Reciprocal Informants: Using Fungal Bioinformatics, Genomics, and Ecology to tie Mechanisms to Ecosystems

A DISSERATION

SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

BY

Lotus Alicia Lofgren

IN PARTIAL FULFILLMENT OF THE REQUIERMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advised by
Dr. Peter G. Kennedy

AUGUST 2019

Acknowledgments

First, I want to acknowledge my family: My mom Char who taught me to question everything. My sister Ali who makes me laugh and inspires me daily with her tenacity, humor, and love. My late father Brendon, who supported me without question and was, without question, the bravest human I have ever known. Especially, I want to thank my amazing husband Clayton Summitt. I could not have gotten through the last few years without Clayton's ability to calm me with both logic and love, without him picking up slack in our domestic lives, without his flexibility, bad puns, and statistics lessons. You are my sunshine Clay.

I'm grateful to my friend Sean Baker and my undergraduate mentors Drs., Kabir Peay, Jenny Bhatnagar, Dave McLaughlin, Rosanne Healy, my late undergraduate advisor Bud Markhart, who collectively set me on this path in the first place. To the University of Minnesota, my cohort, the CBS grad program, and the Plant and Microbial Biology Department, including a multitude of professors who have challenged me intellectually and academically, as well as support staff like Sara Eliason and Gail Kalli who made my life infinitely easier. Thanks to all of the officers, volunteers, members, and public participants involved with Mycology Club at the University of Minnesota. It has been my extreme pleasure to be involved in the founding of such a tenacious group with such a tangible positive impact on campus and in the community. I've learned more about leadership, organization, and education from you all than I can say. I'm forever grateful to the Mycology community at large for welcoming me since day one, and for

rewriting my preconceived notions about what the academy is and could be. Special thanks to Sara Bronco for her personal and career mentorship and to Jessie Uehling who first encouraged me to learn bioinformatics and remains a constant inspiration to me.

To my follow Kennedy Lab members past and present; Dr.'s Christopher Fernandez, Luke McCormack, Nhu Nguyen, and Lauren Cline, as well as Craig See, Louis Mielke, Erin Andrews, with special thanks to Amanda Certano, and many others who have made the Kennedy lab a supportive, rigorous, and collaborative community to learn and grow in. To my committee members, Dr.'s Jonathan Schilling and Kathryn Bushley, who have given me invaluable feedback throughout this process, with special thanks to Dr. H. Corby Kistler, who was also my postbac advisor before starting graduate school, and who has deeply influenced the way that I think about science and the world. Finally, I want to thank my primary advisor, Dr. Peter G. Kennedy, who took a chance on me when there were many reasons to give pause, and who has supported, advocated for, challenged and inspired me every day for the last five years. Thank you for choosing me as your first graduate student, giving me the opportunity to learn here, and the freedom to fail, grow, and find my own path. It's been an honor to learn from you.

Chapter Contributions:

Chapter 1:

Thanks to D. McLaughlin for his documentation of the *S. subaureus* population at Lake Alexander SNA, Y. Han for assistance in processing the bioassays, G. Celio at

the UNM Imaging Center for assistance with the Hartig net micrographs, members of the Kennedy lab, and F. Martin for constructive comments on previous versions of the manuscript, and N. Nguyen and P. Kennedy, who contributed to writing a published version of this research.

Chapter 2:

Thanks to R. Vilgalys, N. Nguyen, J. Ruytinx, H-L Liao, and S. Bronco with The *Suillus Consortium* for use of genome sequences, encouragement and project feedback, staff at the Minnesota Super Computing Institute for computational assistance, H. Nielsen with the Signal P team for his bug fix mastery, and K.Barry, A. Kuo, and I.Grigoriev at the Joint Genome Institute for their feedback on project design and assistance accessing genome data.

Chapter 3:

Thanks to R. Vilgalys for sharing cultures and access to the unpublished genomic data of four fungal species, A. Gryganskyi for use of unpublished genomic data for three species of early diverging fungi, R. de Vries, M. Mäkelä and K. Hildén for use of unpublished genomic data for the *D. squalens* sequencing projects, I. Grigoriev and the JGI MycoCosm team for assistance accessing fungal genomes, F. Keck for assistance with using Phylosignal, and J.Uehling, S. Branco, T. Bruns, F. Martin and P. Kennedy, who contributed to a published version of this research.

