). In comparison, when analyzed as whole the full PNW isolate collection did not indicate a pattern consistent with recombination (Fig 15A) in Appendix, nor did the combined Snohomish and White River sites respectively (Fig 15C, E in Appendix).

Such hyperlocal and restricted evidence of recombination might be consistent with an expected pattern for a geographically narrow cryptic species with limited sexual reproduction. Alternatively however, an absence or rarity of population-level sexual recombination could be indicative of hybridization events followed by diversification within the local populations [20, 185–188, 213, 214]. Other examples such as the large widespread clade RDN1 may indicate potential local-level adaptation (Fig 12). While evidence of recombination was not found through a linkage disequilibrium analysis for the overall RDN1 clade, there are clear patterns of sub-group segregation between site-specific isolates from Willamette River site WI1 (95.8%), White River site WH5 (98.6%), and Willamette River site WI7 (90.5%) (Fig 12 in Appendix). Clade RDN9 also contains a sub-group composed only of six Toutle River site TO3 isolates (100%), and another sub-group consisting only of five isolates from Lewis River site ParPoint (100%) segregated from four isolates from

Columbia River site POR4 (100% and 99.8%) (Fig 12 in Appendix). Such patterns are consistent with expectations of local population genetic drift or adaptation within a species, and suggestive of cryptic species boundaries between such groups (RDN1 vs RDN9).

Another possible explanation for the absence of evidence for population-level sexual recombination lies within a possible limitation of the RADseq technique itself as restriction enzyme sites are not uniformly and randomly distributed throughout the genome [215, 216]. Recombination will most often occur in “hot spots” across the genome, leading to non-universal linkage disequilibrium across the entire genome [215, 216], which may have decreased the number of recombination hot spots captured in this dataset. While this possibility is worth mentioning due to the potential impact on our results, there is some disagreement about the frequency and severity of these issues when using RADseq data [115].

3.9 Conclusions

The application of genome-wide phylogenomic information overall shows that our large Pacific Northwest collection of this important ECM fungus *Cenococcum geophilum* is highly diverse. These approaches show when considered overall there initially appears to be little evidence of gene flow. However, when considered on a group specific basis and/or local watershed or site basis, we can observe potential signals consistent with recombination events. Taken together these patterns are what would be expected given potentially cryptic species barriers and with infrequent, low-level, recombination or hybridization events. It is clear however that even with 187 isolates in our analyses, we are likely under-sampling the diversity in this group of fungi, making conclusive inferences difficult. Further isolate representatives and data at regional, watershed and site levels may provide a more conclusive picture of these patterns. Additional studies which incorporate the Pacific Northwest RADseq data in conjunction with previously characterized isolates from the global *C. geophilum* collection could also further support both the patterns observed within the PNW population and help us to understand these intriguing fungi more broadly. Future research directions for these systems and techniques could include genome-wide association studies using such data to

determine if the SNPs indicated using these RADseq data are co-associated with measurable phenotypic traits.

Chapter 4

Responses of *Cenococcum geophilum* to heavy metals and identification of associated genomic markers

4.1 Preface

A version of this chapter will be submitted for publication by Jessica M. Vélez^{1,2}, Scott J. Emrich³, Wellington Muchero¹, Rachel Andrews¹, Jessy Labbé^{1,5}, and Christopher W. Schadt^{1,2,4,*}.

4.2 Author Affiliations

¹ Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

² The Bredezen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN 37996-3394, USA

³ Min H. Kao Dept of Electrical Engineering and Computer Science, University of Tennessee, Knoxville TN 37996-0845, USA

⁴ Dept of Microbiology, University of Tennessee, Knoxville TN 37996-0845, USA

⁵ Dept. of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996,

USA

* Corresponding Author: Email: schadtcw@ornl.gov; Phone: 865.576.3982

4.3 Author Contributions

In addition to roles detailed in Chapters 1 and 2, Jessica M. Vélez completed the genome-wide association study on the *de novo* and reference-based RADseq datasets and primarily wrote this chapter. Wellington Muchero provided invaluable guidance in regards to association mapping protocols and procedures. Revisions were primarily completed by Wellington Muchero and Christopher W. Schadt.

4.4 Abstract

The growth of agriculture has caused the demand for arable lands to rise globally, but a significant portion of otherwise promising land is contaminated with heavy metal substrates. Biofuel crops have the potential to aid in remediating soils and allow agricultural development of these lands. While some biofuel crops such as *Populus* are generally tolerant of heavy metal contaminants, this tolerance may be further improved with the introduction of a microbial associate such as the ectomycorrhizal fungus *Cenococcum geophilum* as it is generally tolerant of heavy metals. Here we measured the radial growth rates of 56 *C. geophilum* isolates including 55 collected from the United States Pacific Northwest (PNW) in the presence of six metals in toxic concentrations: the heavy metals copper (Cu), cadmium (Cd), lead (Pb), strontium (Sr), and zinc (Zn), with sodium (Na) serving as a non-heavy metal comparative treatment. The average radial growth across all of the heavy metal treatments except Sr was significantly decreased ($p < 0.05$) in comparison to the control condition, with the presence of Sr having no significant effect ($p > 0.05$) on 54 of the isolates and significantly increasing ($p < 0.05$) average radial growth for two isolates when compared to the control condition. Association mapping to determine potential gene markers was completed using either a reference-based or *de novo* RADseq data assembly to map associations of metals tolerance (> 1 cm average radial growth in 70 days) back to regions of the *C. geophilum*

genome. Twenty significant associations were found in the presence of Cd ($p < 0.05$), which had the strongest effect on growth, whereas none of the other individual metals effects on growth showed significant genomic associations. Cadmium growth associations with strong isolate growth (approx. 2 cm in 70 days) in the presence of Cd but not Cu or Zn included two potential markers for unique single nucleotide polymorphisms (SNPs) linked to proteins glucooligosaccharide oxidase and fructose-1,6-bisphosphatase. Of the 20 significant markers, two SNPs were associated with iron ion binding proteins ($p = 0.02$), at least five were associated with metabolic functions including nucleic transcription and translation, and one was associated with carbohydrate membrane transport. Collectively our results show that RADseq based association mapping can be used to identify rapidly candidate isolates with heavy metal tolerance for further investigation and development of agricultural applications.

4.5 Introduction

The growth of the agricultural industry during the 20th century has drastically increased the demand for arable lands [1]. More efficient farming combined with population growth have put significant pressures on land use to ensure the economic security of developed and developing nations [4]. This increased demand and economic dependency shines a light on current limitations of available lands. Among current limitations for agricultural growth is land contaminated with organic and inorganic substances [9, 13], with a significant portion containing severe heavy metal contamination within the soil [6, 8–10]. Approximately 1.73 million hectares of closed landfills, abandoned mining lands and other areas are closed to future development due to contamination in the United States alone [7] and there has been growing interest in bio- and phytoremediation strategies to bring these lands back into agricultural production in the United States and worldwide [13, 217–220] (Fig 2 in Appendix).

Biofuel crops such as *Populus* spp. trees are of particular interest for reforestation, phytoremediation, and recovery of contaminated sites due to their fast growth [22, 140] and overall hardiness in the presence of heavy metal contaminants [24, 29], and processing

of harvested biomass through pyrolysis leads to negligible heavy metal contaminant carry-over [16, 33, 35] as heavy metal contaminants are captured within the char byproduct [33, 35], which represents only 10% of the total yield from the pyrolysis of biomass (Fig 4 in Appendix).

Biofuel crops serve as host to a diverse community of microbes including fungi which are capable of metabolizing and/or immobilizing soil compounds which the plant cannot [2, 52–54, 209, 221–225]. Previous studies have sought to aid growth and increase the general tolerance of heavy metal contaminants by biofuel crops through microbial community manipulations [37–39]. These manipulations could increase biofuel crop tolerance of otherwise toxic soil conditions [38, 43], encouraging the agricultural development of these lands using biofuel crops. Specifically, fungal-plant symbioses may increase the ability of a biofuel crop to thrive in soils containing high heavy metal concentrations. Some mycorrhizal fungi directly interact with heavy metal contaminants within the soil [37, 39], and a variety of fungal isolates from biofuel crop roots have previously been shown to have increased resistance to heavy metal contaminants [16, 55, 73, 224–226].

The fungal species *Cenococcum geophilum* is a ubiquitously distributed ectomycorrhizal fungus that has been positively associated with plant health, growth, and increased contaminant resistance [49, 74]. This versatile fungus is known to associate with both angiosperm and gymnosperm tree species across 40 genera, representing over 200 total tree species [77], and is generally tolerant of high salinity [78], water stress conditions [79], and extreme heavy metal contamination [46]. This wide-ranging resistance to stress conditions in a soil environment is often attributed to a high melanin content and complex cell wall structure [78, 79, 81, 82], which is also thought to play a role in increased resistance to heavy metal contaminants [39, 45, 81].

A full genome of a representative *C. geophilum* isolate was first reported in 2016, and was mapped at approximately 178 Mbp [84], making this one of the largest known fungal genomes. This extreme size is attributed to repeated sequences throughout the genome composed primarily of transposable elements [84], a reality which increases the difficulty of additional wider-scale sequencing applications [75]. Further complications arise due to the wide genetic and phenotypic variation of fungi within the *C. geophilum* taxonomic definition, indicating

it may be a complex of related species [75, 77, 84, 87, 90] and necessitating thorough genetic characterizations to investigate the population structure of any given collection. Combined these complexities suggest the need for broad scale sequencing approaches in *C. geophilum* which could provide robust *de novo* polymorphism detection across diverse populations of isolates for a fraction of the cost of whole genome sequencing.

The restriction-associated DNA sequencing technique (RADseq) provides a cost-efficient opportunity to characterize variation across the genome of *C. geophilum* [116, 119, 123, 227]. This technique was previously applied to a subset of a newly isolated collection of *C. geophilum* from the Pacific Northwest region (PNW) to understand the broader patterns of biogeographical variation within this species complex, where *de novo* assembly provided the more robust phylogenetic analyses (Chapter 2). Fungal heavy metal resistance has been shown to vary across species [37, 46, 55], as well as across individual species isolates and phylogenetic groups [228–230].

In this chapter, we therefore explore the use of such datasets for genome-wide association mapping in order to determine if particular SNPs detected with the *C. geophilum* isolate collection are associated with growth in association with a variety of heavy metals. In order to determine which PNW isolates are the most and least susceptible to heavy metals, a phylogenetically representative collection of 50+ *C. geophilum* isolates were screened against a panel of common heavy metal contaminants including cadmium (Cd), lead (Pb), copper (Cu), strontium (Sr), and zinc (Zn) as well as the non-heavy metal sodium (Na) and controls without added metals. Radial growth was measured to determine which PNW isolates displayed the greatest heavy metal tolerance as defined by an average radial growth > 1 cm at 70 days under each condition. These traits were then mapped back to the *C. geophilum* genome using both a reference-based and *de novo* RADseq data assembly to 1) compare the success of using either assembly in order to determine pros and cons of each when investigating *C. geophilum* using RADseq data; and 2) to detect particular SNPs associated with particular phenotypes. Our comparisons led to the conclusion that only the *de novo* assembly yielded significant associations with a total of 20 significant SNP markers detected in the Cd treatment. The majority of the identified proteins are involved in metabolic processes including protein transport, nucleic transcription and translation, and

heme binding. We additionally found that strains with contrasting growth patterns in which they grew well (approx. 2 cm in 70 days) in the presence of Cd but not in the presence of Zn and Cu contained two unique SNPs associated with proteins involved in metal ion binding that have been associated with increased heavy metal resistance in other fungal and bacterial species, although these associations were not significant in either the Zn or Cu treatments, suggesting there may be trade-offs in the effects of these genetic changes under different conditions.

4.6 Methods

4.6.1 Growth assay development and strain selection

The reference strain 1.58 was used in an initial growth assay to determine the optimal parts per million (ppm) for cadmium (Cd), copper (Cu), lead (Pb), strontium (Sr), sodium (Na), and zinc (Zn) where noted growth inhibition occurred but without complete inhibition or death. *C. geophilum* 1.58 was grown on Modified Melin-Norkrans (MMN) media [152] composed of 3 g l⁻¹ malt extract, 1.25 g l⁻¹ glucose, 0.25 g l⁻¹ (NH₄)₂HPO₄, 0.5 g l⁻¹ KH₂PO₄, 0.15 g l⁻¹ MgSO₄·7H₂O, 0.05 g l⁻¹ CaCl₂, 1 mL⁻¹ FeCl₃ of 1% aqueous solution, and 10 g l⁻¹ agar, adjusted to 7.0 pH using 1N NaOH. All heavy metals were tested separately as additives to the MMN media using the following concentrations: 100, 50, 25, and 15 ppm. Cadmium was additionally tested at 10, 5, 1, and 0.5 ppm. Based on results, all metals except for Cd were screened at 25 ppm concentration in the final assay, with Cd at 0.5 ppm. Representative strains were down selected from the approximately 200 isolates in the PNW collection with an effort to maximize geographic and phylogenetic diversity as determined in Chapter 2 [174]. Selections were made in a manner to ensure inclusion of representative isolates from each rivershed and each overall phylogenetic clade, and a total of 55 PNW isolates were ultimately included in the growth experiments along with the reference 1.58 isolate (Table 4 in Appendix).

4.6.2 Radial growth rate assay using 4 quadrant plates

Once sufficient fungal material was present for each isolate, a sterilized metal cork borer was used to extract a five mm plug which was centered and quartered in a Fisherbrand compartmentalized Petri dish (FB087582) to test four conditions simultaneously. Each Petri dish held three heavy metal conditions and a control for direct visual comparison across conditions in the following two groupings: 1) Cu, Sr, Zn, control, and 2) Pb, Cd, Na, control. Each grouping had five replicates per isolate, leading to ten total plates per isolate. Radial growth was measured from the center to the furthestmost hyphal tip in millimeters using a Bel-Art SP Scienceware Dial-Type caliper (Fisher 12-122) every two weeks and marked with an extra fine metallic silver oil-based Sharpie. Growth was measured for up to 70 days or until any condition of the four conditions on the plate reached the edge of the Petri dish quadrant, after which the plate was stored at 4°C to halt further growth until harvest of biomass for later analyses.

4.6.3 Statistical analyses of radial growth

A multivariate analysis of variance (MANOVA) was performed using R Statistical Software v. 4.0.2 [156] to determine whether the presence of a heavy metal within each growth treatment significantly impacted radial growth after 70 days of consecutive growth. Additionally, Tukey’s Significantly Different (TSD) test was performed using R to determine whether the presence of each heavy metal significantly impacted radial growth of *C. geophilum* isolates across all treatments for the same growth period.

4.6.4 Determination of significant phenotype-associated SNPs

DNA extraction and RADseq dataset analyses and were completed as detailed in Chapter 2. A phenotype file was created indicating if the average radial growth for each respective isolate in each growth treatment was > 1 cm radial diameter at 70 days of growth with “1” indicating yes and “2” indicating no. This phenotype file was correlated using the Firth logistic regression [231] to the reference-based and *de novo* assemblies using PLINK v.2 (www.cog-genomics.org/plink/2.0/) [232, 233]. Multiple testing corrections were completed

on the SNPs yielded using PLINK v.2, and the Benjamini-Hochberg False Discovery Rate Adjustment [234, 235] was used to determine significance ($p < 0.05$). Manhattan plots were generated for each phenotype using the qqman v. 0.1.4 package [236] in R. Datasets were manually reviewed to document SNPs of interest and their corresponding reference-based scaffold number or *de novo* assembly contig across all phenotypes. Reference-based and *de novo* contig sequences associated with phenotypes were then analyzed against the *Cenococcum geophilum* isolate 1.58 genome assembly v.2 [84] using the MycoCosm BLAST feature within the Joint Genome Institute portal [158, 159] and the results, including scaffold ID, sequence start, and sequence end were recorded (Supplemental Table 4.1). Significant SNP markers were manually evaluated using the search function in MycoCosm to determine possible protein identities and functions associated with each SNP location (Supplemental Table 4.1). Results were compared across measured phenotypes and between the reference-based and *de novo* assembly results (Supplemental Table 4.1).

4.7 Results

4.7.1 Heavy metal tolerance across the Pacific Northwest isolates

Four of the 56 isolates from either White, Toutle, Snohomish, or Puyallup rivershed sites did not grow in any condition (17), perhaps due to problems in transfer. A MANOVA indicated that the presence of a heavy metal significantly impacted growth across all *Cenococcum geophilum* isolates ($p < 0.001$), and a TSD revealed that all heavy metal conditions significantly impacted growth of each isolate relative to the isolate control condition ($p < 0.001$). The TSD additionally revealed that nearly all elemental conditions were statistically significant both in contrast to the control condition and to each other ($p < 0.05$) with the exceptions of Cu to Cd ($p = 0.052$) and Zn to Pb ($p = 0.99783$). When investigated further, we observed that the majority of isolates averaged less growth in the presence of Cu, Pb, and Zn when compared to the control condition ($p < 0.05$), and were unaffected by the presence of Sr when compared to the control (Fig 16-17). While the presence of Na did not appear to impact isolate growth as strongly as the heavy metals, there was still a significant difference

in total radial growth observed ($p = 0.009$) (Fig 16). As a collection, isolates averaged radial growth > 1 cm in the control condition after 70 days as well as in the presence of 25 ppm Na and Sr (Fig 16). Isolates grew an average of 0.7 cm in the presence of both 25 ppm Zn and Pb, 0.5 in the presence of 25 ppm Cu, and 0.4 in the presence of 0.5 ppm Cd (Fig 16). Forty-nine isolates grew very little in the presence of Cd (Fig 16-17), and 7 isolates which grew well (approx. 2 cm at 70 days) in the presence of Cd did not also grow well in the presence of Cu and Zn, and vice-versa (Fig 17). In particular, isolates from sites White River 5 (WH5) and Skagit 5 (SK5) consistently grew above the average of the control condition (> 1 cm in 70 days) in the presence of both Cd and Pb for an average of 2 cm and 1.5 cm respectively but grew < 1 cm in 70 days in the presence Cu and Zn (Fig 18).

4.7.2 SNP detection and phenotype association

The *de novo* assembly contained a total of 327,884 contigs of up to 65 base pairs in length, accounting for approximately 1% of the total *C. geophilum* genome. These contigs accounted for 3,264 SNPs that were then compared to measured phenotypes. In all, 185 *de novo* contigs that contained individually significant associations prior to False Discovery Rate correction (FDR) were mapped back to the *C. geophilum* isolate 1.58 reference genome in MycoCosm with 124 protein-coding regions identified (Supplemental Table 4.1). The reference-based assembly accounted for 13,963 SNP associations to measured phenotypes (Supplemental Table 4.1). In all, 137 reference-based sequences were analyzed against the *C. geophilum* 1.58 reference genome, with 102 MycoCosm hits and 88 protein-coding regions identified (Supplemental Table 4.1). After FDR correction for multiple comparisons, no significant associations were detected in any measured phenotypes in the reference-based assembly, and 20 were detected in the Cd treatment only from the *de novo* assembly (Supplemental Table 4.1).

Two significant SNP markers detected in the Cd treatment are associated with the gene coding regions of glucooligosaccharide oxidase, located in scaffold 46 from base pair 951514-951573 (*de novo* contig 860:7/C/T), and fructose-1,6-bisphosphatase, located in scaffold 2 from base pair 1257909-1257968 (*de novo* contig 3682:10/G/A, Table 5, Supplemental Table 4.1). There was a significant association to the gene coding region for the sphingolipid

catabolizing gene acid sphingomyelinase [237] found in scaffold six between base pairs 322,809 and 323,780 (*de novo* contig 33099:34/C/T, Table 5). Five significant associations were linked to proteins notated as a general function prediction only, and one located in scaffold 14 from base pair 1331013-1331072 (*de novo* contig 16995:14/T/C) was simply notated as potentially a binding protein (Table 5).

While not significant after FDR correction ($p \geq 0.05$), 14 of the *de novo* assembly markers discovered in the Cu, Pb, Na, Sr, and Zn treatments are associated with protein coding gene regions for ion binding proteins, including the cation binding protein alpha-amylase found in *C. geophilum* scaffold 50 from base pair 1308270-1308329 (Supplemental Table 4.1). Of these associations, 10 SNP markers associated with zinc ion binding protein coding gene regions were detected in the Cu, Pb, Na, Sr, and Zn treatments but not the Cd treatment (Supplemental Table 4.1). Two annotated iron-binding proteins were determined to be significant ($p < 0.05$) in the Cd treatment (Table 5). Both are found within the cytochrome P450 CYP4/CYP19/CYP26 protein subfamily in scaffold 13 starting from base pair 1829642-1829701. The first, found in *de novo* contig 2968 at base pair 46, is a substitution of a cytosine for an adenine; the second is a substitution of an adenine for a guanine at base pair 50 within the same contig (Table 5).

4.8 Discussion

Cenococcum geophilum has been proposed as a new model organism for use in laboratory studies [75], making the establishment of robust phylogenetic and genomic analyses protocols critical to its continued development in this role. While large amounts of variation in this species are present in nature that can be mined for traits of interest, the large size of the *C. geophilum* genome presents unique challenges to analyses of the role this genetic diversity plays in the functioning of this ubiquitous ectomycorrhizal species. Here we compared the results of association mapping using either a reference-based or *de novo* assembly generated from RADseq data and found that the *de novo* assembly provided a more robust analysis, with 20 significant association markers ($p < 0.05$) detected in the Cd treatment (Table 5).

Past studies of other organisms have recorded a larger number of SNPs called from reference-based assemblies, challenging our observations [124, 238]. These same studies concluded that downstream processing of RADseq data has a large impact on the final outcome of the analyses performed. To this point, the increased robustness we observed from the *de novo* assembly is likely due to the improved depth per site in the *de novo* versus reference-based RADseq assembly based on VCFtools quality filtering, as well as the likelihood of capturing transposable repeat sites in the reference-based alignment which would ultimately be designated as non-significant ($p > 0.05$) associations.

Our analyses detected 20 SNP markers from the *de novo* assembly which were significantly associated ($p < 0.05$) with the presence of Cd (Table 5) after FDR corrections. This may reflect the heightened toxicity of Cd [239] as compared to the other heavy metal conditions, as Cd appears to be the only heavy metal which produced a strong enough reaction for significant responses in the tested isolates (Supplemental Table 4.1). Interestingly and in contrast to the results on Cd, most isolates grew well (approx. 1.4 cm in 70 days) in the presence of 25 ppm Sr, with two isolates POR4_12.6 ($p = 0.02$) and LewPark3_26.2 ($p = 0.03$) significantly surpassing the average radial growth of their respective control (Fig 17). The majority of identified association markers were detected in the presence of Sr, with a total of 119 out of the 124 detected from the *de novo* assembly and 58 of the 88 detected from the reference-based assembly (Supplemental Table 4.1). None of these were determined to be significant after FDR correction. There is no singular pattern in the protein functions associated with the SNP markers detected, which range from metabolic functions to membrane transporters and ion binding proteins (Supplemental Table 4.1).

Contrasted to the observed growth in Sr treatments, many of the isolates did not grow in the presence of 0.5 ppm Cd, which seem to have proved toxic for 49 of the isolates included in our study group (Fig 1-2). Despite this toxicity, we observed an interesting previously undocumented inverse relationship in isolates which were capable of strong growth (approx. 2 cm in 70 days) in the presence of Cd but not in 25 ppm Cu and Zn, and vice-versa. In summary, isolates which grew strongly in the presence of Cd did not also grow well in the presence of Cu and Zn, and this inverse phenotype was significantly associated with 20 SNP markers. Of the detected markers, at least five are associated with gene coding regions of

proteins involved in metabolic functions, including the amidohydrolase protein superfamily involved in hydrolysis of multiple substrates [240] and a permease of the major facilitator superfamily [241] (Supplemental Table 4.1). This permease may be functionally coupled with an additionally detected SNP marker associated with an ATP binding serine/threonine protein kinase, as both protein families are necessarily linked for carbohydrate permease functionality [242, 243].