Funding and Resource Contributions :

I am grateful to have received funding for this work in the form of a National Science Foundation Graduate Research Fellowship (Grant No. 00039202), a Doctoral Dissertation Fellowship from the University of Minnesota's graduate school, Summer funding from the Department of Plant and Microbial Biology, sequencing support from The Joint Genome Institute's 1000 Fungal Genomes Project, and travel support from the Department of Plant and Microbial Biology, the Microbial and Plant Genomics Institute, The Mycological Society of America, and the International Mycological Congress.

Dedication

I dedicate this work to the feeling of being very small and inconsequential when confronted with something beautiful, complicated, or vast. I do not have a word for this feeling, but doing science provides it regularly. It is my favorite feeling next to love.

Abstract

Across both wild and human-structured ecosystems, fungi interact with every plant species on earth. From mycorrhizal mutualisms, harmless endophytes, and deadly pathogens, the results of these interactions can mean the difference between a plant's ability to grow and flourish, or languish and expire. Fungal-host dynamics are not static traits, either over evolutionarily time or during the lifetime of individuals where ecological context dependency shapes the outcomes of fungal-host interactions. Understanding the ecological and genetic factors that structure plant-fungal relationships has wide ranging consequences for ecosystems, agro-ecosystems, and human health. However, it's not well understood how complex genetic mechanisms and ecological pressures work in concert to structure the outcomes of fungal-host interactions, particularly among fungal mutualists. This dissertation contributes to this understanding by investigating how fungal-host relationships are regulated at two levels: **broadly**, investigating the ecology of fungal-host systems, and **specifically**, investigating the genetic and genomic basis of how these interactions are mediated.

I begin Chapter 1 from the perspective of fungal ecology, investigating the influence of neighborhood (the surrounding plant community) on host specificity patterns using the host-specialist ectomycorrhizal (ECM) genus *Suillus*. The number of host species that a given fungal species will associate with, and how closely related these host species are, is the study of fungal host specificity. While some fungi associate with only a single species of host (high host specificity), most associate with tens or hundreds of host species (low host specificity). Fungi in the genus *Suillus* are famous for their high host

specificity, primarily associating with plants in the family Pineaceae (particularly White Pines, Red Pines and Larchs). Using a combination of field sampling, sequencing, and colonization bioassays, I present evidence that one species, *S. subaureus*, has undergone a novel host-expansion onto Angiosperms, and argue that neighborhood effects influence ECM colonization outcomes over both space and time. In Chapter 2, I expand from fungal ecology into fungal genomes. Using genome mining and comparative genomics, I look for signatures of ECM host specificity using 19 genome sequenced *Suillus* species in relation to 1) other (non-*Suillus*) ECM fungi and 2) an intrageneric comparison between *Suillus* that specialize on Red Pine, White Pine or Larch. I present evidence for the involvement of several molecular classes in regulating *Suillus* host specificity including species specific small secreted proteins, G-protein coupled receptors, and terpene secondary metabolites. Finally, in Chapter 3, I use the genomic and bioinformatic tool sets developed in Chapters 1 and 2, to expand my analysis across the fungal phylogeny and ask questions about a potential molecular correlate of fungal guild and trophic mode: ribosomal DNA (rDNA) copy number. To do this, I developed a bioinformatic pipeline to estimate rDNA copy number variation from whole genome sequence data, and applied it to a phylogenetically and ecologically diverse set of 91 fungal genomes. I present evidence that rDNA copy number is inversely associated phylogenetic distance, but displays a high level of variation, spanning an order of magnitude in *Suillus* alone, with no detectable correlation to guild occupation or genome size. Taken together, the work presented here shows that genomic and bioinformatic approaches used in concert with classical ecological methodologies, offer great potential to expand our understanding of

the two-way influence of ecosystem-level processes and gene-level mechanisms in structuring plant-fungal interactions.