Two separately detected SNPs of a thiamine in place of a cytosine are associated with gene coding regions for proteins involved in either transcription via RNA processing and modification or translation via biogenesis and ribosomal structure (Table 5). A third replacement of a thiamine for a cytosine was detected in the nucleic protein transporter karyopherin beta 3 [244], and combined with the previous two SNP markers may indicate increased protein translation in response to Cd toxicity (Table 5). Of note, many of the discovered SNP markers are associated with gene coding regions of proteins linked to energy production, including ATP binding, carbohydrate catabolism, and lipid transport and metabolism (Table 5). The increased association with these particular proteins may reflect the significantly increased radial growth observed in the WH5 and SK5 isolates in the presence of Cd, even surpassing their respective growth under control conditions (Fig 3).

Our comparisons across treatments indicated two particularly interesting significant ($p < 0.05$) SNPs linked to the proteins glucooligosaccharide oxidase and fructose-1,6-bisphosphatase. When reviewed, these SNP markers were specifically associated with strong isolate growth (approx. 2 cm in 70 days) in the presence of Cd but not Cu or Zn (Supplemental Table 4.1). Both proteins have been previously shown to increase heavy metal resistance in other species. Glucooligosaccharide oxidase contains the flavin adenine dinucleotide (FAD) binding domain [245, 246], which is a coenzyme involved in substrate redox reactions [247]. Flavoenzymes have been shown to increase bacterial resistance to heavy metals, particularly mercury in the form of mercuric reductase, which catalyzes the highly toxic Hg^{2+} to the less-toxic Hg^0 [248, 249], but to the best of our knowledge such relationships have not yet been documented in any fungal species.

Fructose-1,6-bisphosphatase catalyzes the enzymatic conversion of fructose-1,6-bisphosphate to fructose-6-phosphate and requires the binding of both a structural and catalytic metal

ion [250]. However, Kuznetsova et al. (2010) [251] described a metal-independent family of FBPases, YK23, in the single-celled yeast *Saccharomyces cerevisiae*, which shared 31% identity with another metal-independent FBPase described by Ganapathy et al. (2015) [252] in *Mycobacterium tuberculosis*. While the FBPase identified in this study does not appear to share similarities with the YK23 FBPase family, FBPase is known to either rely directly upon or be completely independent of metal ion binding in order to function, and in at least one study is hypothesized to create lithium resistance due to this independence [252].

4.9 Conclusion

Within this study, we have shown that RADseq techniques may be used in combination with association mapping approaches to statistically detect significant SNP markers. This technique was used to implicate genes for proteins of interest within the *Cenococcum geophilum* genome associated with a measured phenotype, in this case radial growth responses to the presence of heavy metals. These results add strong support to the use of *C. geophilum* as a model species in both laboratory-based studies of its biological functioning, as well as indicate potential for the development of future agricultural applications relevant for their use with biofuel crop systems in contaminated lands. The further study and development of these approaches including isolates from the global *C. geophilum* collection would allow us to greatly increase our understanding of the significant phenotypic associations which may be unique to the Pacific Northwest isolates as well as to the *C. geophilum* species complex. Follow on work should address gene-specific manipulations using the implicated proteins of this study. Additionally, greenhouse applications using a combination of isolates from sites which are generally tolerant of multiple heavy metal contaminants in combination with biofuel crops such as *Populus* could serve to elucidate whether these isolates and the significant associations are indeed correlated to heavy metal tolerance both with regards to fungal growth and in symbiosis with the plant host.

Chapter 5

Conclusions and Future Directions

5.1 Phylogenetic complexity of *Cenococcum geophilum*

The initial study of >200 Pacific Northwest (PNW) isolates revealed the existence of multiple cryptic clades of *C. geophilum* as well as distinct phylogenetic groups from the PNW which may be uniquely associated with *Populus trichocarpa*, and confirms the common view of this species as a hyper-diverse group of ectomycorrhizal fungi on both regional and global scales. The genetic diversity present within both local and global isolate collections of *C. geophilum* isolates is striking, and strongly indicates patterns consistent with recombination within the PNW isolate collection based on analyses performed on the *GAPDH* and ITS gene regions. However, restriction-associated DNA sequencing (RADseq) approaches expand the genomic information analyzed through rigorous *de novo* single nucleotide polymorphism (SNP) calling allows for more in-depth phylogenetic analyses on this ubiquitous fungal taxon which has proven historically difficult to classify both physiologically and phylogenetically. To determine whether this represents one global, hyper-diverse species, or a myriad of cryptic species, a more robust analysis with greater population-level resolution across many loci to more accurately quantify gene flow was performed. This allowed the opportunity to delineate the individual species which may currently be included in this greater *C. geophilum* species complex, or the other mechanisms by which it may maintain such diversity.

The application of RADseq to >180 PNW *C. geophilum* isolates shows that when investigated across multiple loci in the *C. geophilum* genome, the PNW collection continues

to represent a highly diverse collection. These approaches showed when considered on a group specific basis and/or local rivershed or site basis, potential signals consistent with recombination events are observed, despite little initial evidence of gene flow at the genome-wide scale. These patterns match expected with low-level, infrequent recombination or hybridization events and potential barriers within a cryptic species. It is worth noting that based on a single gene scale, recombination is likely occurring in localized populations at specific genomic locations. This indicates that even an analysis including >180 isolates under-samples the diversity of the PNW collection, and additional isolates representing a more comprehensive regional, site, and watershed level of groupings may provide stronger localized evidence of these patterns.

The analyses of these RADseq data in an association mapping yielded 20 significant markers for further study, providing candidate isolates for future research into possible agricultural relevance in greenhouse studies and adding strong support to the use of *C. geophilum* as a model species within laboratory studies. Significant associations were detected in both the reference-based and *de novo* RADseq assemblies, with specific proteins identified which may serve important roles in *C. geophilum* environmental stressor and particularly cadmium tolerance within specific isolate collections.

5.2 Future directions

The incorporation of RADseq data from previously characterized isolates from the global *C. geophilum* collection which incorporate the PNW RADseq data in conjunction with the PNW isolate collection could provide stronger support for both localized recombination or hybridization events and increase understanding of intriguing species on a broader spectrum. Such expanded RADseq datasets designating location-based populations which are separated may reveal additional speciation and sexual recombination evidence within the global *C. geophilum* population. Future research directions for these systems and techniques could include expanded genome-wide association studies using a myriad of phenotypes including isolate color, isolation site characteristics, and responses to additional stressors to determine if the SNPs indicated using these RADseq data are co-associated with

other measurable phenotypic traits. Moreover, these analyses can be used to direct gene-specific manipulations in future research, and greenhouse applications using a combination of isolates from sites which are generally resistant to multiple heavy metal contaminants in combination with a biofuel crop such as *Populus* can serve to determine whether these isolates have potential agricultural applications in contaminated lands and whether the specified significant associations are indeed correlated to heavy metal resistance, particularly in the case of the more toxic Cd in comparison to Cu and Zn. This proof of concept for the use of RADseq techniques and applications in to detect SNPs and markers for proteins of interest within the *Cenococcum geophilum* genome opens multiple avenues of future research which will broaden our understanding of this important and fascinating ectomycorrhizal fungus.

Bibliography

- [1] Carolyn Dimitri, Anne Effland, and Neilson Conklin. The 20th century transformation of U.S. agriculture and farm policy. *Transformation*, 3(Economic Information Bulletin Number 3):17, 2005. [1](#), [50](#)
- [2] Amanda J. Bennett, Gary D. Bending, David Chandler, Sally Hilton, and Peter Mills. Meeting the demand for crop production: The challenge of yield decline in crops grown in short rotations. *Biological Reviews*, 87(1):52–71, 2012. [1](#), [51](#)
- [3] Amitava Rakshit, Purushothaman Chirakuzhyil Abhilash, Harikesh Bahadur Singh, and Subhadip Ghosh, editors. *Adaptive Soil Management : From Theory to Practices*. Springer Nature Singapore, 2017. [1](#)
- [4] M. S. Swaminathan. *Sustainable agriculture towards food security*. Springer, 1996. [1](#), [50](#)
- [5] Fabrício Ângelo Gabriel, Emmanuelle Maria Gonçalves Lorena, Ana Paula Xavier de Gondra Bezerra, Ítala Gabriela Sobral Santos, Alex Souza Moraes, and Fernando Cartaxo Rolim Neto. Pollution by Heavy Metals: Environmental Implications and Key Strategies for Remediation. *Revista Geama*, 7(1):76–85, 2016. [1](#)
- [6] Claudia Rojas, Rosemary M. Gutierrez, and Mary Ann Bruns. Bacterial and eukaryal diversity in soils forming from acid mine drainage precipitates under reclaimed vegetation and biological crusts. *Applied Soil Ecology*, 105:57–66, 2016. [1](#), [50](#)
- [7] Briana Niblick and Amy E. Landis. Assessing renewable energy potential on United States marginal and contaminated sites. *Renewable and Sustainable Energy Reviews*, 60:489–497, 2016. [xii](#), [1](#), [50](#), [102](#)
- [8] Tomohito Arao, Satoru Ishikawa, Masaharu Murakami, Kaoru Abe, Yuji Maejima, and Tomoyuki Makino. Heavy metal contamination of agricultural soil and countermeasures in Japan. *Paddy and Water Environment*, 8(3):247–257, 2010. [1](#), [50](#)

- [9] Osiel González Dávila, Juan Miguel Gómez-bernal, and Esther Aurora Ruíz-huerta. Plants and Soil Contamination with Heavy Metals in Agricultural Areas. *Environmental Contamination*, pages 37–50, 2011. [50](#)
- [10] Manoel Lago-Vila, Andrés Rodríguez-Seijo, Daniel Arenas-Lago, Luisa Andrade, and Maria Flora Alonso Vega. Heavy metal content and toxicity of mine and quarry soils. *Journal of Soils and Sediments*, 2016. [1](#), [50](#)
- [11] Marina Efremova and Alexandra Izosimova. Contamination of Agricultural Soils with Heavy Metals. *Combating Soil Degradation*, pages 250–485, 2000. [1](#)
- [12] M. Puschenreiter, O. Horak, W. Friesl, and W. Hartl. Low-cost agricultural measures to reduce heavy metal transfer into the food chain - A review. *Plant, Soil and Environment*, 51(1):1–11, 2005. [1](#)
- [13] Raymond a. Wuana and Felix E. Okieimen. Heavy Metals in Contaminated Soils: A Review of Sources, Chemistry, Risks and Best Available Strategies for Remediation. *ISRN Ecology*, 2011:1–20, 2011. [1](#), [50](#)
- [14] A K Chopra, Chakresh Pathak, and G Prasad. Scenario of heavy metal contamination in agricultural soil and its management. *Journal of Applied Natural Science*, 1(2007):99–108, 2009. [1](#)
- [15] Ye Tao Tang, Teng Hao Bo Deng, Qi Hang Wu, Shi Zhong Wang, Rong Liang Qiu, Ze Bin Wei, Xiao Fang Guo, Qi Tang Wu, Mei Lei, Tong Bin Chen, G. Echevarria, T. Sterckeman, M. O. Simonnot, and J. L. Morel. Designing Cropping Systems for Metal-Contaminated Sites: A Review. *Pedosphere*, 22(4):470–488, 2012. [xii](#), [1](#), [2](#), [103](#)
- [16] Muhammad Aamer Mehmood, Muhammad Ibrahim, Umer Rashid, Muhammad Nawaz, Shafaqat Ali, Athar Hussain, and Munazza Gull. Biomass production for bioenergy using marginal lands. *Sustainable Production and Consumption*, 9(August 2016):3–21, 2017. [2](#), [51](#)
- [17] Abin Sebastian and Majeti Narasimha Vara Prasad. Cadmium minimization in rice. A review. *Agronomy for Sustainable Development*, 34(1):155–173, 2014. [2](#)

- [18] Annette M. Fahrenkrog, Leandro G. Neves, Márcio F.R. Resende, Christopher Dervinis, Ruth Davenport, W. Brad Barbazuk, and Matias Kirst. Population genomics of the eastern cottonwood (*Populus deltoides*). *Ecology and Evolution*, 7(22), 2017. [2](#)
- [19] Stefan Jansson, Rishikesh P. Bhalerao, and Andrew T Groover. *Genetics and Genomics of Populus*, volume 8. Springer Science + Business Media, 2010. [11](#)
- [20] T. Lin, X. Wan, and F. Zhang. The short-term responses of glutathione and phytochelation synthetic pathways genes to additional nitrogen under cadmium stress in poplar leaves. *Russian Journal of Plant Physiology*, 63(6):754–762, 2016. [44](#), [45](#)
- [21] Populus Torr, N Putnam, S Ralph, S Rombauts, A Salamov, J Schein, L Sterck, and A Aerts. The Genome of Black Cottonwood *Populus trichocarpa* (Torr. & Gray). *Science*, 313(September):1596–1604, 2006. [2](#), [11](#)
- [22] Agnieszka Szuba. Ectomycorrhiza of *Populus*. *Forest Ecology and Management*, 347:156–169, 2015. [2](#), [3](#), [50](#)
- [23] Fernando Guerra, Felipe Gainza, Ramón Pérez, and Francisco Zamudio. Phytoremediation of heavy metals using poplars (*Populus* spp.): A glimpse of the plant responses to copper, cadmium and zinc stress. *Handbook of Phytoremediation*, pages 387–413, 2011. [2](#)
- [24] Zarati Houda, Zoubeir Bejaoui, Ali Albouchi, Dharmendra K. Gupta, and Francisco J. Corpas. Comparative study of plant growth of two poplar tree species irrigated with treated wastewater, with particular reference to accumulation of heavy metals (Cd, Pb, As, and Ni). *Environmental monitoring and assessment*, 188(2):99, 2016. [50](#)
- [25] David Macaya-Sanz, Jin Gui Chen, Udaya C. Kalluri, Wellington Muchero, Timothy J. Tschaplinski, Lee E. Gunter, Sandra J. Simon, Ajaya K. Biswal, Anthony C. Bryan, Raja Payyavula, Meng Xie, Yongil Yang, Jin Zhang, Debra Mohnen, Gerald A. Tuskan, and Stephen P. Difazio. Agronomic performance of *Populus deltoides* trees engineered for biofuel production. *Biotechnology for Biofuels*, 10(1), 2017. [2](#)

- [26] Ilga Porth and Yousry A. El-Kassaby. Using Populus as a lignocellulosic feedstock for bioethanol. *Biotechnology Journal*, 10(4):510–524, 2015. [2](#), [11](#)
- [27] Ivana Radojčić Redovniković, Alessandra De Marco, Chiara Proietti, Karla Hanousek, Marija Sedak, Nina Bilandžić, and Tamara Jakovljević. Poplar response to cadmium and lead soil contamination. *Ecotoxicology and Environmental Safety*, 144(May):482–489, 2017.
- [28] M. Rafati, N. Khorasani, F. Moattar, a. Shirvany, F. Moraghebi, and S. Hosseinzadeh. Phytoremediation potential of populus alba and morus alba for cadmium, chromium and nickel absorption from polluted soil. *International Journal of Environmental Research*, 5(4):961–970, 2011.
- [29] Jin Zhang, Yongil Yang, Kaijie Zheng, Meng Xie, Kai Feng, Sara S. Jawdy, Lee E. Gunter, Priya Ranjan, Vasanth R. Singan, Nancy Engle, Erika Lindquist, Kerrie Barry, Jeremy Schmutz, Nan Zhao, Timothy J. Tschaplinski, Jared LeBoldus, Gerald A. Tuskan, Jin Gui Chen, and Wellington Muchero. Genome-wide association studies and expression-based quantitative trait loci analyses reveal roles of HCT2 in caffeoylquinic acid biosynthesis and its regulation by defense-responsive transcription factors in Populus. *New Phytologist*, 220(2):502–516, 2018. [2](#), [12](#), [50](#)
- [30] Poulomi Sannigrahi, Arthur J Ragauskas, and Gerald A Tuskan. Poplar as a feedstock for biofuels: A review of compositional characteristics. *Biofuels, Bioproducts and Biorefining*, 4(2):209–226, 2012. [2](#), [11](#)
- [31] Valérie Bert, Julienne Allemon, Philippe Sajet, Sébastien Dieu, Arnaud Papin, Serge Collet, Rodolphe Gaucher, Michel Chalot, Boudewijn Michiels, and Cécile Raventos. Torrefaction and pyrolysis of metal-enriched poplars from phytotechnologies: Effect of temperature and biomass chlorine content on metal distribution in end-products and valorization options. *Biomass and Bioenergy*, 96:1–11, 2017. [2](#)
- [32] Chenting Zhang, Xun Hu, Hongyu Guo, Tao Wei, Dehua Dong, Guangzhi Hu, Song Hu, Jun Xiang, Qing Liu, and Yi Wang. Pyrolysis of poplar, cellulose and lignin:

- Effects of acidity and alkalinity of the metal oxide catalysts. *Journal of Analytical and Applied Pyrolysis*, 134(August):590–605, 2018. [2](#)
- [33] A V Bridgwater, D Meier, and D Radlein. An overview of fast pyrolysis of biomass. *Organic Chemistry*, 30:1479–1493, 1999. [xii](#), [2](#), [51](#), [104](#)
- [34] Khan Muhammad Qureshi, Andrew Ng Kay Lup, Saima Khan, Faisal Abnisa, and Wan Mohd Ashri Wan Daud. A technical review on semi-continuous and continuous pyrolysis process of biomass to bio-oil. *Journal of Analytical and Applied Pyrolysis*, 2018.
- [35] M. Stals, E. Thijssen, J. Vangronsveld, R. Carleer, S. Schreurs, and J. Yperman. Flash pyrolysis of heavy metal contaminated biomass from phytoremediation: Influence of temperature, entrained flow and wood/leaves blended pyrolysis on the behaviour of heavy metals. *Journal of Analytical and Applied Pyrolysis*, 87(1):1–7, 2010. [2](#), [51](#)
- [36] Jing Jing Zhao, Xin Jie Shen, Xavier Domene, Josep Maria Alcañiz, Xing Liao, and Cristina Palet. Comparison of biochars derived from different types of feedstock and their potential for heavy metal removal in multiple-metal solutions. *Scientific Reports*, 9(1):1–12, 2019. [2](#)
- [37] Jan V. Colpaert, Jan H L Wevers, Erik Krzmaric, and Kristin Adriaensen. How metal-tolerant ecotypes of ectomycorrhizal fungi protect plants from heavy metal pollution. *Annals of Forest Science*, 68(1):17–24, 2011. [2](#), [51](#), [52](#)
- [38] Yves Dessaux, Catherine Grandclément, and Denis Faure. Engineering the Rhizosphere. *Trends in Plant Science*, xx:1–13, 2016. [2](#), [51](#)
- [39] U Galli, H Schuepp, and C Brunold. Heavy-Metal Binding By Mycorrhizal Fungi. *Physiologia Plantarum*, 92(2):364–368, 1994. [2](#), [3](#), [51](#)
- [40] Álvarez S.P., Tapia M.A.M., and Duarte B.N.D. Fungal Bioremediation as a Tool for Polluted Agricultural Soils. In Ram Prasad, editor, *Mycoremediation and Environmental Sustainability*, chapter 1. Springer, 2017. [2](#)

- [41] Naser A Anjum, Sarvajeet Singh, and Narendra Tuteja. *Enhancing Cleanup of Environmental Pollutants*, volume 1. Springer, 2017.
- [42] Katarzyna Bandurska, Agnieszka Berdowska, and Piotr Krupa. The Protective Role of Ectomycorrhizas in Relation to *Pinus sylvestris* L. Growing in Areas Contaminated with Heavy Metals. *Polish Journal of Environmental Studies*, 29(5), 2020. [2](#)
- [43] Ian D. Pulford and C. Watson. Phytoremediation of heavy metal-contaminated land by trees - A review. *Environment International*, 29(4):529–540, 2003. [51](#)
- [44] Joachim Sell, Achim Kayser, Rainer Schulin, and Ivano Brunner. Contribution of ectomycorrhizal fungi to cadmium uptake of poplars and willows from a heavily polluted soil. *Plant and Soil*, 277(1-2):245–253, 2005. [2](#)
- [45] Geoffrey M. Gadd and Louise Rome. Biosorption of copper by fungal melanin. *Applied Microbiology and Biotechnology*, 29:610–617, 1988. [3](#), [51](#)
- [46] Doris Krpata, Ursula Peintner, Ingrid Langer, Walter J. Fitz, and Peter Schweiger. Ectomycorrhizal communities associated with *Populus tremula* growing on a heavy metal contaminated site. *Mycological Research*, 112(9):1069–1079, 2008. [3](#), [11](#), [51](#), [52](#)
- [47] Cunxian Ma, Guangshu Zhai, Huimin Wu, Izabela Kania-Korwel, Hans Joachim Lehmler, and Jerald L. Schnoor. Identification of a novel hydroxylated metabolite of 2,2,3,5,6-pentachlorobiphenyl formed in whole poplar plants. *Environmental Science and Pollution Research*, 23(3):2089–2098, 2016. [3](#)
- [48] Muhammad Riaz, Muhammad Kamran, Yizeng Fang, Qianqian Wang, Huayuan Cao, Guoling Yang, Lulu Deng, Youjuan Wang, Yaoyu Zhou, Ioannis Anastopoulos, and Xiurong Wang. Arbuscular mycorrhizal fungi-induced mitigation of heavy metal phytotoxicity in metal contaminated soils: A critical review. *Journal of Hazardous Materials*, 402(September):123919, 2021.
- [49] Kun Zong, Jian Huang, Kazuhide Nara, Yahua Chen, Zhenguo Shen, and Chunlan Lian. Inoculation of ectomycorrhizal fungi contributes to the survival of tree seedlings

- in a copper mine tailing. *Journal of Forest Research*, 20(6):493–500, 2015. [2](#), [3](#), [11](#), [30](#), [51](#)
- [50] Bram Beckers, Michiel Op De Beeck, Nele Weyens, Wout Boerjan, and Jaco Vangronsveld. Structural variability and niche differentiation in the rhizosphere and endosphere bacterial microbiome of field-grown poplar trees. *Microbiome*, 5(1):25, 2017. [2](#)
- [51] Gregory Bonito, Khalid Hameed, Rafael Ventura, Jay Krishnan, Christopher W. Schadt, and Rytas Vilgalys. Isolating a functionally relevant guild of fungi from the root microbiome of Populus. *Fungal Ecology*, 22, 2016.
- [52] M. A. Cregger, A. M. Veach, Z. K. Yang, M. J. Crouch, R. Vilgalys, G. A. Tuskan, and C. W. Schadt. The Populus holobiont: dissecting the effects of plant niches and genotype on the microbiome. *Microbiome*, 6(1):31, 2018. [2](#), [3](#), [11](#), [51](#)
- [53] Roeland L. Berendsen, Corné M.J. J Pieterse, and Peter A.H.M. H M Bakker. The rhizosphere microbiome and plant health. *Trends in plant science*, 17(8):478–486, 8 2012. [2](#), [11](#)
- [54] Sharon L. Doty, Andrew W. Sher, Neil D. Fleck, Mahsa Khorasani, Roger E. Bumgarner, Zareen Khan, Andrew W. K. Ko, Soo-Hyung Kim, and Thomas H. DeLuca. Variable Nitrogen Fixation in Wild Populus. *Plos One*, 11(5):e0155979, 2016. [51](#)
- [55] Laurence Lacercat-Didier, Charlotte Berthelot, Julie Foulon, Audrey Errard, Elena Martino, Michel Chalot, and Damien Blaudez. New mutualistic fungal endophytes isolated from poplar roots display high metal tolerance. *Mycorrhiza*, pages 1–15, 2016. [2](#), [3](#), [51](#), [52](#)
- [56] Dongfeng Long, Jianjun Liu, Qisheng Han, Xiaobing Wang, and Jian Huang. Ectomycorrhizal fungal communities associated with Populus simonii and Pinus tabuliformis in the hilly-gully region of the Loess Plateau, China. *Scientific Reports*, 6(1):24336, 2016.

- [57] Migun Shakya, Neil Gottel, Hector Castro, Zamin K Yang, Lee Gunter, Jessy Labbé, Wellington Muchero, Gregory Bonito, Rytas Vilgalys, Gerald Tuskan, Mircea Podar, and Christopher W Schadt. A multifactor analysis of fungal and bacterial community structure in the root microbiome of mature *Populus deltoides* trees. *PloS one*, 8(10):e76382, 1 2013. [2](#), [3](#)
- [58] Ramya Narendrula-Kotha and Kabwe K. Nkongolo. Bacterial and fungal resilience to long-term metal exposure in a mining region revealed by metagenomics sequencing. *Ecological Genetics and Genomics*, In Press:13–24, 2016. [2](#)
- [59] Klaus Butterbach-Bahl, Elizabeth M Baggs, Michael Dannenmann, Ralf Kiese, and Sophie Zechmeister-Boltenstern. Nitrous oxide emissions from soils: how well do we understand the processes and their controls? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368(1621):20130122, 2013. [3](#)
- [60] B. Karthikeyan, C. Abdul Jaleel, G. M A Lakshmanan, and M. Deiveekasundaram. Studies on rhizosphere microbial diversity of some commercially important medicinal plants. *Colloids and Surfaces B: Biointerfaces*, 62(1):143–145, 2008.
- [61] Susanne Schreiter, Guo-Chun Ding, Holger Heuer, GÄnter Neumann, Martin Sandmann, Rita Grosch, Siegfried Kropf, and Kornelia Smalla. Effect of the soil type on the microbiome in the rhizosphere of field-grown lettuce. *Frontiers in Microbiology*, 5(April):1–13, 2014.
- [62] M. Toro, R. Azcón, and J. M. Barea. The use of isotopic dilution techniques to evaluate the interactive effects of *Rhizobium* genotype, mycorrhizal fungi, phosphate-solubilizing rhizobacteria and rock phosphate on nitrogen and phosphorus acquisition *Medicago sativa*. *New Phytologist*, 138:265–273, 1998. [3](#)
- [63] Harsh P Bais, Tiffany L Weir, Laura G Perry, Simon Gilroy, and Jorge M Vivanco. The role of root exudates in rhizosphere interactions with plants and other organisms. *Annual review of plant biology*, 57(July):233–266, 2006. [3](#)

- [64] D. S. Jenkinson, H. C. Harris, J. Ryan, A. M. McNeill, C. J. Pilbeam, and K. Coleman. Organic matter turnover in a calcareous clay soil from Syria under a two-course cereal rotation. *Soil Biology and Biochemistry*, 31(5):687–693, 1999.
- [65] P. P. Motavalli, R. J. Kremer, M. Fang, and N. E. Means. Impact of genetically modified crops and their management on soil microbially mediated plant nutrient transformations. *Journal of environmental quality*, 33(3):816–824, 2005.
- [66] Günter Neumann, Timothy S. George, and Claude Plassard. Strategies and methods for studying the rhizosphere-the plant science toolbox. *Plant and Soil*, 321(1-2):431–456, 2009.
- [67] Jos M. Raaijmakers, Timothy C. Paulitz, Christian Steinberg, Claude Alabouvette, and Yvan Moënne-nne Looco. The rhizosphere: A playground and battlefield for soilborne pathogens and beneficial microorganisms. *Plant and Soil*, 321(1-2):341–361, 2009.
- [68] Michael Wyrebek, Cristina Huber, Ramanpreet Kaur Sasan, and Michael J. Bidochka. Three sympatrically occurring species of *Metarhizium* show plant rhizosphere specificity. *Microbiology*, 157:2904–2911, 2011. 3
- [69] S D Garrett. Ecological Groups of Soil Fungi: A Survey of Substrate Relationships. *New Phytologist*, 50(2):149–166, 1951. 3
- [70] G. T. Hill, N. A. Mitkowski, L. Aldrich-Wolfe, L. R. Emele, D. D. Jurkonie, A. Ficke, S. Maldonado-Ramirez, S. T. Lynch, and E. B. Nelson. Methods for assessing the composition and diversity of soil microbial communities. *Applied Soil Ecology*, 15(1):25–36, 2000.
- [71] Muhammad Ikram, Niaz Ali, Gul Jan, Farzana Gul Jan, Inayat Ur Rahman, Amjad Iqbal, and Muhammad Hamayun. IAA producing fungal endophyte *Penicillium roqueforti* Thom., enhances stress tolerance and nutrients uptake in wheat plants grown on heavy metal contaminated soils. *PLoS ONE*, 13(11):1–22, 2018.

- [72] Jonathan M. Plett, Hengfu Yin, Ritesh Mewalal, Rongbin Hu, Ting Li, Priya Ranjan, Sara Jawdy, Henrique C. De Paoli, George Butler, Tessa Maureen Burch-Smith, Hao-Bo Guo, Chun Ju Chen, Annegret Kohler, Ian C. Anderson, Jessy L. Labbé, Francis Martin, Gerald A. Tuskan, and Xiaohan Yang. *Populus trichocarpa* encodes small, effector-like secreted proteins that are highly induced during mutualistic symbiosis. *Scientific Reports*, 7(1):382, 2017. [3](#)
- [73] Shahid Mahmood, Roger D. Finlay, Susanne Erland, and Håkan Wallander. Solubilisation and colonisation of wood ash by ectomycorrhizal fungi isolated from a wood ash fertilised spruce forest. *FEMS Microbiology Ecology*, 35(2):151–161, 2001. [3](#), [51](#)
- [74] Vijay Singh Meena, Pankaj Kumar Mishra, Jaideep Kumar Bisht, and Arunava Pattanayak. *Agriculturally Important Microbes for Sustainable Agriculture*, volume I. Springer Nature Singapore, 2017. [11](#), [51](#)
- [75] Keisuke Obase, Greg W Douhan, Yosuke Matsuda, and Matthew E Smith. Progress and Challenges in Understanding the Biology, Diversity, and Biogeography of *Cenococcum geophilum*. In *Biogeography of Mycorrhizal Symbiosis*, volume 230, pages 299–317. Springer International Publishing, 2017. [3](#), [4](#), [6](#), [7](#), [12](#), [23](#), [30](#), [31](#), [32](#), [42](#), [51](#), [52](#), [57](#)
- [76] J. W. Spatafora, C. a. Owensby, G. W. Douhan, E. W. a. Boehm, and C. L. Schoch. Phylogenetic placement of the ectomycorrhizal genus *Cenococcum* in Gloniaceae (Dothideomycetes). *Mycologia*, 104(3):758–765, 2012. [3](#), [4](#), [44](#)
- [77] Greg W. Douhan, Karyn L. Huryn, LeAnn I. Douhan, Source Mycologia, No Nov Dec, Greg W. Douhan, Karyn L. Huryn, and LeAnn I. Douhan. Significant diversity and potential problems associated with inferring population structure within the *Cenococcum geophilum* species complex. *Mycologia*, 99(6):812–819, 2007. [3](#), [22](#), [23](#), [25](#), [30](#), [31](#), [51](#), [52](#)

- [78] Yosuke Matsuda, Mai Yamakawa, Tomomi Inaba, Keisuke Obase, and Shin ichiro Ito. Intraspecific variation in mycelial growth of *Cenococcum geophilum* isolates in response to salinity gradients. *Mycoscience*, 58(5):369–377, 2017. [3](#), [11](#), [30](#), [51](#)
- [79] Christopher W. Fernandez and Roger T. Koide. The function of melanin in the ectomycorrhizal fungus *Cenococcum geophilum* under water stress. *Fungal Ecology*, 6(6):479–486, 2013. [3](#), [11](#), [30](#), [51](#)
- [80] Paul C. F. Tam. Heavy-Metal Tolerance By Ectomycorrhizal Fungi and Metal Amelioration By *Pisolithus-Tinctorius*. *Mycorrhiza*, 5(3):181–187, 1995. [3](#)
- [81] Radames J.B. Cordero and Arturo Casadevall. Functions of fungal melanin beyond virulence. *Fungal Biology Reviews*, pages 1–14, 2017. [3](#), [11](#), [30](#), [51](#)
- [82] Andrea Vanesa Toledo, Mario Emilio Ernesto Franco, Silvina Marianela Yanil Lopez, María Inés Troncozo, Mario Carlos Nazareno Saparrat, and Pedro Alberto Balatti. Melanins in fungi: Types, localization and putative biological roles. *Physiological and Molecular Plant Pathology*, 2017. [3](#), [11](#), [51](#)
- [83] Elizabeth C. Bourne, Diogo Mina, Susana C. Gonçalves, João Loureiro, Helena Freitas, and Ludo A H Muller. Large and variable genome size unrelated to serpentine adaptation but supportive of cryptic sexuality in *Cenococcum geophilum*. *Mycorrhiza*, 24(1):13–20, 2014. [3](#), [12](#), [23](#), [31](#), [44](#)
- [84] Martina Peter, Annegret Kohler, Robin A. Ohm, Alan Kuo, Jennifer Krützmann, Emmanuelle Morin, Matthias Arend, Kerrie W. Barry, Manfred Binder, Cindy Choi, Alicia Clum, Alex Copeland, Nadine Grisel, Sajeet Haridas, Tabea Kipfer, Kurt LaButti, Erika Lindquist, Anna Lipzen, Renaud Maire, Barbara Meier, Sirma Mihaltcheva, Virginie Molinier, Claude Murat, Stefanie Pöggeler, C. Alisha Quandt, Christoph Sperisen, Andrew Tritt, Emilie Tisserant, Pedro W. Crous, Bernard Henrissat, Uwe Nehls, Simon Egli, Joseph W. Spatafora, Igor V. Grigoriev, and Francis M. Martin. Ectomycorrhizal ecology is imprinted in the genome of the dominant symbiotic fungus *Cenococcum geophilum*. *Nature Communications*, 7:12662, 2016. [4](#), [6](#), [11](#), [12](#), [30](#), [31](#), [33](#), [36](#), [42](#), [44](#), [51](#), [52](#), [55](#)

- [85] Pedro Talhinhos, Daniela Tavares, Ana Paula Ramos, Susana Gonçalves, and João Loureiro. Validation of standards suitable for genome size estimation of fungi. *Journal of Microbiological Methods*, 142(September):76–78, 2017. [3](#), [12](#)
- [86] Greg W. Douhan, Darren P Martin, and Dave M Rizzo. Using the putative asexual fungus *Cenococcum geophilum* as a model to test how species concepts influence recombination analyses using sequence data from multiple loci. *Current genetics*, 52(5-6):191–201, 2007. [4](#), [11](#), [12](#), [13](#), [23](#), [31](#), [44](#), [45](#)
- [87] Jean-luc Jany, Jean Garbaye, and Francis Martin. *Cenococcum geophilum* populations show a high degree of genetic diversity in beech forests. *New Phytologist*, 154(3):651–659, 2002. [4](#), [12](#), [52](#)
- [88] Mohammad Bahram, Sergei Põlme, Urmas Kõljalg, and Leho Tedersoo. A single European aspen (*Populus tremula*) tree individual may potentially harbour dozens of *Cenococcum geophilum* ITS genotypes and hundreds of species of ectomycorrhizal fungi. *FEMS Microbiology Ecology*, 75(2):313–320, 2011. [4](#), [12](#)
- [89] Yosuke Matsuda, Kosuke Takeuchi, Keisuke Obase, and Shin-ichiro Ito. Spatial distribution and genetic structure of *Cenococcum geophilum* in coastal pine forests in Japan. *FEMS Microbiology Ecology*, 91(10):fiv108, 2015. [4](#), [11](#), [12](#), [13](#), [22](#), [24](#), [25](#), [31](#)
- [90] G. W. Douhan and D. M. Rizzo. Phylogenetic divergence in a local population of the ectomycorrhizal fungus *Cenococcum geophilum*. *New Phytologist*, 166(1):263–271, 2005. [4](#), [12](#), [22](#), [25](#), [31](#), [52](#)
- [91] Keisuke Obase, Greg W. Douhan, Yosuke Matsuda, and Matthew E. Smith. Revisiting phylogenetic diversity and cryptic species of *Cenococcum geophilum* sensu lato. *Mycorrhiza*, 26(6):529–540, 3 2016. [4](#), [5](#), [11](#), [12](#), [16](#), [22](#), [31](#), [32](#), [33](#)
- [92] Maíra de Freitas Pereira, Claire Veneault-Fourrey, Patrice Vion, Frédéric Guinet, Emmanuelle Morin, Kerrie W. Barry, Anna Lipzen, Vasanth Singan, Stephanie Pfister, Hyunsoo Na, Megan Kennedy, Simon Egli, Igor Grigoriev, Francis Martin, Annegret

- Kohler, and Martina Peter. Secretome Analysis from the Ectomycorrhizal Ascomycete *Cenococcum geophilum*. *Frontiers in Microbiology*, 9(February):1–17, 2018. [11](#), [13](#), [22](#), [25](#), [26](#), [31](#)
- [93] Keisuke Obase, Greg W. Douhan, Yosuke Matsuda, and Matthew E. Smith. Isolation source matters: Sclerotia and ectomycorrhizal roots provide different views of genetic diversity in *cenococcum geophilum*. *Mycologia*, 110(3):473–481, 2018. [4](#)
- [94] Nicholas A Bokulich and David A Mills. Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. *Applied and environmental microbiology*, 79(8):2519–26, 4 2013. [4](#), [15](#)
- [95] Kendall J Martin and Paul T Rygielwicz. Fungal-specific PCR primers developed for analysis of the ITS region of environmental DNA extracts. *BMC microbiology*, 5:28, 5 2005.
- [96] Mari L Shinohara; Katherine F LoBuglio; Scott O Rogers. Comparison of ribosomal DNA ITS regions among geographic isolates of *Cenococcum geophilum*. *Curr Genet*, 35:527–535, 1999.
- [97] Conrad L. Schoch, Keith a. Seifert, S. Huhndorf, V. Robert, John L. Spouge, C. a. Levesque, W. Chen, Fungal Barcoding Consortium, E. Bolchacova, Kerstin Voigt, P. W. Crous, a. N. Miller, M. J. Wingfield, M. C. Aime, K.-D. An, F.-Y. Bai, R. W. Barreto, D. Begerow, M.-J. Bergeron, M. Blackwell, T. Boekhout, M. Bogale, N. Boonyuen, a. R. Burgaz, B. Buyck, L. Cai, Qing Cai, Gianluigi Cardinali, P. Chaverri, B. J. Coppins, Ana Crespo, P. Cubas, C. Cummings, U. Damm, Z. Wilhelm De Beer, G. S. de Hoog, R. Del-Prado, B. Dentinger, J. Dieguez-Uribeondo, P. K. Divakar, B. Douglas, M. Duenas, T. a. Duong, U. Eberhardt, J. E. Edwards, M. S. Elshahed, K. Fliegerova, M. Furtado, M. a. Garcia, Z.-W. Ge, G. W. Griffith, K. Griffiths, J. Z. Groenewald, M. Groenewald, M. Grube, M. Gryzenhout, L.-D. Guo, F. Hagen, S. Hambleton, Richard C. Hamelin, Karen Hansen, P. Harrold, G. Heller, C. Herrera, K. Hirayama, Y. Hirooka, H.-M. Ho, K. Hoffmann, V. Hofstetter, F. Hognabba, P. M. Hollingsworth, S.-B. Seung-Beom S.-B. Seung-Beom Hong,

- K. Hosaka, J. Houbraken, Karen Hughes, S. Huhtinen, Kevin D. Hyde, T. James, E. M. Johnson, J. E. Johnson, P. R. Johnston, E. B. G. Jones, L. J. Kelly, P. M. Kirk, D. G. Knapp, U. Koljalg, G. M. Kovacs, C. P. Kurtzman, S. Landvik, S. D. Leavitt, a. S. Liggenstoffer, K. Liimatainen, L. Lombard, J. J. Luangsa-ard, H. T. Lumbsch, H. Maganti, Sajeewa S. N. Maharachchikumbura, M. P. Martin, T. W. May, a. R. McTaggart, a. S. Methven, W. Meyer, J.-M. Moncalvo, S. Mongkolsamrit, L. G. Nagy, R. H. Nilsson, T. Niskanen, I. Nyilasi, G. Okada, I. Okane, I. Olariaga, J. Otte, T. Papp, D. Park, Tamás Petkovits, R. Pino-Bodas, W. Quaedvlieg, H. a. Raja, D. Redecker, T. L. Rintoul, C. Ruibal, J. M. Sarmiento-Ramirez, Imke Schmitt, A. Schussler, C. Shearer, K. Sotome, F. O. P. Stefani, S. Stenroos, B. Stielow, H. Stockinger, S. Suetrong, S.-O. Suh, G.-H. Sung, M. Suzuki, K. Tanaka, L. Tedersoo, M. Teresa Telleria, E. Tretter, Wendy a. Untereiner, H. Urbina, C. Vagvolgyi, A. Vialle, T. D. Vu, G. Walther, Q.-M. Wang, Y. Wang, B. S. Weir, M. Weiss, Merlin M. White, J. Xu, R. Yahr, Zhu L. Yang, A. Yurkov, J.-C. Zamora, Ning Zhang, Wen-Ying W.-Y. Zhuang, and David Schindel. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America*, 109(16):1–6, 2012. [24](#)
- [98] Ying Zhao, Ying Zhao, Chi Ching Tsang, Meng Xiao, Jingwei Cheng, Yingchun Xu, Susanna K.P. Lau, and Patrick C.Y. Woo. Intra-genomic internal transcribed spacer region sequence heterogeneity and molecular diagnosis in clinical microbiology. *International Journal of Molecular Sciences*, 16(10):25067–25079, 2015. [4](#)
- [99] Gábor M. Kovács, Tünde Jankovics, and Levente Kiss. Variation in the nrDNA ITS sequences of some powdery mildew species: Do routine molecular identification procedures hide valuable information? *European Journal of Plant Pathology*, 131(1):135–141, 2011. [4](#), [24](#)
- [100] Uwe K. Simon and Michael Weiß. Intragenomic variation of fungal ribosomal genes is higher than previously thought. *Molecular Biology and Evolution*, 25(11):2251–2254, 2008. [4](#), [24](#)

- [101] Michael S. Blouin. Molecular prospecting for cryptic species of nematodes: Mitochondrial DNA versus internal transcribed spacer. *International Journal for Parasitology*, 32(5):527–531, 2002. [4](#), [24](#)
- [102] Lisa A. Stevenson, Robin B. Gasser, and Neil B. Chilton. The ITS-2 rDNA of *Teladorsagia circumcincta*, *T. trifurcata* and *T. davtiani* (Nematoda: Trichostrongylidae) indicates that these taxa are one species. *International Journal for Parasitology*, 26(10):1123–1126, 1996. [4](#), [24](#)
- [103] Keith A. Seifert, Brenda D. Wingfield, and Michael J. Wingfield. A critique of DNA sequence analysis in the taxonomy of filamentous Ascomycetes and ascomycetous anamorphs. *Canadian Journal of Botany*, 73(S1):760–767, 1995. [4](#), [24](#)
- [104] Michael A. Sirover. New insights into an old protein: The functional diversity of mammalian glyceraldehyde-3-phosphate dehydrogenase. *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, 1432(2):159–184, 1999. [5](#)
- [105] Martin C. Harmsen, Frank H.J. Schuren, Serge M. Moukha, Carin M. van Zuilen, Peter J. Punt, and Joseph G.H. Wessels. Sequence analysis of the glyceraldehyde-3-phosphate dehydrogenase genes from the basidiomycetes *Schizophyllum commune*, *Phanerochaete chrysosporium* and *Agaricus bisporus*. *Current Genetics*, 22(6):447–454, 1992. [5](#)
- [106] Juliana O. Lima, Jorge F. Pereira, Johana Rincones, Joan G. Barau, Elza F. Araújo, Gonçalo A.G. Pereira, and Marisa V. Queiroz. The glyceraldehyde-3-phosphate dehydrogenase gene of *Moniliophthora perniciosa*, the causal agent of witches’ broom disease of *Theobroma cacao*. *Genetics and Molecular Biology*, 32(2):362–366, 2009.
- [107] Bertrand Neveu, François Belzile, and Richard R. Bélanger. Cloning of the glyceraldehyde-3-phosphate dehydrogenase gene from *Pseudozyma flocculosa* and functionality of its promoter in two *Pseudozyma* species. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 92(2):245–255, 2007.

- [108] Peter J. Punt, Maria A. Dingemans, Anneke Kuyvenhoven, Ronald D.M. Soede, Peter H. Pouwels, and Cees A.M.J.J. van den Hondel. Functional elements in the promoter region of the *Aspergillus nidulans* *gpdA* gene encoding glyceraldehyde-3-phosphate dehydrogenase. *Gene*, 93(1):101–109, 1990. [5](#)
- [109] D. L. Swofford. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*., 2003. [5](#), [18](#), [19](#), [23](#)
- [110] Kimberly R. Andrews, Jeffrey M. Good, Michael R. Miller, Gordon Luikart, and Paul A. Hohenlohe. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17(2):81–92, 2016. [5](#), [32](#)
- [111] Julian M. Catchen, Paul A. Hohenlohe, Louis Bernatchez, W. Chris Funk, Kimberly R. Andrews, and Fred W. Allendorf. Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources*, 17(3):362–365, 2017.
- [112] Felix Grewe, Jen Pen Huang, Steven D. Leavitt, and H. Thorsten Lumbsch. Reference-based RADseq resolves robust relationships among closely related species of lichen-forming fungi using metagenomic DNA. *Scientific Reports*, 7(1):1–11, 2017. [5](#), [6](#), [32](#), [41](#)
- [113] Florianne Marandel, Grégory Charrier, Jean Baptiste Lamy, Sabrina Le Cam, Pascal Lorange, and Verena M. Trenkel. Estimating effective population size using RADseq: Effects of SNP selection and sample size. *Ecology and Evolution*, 2020.
- [114] Rob Massatti, Anton A. Reznicek, and L. Lacey Knowles. Utilizing RADseq data for phylogenetic analysis of challenging taxonomic groups: A case study in *Carex* sect. *Racemosae*. *American Journal of Botany*, 103(2):337–347, 2016.
- [115] Garrett J. McKinney, Wesley A. Larson, Lisa W. Seeb, and James E. Seeb. RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al. (2016), 5 2017. [5](#), [46](#)

- [116] Natalia J. Bayona-Vásquez, Travis C. Glenn, Troy J. Kieran, Todd W. Pierson, Sandra L. Hoffberg, Peter A. Scott, Kerin E. Bentley, John W. Finger, Swarnali Louha, Nicholas Troendle, Pindaro Diaz-Jaimes, Rodney Mauricio, and Brant C. Faircloth. Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *PeerJ*, 2019(10):1–25, 2019. [xii](#), [5](#), [6](#), [27](#), [32](#), [34](#), [52](#), [105](#)
- [117] Jeffrey M DaCosta and Michael D Sorenson. ddRAD-seq phylogenetics based on nucleotide, indel, and presence-absence polymorphisms: Analyses of two avian genera with contrasting histories. *Molecular phylogenetics and evolution*, 94(Pt A):122–35, 1 2016. [32](#)
- [118] Jahn Davik, Daniel James Sargent, May Bente Brurberg, Sigbjørn Lien, Matthew Kent, and Muath Alsheikh. A ddRAD Based Linkage Map of the Cultivated Strawberry, *Fragaria xananassa*. *PloS one*, 10(9):e0137746, 1 2015.
- [119] Robert J Toonen, Jonathan B Puritz, Zac H Forsman, Jonathan L Whitney, Iria Fernandez-Silva, Kimberly R Andrews, and Christopher E Bird. ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ*, 1:e203, 2013. [5](#), [32](#), [52](#)
- [120] Felix Grewe, Elisa Lagostina, Huini Wu, Christian Printzen, and H. Thorsten Lumbsch. Population genomic analyses of RAD sequences resolves the phylogenetic relationship of the lichen-forming fungal species *Usnea antarctica* and *Usnea aurantiacoatra*. *MycoKeys*, 43:91–113, 2018. [6](#), [32](#), [41](#)
- [121] Rodolfo Salas-Lizana and Ryoko Oono. Double-digest RADseq loci using standard Illumina indexes improve deep and shallow phylogenetic resolution of *Lophodermium*, a widespread fungal endophyte of pine needles. *Ecology and Evolution*, 8(13):6638–6651, 2018.
- [122] Firas Talas and Bruce A. McDonald. Genome-wide analysis of *Fusarium graminearum* field populations reveals hotspots of recombination. *BMC Genomics*, 16(1):1–12, 2015. [6](#), [7](#)