Table of Contents

List of Figures

Chapter 1: Ectomycorrhizal host specificity in a changing world: can legacy effects explain anomalous current associations?

Synopsis

Despite the importance of ectomycorrhizal (ECM) fungi in forest ecosystems, knowledge about the ecological and co-evolutionary mechanisms underlying ECM host associations remains limited. Using a widely distributed group of ECM fungi known to form tight associations with trees in the family Pinaceae, we characterized host specificity among three unique *Suillus*-host species pairs using a combination of field root tip sampling and experimental bioassays. We demonstrate that the ECM fungus *S. subaureus* can successfully colonize *Quercus* hosts in both field and glasshouse settings, making this species unique in an otherwise Pinaceae specific clade. Importantly, however, we found that the colonization of *Quercus* by *S. subaureus* required co-planting with a Pinaceae host. While our experimental results indicate that gymnosperms are required for the establishment of new *S. subaureus* colonies, Pineaceae hosts are locally absent at both our field sites. Given the historical presence of Pineaceae hosts before human alteration, it appears the current *S. subaureus* - *Quercus* associations represent carryover from past host presence. Collectively, our results suggest that patterns of ECM specificity should be viewed not only in light of current forest community composition, but also as a legacy effect of host community change over time.

Introduction

There is widespread recognition that both plant health and ecosystem functioning are strongly influenced by symbiotic interactions with microorganisms (Van Der Heijden

et al., 2008). In many forest soils, trees form close associations with ectomycorrhizal (ECM) fungi, which facilitate nutrient and water acquisition in exchange for photosynthetically derived sugars (Smith & Read, 2008). Unlike associations between plants and other microbial groups (e.g. arbuscular mycorrhizal fungi or nitrogen fixing bacteria), a considerable number of ECM fungi exhibit strong patterns of host specificity (Molina *et al.*, 1992a). This specificity is most often observed at the level of host family or genus and involves a diverse array of plant lineages (e.g. *Alnus* (Molina, 1979), *Pisonia* (Hayward & Horton, 2012), *Gnetum* (Tedersoo & Põlme, 2012), Pinaceae (Bruns *et al.*, 2002)). Despite some informed speculation (Kropp & Trappe, 1982; Bruns *et al.*, 2002; Walker *et al.*, 2014), our current understanding of the ecological and co-evolutionary mechanisms underlying ECM host association patterns remains limited. Plant control of colonization by ECM fungi may take place at multiple stages of mycorrhization, including spore germination, directed mycelial growth, plant-fungal contact, during formation of the Hartig net, or post mycorrhization in response to nutrient transfer (Fries, 1984; Duddridge, 1986; Ditengou *et al.*, 2015; Hortal *et al.*, 2017). Although both spore and mycelial colonization are thought to occur in response to host-initiated molecular triggers, spores and mycelia likely require distinct molecular signals in order for colonization to occur and it is likely that a plant's ability to trigger spore germination is independent of the ability to ultimately form functional mycorrhizas with a given fungal species (Palm & Stewart, 1984; Kikuchi *et al.*, 2007; Ishida *et al.*, 2008). Because signaling molecule quantity and quality are dependent on host identity (Palm & Stewart, 1984; Massicotte *et al.*, 1994), forest community composition has important ecological consequences for host specificity.

The process of mycorrhization often occurs in the context of many potential host plants. Deviations from expected host specificity patterns may be mediated by either a mycelial- or spore-based mechanism, both of which may be influenced by third-party organisms. For example, the potential for alternative host associations can arise when ECM fungi already established on a primary host simultaneously colonize a second host via mycelial networks, or when proximal plants or microbial organisms trigger spore germination in ECM species that would otherwise exhibit dormancy (Fries, 1984; Hubert & Gehring, 2008; Bogar & Kennedy, 2013; Bogar *et al.*, 2015). The ability of proximal trees to influence ECM community composition has already been documented as an example of how neighborhood effects can act as an important mediator of host-symbiont interactions (Bogar & Kennedy, 2013). However, extending the context of plant-microbe interactions to encompass all extant community members, may still fall short of encompassing the causal agents responsible for patterns of ECM host association.