- [123] Brant K Peterson, Jesse N Weber, Emily H Kay, Heidi S Fisher, and Hopi E Hoekstra. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS one*, 7(5):e37135, 1 2012. [6](#), [52](#)
- [124] Aaron B.A. Shafer, Claire R. Peart, Sergio Tusso, Inbar Maayan, Alan Brelsford, Christopher W. Wheat, and Jochen B.W. Wolf. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, 8(8), 2017. [6](#), [32](#), [58](#)
- [125] Santiago Herrera and Timothy M. Shank. RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. *Molecular Phylogenetics and Evolution*, 100:70–79, 2016. [6](#)
- [126] Andrew W. Wilson, Norman J. Wickett, Paul Grabowski, Jeremie Fant, Justin Borevitz, and Gregory M. Mueller. Examining the efficacy of a genotyping-by-sequencing technique for population genetic analysis of the mushroom *Laccaria bicolor* and evaluating whether a reference genome is necessary to assess homology. *Mycologia*, 107(1):217–226, 2015. [6](#), [31](#)
- [127] Matthew Combs, Emily E. Puckett, Jonathan Richardson, Destiny Mims, and Jason Munshi-South. Spatial population genomics of the brown rat (*Rattus norvegicus*) in New York City. *Molecular Ecology*, 27(1):83–98, 2018. [6](#)
- [128] Catherine S. McFadden, Roxanne Haverkort-Yeh, Alexandra M. Reynolds, Anna Halász, Andrea M. Quattrini, Zac H. Forsman, Yehuda Benayahu, and Robert J. Toonen. Species boundaries in the absence of morphological, ecological or geographical differentiation in the Red Sea octocoral genus *Ovabunda* (Alcyonacea: Xeniidae). *Molecular Phylogenetics and Evolution*, 112:174–184, 2017.
- [129] Oihana Razkin, Gontran Sonet, Karin Breugelmans, María José Madeira, Benjamín Juan Gómez-Moliner, and Thierry Backeljau. Species limits, interspecific hybridization and phylogeny in the cryptic land snail complex *Pyramidula*: The power of RADseq data. *Molecular Phylogenetics and Evolution*, 101:267–278, 2016. [6](#)

- [130] Deren A.R. Eaton and Isaac Overcast. Ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, 36(8):2592–2594, 2020. [6](#), [36](#)
- [131] Julian M. Catchen, Angel Amores, Paul Hohenlohe, William Cresko, and John H. Postlethwait. Stacks: Building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, 1(3):171–182, 2011. [6](#), [37](#)
- [132] Julian Catchen, Paul A Hohenlohe, Susan Bassham, Angel Amores, and William A Cresko. Stacks: an analysis tool set for population genomics. *Molecular ecology*, 22(11):3124–40, 2013.
- [133] Nicolas C. Rochette, Angel G. Rivera-Colón, and Julian M. Catchen. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, 28(21):4737–4754, 2019. [6](#), [37](#)
- [134] Arpita Konar, Olivia Choudhury, Rebecca Bullis, Lauren Fiedler, Jacqueline M. Kruser, Melissa T. Stephens, Oliver Gailing, Scott Schlarbaum, Mark V. Coggeshall, Margaret E. Staton, John E. Carlson, Scott Emrich, and Jeanne Romero-Severson. High-quality genetic mapping with ddRADseq in the non-model tree *Quercus rubra*. *BMC Genomics*, 18(1):1–12, 2017. [6](#)
- [135] Davoud Torkamaneh, Jérôme Laroche, and François Belzile. Genome-wide SNP calling from genotyping by sequencing (GBS) data: A comparison of seven pipelines and two sequencing technologies. *PLoS ONE*, 11(8):1–14, 2016. [7](#)
- [136] Gregory Bonito, Hannah Reynolds, Michael S. Robeson, Jessica Nelson, Brendan P. Hodkinson, Gerald Tuskan, Christopher W. Schadt, and Rytas Vilgalys. Plant host and soil origin influence fungal and bacterial assemblages in the roots of woody plants. *Molecular Ecology*, 23(13):3356–3370, 2014. [10](#), [11](#)
- [137] S J Grayston, S Q Wang, C D Campbell, and A C Edwards. Selective influence of plant species on microbial diversity in the rhizosphere, 1998.

- [138] Peter G Kennedy, Logan M Higgins, Rachel H Rogers, and Marjorie G Weber. Colonization-competition tradeoffs as a mechanism driving successional dynamics in ectomycorrhizal fungal communities. *PloS one*, 6(9):e25126, 2011.
- [139] Xavier Raynaud, Benoît Jaillard, and Paul W. Leadley. Plants May Alter Competition by Modifying Nutrient Bioavailability in Rhizosphere: A Modeling Approach. *The American Naturalist*, 171(1):44–58, 2008. [10](#)
- [140] G. A. Tuskan, S. P. Difazio, and T. Teichmann. Poplar Genomics is Getting Popular: The Impact of the Poplar Genome Project on Tree Research. *Plant Biology*, 6(1):2–4, 2004. [11](#), [50](#)
- [141] Stéphane Hacquard and Christopher W. Schadt. Towards a holistic understanding of the beneficial interactions across the Populus microbiome. *New Phytologist*, 205(4):1424–1430, 2015. [11](#)
- [142] Keisuke Obase, Greg W Douhan, Yosuke Matsuda, and Matthew E Smith. Culturable fungal assemblages growing within Cenococcum sclerotia in forest soils. *FEMS microbiology ecology*, 90:708–717, 2014. [11](#), [12](#), [13](#), [14](#), [30](#)
- [143] James M Trappe. Studies on Cenococcum graniforme. I. an efficient method for isolation from sclerotia. *Can. J. Bot.*, 47:1389–1390, 1969. [11](#), [30](#)
- [144] Katherine F. LoBuglio and John W. Taylor. Recombination and genetic differentiation in the mycorrhizal fungus Cenococcum geophilum Fr. *Mycologia*, 94(5):772–780, 2002. [12](#), [23](#), [31](#), [44](#)
- [145] Roba Bdeir, Wellington Muchero, Yordan Yordanov, Gerald A. Tuskan, Victor Busov, and Oliver Gailing. Genome-wide association studies of bark texture in Populus trichocarpa. *Tree Genetics and Genomes*, 15(1), 2019. [12](#)
- [146] Luke M Evans, Gancho T Slavov, Eli Rodgers-Melnick, Joel Martin, Priya Ranjan, Wellington Muchero, Amy M Brunner, Wendy Schackwitz, Lee Gunter, Jin-Gui Chen, Gerald a Tuskan, and Stephen P DiFazio. Population genomics of Populus trichocarpa

- identifies signatures of selection and adaptive trait associations. *Nature Genetics*, advance on(10):1089–1096, 2014. [13](#)
- [147] Juan Franco-Coronado, Kerrie Barry, Priya Ranjan, Gerald A. Tuskan, Yongil Yang, Jin-Gui Chen, Kelsey L. Sondreli, Jeong-Yeh Yang, Breeanna R. Urbanowicz, Erika Lindquist, Wellington Muchero, Jeff H. Chang, Jared M. LeBoldus, Sara Jawdy, Jeremy Schmutz, Robert S. Brueggeman, Jin Zhang, Vasanth Singan, Alexandra J. Weisberg, Nivi Abraham, and Kelley W. Moremen. Association mapping, transcriptomics, and transient expression identify candidate genes mediating plant–pathogen interactions in a tree. *Proceedings of the National Academy of Sciences*, 115(45):11573–11578, 2018.
- [148] Athena D. Mckown, Jaroslav Klápště, Robert D. Guy, Armando Geraldes, Ilga Porth, Jan Hannemann, Michael Friedmann, Wellington Muchero, Gerald A. Tuskan, Jürgen Ehrling, Quentin C.B. Cronk, Yousry A. El-Kassaby, Shawn D. Mansfield, and Carl J. Douglas. Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*. *New Phytologist*, 203(2):535–553, 2014.
- [149] Athena D. McKown, Jaroslav Klápště, Robert D. Guy, Raju Y. Soolanayakanahally, Jonathan La Mantia, Ilga Porth, Oleksandr Skyba, Faride Unda, Carl J. Douglas, Yousry A. El-Kassaby, Richard C. Hamelin, Shawn D. Mansfield, and Quentin C.B. Cronk. Sexual homomorphism in dioecious trees: Extensive tests fail to detect sexual dimorphism in *Populus*. *Scientific Reports*, 7(1):1–14, 2017.
- [150] Wellington Muchero, Ranjan Pryia, Gancho Slavov, Steve DiFazio, Wendy Schackwitz, Joel Martin, Michael Studer, Charles Wyman, Mark Davis, Robert Sykes, Daniel Rokhsar, and Gerald Tuskan. *Populus* resequencing: towards genome-wide association studies. *BMC Proceedings*, 5(S7):I21, 2011.
- [151] Gerald A. Tuskan, Ritesh Mewalal, Lee E. Gunter, Kaitlin J. Palla, Kelsey Carter, Daniel A. Jacobson, Piet C. Jones, Benjamin J. Garcia, Deborah A. Weighill, Philip D. Hyatt, Yongil Yang, Jin Zhang, Nicholas Reis, Jin Gui Chen, and Wellington Muchero.

- Defining the genetic components of callus formation: A GWAS approach. *PLoS ONE*, 13(8):1–18, 2018. [12](#)
- [152] D. H. Marx. The influence of ectotrophic mycorrhizal fungi on the resistance of pine roots to pathogenic infections. II. Production, identification, and biological activity of antibiotics produced by *Leucopaxillus cerealis* var. *piceina*. *Phytopathology*, 59(4):411–417, 4 1969. [14](#), [33](#), [53](#)
- [153] Laurel A. Kluber, Sarah R. Carrino-Kyker, Kaitlin P. Coyle, Jared L. DeForest, Charlotte R. Hewins, Alanna N. Shaw, Kurt A. Smemo, and David J. Burke. Mycorrhizal Response to Experimental pH and P Manipulation in Acidic Hardwood Forests. *PLoS ONE*, 7(11):1–10, 2012. [14](#)
- [154] M. L. Berbee, Mona Pirseyedi, S Hubbard, and Mona Pirseyedi. Cochliobolus Phylogenetics and the Origin of Known, Highly Virulent Pathogens, Inferred from ITS and Glyceraldehyde-3-Phosphate Dehydrogenase Gene Sequences. *Mycologia*, 91(6):964–977, 1999. [15](#)
- [155] Matthew Kearse, Richard Moir, Amy Wilson, Steven Stones-Havas, Matthew Cheung, Shane Sturrock, Simon Buxton, Alex Cooper, Sidney Markowitz, Chris Duran, Tobias Thierer, Bruce Ashton, Peter Meintjes, and Alexei Drummond. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12):1647–1649, 2012. [15](#), [38](#)
- [156] R Core Team. R: A language and environment for statistical computing, 2017. [16](#), [37](#), [54](#)
- [157] M. C. Freitas and A. S. Nobre. Bioaccumulation of heavy metals using *Parmelia sulcata* and *parmelia caperata* for air pollution studies. *Journal of Radioanalytical and Nuclear Chemistry*, 217(1):17–20, 1997. [16](#)
- [158] Igor V. Grigoriev, Roman Nikitin, Sajeet Haridas, Alan Kuo, Robin Ohm, Robert Otilar, Robert Riley, Asaf Salamov, Xueling Zhao, Frank Korzeniewski, Tatyana Smirnova, Henrik Nordberg, Inna Dubchak, and Igor Shabalov. MycoCosm portal:

- Gearing up for 1000 fungal genomes. *Nucleic Acids Research*, 42(D1):699–704, 2014. [16](#), [31](#), [55](#)
- [159] Igor V. Grigoriev, Henrik Nordberg, Igor Shabalov, Andrea Aerts, Mike Michael Cantor, David Goodstein, Alan Kuo, Simon Minovitsky, Roman Nikitin, Robin A. Ohm, Robert Otiillar, Alexander Poliakov, Igor Ratnere, Robert Riley, Tatyana Smirnova, Daniel Rokhsar, Inna Dubchak, Serge Dusheyko, Susan Hua, Alexander Poliakov, Igor Shabalov, Tatyana Smirnova, Igor V. Grigoriev, and Inna Dubchak. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Research*, 42(D1):26–32, 2014. [16](#), [55](#)
- [160] Sudhir Kumar, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, 35(6):1547–1549, 2018. [16](#), [18](#), [38](#)
- [161] H. Akaike. Information theory and an extension of the maximum likelihood principle. In B N Petrov and F Csaki, editors, *Second International Symposium on Information Theory*, pages 267–281, Budapest, 1973. Akadémiai Kiado. [16](#), [24](#)
- [162] Gideon Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, 1978. [16](#), [24](#)
- [163] John P. Huelsenbeck and Fredrik Ronquist. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8):754–755, 2001. [17](#), [18](#)
- [164] Ben C Stöver and Kai F Müller. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics*, 11:7, 2010. [17](#)
- [165] Donovan H. Parks, Timothy Mankowski, Somayyeh Zangoeei, Michael S. Porter, David G. Armanini, Donald J. Baird, Morgan G.I. Langille, and Robert G. Beiko. GenGIS 2: Geospatial Analysis of Traditional and Genetic Biodiversity, with New Gradient Algorithms and an Extensible Plugin Framework. *PLoS ONE*, 8(7), 2013. [17](#)
- [166] Hadley Wickam. *ggplot2: Elegant Graphics for Data Analysis*. Springer, Verlage, New York., 2016. [18](#)

- [167] Y. Inoue, K. Kawasaki, M. Watanabe, H. Ohta, O. Bolormaa, N. Sakagami, N. Fujitake, and S. Hiradate. Characterization of major and trace elements in sclerotium grains. *European Journal of Soil Science*, 58(3):786–793, 2006. [22](#)
- [168] M. Watanabe, A. Genseki, N. Sakagami, Y. Inoue, Ohta H., and N. Fujitake. Aluminum oxyhydroxide polymorphs and some micromorphological characteristics in sclerotium grains. *Soil Science and Plant Nutrition*, 50(8):1205–1210, 2004.
- [169] Makiko Watanabe, Nobuo Sakagami, Yudzuru Inoue, Akira Genseki, Hiroyuki Ohta, and Nobuhide Fujitake. The relation between distribution of sclerotium grain and chemical properties in nonallophanic Andosols. *Pedologist*, 48(1):24–32, 1 2004. [22](#)
- [170] Robert Fogel and Gary Hunt. Fungal and arboreal biomass in a western Oregon Douglas-fir ecosystem: distribution patterns and turnover. *Canadian Journal of Forest Research*, 9(2):245–256, 1979. [23](#)
- [171] Austin Burt, Deidre a Cartert, Gina L Koenigt, Thomas J Whites, and John W Taylor. Molecular markers reveal cryptic *Coccidioides immitis* in the human pathogen. *Mycotaxon*, 93(January 1996):770–773, 1996. [23](#)
- [172] Miguel Jurado, Patricia Marín, Covadonga Vázquez, and M. Teresa González-Jaén. Divergence of the IGS rDNA in *Fusarium proliferatum* and *Fusarium globosum* reveals two strain specific non-orthologous types. *Mycological Progress*, 11(1):101–107, 2012. [23](#)
- [173] Salvador Mirete, Belén Patiño, Miguel Jurado, Covadonga Vázquez, and María Teresa González-Jaén. Structural variation and dynamics of the nuclear ribosomal intergenic spacer region in key members of the *Gibberella fujikuroi* species complex. *Genome*, 56(4):205–213, 2013. [23](#)
- [174] Jessica M Vélez, Reese M Morris, Rytas Vilgalys, Jessy Labbé, and Christopher W Schadt. Phylogenetic diversity of 200+ isolates of the ectomycorrhizal fungus *Cenococcum geophilum* associated with *Populus trichocarpa* soils in the Pacific

- Northwest, USA and comparison to globally distributed representatives. *bioRxiv*, page 2020.03.24.005348, 1 2020. [30](#), [31](#), [32](#), [33](#), [42](#), [44](#), [53](#)
- [175] Christopher W. Fernandez, M. Luke McCormack, Jason M. Hill, Seth G. Pritchard, and Roger T. Koide. On the persistence of *Cenococcum geophilum* ectomycorrhizas and its implications for forest carbon and nutrient cycles. *Soil Biology and Biochemistry*, 65:141–143, 2013. [30](#)
- [176] James M. Trappe. *Cenococcum graniforme* its distribution, ecology, mycorrhiza formation and inherent variation, 1962. [30](#)
- [177] Harris A. Lewin, Gene E. Robinson, W. John Kress, William J. Baker, Jonathan Coddington, Keith A. Crandall, Richard Durbin, Scott V. Edwards, Félix Forest, M. Thomas P. Gilbert, Melissa M. Goldstein, Igor V. Grigoriev, Kevin J. Hackett, David Haussler, Erich D. Jarvis, Warren E. Johnson, Aristides Patrinos, Stephen Richards, Juan Carlos Castilla-Rubio, Marie Anne Van Sluys, Pamela S. Soltis, Xun Xu, Huanming Yang, and Guojie Zhang. Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences of the United States of America*, 115(17):4325–4333, 2018. [31](#)
- [178] Charley G.P. McCarthy and David A. Fitzpatrick. *Multiple Approaches to Phylogenomic Reconstruction of the Fungal Kingdom*. Elsevier Inc., 1 edition, 2017. [31](#)
- [179] Christina A. Cuomo. Harnessing Whole Genome Sequencing in Medical Mycology. *Current Fungal Infection Reports*, 11(2):52–59, 2017. [31](#)
- [180] Martha Malapi-Wight, Catalina Salgado-Salazar, Jill E. Demers, David L. Clement, Karen K. Rane, and Jo Anne Crouch. Sarcococca blight: Use of whole-genome sequencing for fungal plant disease diagnosis. *Plant Disease*, 100(6):1093–1100, 2016.
- [181] David Pizarro, Francesco Dal Grande, Steven Don Leavitt, Paul Stanley Dyer, Imke Schmitt, Ana Crespo, Helge Thorsten Lumbsch, Pradeep Kumar Divakar, and Marta Barluenga. Whole-Genome Sequence Data Uncover Widespread Heterothallism in the

- Largest Group of Lichen-Forming Fungi. *Genome Biology and Evolution*, 11(3):721–730, 2019. [31](#)
- [182] Ying Yang, Min Chen, Zongwei Li, Abdullah M.S. Al-Hatmi, Sybren de Hoog, Weihua Pan, Qiang Ye, Xiaochen Bo, Zhen Li, Shengqi Wang, Junzhi Wang, Huipeng Chen, and Wanqing Liao. Genome sequencing and comparative genomics analysis revealed pathogenic potential in *Penicillium capsulatum* as a novel fungal pathogen belonging to Eurotiales. *Frontiers in Microbiology*, 7(OCT):1–14, 2016. [31](#)
- [183] Ricardo Araujo. Towards the Genotyping of Fungi: Methods, Benefits and Challenges. *Current Fungal Infection Reports*, 8(3):203–210, 2014. [31](#)
- [184] Ralph Dean, Thomas Mitchell, Nicole Donofrio, Jun Seop Jeong, and Amy Powell. Fungal biology reaps the benefit of genomics. *Genome Biology*, 6(8):6–8, 2005. [31](#)
- [185] André Drenth, Alistair R. McTaggart, and Brenda D. Wingfield. Fungal clones win the battle, but recombination wins the war. *IMA Fungus*, 10(1):1–6, 2019. [31](#), [44](#), [45](#)
- [186] Michael G. Milgroom, María Del Mar Jiménez-Gasco, Concepción Olivares-García, Milton T. Drott, and Rafael M. Jiménez-Díaz. Recombination between clonal lineages of the asexual fungus *Verticillium dahliae* detected by genotyping by sequencing. *PLoS ONE*, 9(9), 2014.
- [187] J.W. Taylor, D.J. Jacobson, and M.C. Fisher. The Evolution of Asexual Fungi: Reproduction, Speciation and Classification. *Annual Review of Phytopathology*, 37(37):197–246, 1999.
- [188] John W. Taylor, Christopher Hann-Soden, Sara Branco, Iman Sylvain, and Christopher E. Ellison. Clonal reproduction in fungi. *Proceedings of the National Academy of Sciences of the United States of America*, 112(29):8901–8908, 2015. [31](#), [44](#), [45](#)
- [189] Wataru Kai, Kazuharu Nomura, Atushi Fujiwara, Yoji Nakamura, Motoshige Yasuike, Nobuhiko Ojima, Tetsuji Masaoka, Akiyuki Ozaki, Yukinori Kazeto, Koichiro Gen, Jiro

- Nagao, Hideki Tanaka, Takanori Kobayashi, and Mitsuru Ototake. A ddRAD-based genetic map and its integration with the genome assembly of Japanese eel (*Anguilla japonica*) provides insights into genome evolution after the teleost-specific genome duplication. *BMC genomics*, 15:233, 1 2014. [32](#)
- [190] Tony Kess, Jeffrey Gross, Fiona Harper, and Elizabeth G. Boulding. Low-cost ddRAD method of SNP discovery and genotyping applied to the periwinkle *Littorina saxatilis*. *Journal of Molluscan Studies*, 82(September 2015):104–109, 2015.
- [191] Kenta Shirasawa, Masaru Tanaka, Yasuhiro Takahata, Daifu Ma, Qinghe Cao, Qingchang Liu, Hong Zhai, Sang Soo Kwak, Jae Cheol Jeong, Ung Han Yoon, Hyeong Un Lee, Hideki Hirakawa, and Sachiko Isobe. A high-density SNP genetic map consisting of a complete set of homologous groups in autohexaploid sweetpotato (*Ipomoea batatas*). *Scientific Reports*, 7(March):1–8, 2017. [32](#)
- [192] Claudio G Ametrano, Felix Grewe, Pedro W Crous, Stephen B Goodwin, Chen Liang, Laura Selbmann, H Thorsten Lumbsch, Steven D Leavitt, and Lucia Muggia. Genome-scale data resolve ancestral rock-inhabiting lifestyle in Dothideomycetes (Ascomycota). *IMA Fungus*, 19(10):1–12, 2019. [32](#), [41](#)
- [193] Carly F. Graham, Travis C. Glenn, Andrew G. McArthur, Douglas R. Boreham, Troy Kieran, Stacey Lance, Richard G. Manzon, Jessica A. Martino, Todd Pierson, Sean M. Rogers, Joanna Y. Wilson, and Christopher M. Somers. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources*, 15(6):1304–1315, 2015. [32](#)
- [194] Sandra L. Hoffberg, Troy J. Kieran, Julian M. Catchen, Alison Devault, Brant C. Faircloth, Rodney Mauricio, and Travis C. Glenn. RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Molecular Ecology Resources*, 16(5):1264–1278, 2016. [32](#)
- [195] Nadin Rohland and David Reich. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, 22(5):939–946, 2012. [33](#)

- [196] Anthony M. Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120, 2014. [36](#)
- [197] Heng Li and Richard Durbin. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14):1754–1760, 2009. [36](#)
- [198] Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, 2009. [36](#)
- [199] Heng Li. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21):2987–2993, 2011. [36](#)
- [200] Petr Danecek, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, Gerton Lunter, Gabor T. Marth, Stephen T. Sherry, Gilean McVean, and Richard Durbin. The variant call format and VCFtools. *Bioinformatics*, 27(15):2156–2158, 2011. [36](#)
- [201] Nicolas C. Rochette and Julian M. Catchen. Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, 12(12):2640–2659, 2017. [37](#)
- [202] Enis Afgan, Dannon Baker, B er enice Batut, Marius Van Den Beek, Dave Bouvier, Martin Ech, John Chilton, Dave Clements, Nate Coraor, Bj orn A. Gr uning, Aysam Guerler, Jennifer Hillman-Jackson, Saskia Hiltemann, Vahid Jalili, Helena Rasche, Nicola Soranzo, Jeremy Goecks, James Taylor, Anton Nekrutenko, and Daniel Blankenberg. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, 46(W1):W537–W544, 2018. [37](#)
- [203] Francios Balloux Thibaut Jombart, Sebastien Devillard. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11(94), 2010. [37](#)

- [204] Brian J. Knaus and Niklaus J. Grünwald. vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, 17(1):44–53, 2017. [37](#), [38](#)
- [205] Thibaut Jombart. Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11):1403–1405, 2008. [37](#)
- [206] Zhian N. Kamvar, Javier F. Tabima, and Niklaus J. Grünwald. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2014(1):1–14, 2014. [38](#)
- [207] Zhian N. Kamvar, Jonah C. Brooks, and Niklaus J. Grünwald. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, 6(JUN):1–10, 2015. [38](#)
- [208] Cindy F. Verdu, Erwan Guichoux, Samuel Quevauvillers, Olivier De Thier, Yec’han Laizet, Adline Delcamp, Frédéric Gévaudant, Arnaud Monty, Annabel J. Porté, Philippe Lejeune, Ludivine Lassois, and Stéphanie Mariette. Dealing with paralogy in RADseq data: in silico detection and single nucleotide polymorphism validation in *Robinia pseudoacacia* L. *Ecology and Evolution*, 6(20):7323–7333, 2016. [38](#)
- [209] D.L. Hawksworth, P.M. Kirk, B.C. Sutton, and D.N. Pegler. *Ainsworth and Bisby’s Dictionary of the Fungi*. CABI, Wallingford, United Kingdom, 8th edition, 1996. [44](#), [51](#)
- [210] Erica M. Goss, Javier F. Tabima, David E.L. Cooke, Silvia Restrepo, William E. Frye, Gregory A. Forbes, Valerie J. Fieland, Martha Cardenas, and Niklaus J. Grünwald. The Irish potato famine pathogen *Phytophthora infestans* originated in central Mexico rather than the Andes. *Proceedings of the National Academy of Sciences of the United States of America*, 111(24):8791–8796, 2014. [44](#), [45](#)
- [211] Bret A. Payseur and Loren H. Rieseberg. A genomic perspective on hybridization and speciation. *Molecular ecology*, 25(11):2337–2360, 2016.
- [212] Annabel L. Smith, Trevor R. Hodkinson, Jesus Villellas, Jane A. Catford, Anna Mária Csergő, Simone P. Blomberg, Elizabeth E. Crone, Johan Ehrlén, Maria B. Garcia,

- Anna Liisa Laine, Deborah A. Roach, Roberto Salguero-Gómez, Glenda M. Wardle, Dylan Z. Childs, Bret D. Elderd, Alain Finn, Sergi Munné-Bosch, Maude E.A. Baudraz, Judit Bódis, Francis Q. Brearley, Anna Bucharova, Christina M. Caruso, Richard P. Duncan, John M. Dwyer, Ben Gooden, Ronny Groenteman, Liv Norunn Hamre, Aveliina Helm, Ruth Kelly, Lauri Laanisto, Michele Lonati, Joslin L. Moore, Melanie Morales, Siri Lie Olsen, Meelis Pärtel, William K. Petry, Satu Ramula, Pil U. Rasmussen, Simone Ravetto Enri, Anna Roeder, Christiane Roscher, Marjo Saastamoinen, Ayco J.M. Tack, Joachim Paul Töpper, Gregory E. Vose, Elizabeth M. Wandrag, Astrid Wingler, and Yvonne M. Buckley. Global gene flow releases invasive plants from environmental constraints on genetic diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 117(8):4218–4227, 2020. [44](#)
- [213] C. L. Halliday and D. A. Carter. Clonal reproduction and limited dispersal in an environmental population of *Cryptococcus neoformans* var. *gattii* isolates from Australia. *Journal of Clinical Microbiology*, 41(2):703–711, 2003. [45](#)
- [214] Bart P.S. Nieuwenhuis and Timothy Y. James. The frequency of sex in fungi. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1706), 2016. [45](#)
- [215] David B. Lowry, Sean Hoban, Joanna L. Kelley, Katie E. Lotterhos, Laura K. Reed, Michael F. Antolin, and Andrew Storfer. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular ecology resources*, 17(2):142–152, 2017. [46](#)
- [216] Daniel Ortiz-Barrientos, Jan Engelstädter, and Loren H. Rieseberg. Recombination Rate Evolution and the Origin of Species. *Trends in Ecology and Evolution*, 31(3):226–236, 2016. [46](#)
- [217] Yllwl Liu and Chz Li. Heavy Metal Contamination and Remediation in Asian Agricultural Land. *Niaes.Affrc.Go.Jp*, pages 1–9, 2001. [50](#)
- [218] Amanullah Mahar, Ping Wang, Amjad Ali, Mukesh Kumar Awasthi, Altaf Hussain Lahori, Quan Wang, Ronghua Li, and Zengqiang Zhang. Challenges and opportunities

- in the phytoremediation of heavy metals contaminated soils: A review. *Ecotoxicology and Environmental Safety*, 126:111–121, 2016.
- [219] Nadeem Sarwar, Muhammad Imran, Muhammad Rashid Shaheen, Wajid Ishaq, Asif Kamran, Amar Matloob, Abdur Rehim, and Saddam Hussain. Phytoremediation strategies for soils contaminated with heavy metals: modifications and future perspectives. *Chemosphere*, 2016.
- [220] Ugya A.Y., T. S. Imam, and S. M. Tahir, editors. *Emerging Trends in the Remediation of Pollution*. Number November. Lambert Academic Publishing, 2017. 50
- [221] Bram Beckers, Michiel Op De Beeck, Nele Weyens, Rebecca Van Acker, Marc Van Montagu, Wout Boerjan, and Jaco Vangronsveld. Lignin engineering in field-grown poplar trees affects the endosphere bacterial microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, 113(8):1523264113–, 2016. 51
- [222] Kumar Gupta Dharmendra, editor. *Plant-Based Remediation Processes*, volume 35. Springer-Verlag Berlin Heidelberg, 2013.
- [223] Neil R Gottel, Hector F Castro, Marilyn Kerley, Zamin Yang, Dale a Pelletier, Mircea Podar, Tatiana Karpinets, Ed Uberbacher, Gerald a Tuskan, Rytas Vilgalys, Mitchel J Doktycz, and Christopher W Schadt. Distinct microbial communities within the endosphere and rhizosphere of *Populus deltoides* roots across contrasting soil types. *Applied and environmental microbiology*, 77(17):5934–44, 9 2011.
- [224] D. S. Powlson, P. J. Gregory, W. R. Whalley, J. N. Quinton, D. W. Hopkins, A. P. Whitmore, P. R. Hirsch, and K. W T Goulding. Soil management in relation to sustainable agriculture and ecosystem services. *Food Policy*, 36(SUPPL. 1):S72–S87, 2011. 51
- [225] Amrita Sengupta, Sunil Kumar Gunri, and Tapas Biswas. *Plant-Microbe Interactions in Agro-Ecological Perspectives*, volume 2. Springer Nature Singapore, 2017. 51

- [226] Antoine Harfouche, Richard Meilan, and Arie Altman. Tree genetic engineering and applications to sustainable forestry and biomass production. *Trends in biotechnology*, 29(1):9–17, 2011. [51](#)
- [227] Adam D. Leaché and Jamie R. Oaks. The Utility of Single Nucleotide Polymorphism (SNP) Data in Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, 48(1):69–84, 2017. [52](#)
- [228] M. Matsubara, J. M. Lynch, and F. a a M De Leij. A simple screening procedure for selecting fungi with potential for use in the bioremediation of contaminated land. *Enzyme and Microbial Technology*, 39(7):1365–1372, 2006. [52](#)
- [229] Prasun Ray, Richa Tiwari, U. Gangi Reddy, and Alok Adholeya. Detecting the heavy metal tolerance level in ectomycorrhizal fungi in vitro. *World Journal of Microbiology and Biotechnology*, 21(3):309–315, 2005.
- [230] Al-Kadeeb A. Siham. Effect of Lead and Copper on the Growth of Heavy Metal Resistance Fungi Isolated from Second Industrial City in Riyadh, Saudi Arabia. *Journal of Applied Sciences*, 7:1019–1024, 2007. [52](#)
- [231] Xuefeng Wang. Firth logistic regression for rare variant association tests. *Frontiers in Genetics*, 5(187), 2014. [54](#)
- [232] Christopher C. Chang, Carson C. Chow, Laurent C.A.M. Tellier, Shashaank Vattikuti, Shaun M. Purcell, and James J. Lee. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 4(1):1–16, 2015. [54](#)
- [233] Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A.R. Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul I.W. De Bakker, Mark J. Daly, and Pak C. Sham. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3):559–575, 2007. [54](#)
- [234] B. Devlin and Kathryn Roeder. Genomic control for association studies. *Biometrics*, 55(4):997–1004, 1999. [55](#)

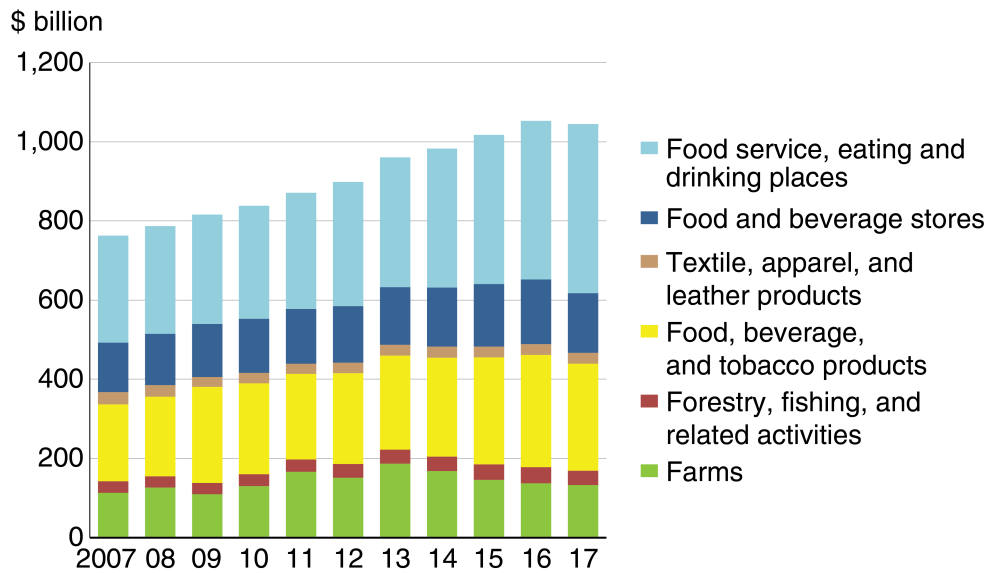
- [235] Gerard Manley. How Does Multiple Testing Correction Work? *Nature Biotechnology*, 27(12):1135–1137, 2009. [55](#)
- [236] Stephen D. Turner. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software*, 3(25):731, 2018. [55](#)
- [237] Russell W. Jenkins, Daniel Canals, and Yusuf A. Hannun. Roles and regulation of secretory and lysosomal acid sphingomyelinase. *Cell Signal*, 21(6):836–846, 2009. [57](#)
- [238] Sorel Fitz-Gibbon, Andrew L. Hipp, Kasey K. Pham, Paul S. Manos, and Victoria L. Sork. Phylogenomic inferences from reference-mapped and de novo assembled short-read sequence data using RADseq sequencing of California white oaks (*Quercus* section *Quercus*). *Genome*, 60(9):743–755, 2017. [58](#)
- [239] Muhammad Shahid, Camille Dumat, Sana Khalid, Nabeel Khan Niazi, and Paula M C Antunes. Cadmium Bioavailability, Uptake, Toxicity and Detoxification in Soil-Plant System. In Pim de Voogt, editor, *Reviews of Environmental Contamination and Toxicology Volume 241*, pages 73–137. Springer International Publishing, Cham, 2017. [58](#)
- [240] Clara M. Seibert and Frank M. Raushel. Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry*, 44(17):6383–6391, 2005. [59](#)
- [241] Milton H. Saier, J. Thomas Beatty, Andre Goffeau, Kevin T. Harley, Wilbert H.M. Heijne, Su Chi Huang, Donald L. Jack, Peter S. Jähn, Katharine Lew, Jia Liu, Stephanie S. Pao, Ian T. Paulsen, Tsai Tien Tseng, and Pritbir S. Virk. The major facilitator superfamily. *Journal of Molecular Microbiology and Biotechnology*, 1(2):257–279, 1999. [59](#)
- [242] Ralph Bertram, Maximilian Schlicht, Kerstin Mahr, Harald Nothhaft, Milton H. Saier, and Fritz Titgemeyer. In Silico and Transcriptional Analysis of Carbohydrate Uptake Systems of *Streptomyces coelicolor* A3(2). *Journal of Bacteriology*, 186(5):1362–1373, 2004. [59](#)

- [243] Elizabeth Fullam, Ivan Prokes, Klaus Fütterer, and Gurdyal S. Besra. Structural and functional analysis of the solute-binding protein UspC from *Mycobacterium tuberculosis* that is specific for amino sugars. *Open Biology*, 6(6), 2016. [59](#)
- [244] Amanda J. O'Reilly, Joel B. Dacks, and Mark C. Field. Evolution of the karyopherin- β family of nucleocytoplasmic transport factors; ancient origins and continued specialization. *PLoS ONE*, 6(4), 2011. [59](#)
- [245] Zhiliang Fan, Gbemeloluwa B. Oguntimein, and Peter J. Reilly. Characterization of kinetics and thermostability of *Acremonium strictum* glucooligosaccharide oxidase. *Biotechnology and Bioengineering*, 68(2):231–237, 2000. [59](#)
- [246] Anthony Levasseur, Elodie Drula, Vincent Lombard, Pedro M. Coutinho, and Bernard Henrissat. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels*, 6(1):1–14, 2013. [59](#)
- [247] Christopher T. Walsh and Timothy A. Wencewicz. Flavoenzymes: Versatile catalysts in biosynthetic pathways. *Natural Product Reports*, 30(1):175–200, 2013. [59](#)
- [248] Andréa M.A. Nascimento and Edmar Chartone-Souza. Operon mer: Bacterial resistance to mercury and potential for bioremediation of contaminated environments. *Genetics and Molecular Research*, 2(1):92–101, 2003. [59](#)
- [249] Simon Silver and Le T. Phung. Bacterial heavy metal resistance: New surprises. *Annual Review of Microbiology*, 50:753–789, 1996. [59](#)
- [250] L J Biol Chem. Fructose 1,6-Bisphosphate. *The Journal of Biological Chemistry*, 253(16):5712–5718, 1977. [60](#)
- [251] Ekaterina Kuznetsova, Linda Xu, Alexander Singer, Greg Brown, Aiping Dong, Robert Flick, Hong Cui, Marianne Cuff, Andrzej Joachimiak, Alexei Savchenko, and Alexander F. Yakunin. Structure and activity of the metal-independent fructose-1,6-bisphosphatase YK23 from *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 285(27):21049–21059, 2010. [60](#)

- [252] Uday Ganapathy, Joeli Marrero, Susannah Calhoun, Hyungjin Eoh, Luiz Pedro Sorio De Carvalho, Kyu Rhee, and Sabine Ehrt. Two enzymes with redundant fructose bisphosphatase activity sustain gluconeogenesis and virulence in *Mycobacterium tuberculosis*. *Nature Communications*, 6, 2015. [60](#)

Appendix

Value added to GDP by agriculture and related industries, 2007-17



Note: GDP refers to gross domestic product.

Source: USDA, Economic Research Service using data from U.S. Department of Commerce, Bureau of Economic Analysis, Value Added by Industry series.

Figure 1: Agricultural security is critical for economic security, which includes food security, sovereign independence, and increased general prosperity of the population. In the United States, agriculture accounted for \$992 billion of the gross domestic product in 2015. Source: [USDA Economic Research Service](#)

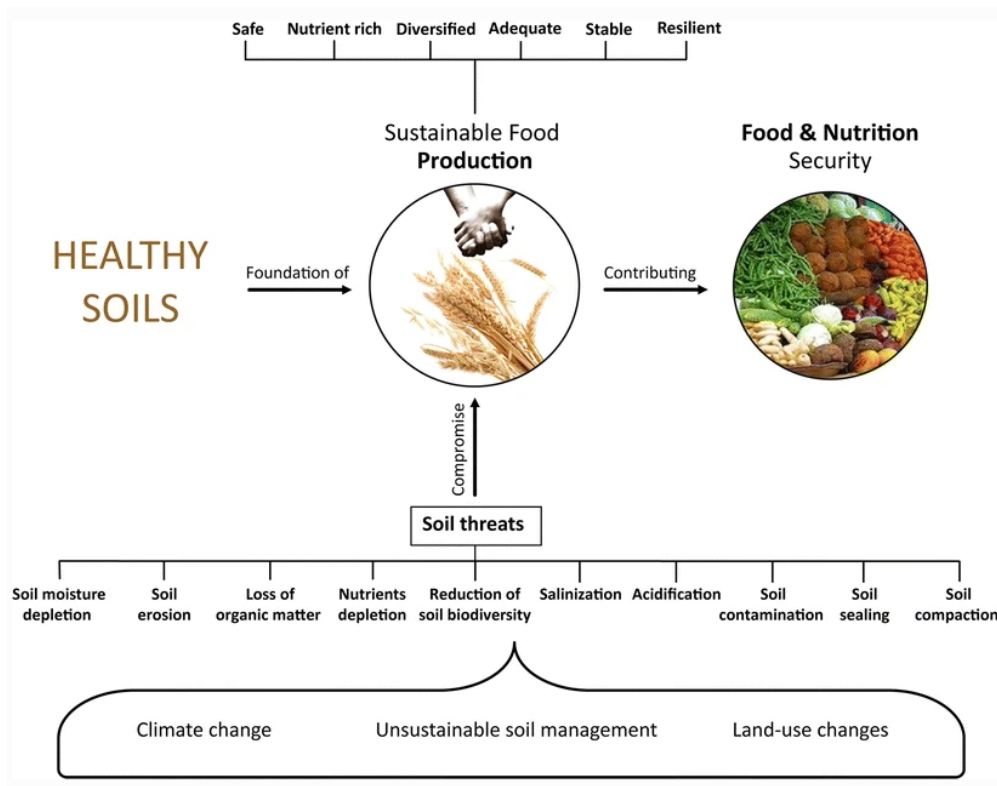


Figure 2: There are up to 1.73 million hectares of arable land in the United States, increasing the need for sustainable agricultural practices which maintain and increase soil health [7].

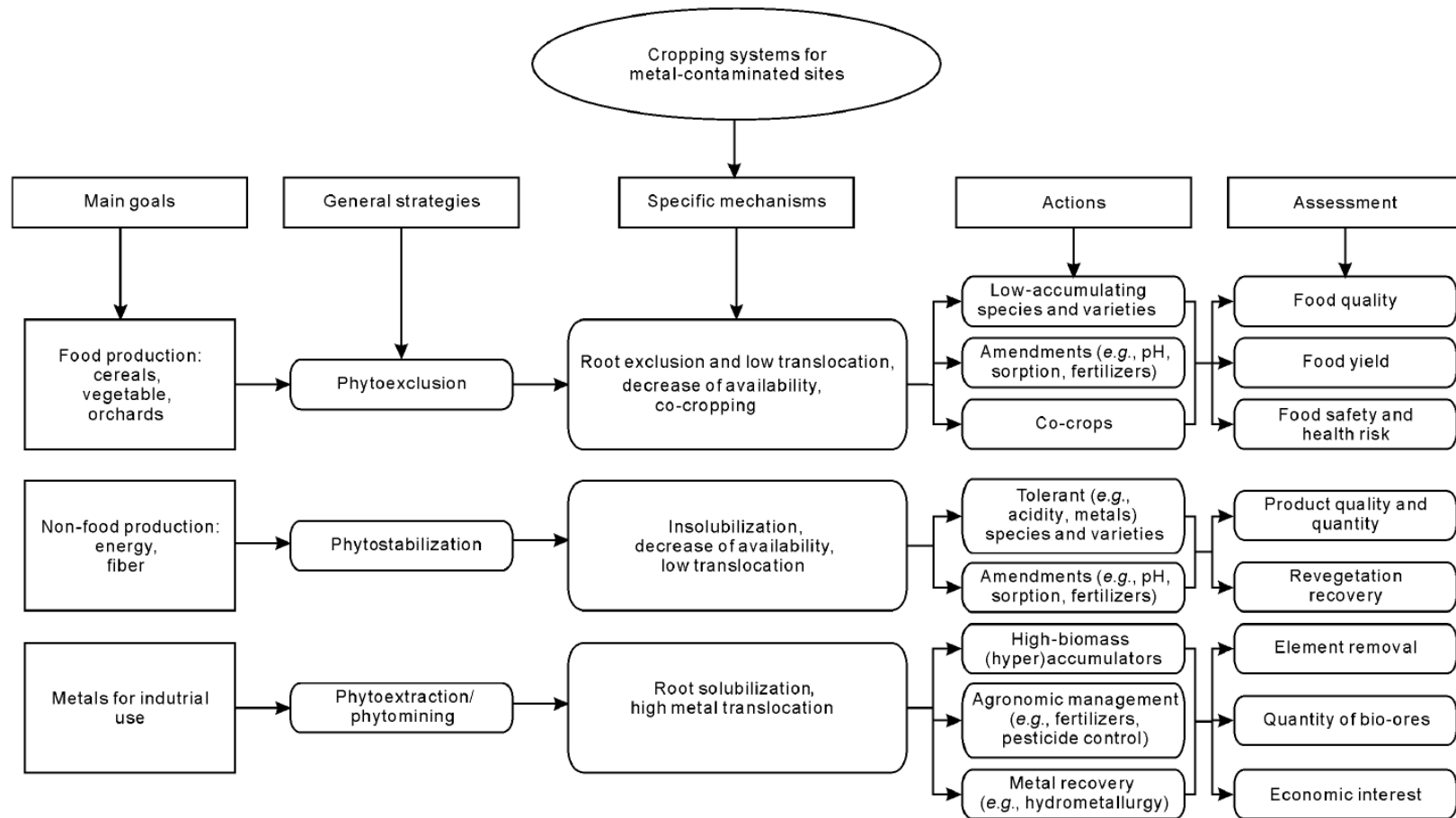


Figure 3: A cropping system involving planting biofuel crops allows for remediation of agricultural soils without extreme negative impact on the quality of the product produced [15].

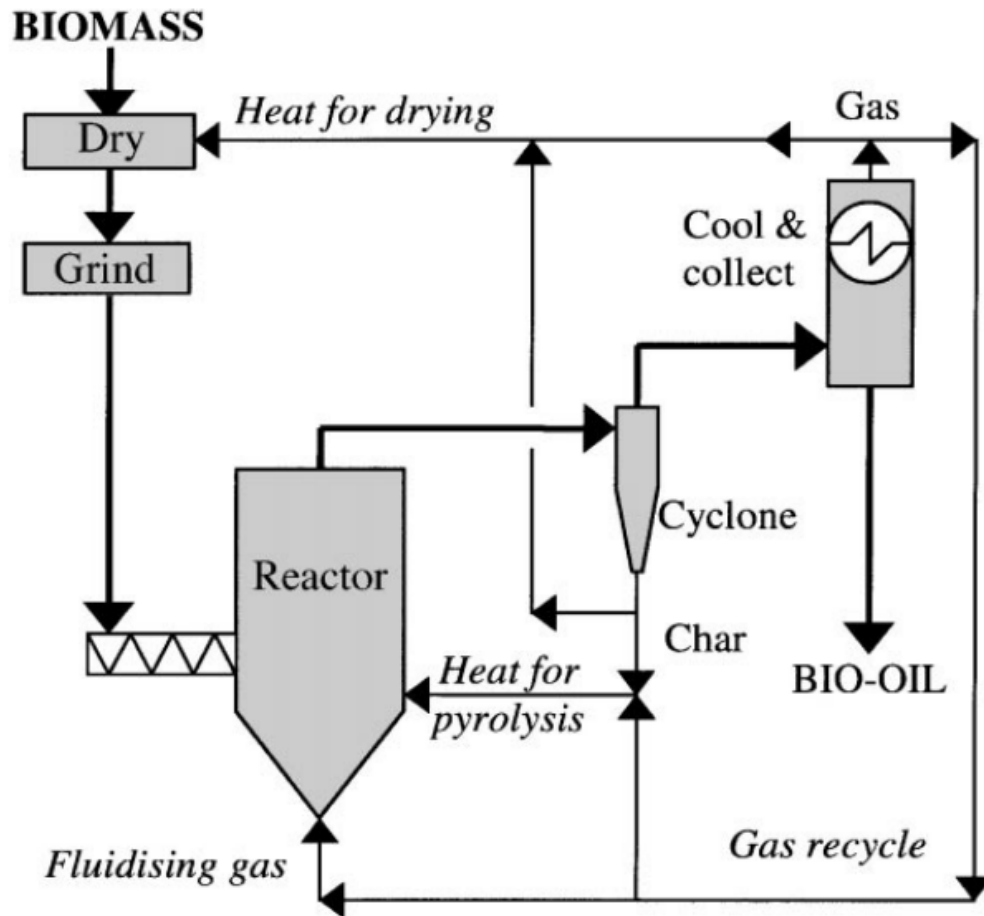


Figure 4: Pyrolysis of harvested biomass captures heavy metal contaminants in the char byproduct which accounts for 10 of the total product [33].