Legacy effects (defined here as the long-term influence of a species after its local extinction) include anthropogenic disturbance events which can alter community dynamics many years after an event took place (Cuddington, 2013). In forest ecosystems, land use histories are important determinants of both community structure and function, with far reaching effects on both plants and microbes (Goodale & Aber, 2001; Foster, 2006; Fraterrigo *et al.*, 2006). In multi-host stands, disturbance events such as fire, disease, and logging can facilitate the asymmetric removal of a given host species (Metz *et al.*, 2012; Hollingsworth *et al.*, 2013; Covey *et al.*, 2015) which may open new niche

space for existing hosts as well as put significant pressure on host-specific fungi to associate with non-primary host trees.

Suillus is one of the most well-known examples of an ECM fungal lineage that exhibits a high degree of host specificity (Dahlberg & Finlay, 1999). Suilloid fungi are noted for their close associations with trees in the family Pinaceae (Molina *et al.*, 1992b; Kretzer *et al.*, 1996; Horton & Bruns, 1998; Horton *et al.*, 2005; Nguyen *et al.*, 2016a). Unlike many other ECM fungal lineages, *Suillus* species possess both reactive spores (i.e. those that readily germinate in the presence of compatible host roots) (Fries, 1978) and long-distance rhizomorphic mycelium (Agerer, 2001). This combination of traits makes them capable of readily colonizing host roots using either spore germination or mycelial extension from established ectomycorrhizas. Control of host specificity in *Suillus* spp. may occur at both of these stages, although most experimental tests have only been conducted via mycelial colonization (Molina & Trappe, 1982; Duddridge, 1986; Finlay, 1989; but see Liao *et al.*, 2016). For example, in field settings, *S. grevillei* and *S. cavipes* associate exclusively with *Larix*, but will form ectomycorrhizas with *Pinus* hosts in laboratory settings (Finlay, 1989). However, the interaction with novel hosts in laboratory settings has been associated with abnormal cellular development and the accumulation of phenolic compounds as well as anomalies in host nutrient provisioning (Molina, 1979; Malajczuk *et al.*, 1982; Duddridge, 1986; Finlay *et al.*, 1988).

A single species of *Suillus*, *S. subaureus*, has long been cited as a possible exception to the tightly coupled relationship between *Suillus* and the Pinaceae. Sporocarp

collection records of *S. subaureus* often include site descriptions that note the absence of known Pinaceae hosts and, instead, the presence of angiosperm trees such as *Quercus* and *Populus* (Smith, A.H., Thiers, 1964; Homola & Mistretta, 1977; Kuo & Methven, 2010). Despite much speculation, to our knowledge, the natural host(s) of *S. subaureus* has never been confirmed. If *S. subaureus* associates with hosts outside the Pinaceae, it would represent either host switching or host expansion for a species deeply nested within a clade of host specialists (Kretzer *et al.*, 1996; Nguyen *et al.*, 2016a). Such an exception would provide an ideal system for inquiry into the genetic and molecular mechanisms mediating ECM specificity, including the level at which symbiosis is regulated (such as genetic vs epigenetic factors). Finally, understanding the ecological drivers of changing ECM host associations (including host switching or host expansion from specialist to generalist fungi or gymnosperm to angiosperm associates) could have important implications for understanding and predicting plant and fungal range shifts related to anthropogenic disturbance and global change (Dickie *et al.*, 2010; Pickles *et al.*, 2012; Hayward *et al.*, 2015).