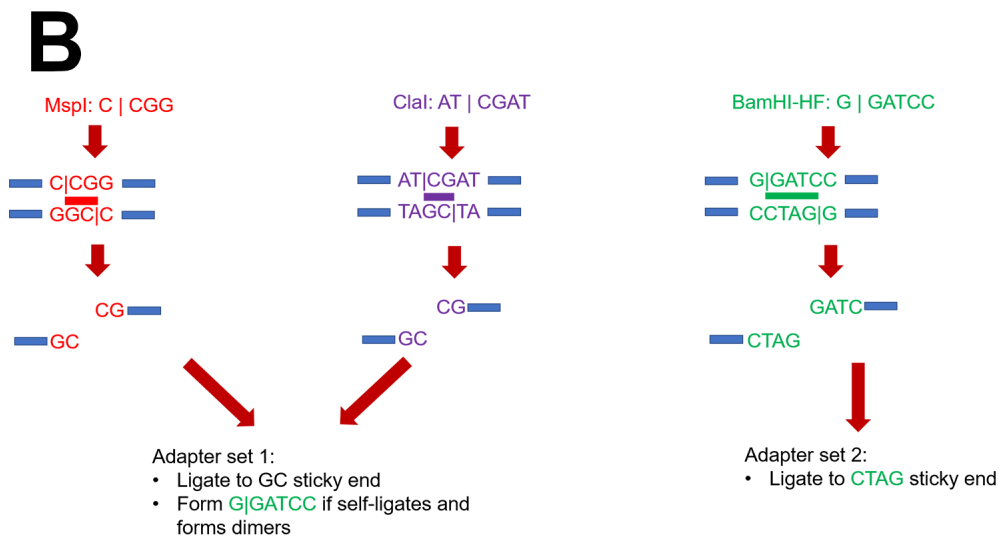
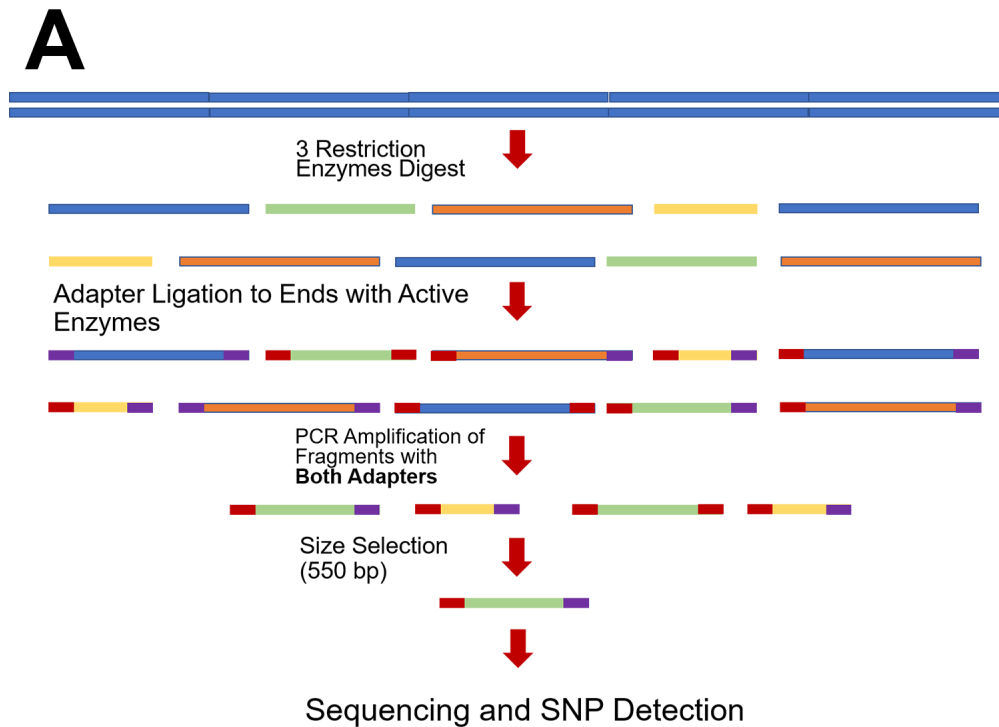


Figure 5: The 3RAD protocol developed by Bayona-Vasquez et al. (2019) [116] includes digestion of genomic DNA with three restriction enzymes, two of which share a cut site sticky end (A). Internal adapters are designed to ligate to one of the two sticky ends, and one internal adapter is also designed to self-ligate at the cut site for the 3rd restriction enzyme, leading to lesser chances of self-ligation and more efficient ligation of internal adapters (B).

Table 1: Model with best fit analysis, Bayesian Information Criterion, and Akaike Information Criterion (AICc) for each alignment per an analysis using MEGA X software. Lower BIC and AICc values indicate the best model fit for use in analyses.

Alignment	Best Model	AICc	BIC
PNW ITS	K2 + G	2873.433	7082.611
PNW <i>GAPDH</i>	K2 + G	5468.328	9860.315
PNW ITS + <i>GAPDH</i>	K2 + G	8322.607	12970.723
GP ITS	K2 + G	7258.62	21156.093
GP <i>GAPDH</i>	K2 + G	12807.63	29779.549
GP ITS + <i>GAPDH</i>	K2 + G + I	15623	26550.792

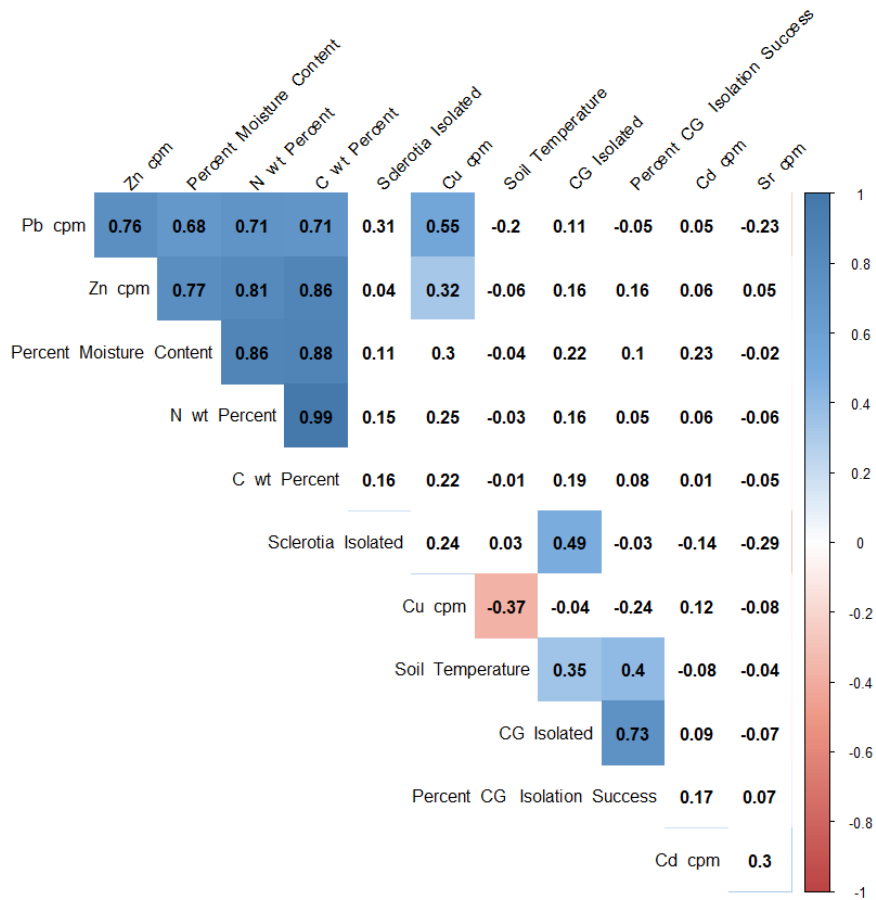


Figure 6: Correlogram of PNW clade, rivershed of origin, site latitude, soil percent moisture content, C/N weight percent, temperature, elemental metal (lead, zinc, copper, cadmium, strontium) counts per minute, total sclerotia isolated, and total *C. geophilum* isolation success of 105 PNW soil samples. Positive correlations are highlighted in blue and negative correlations are highlighted in red, with color intensity proportional to the correlation coefficient. Only those correlation coefficients with $p < 0.05$ are shown in color. No measured soil conditions or qualities were determined to correlate with the total sclerotia obtained or *C. geophilum* successfully isolated.

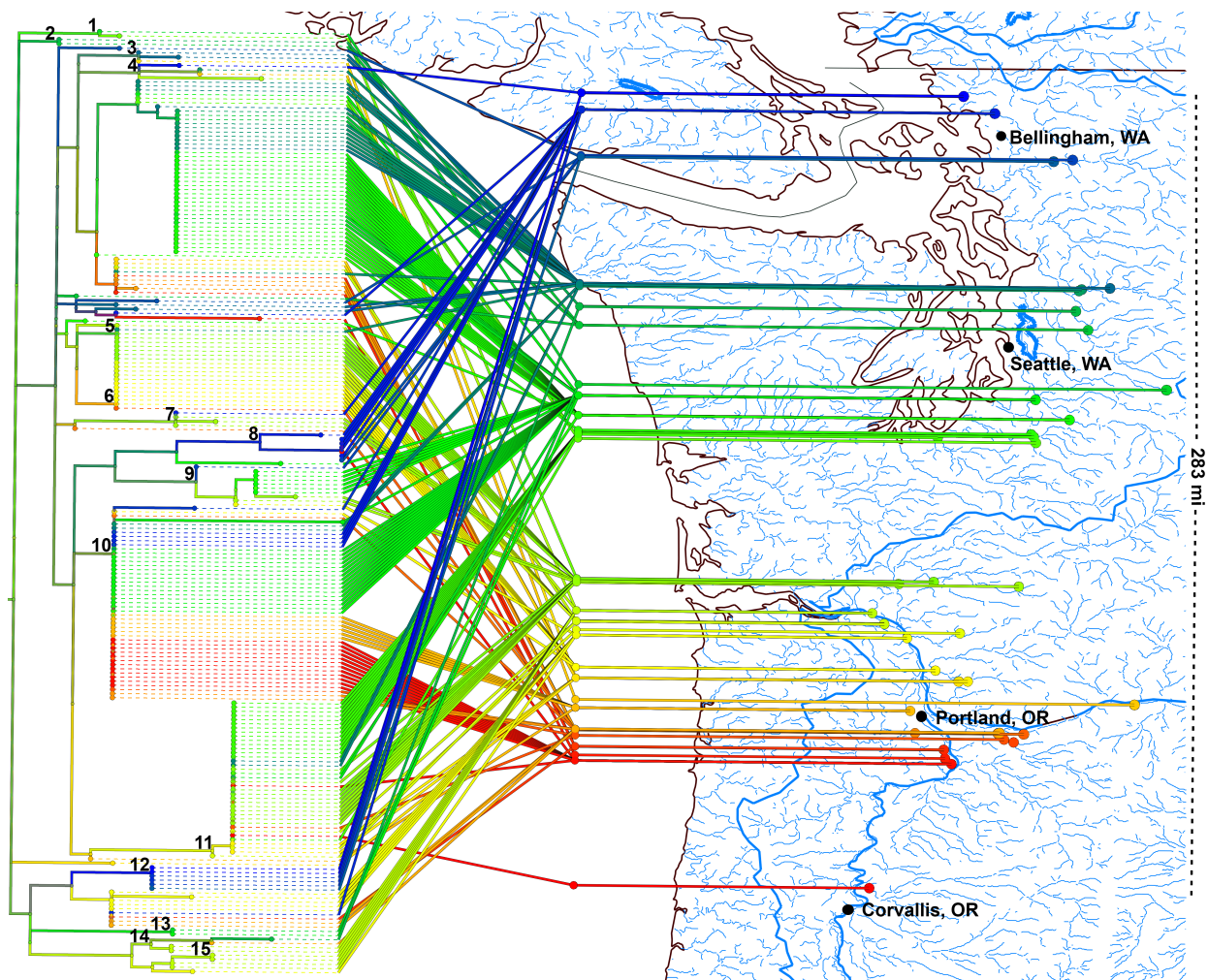


Figure 7: The Pacific Northwest (PNW) isolate *GAPDH* RAxML phylogeny mapped by latitude to the PNW region. Clades with >80% bootstrap support values are indicated above the associated clade. A total of 15 clades encompassing 155 isolates were identified in the PNW isolate collection, with 74 isolates remaining unresolved. Isolates from smaller clades tended to group by region of origin with a few notable exceptions present over large north-south distances.

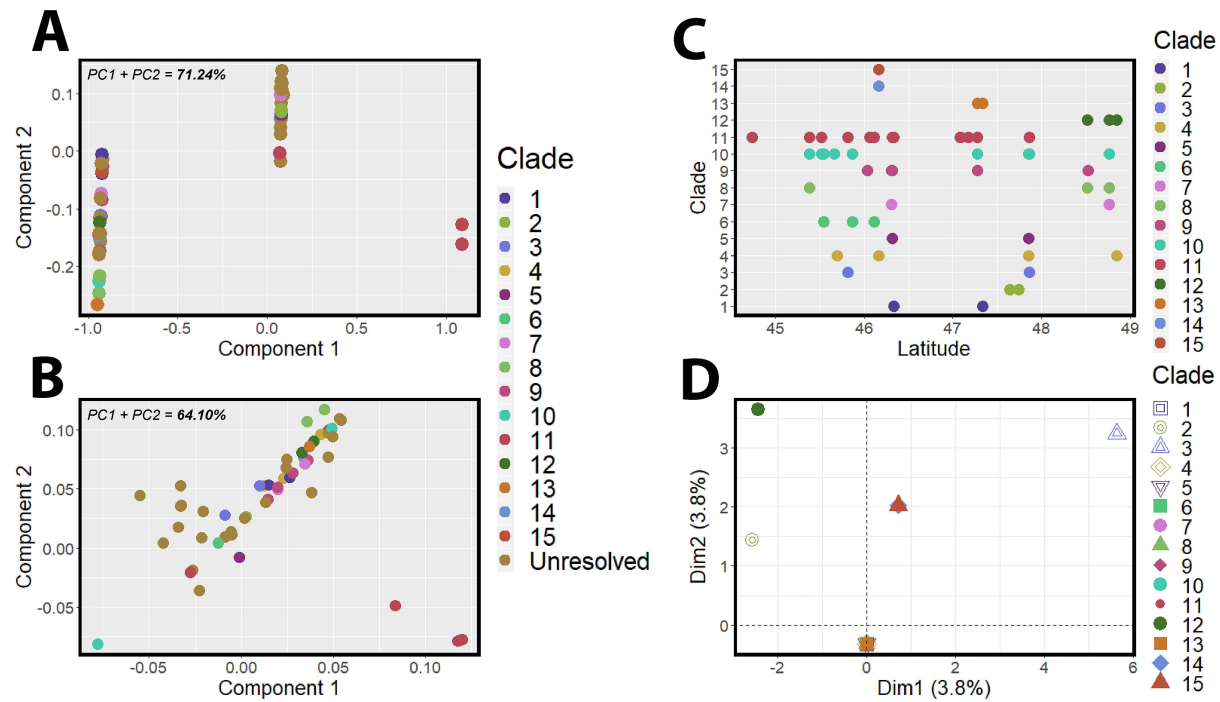


Figure 8: Principle components analysis (PCoA) of the PNW *GAPDH* RAxML phylogeny revealed three distinct isolate clusters, with clade 11 segregating as a separate cluster from the remaining PNW collection (A). A PCoA of the PNW *GAPDH* alignment also revealed three distinct clusters, with clades 10 and 11 segregating from the remaining PNW collection (B). A scatterplot of clade versus latitude did not reveal distinct patterns within the larger groups of the PNW collection (C), and a multiple components analysis showed some differentiation with clades 3, 12, 14, and 15 from the remaining PNW isolate collection, but revealed weak associations overall between latitude, isolate clade, and phylogenetic differentiation (D).

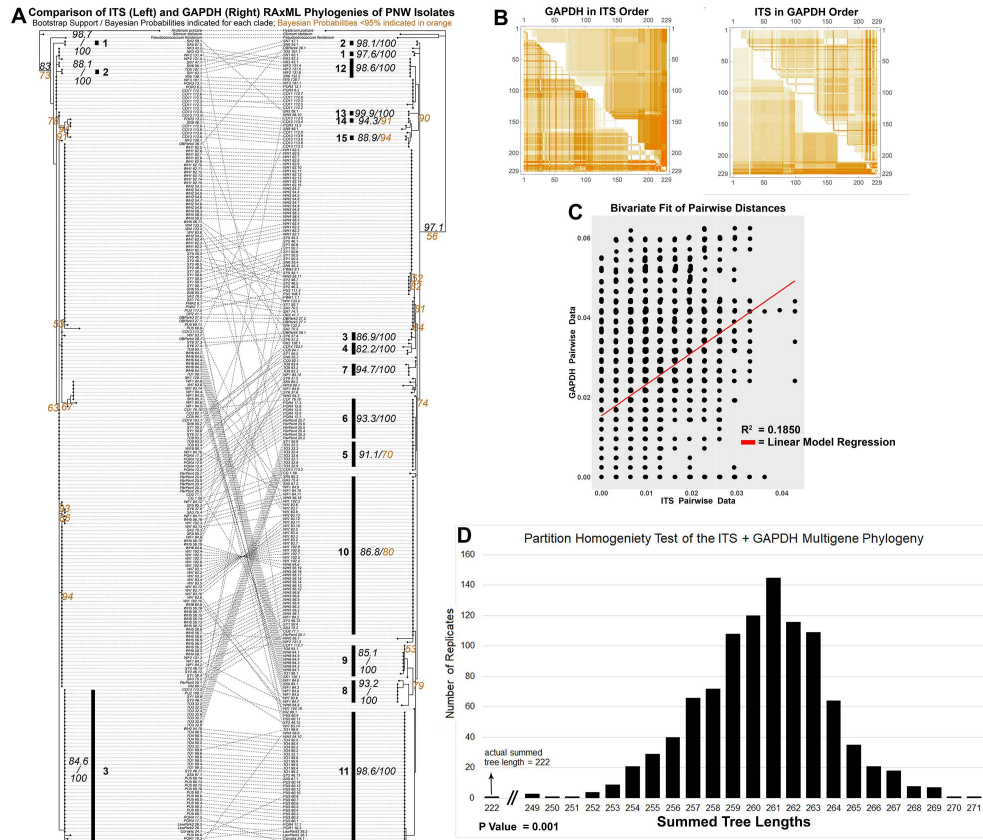


Figure 9: Phylogenetic incongruence between the ITS (left) and GAPDH (right) RAxML phylogenies (A). Heatmap showing pairwise distances for ITS and GAPDH genes from *Cenococcum geophilum*. Genetic distances (HKY85) were calculated among all sequence pairs for ITS (above diagonal) and GAPDH (below diagonal). Darker color indicates higher sequence divergence. The first is sequence order based on UPGMA clustering of ITS distances, the second is sequence order based on UPGMA clustering of GAPDH distances (B). A scatterplot of pairwise HKY85 distances for ITS vs GAPDH datasets shows low correlation between the ITS and GAPDH gene regions (C). The parsimony-based interpartition length difference (PTP-ILD) test indicated that the ITS and GAPDH data sets are incongruent due to no significant difference between the observed data set parsimony trees and the distribution of the randomized tree length calculations ($p = 0.001$) (D).

Legend (Study / Geographic Region):

- This Study / Pacific Northwest, USA
- de Freitas Pereira et al. 2018 / Western Europe
- Obase et al. 2016 / Florida & Georgia, USA
- Matsuda et al. 2015 / Japan
- Douhan & Rizzo 2005 / Douhan 2007 / Browns Valley, CA, USA

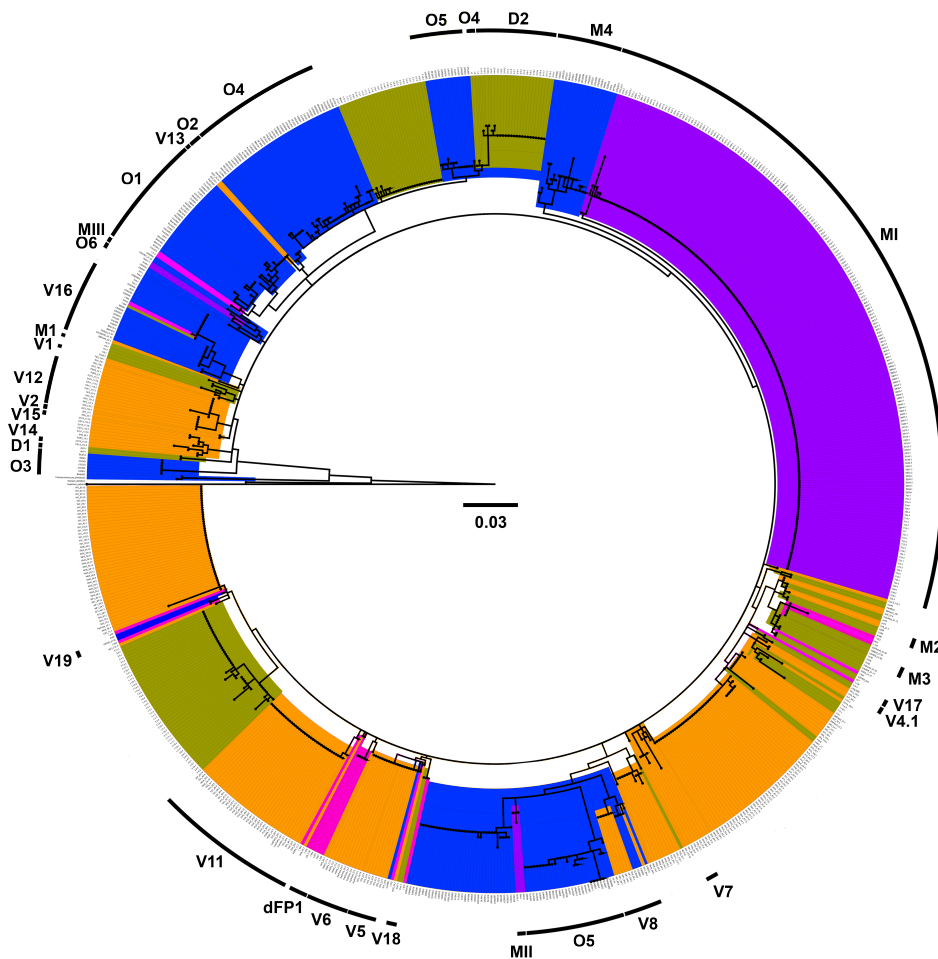


Figure 10: Global isolate collection GAPDH RAXML phylogenetic tree. Clades with strong bootstrap support (>80%) are labeled on the outer ring. Strongly supported clades implicated in this study are highlighted in orange and designated with V. Clades V16-V19 represent newly designated clades within the global *C. geophilum* isolate collection. The PNW isolates are highlighted in orange. Isolates are highlighted per the most recent published study of origin as follows: D (Douhan et al., 2007b; Douhan and Rizzo, 2005) highlighted in gold; dFP (de Freitas Pereira et al., 2018) highlighted in pink; M (Matsuda et al., 2015) highlighted in purple; and O (Obase et al., 2016) highlighted in blue.

Table 2: Previously resolved clades per study, number of isolates, geographic region, and host plant. The Pacific Northwest (PNW) isolate collection resolved 15 clades which appear to be uniquely associated with *Populus trichocarpa* within its host range in the PNW.

Study	Geographic Region	Total No. Isolates	Host Plant	Gene(s) Used for Phylogenetic Analyses	No. clades resolved at >80% bootstrap support
Douhan & Rizzo 2005	Browns Valley, California	103	Oak	GAPDH	3
	Non-California (location not included)	7	Not indicated		
Douhan & Huryn 2007	Browns Valley, California	74	Mixed host species	GAPDH	9c
	Pacific Northwest, USA				
	Alabama, USA				
Matsuda et al. 2015	Alaska, USA	225	<i>Pinus thunbergii</i>	GAPDH	3
	Europe				
	Japan				
Obase et al. 2016	Florida and Georgia, USA	242a	<i>Pinus elliotii</i>	ITSb	7d
	Japan		<i>Pinus taeda</i>	<i>GAPDHb</i>	
	Europe				
de Freitas et al. 2018	North America	16	<i>Picea abies</i>	ITSb	3e
			<i>Pinus sylvestris</i>	<i>GAPDHb</i>	
			<i>Fagus sylvatica</i>		
PNW Collection	Europe	231	Mixed forest	<i>GAPDH</i>	15
	Pacific Northwest, USA		<i>Populus trichocarpa</i>		
	California, USA				
Global Collection	Florida and Georgia, USA	790	Mixed host species	<i>GAPDH</i>	34
	Japan				
	Europe				
	North America				

Table 3: Pacific Northwest *Cenococcum geophilum* isolates used in this study with the rivershed of origin, *de novo* assembly cluster, *de novo* assembly clade and bootstrap support value, and single gene (GPD) clade and bootstrap support value. Unresolved isolates are designated with a dash (-).

Pacific Northwest Isolate	Rivershed	<i>De novo</i> Assembly Cluster	<i>De novo</i> Assembly Clade	<i>De novo</i> Assembly Support	Single Gene Clade	Single Gene Clade Support
1_58	-	2	1	99.4		-
CO11_112_1	Columbia River	3	4	83.4	2	84.1
CO13_113_4	Columbia River	3	2	98.4	4	82.6
CO13_113_5	Columbia River	3	2	98.4	4	82.6
CO11_112_2	Columbia River	3	4	83.4	4	82.6
CO11_112_3	Columbia River	3	4	83.4	4	82.6
CO11_112_5	Columbia River	3	4	83.4	4	82.6
CO11_112_7	Columbia River	3	4	83.4	4	82.6
Corvalis_24_1	Columbia River	1	9	81.1	6	100
POR4_17_1	Columbia River	1	9	81.1	6	100
CO1_76_10	Columbia River	3	9	81.1	8	91.6
POR4_12_2	Columbia River	3	9	81.1	8	91.6
POR4_12_5	Columbia River	3	9	81.1	8	91.6
POR4_12_6	Columbia River	3	9	81.1	8	91.6
POR4_17_3	Columbia River	3	9	81.1	8	91.6
CO14_103_1	Columbia River	3		-	10	84.4
CO13_113_3	Columbia River	3	3	100		-
CO13_113_6	Columbia River	3	3	100		-
CO13_113_2	Columbia River	3	9	81.1		-
CO3_82_1	Columbia River	3	9	81.1		-
CO5_94_1	Columbia River	3		-		-
OR2_41_1	Columbia River	4		-		-
WF1_84_7	Lake Whatcom	2	1	99.4	1	95.9
WF1_84_4	Lake Whatcom	4	10	100	1	95.9
WF1_84_5	Lake Whatcom	4	10	100	1	95.9
WF1_84_6	Lake Whatcom	4	10	100	1	95.9
WF1_84_10	Lake Whatcom	3	7	94.2	3	84.1
WF1_84_9	Lake Whatcom	3	8	95.9	3	84.1
WF1_84_11	Lake Whatcom	4	10	100	3	84.1
WF3_101_1	Lake Whatcom	3	4	83.4	4	82.6
WF2_131_4	Lake Whatcom	3	5	100	4	82.6
WF2_131_6	Lake Whatcom	3	5	100	4	82.6
WF1_84_12	Lake Whatcom	3	8	95.9	5	97.6
WF2_131_3	Lake Whatcom	2	1	99.4		-
WF1_84_8	Lake Whatcom	4	10	100		-
ParPoint_25_1	Lewis River	2	1	99.4	3	84.1
LewPark2_26_1	Lewis River	1	9	81.1	6	100
LewPark3_26_2	Lewis River	1	9	81.1	6	100

Table 3 Continued.

ParPoint_25_2	Lewis River	3	9	81.1	8	91.6
ParPoint_25_3	Lewis River	3	9	81.1	8	91.6
ParPoint_25_4	Lewis River	3	9	81.1	8	91.6
ParPoint_25_5	Lewis River	3	9	81.1	8	91.6
ParPoint_25_7	Lewis River	3	9	81.1	8	91.6
DBPark4_28_1	Lewis River	3	11	100	9	87.4
DBPark3_27_1	Lewis River	4	15	100		-
DBPark3_27_2	Lewis River	4	15	100		-
DBPark4_36_1	Lewis River	3		-		-
NI2_89_1	Nisqualla River	1	9	81.1	6	100
NK3_43_1	Nooksack River	3	5	100	4	82.6
NK3_43_2	Nooksack River	3	5	100	4	82.6
NK2_128_1	Nooksack River	3		-	10	84.4
PU5_60_1	Puyallup River	1	9	81.1	6	100
PU5_60_10	Puyallup River	1	9	81.1	6	100
PU5_60_11	Puyallup River	1	9	81.1	6	100
PU5_60_12	Puyallup River	1	9	81.1	6	100
PU5_60_13	Puyallup River	1	9	81.1	6	100
PU5_60_14	Puyallup River	1	9	81.1	6	100
PU5_60_2	Puyallup River	3	9	81.1	6	100
PU5_60_3	Puyallup River	1	9	81.1	6	100
PU5_60_4	Puyallup River	1	9	81.1	6	100
PU5_60_6	Puyallup River	1	9	81.1	6	100
PU2_108_1	Puyallup River	1	9	81.1		-
PU3_117_2	Puyallup River	4		-		-
PWA1_7_1	Puyallup River	4		-		-
SA5_67_2	Sandy River	3		-	3	84.1
POR3_13_1	Sandy River	3	4	83.4	4	82.6
POR3_6_2	Sandy River	3	4	83.4	4	82.6
SA5_67_1	Sandy River	1	9	81.1	6	100
POR3_13_2	Sandy River	3		-		-
SA2_70_2	Sandy River	4		-		-
POR1_10_3	Santiam River	1	9	81.1	6	100
SK5_85_1	Skagit River	4	10	100	1	95.9
SK1_120_1	Skagit River	3		-	2	84.1
SK5_85_3	Skagit River	3	6	99.6		-
SK5_85_2	Skagit River	4	10	100		-
SY1_50_4	Skykomish River	2	1	99.4	3	84.1
SY2_46_13	Skykomish River	4	12	99.6	3	84.1
SN2_59_1	Skykomish River	3		-	4	82.6
SY2_46_12	Skykomish River	1	9	81.1	6	100
SY6_37_3	Skykomish River	3	11	100	9	87.4
SY1_50_2	Skykomish River	4		-	10	84.4

Table 3 Continued.

SN6_55_2	Skykomish River	3	6	99.6	-	-
SY6_37_5	Skykomish River	3	7	94.2	-	-
SY6_37_6	Skykomish River	1	9	81.1	-	-
SY2_46_3	Skykomish River	4	12	99.6	-	-
SY2_46_5	Skykomish River	4	12	99.6	-	-
SY2_46_7	Skykomish River	4	12	99.6	-	-
SY1_50_5	Skykomish River	4	13	100	-	-
SY1_50_6	Skykomish River	4	13	100	-	-
SY1_50_7	Skykomish River	4	13	100	-	-
PWA2_8_1	Skykomish River	4		-	-	-
SY1_50_1	Skykomish River	4		-	-	-
SY1_50_3	Skykomish River	4		-	-	-
SY5_45_1	Skykomish River	4		-	-	-
SN6_55_1	Skykomish River	3		-	-	-
SN6_55_4	Skykomish River	4		-	-	-
TO1_99_1	Toutle River	4		-	2	84.1
TO9_93_1	Toutle River	4		-	2	84.1
TO9_93_4	Toutle River	3	8	95.9	5	97.6
TO1_99_3	Toutle River	1	9	81.1	6	100
TO1_99_4	Toutle River	1	9	81.1	6	100
TO1_99_5	Toutle River	1	9	81.1	6	100
TO1_99_6	Toutle River	1	9	81.1	6	100
TO1_99_8	Toutle River	1	9	81.1	6	100
TO1_99_9	Toutle River	1	9	81.1	6	100
TO3_32_1	Toutle River	1	9	81.1	6	100
TO4_90_2	Toutle River	1	9	81.1	6	100
TO4_90_3	Toutle River	1	9	81.1	6	100
TO4_90_4	Toutle River	1	9	81.1	6	100
TO4_90_5	Toutle River	3	9	81.1	6	100
TO3_32_2	Toutle River	1	9	81.1	7	88.9
TO3_32_3	Toutle River	1	9	81.1	7	88.9
TO3_32_7	Toutle River	1	9	81.1	7	88.9
TO3_32_8	Toutle River	1	9	81.1	7	88.9
TO3_32_9	Toutle River	1	9	81.1	7	88.9
TO5_107_1	Toutle River	3		-		-
WH6_64_1	White River	3		-	2	84.1
WH6_64_2	White River	4		-	2	84.1
WH6_64_4	White River	4		-	2	84.1
WH6_64_5	White River	4		-	2	84.1
WH5_56_13	White River	2	1	99.4	3	84.1
WH5_56_14	White River	2	1	99.4	3	84.1
WH5_56_19	White River	2	1	99.4	3	84.1
WH5_56_2	White River	2	1	99.4	3	84.1

Table 3 Continued.

WH5_56_3	White River	2	1	99.4	3	84.1
WH5_56_4	White River	2	1	99.4	3	84.1
WH5_56_5	White River	2	1	99.4	3	84.1
WH5_56_6	White River	2	1	99.4	3	84.1
WH5_56_8	White River	2	1	99.4	3	84.1
WH6_64_6	White River	2	1	99.4	3	84.1
WH4_58_6	White River	1	9	81.1	6	100
WH2_54_10	White River	4		-	6	100
WH4_58_4	White River	2	1	99.4		-
WH5_56_11	White River	2	1	99.4		-
WH5_56_7	White River	2	1	99.4		-
WH5_56_9	White River	2	1	99.4		-
WH1_62_11	White River	4	14	100		-
WH1_62_13	White River	4	14	100		-
WH1_62_1	White River	4		-		-
WH1_62_10	White River	4		-		-
WH1_62_12	White River	3		-		-
WH1_62_14	White River	4		-		-
WH1_62_15	White River	4		-		-
WH1_62_2	White River	4		-		-
WH1_62_5	White River	4		-		-
WH1_62_6	White River	4		-		-
WH1_62_9	White River	4		-		-
WH2_54_4	White River	4		-		-
WH2_54_8	White River	4		-		-
WH4_58_3	White River	4		-		-
WH6_64_8	White River	4		-		-
WI7_83_9	Willamette River	2	1	99.4	1	95.9
WI1_102_3	Willamette River	2	1	99.4	3	84.1
WI1_102_4	Willamette River	2	1	99.4	3	84.1
WI1_102_7	Willamette River	2	1	99.4	3	84.1
WI1_102_8	Willamette River	2	1	99.4	3	84.1
WI1_102_9	Willamette River	2	1	99.4	3	84.1
WI7_83_10	Willamette River	2	1	99.4	3	84.1
WI7_83_11	Willamette River	2	1	99.4	3	84.1
WI7_83_13	Willamette River	2	1	99.4	3	84.1
WI7_83_3	Willamette River	2	1	99.4	3	84.1
WI7_83_6	Willamette River	2	1	99.4	3	84.1
WI7_83_7	Willamette River	2	1	99.4	3	84.1
WI7_83_8	Willamette River	2	1	99.4	3	84.1
WI5_138_1	Willamette River	3		-	4	82.6
WI1_102_10	Willamette River	3	1	99.4		-
WI7_83_2	Willamette River	2	1	99.4		-

Table 3 Continued.

WI4_133_2	Willamette River	4	4	83.4	-
WI18_80_1	Willamette River	2	8	95.9	-

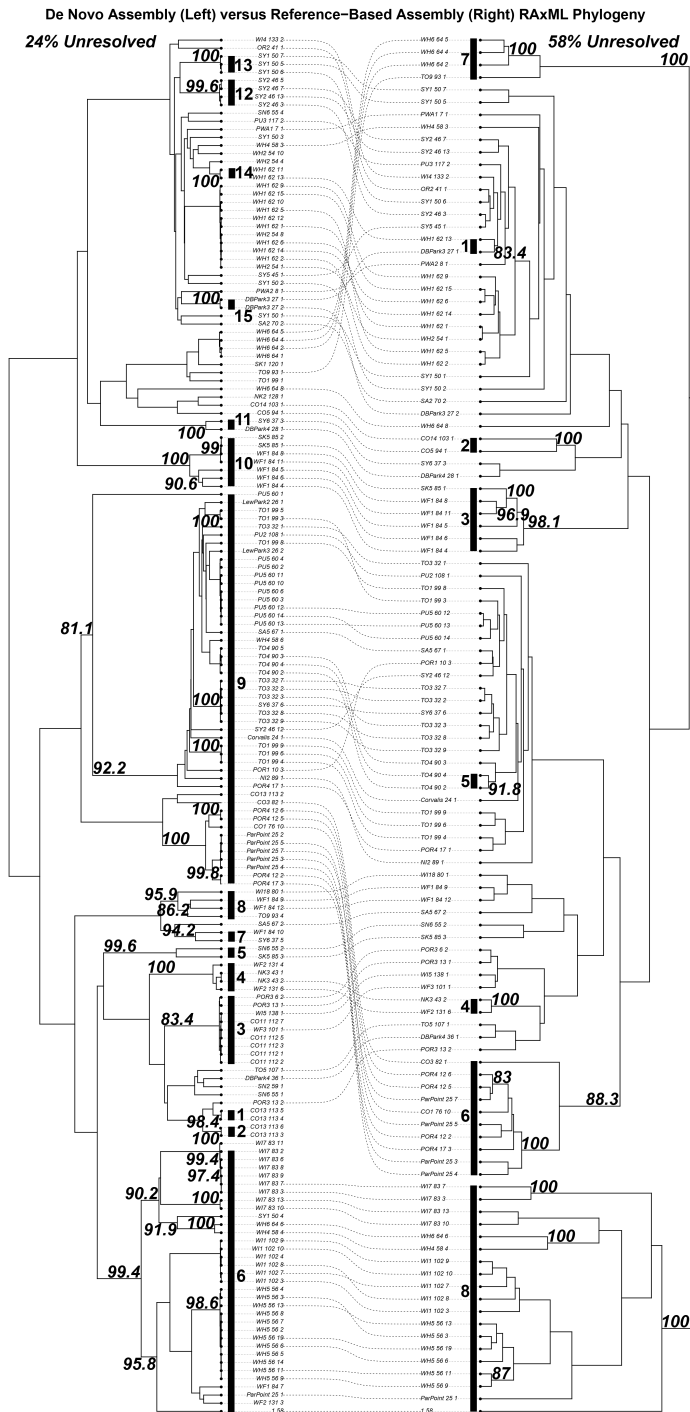


Figure 11: The RADseq *de novo* RAxML phylogeny directly compared to the reference-based RADseq phylogeny. Neither assembly retained the *Pseudococococcus floridanum* isolate BA4b018 due to insufficient read depth. The *de novo* assembly strongly resolved (>80% support) 76% of the retained 171 isolates into 15 clades with strongly supported substructure. The reference-based assembly resolved 42% of the retained 112 isolates into 8 clades.

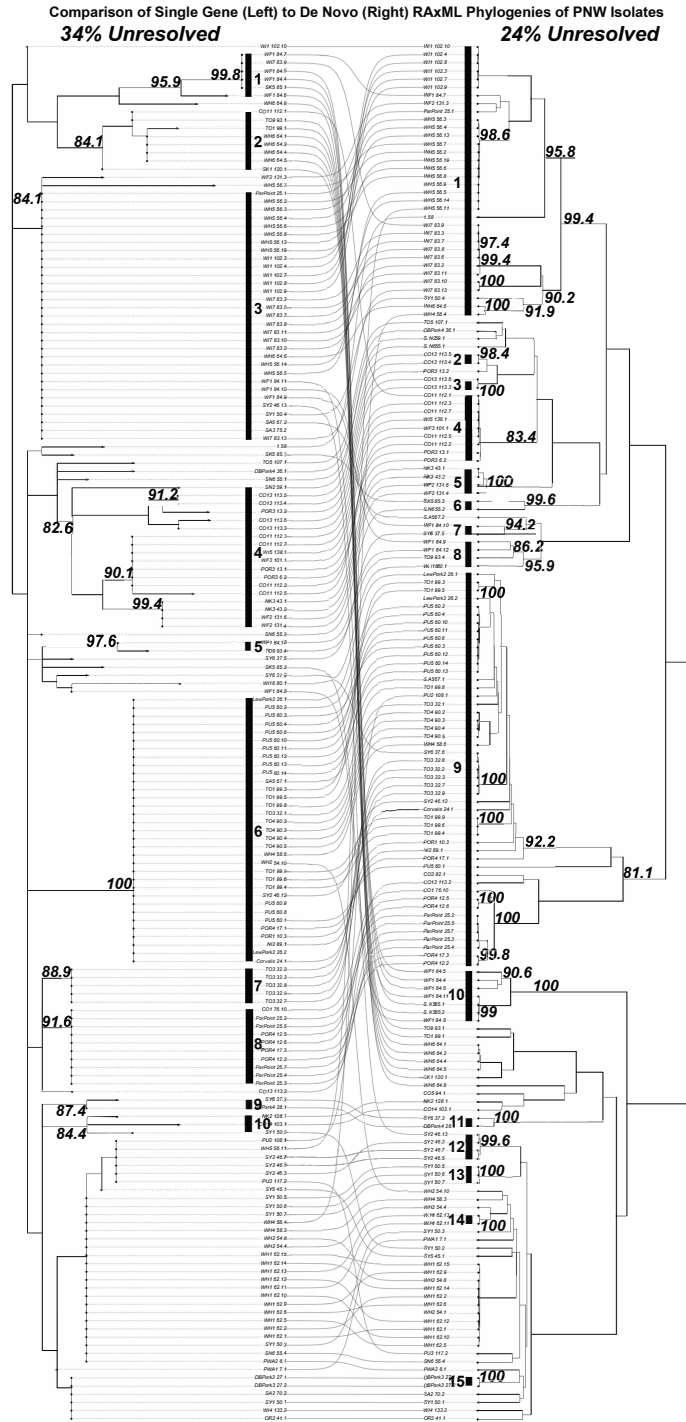


Figure 12: The single gene (GPD) RAxML phylogeny directly compared to the RADseq *de novo* assembly RAxML phylogeny. Of 171 retained isolates, a total of 34% were unresolved using GPD versus 24% unresolved using RADseq data. A greater level of sub-structure was also observed within clades DN1 and DN9 of the *de novo* assembly phylogeny when compared to the largest GPD3 and GPD6 clades. Clade GPD3 is predominantly grouped into clade DN1, and clades GPD6-8 are predominantly grouped into clade DN9.

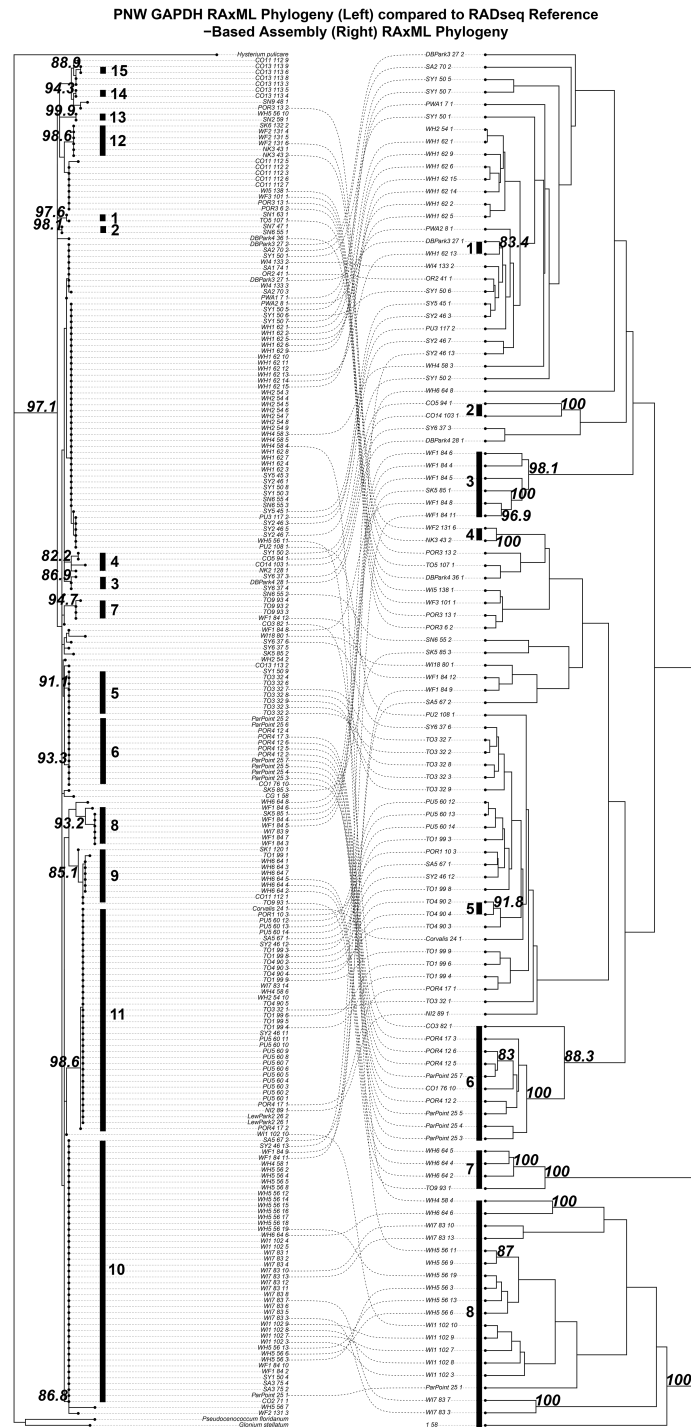


Figure 13: The single gene (GPD) RAxML phylogeny compared to the reference-based RADseq RAxML phylogeny. The GPD phylogeny strongly resolved (>80%) 15 clades and 66% of 171 isolates. The reference-based assembly only resolved 42% of the retained 112 isolates into 8 clades.