In this study, we first investigated the hosts of *S. subaureus* observed at two geographically distant field sites and then, based on those associations, tested four hypotheses in a series of glasshouse bioassays. The first two bioassays, referred to as the *Angiosperm Spore Colonization Bioassay* and the *Gymnosperm Spore Colonization Bioassay*, were conducted to provide experimental evidence of either host expansion (i.e. colonization of multiple phylogenetically distant hosts) or host switching (i.e. colonization of hosts only in specific phylogenetic lineages) for *S. subaureus*. Based on

our working knowledge about this study system, we hypothesized 1) that the presence of angiosperm hosts alone would not be sufficient to trigger *S. subaureus* spore germination and thereby prevent mycorrhization, and 2) that the presence of ancestral Pinaceae hosts would be sufficient and/or necessary to trigger spore germination and thereby lead to mycorrhization. In the third bioassay, referred to as the *Mycelial Colonization Bioassay*, we tested the hypothesis that *S. subaureus* mycorrhization on alternative hosts (angiosperms) is possible, but only when the alternative host is co-planted with the primary host (Pinaceae). Finally, in the fourth bioassay, referred to as the *Primary Host Removal Bioassay*, we tested the hypothesis that removal of the primary host would facilitate angiosperm colonization by *S. subaureus*

Materials and Methods

Site descriptions, field sampling and species identification

Fieldwork was conducted at two locations in the midwestern United States. The first site, Lake Alexander Woods Scientific and Natural Area (SNA), was located in Cushing Minnesota, USA (46.158609N, -94.561718W, elevation c. 400 m). The mean annual temperature at the site is 4 °C (maximum of 33 °C in July and minimum of -32 °C in January) and the mean annual precipitation is *ca.* 700 mm, which comes mostly as rain during the spring and summer months. The predominant soil type is Alstad loam. At the time of sampling, the site was a *ca.* 70-year-old mixed deciduous forest in which conifer trees were locally absent (a single *Pinus strobus* sapling was present in the area, but was located >75 m from the nearest *S. subaureus* sporocarp collection). Overstory trees

included red oak (*Q. rubra*), paper birch (*Betula papyrifera*), big-tooth aspen (*P. grandidentata*), and sugar maple (*Acer saccharum*). The site is located in the 'Pine Moraines and Outwash Plains Subsection' of the Minnesota floristic designation, where *P. strobus* was a canopy dominant prior to intensive logging in the 19th century. Historic aerial photographs of the site (www.lib.umn.edu/borchert) confirm that logging events were common in the general area between 1939 (when the earliest photographs were taken for the area) and 1980. At the exact location where samples were collected, the most recent logging event appeared to have taken place prior to 1955. The second field site was located at Tolleston Dunes National Lakeshore in Hammond Indiana, USA (41.604623N, -87.439874W, elevation c. 180 m). The mean annual temperature is 8 °C (maximum of 34 °C in July and minimum of -22 °C in January) and mean annual precipitation is *ca.* 1128 mm. The predominant soil type is sand-silt from the Oakville-Adrian complex. The forest canopy was dominated by mature black oak (*Q. velutina*), paper birch (*B. papyrifera*) and cottonwood (*P. deltoides*). Conifer trees were locally absent at the time of collection, although they are present as part of mature 'Dune and Swale Complex' characteristic of the Great Lakes shoreline. Historic aerial photographs of the site (https://igs.indiana.edu/IHAPI) confirm that disturbance events (logging or periodic burning) were common in the area prior to 1973.

In late August 2014, nine *S. subaureus* sporocarps were collected from the MN site. Soil cores (15 x 15 x 15 cm) were taken directly under six of the *S. subaureus* sporocarps. Ectomycorrhizal root tips were sieved from the soil and individual ectomycorrhizas exhibiting a suilloid morphology (up to 6 per core) were extracted for

total genomic DNA using the REDExtract-N-Amp plant kit (Sigma-Aldrich St. Louis, MO, USA). From each sample, the fungal rRNA internal transcribed spacer (ITS) region was PCR amplified using the primer pair ITS1-F / ITS4 (White *et al*., 1990; Gardes and Bruns 1993) as well as a portion of the plant *trn*L chloroplast gene using the primers *trn*C / *trn*D (Taberlet *et al.*, 1991). Amplicons were cleaned using ExoSAP-IT (USB Corporation, Cleveland, OH, USA) and sequenced using single-pass Sanger sequencing with either ITS1-F (fungus) or *trn*C (plant) primers at the University of Arizona Genetics Core, USA. In early October 2016, three *S. subaureus* sporocarps were collected from the IN site, along with one soil core taken directly under a sporocarp of *S. subaureus*. Mushrooms and root tips were prepared and sequenced as above.