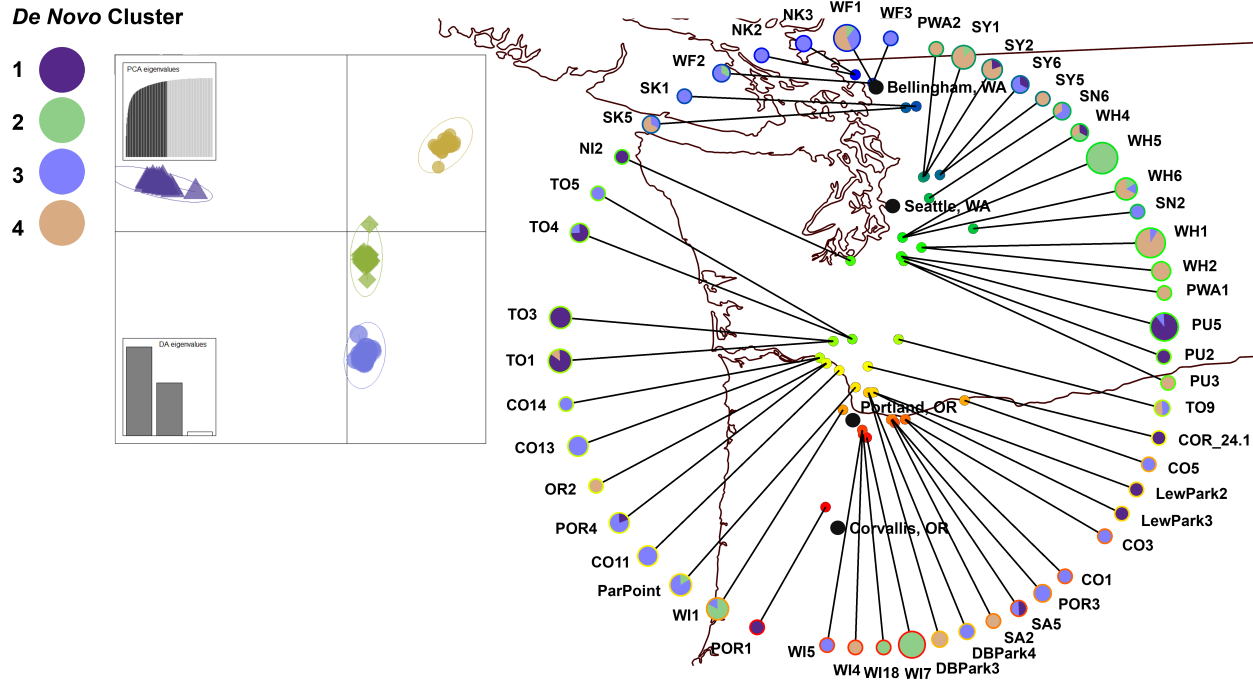


Figure 14: A discriminant analysis of principle components of the *de novo* assembly revealed four distinct phylogenetic clusters. When *Cenococcum geophilum* isolates were mapped to the original Pacific Northwest site of isolation, there were no patterns of segregation found within the clusters based on latitude. Many sites also contained more than one cluster, a pattern which occurred regardless of the size of the site-associated isolate collection. Sites of isolation are coded from blue to red running North to South, and the site-specific color is indicated in the circular outline of each site cluster pie chart.

Index of Association

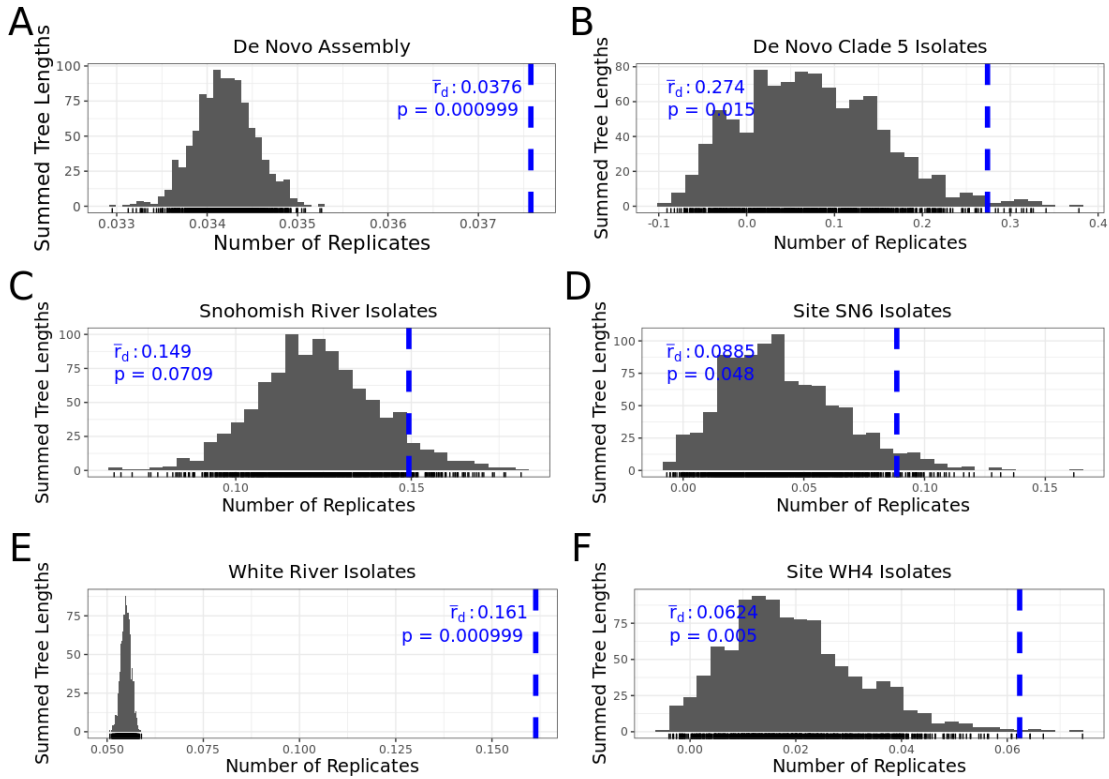


Figure 15: The Index of Association (I_a) was calculated for each *de novo* assembly cluster, clade, and Pacific Northwest site of origin with at least three isolates. There was no evidence of recombination within the full *de novo* assembly ($\bar{r}_d = 0.0376$, $p < 0.001$) (A). *De novo* clade 5 fell within the expected distribution with no allelic linkage ($\bar{r}_d = 0.274$, $p < 0.01$), indicating evidence of recombination (B). The Snohomish rivershed isolate collection fell within the expected distribution, however the results were not significant ($\bar{r}_d = 0.149$, $p = 0.07$) (C). In contrast, Snohomish river site six (SN6) fell within the expected distribution ($\bar{r}_d = 0.0885$, $p < 0.05$) (D). The White River collection fell outside of the expected distribution, ($\bar{r}_d = 0.161$, $p < 0.001$) (E), while White River site WH4 fell within the expected distribution ($\bar{r}_d = 0.0624$, $p < 0.01$), implicating a clonal population with site-specific recombination (F).

Table 4: Pacific Northwest (PNW) *Cenococcum geophilum* isolates used in this study with the rivershed of origin and single gene (GPD) clade. Unresolved isolates are designated with a dash (-). In total, 55 isolates were used representing 10 PNW riversheds and Lake Whatcom, along with the reference 1.58 strain.

PNW Strain ID	Isolation River Site	PNW Clade
1.58	<i>Reference</i>	-
CO5_94.1	Columbia River	4
OR2_41.1	Columbia River	-
POR4_12.5	Columbia River	6
POR4_12.6	Columbia River	6
POR4_17.3	Columbia River	6
WF1_84.12	Lake Whatcom	7
WF1_84.4	Lake Whatcom	8
DBPark3_27.1	Lewis River	-
DBPark3_27.2	Lewis River	-
DBPark4_28.1	Lewis River	3
LewPark2_26.1	Lewis River	11
LewPark3_26.2	Lewis River	11
ParPoint_25.2	Lewis River	6
PU5_60.14	Puyallup River	11
PU5_60.4	Puyallup River	11
POR3_13.1	Sandy River	-
POR3_6.2	Sandy River	-

Table 4 Continued.

SA3_75.2	Sandy River	10
SA5_67.1	Sandy River	11
SK1_120.1	Skagit River	9
SK5_85.3	Skagit River	-
PWA2_8.1	Skykomish River	-
SY1_50.6	Skykomish River	-
SY1_50.7	Skykomish River	-
SY1_50.8	Skykomish River	-
SY2_46.1	Skykomish River	-
SY2_46.12	Skykomish River	11
SY2_46.5	Skykomish River	-
SY6_37.3	Skykomish River	3
SY6_37.5	Skykomish River	-
SN2_59.1	Snohomish River	13
SN6_55.1	Snohomish River	2
SN9_48.1	Snohomish River	-
TO3_32.2	Toutle River	5
TO3_32.6	Toutle River	5
TO3_32.8	Toutle River	5
TO4_90.2	Toutle River	11

Table 4 Continued.

WH1_62.6	White River	-
WH2_54.10	White River	11
WH2_54.4	White River	-
WH2_54.6	White River	-
WH2_54.8	White River	-
WH4_58.4	White River	-
WH4_58.6	White River	11
WH5_56.12	White River	-
WH5_56.19	White River	-
WH5_56.3	White River	10
WH5_56.4	White River	10
WH5_56.5	White River	10
WH5_56.9	White River	10
WH6_64.2	White River	9
WH6_64.4	White River	9
WH6_64.8	White River	-
WI7_83.3	Willamette River	10
WI7_83.9	Willamette River	8

Table 5: Twenty significant associations were detected in the presence of cadmium. The *Cenococcum geophilum* scaffold number and coding region base pair start and end are indicated. The majority of the associated gene coding regions include proteins involved in metabolic processes including protein transport, nucleic transcription and translation, and heme binding

Scaffold	SNP ID	Original / Alternate Nucleotide	Number of Isolates (% of Population)	Pfam Annotation	Benjamin and Hochberg False Discovery Rate
scaffold_46:951514-951573	c860:7\C,\T	C / T	32 (57%)	PF01565FAD binding domain PF08031 Berberine and berberine like	0.03
scaffold_13:1829642-1829701	c2968:50\G,\A	G / A	36 (64%)	PF00067 Cytochrome P450	0.02
scaffold_13:1829642-1829701	c2968:46\A,\C	A / C	36 (64%)	PF00067 Cytochrome P450	0.02
scaffold_10:247053-248702	c3323:26\G,\A	G / A	41 (73%)	PF04909 Amidohydrolase	0.02
scaffold_2:1257909-1257968	c3682:10\G,\A	G / A	33 (59%)	PF00316 Fructose-1-6-bisphosphatase, N-terminal domain	0.03
scaffold_7:3260814-3260873	c3868:22\G,\A	G / A	35 (63%)	PF06985 Heterokaryon incompatibility protein (HET)	0.02
scaffold_110:159394-159453	c4109:25\C,\A	C / A	30 (54%)	PF07690 Major Facilitator Superfamily	0.05
scaffold_15:1880211-1880270	c12480:55\C,\T	C / T	30 (54%)	PF02731 SKIP/SNW domain	0.04
scaffold_46:1241403-1241462	c13048:16\A,\G	A / G	28 (50%)	PF02185 Hrl repeat PF00130 Phorbol esters/diacylglycerol binding domain (C1 domain) PF00069 Protein kinase domain PF00433 Protein kinase C terminal domain	0.02
scaffold_32:1324203-1324260	c15327:6\C,\T	C / T	40 (71%)		0.02
scaffold_14:1331013-1331072	c16995:14\T,\C	T / C	30 (54%)		0.04
scaffold_46:807302-809375	c17801:54\C,\T	C / T	32 (57%)	PF00749 tRNA synthetases class I (E and Q), catalytic domain	0.04
scaffold_28:270362-270421	c18237:53\G,\A	G / A	32 (57%)	PF08123 Histone methylation protein DOT1	0.02
scaffold_88:423132-423191	c22272:42\C,\T	C / T	28 (50%)	PF02985 HEAT repeat	0.02
scaffold_2:4601118-4601177	c22932:39\T,\G	T / G	26 (46%)	PF04109 Autophagy protein Apg9	0.03
scaffold_2:4600251-4601543	c23056:4\T,\C	T / C	26 (46%)	PF06792 Uncharacterised protein family (UPF0261)	0.02
scaffold_48:278028-278085	c27219:37\T,\C	T / C	39 (70%)	PF02809 Ubiquitin interaction motif PF00646 F-box domain	0.04

Table 5 Continued.

scaffold_6:323153-323212	c33099:34\C,\T	C / T	31 (55%)	PF00149 Calcineurin-like phosphoesterase	0.03
scaffold_29:1482286-1482345	c84991:27\A,\G	A / G	44 (79%)	PF00550 Phosphopantetheine attachment site PF07993 Male sterility protein	0.05
scaffold_42:1381529-1381588	c176737:48\T,\C	T / C	28 (50%)	PF00501 AMP-binding enzyme	0.02

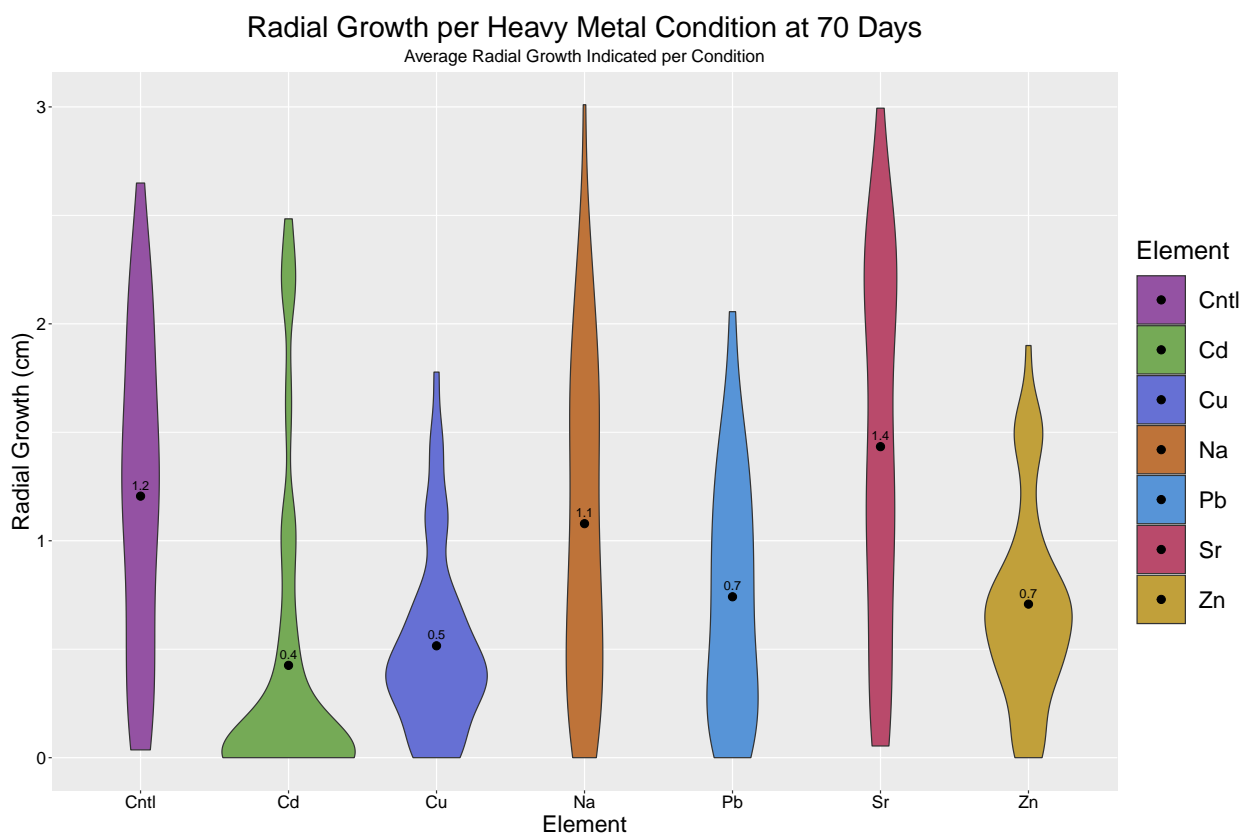


Figure 16: Average radial growth in cm at 70 days of all *Cenococcum geophilum* isolates used in this study per heavy metal condition. The average per condition is indicated with a black dot and value. All heavy metal conditions were at 25 ppm save for cadmium (Cd), which was at 0.5 ppm. All conditions significantly impacted average growth when compared to the control condition of each isolate. Generally isolates grew less in the presence of a heavy metal with cadmium (Cd) having the lowest average at 0.4 cm followed by copper (Cu) at 0.5 cm and lead (Pb) and zinc (Zn) at 0.7 cm. While sodium (Na) did have a significant impact on overall growth, the average was closest to the control at 1.1 cm. Isolates tended to grow as well or better in the presence of strontium (Sr) at 1.4 cm average radial growth.

Radial Growth of PNW Isolates per Treatment at 70 Days

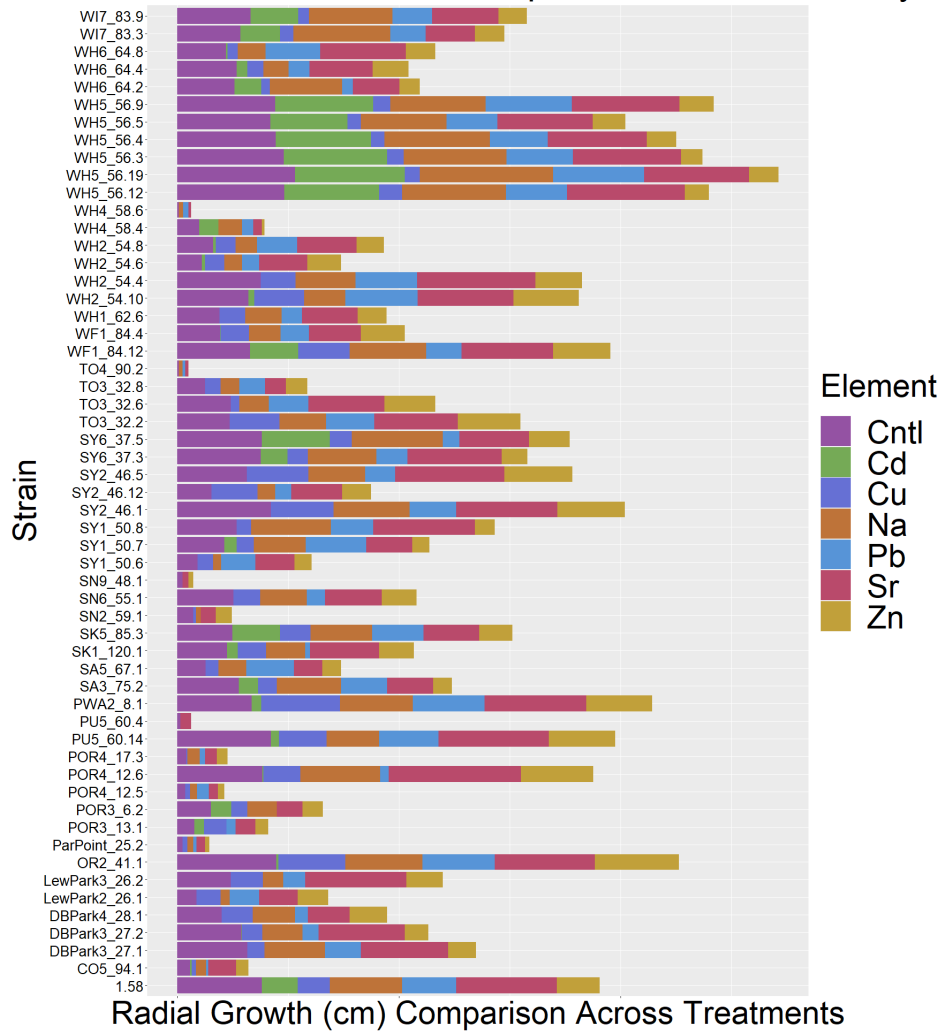


Figure 17: Average radial growth per *Cenococcum geophilum* isolate and heavy metal treatment at 70 days of growth. All heavy metal conditions were at 25 ppm save for cadmium (Cd), which was at 0.5 ppm. Generally isolates grew as much as the control treatment in the sodium (Na) treatment and as much or more in the strontium (Sr) treatment. Most isolates had decreased average radial growth in the cadmium (Cd), copper (Cu), lead (Pb), and zinc (Zn) treatments except for isolates which originated from either White River 5 isolation site (WH5) or Skagit River 5 isolation site (SK5).

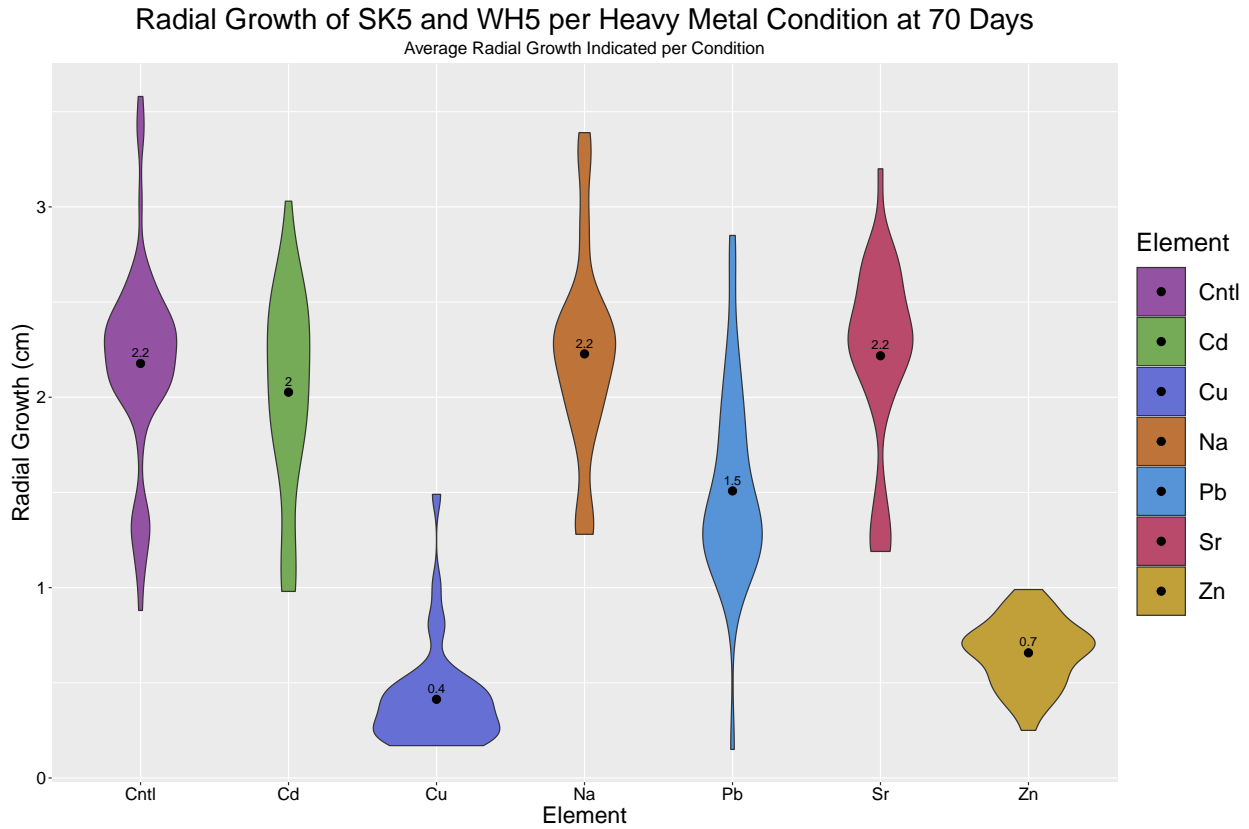


Figure 18: Average radial growth in cm at 70 days of *Cenococcum geophilum* isolates originating from either White River 5 isolation site (WH5) or Skagit River 5 isolation site (SK5). The average per condition is indicated with a black dot and value. All heavy metal conditions were at 25 ppm save for cadmium (Cd), which was at 0.5 ppm. All conditions significantly impacted average growth when compared to the control condition of each isolate. Isolates grew less than the control in both copper (Cu) and zinc (Zn) conditions at 0.4 cm and 0.7 cm respectively, but nearly equal average growth in the 0.5 ppm Cd condition in contrast to the average of the entire isolate collection at 0.4 cm. Higher average radial growth was also seen in the sodium (Na), lead (Pb), and strontium (Sr) conditions at 2.2 cm, 1.5 cm, and 2.2 cm respectively.

Vita

Jessica M. Vélez was born in 1983 in El Paso, Texas to Raúl Vélez and Evelyn Nazario. During her childhood, she lived in North Carolina, Puerto Rico, and Texas. She completed a Bachelor of Arts in English at the University of Texas at Austin in 2004 and moved to Knoxville, Tennessee. Here, she returned to school and completed an second Bachelor of Science in Microbiology at the University of Tennessee Knoxville, and began an internship at Oak Ridge National Laboratory. In 2015, she entered into the Bredesen Center for Interdisciplinary Studies Energy Science and Engineering PhD program at UTK, and completed her doctoral degree in December of 2020.