Angiosperm Spore Colonization Bioassay

Quercus rubra and *Q. macrocarpa* acorns were obtained (Sheffield's Seed Company, Locke, NY, USA), cupules were removed and the acorns were surface-sterilized in 10% bleach for 12 hours prior to being rinsed twice, placed into open plastic bags with moistened medium grade sand (10 ml sand / 30 acorns) and stratified at 4°C for 77 days. In September 2014, *P. tremuloides* and *P. grandidentata* roots were collected from the Cedar Creek Ecosystem Science Reserve in East Bethel, MN, USA. After removing tertiary and secondary roots, the primary root was trimmed to a length of 30 cm and packed in heat-sterilized peat moss. Shoots produced from primary roots (approximately 12 cm tall) were cut at the stem base, dipped in 1.6% indole butyric acid and rooted in sterilized sand for 30 days before transplanting.

Spores from the nine *S. subaureus* sporocarps were collected following the methods outlined in Kennedy *et al.* (2011) and stored in moistened sterile growth media at 4°C until use. Plant growth media consisting of a 2:2:1 mix of peat (no. 0128P, Premier Horticulture, Quakertown, PA, USA): forest soil (from the University of Minnesota St. Paul campus): sand (Monterrey no. 2/16; Cemex, Marina, CA, USA) was autoclaved for 90 minutes at 20 psi and 121°C for two consecutive days prior to adding fungal inoculum. Plant growth media was inoculated with *S. subaureus* spores at a concentration of 5 x 10^5 spores/ml soil. Small cone-tainers (150 ml capacity) were sterilized overnight in 10% NaOCl, rinsed, dried and stuffed with a small amount of synthetic pillow stuffing to keep plant growth media in place. Seedlings were randomly arrayed on benches at the University of Minnesota (UMN) Growth Facilities Greenhouse and grown under a 16 hour photoperiod, 24/21°C day/night, daily watering, and in the absence of fertilization (Fig. 1.1a).

Seedlings of *Q. rubra* and *Q. macrocarpa* (n = 20 per species per time point) were checked for evidence of colonization at three (92 days) and six months (185 days) after planting. *P. tremuloides* (n = 12) and *P. grandidentata* (n = 5) were destructively harvested and checked for evidence of colonization three months (92 days) after planting. The six month time point for the *Populus* species was not taken due to the small number of *Populus* cuttings that successfully rooted.

Gymnosperm Spore Colonization Bioassay

Mushrooms of *S. americanus* and *S. clintonianus* (previously known as *S. grevillei* in North America, (Nguyen *et al*., 2016)) were collected in the fall of 2014 from

multiple forests in Minnesota, USA underneath *Pinus strobus* and *Larix laricina*, respectively. The methods in this second bioassay matched those of the Angiosperm Spore Colonization Bioassay except where specified below. Spores of these two *Suillus* species were prepared from the fresh collections, whereas spores of *S. subaureus* for this second bioassay were from the same stock as above. Seeds of *P. strobus* and *L. laricina* (hereafter referred to as *Pinus* and *Larix*) were sourced from the Badoura State Forest Nursery (Minnesota Department of Natural Resources). A *Q. rubra* treatment was included as a negative control based on the results of the Angiosperm Spore Colonization Bioassay. *Q. rubra* acorns were collected from a parent tree located on the UMN St. Paul campus. *Pinus* and *Larix* seeds were surface-sterilized and stratified for 60 days at 4°C following Mujic *et al.* (2015). Stratified seeds were germinated in sterilized plant growth media and grown for 30 days prior to transplanting into 350 ml capacity cone-tainers. Individual cone-tainers were inoculated with either *S. americanus, S. clintonianus* or *S. subaureus* at a concentration of 5 x 10^5 spores/ml soil. Two seedlings were planted per cone-tainer, representing two plants of the same host (n = 6 pots / treatment = 12 plants / treatment) (Fig. 1.1b).

Plants were grown in a second UMN greenhouse under the following conditions: 16 hour photoperiod, 24/21°C day/night, daily watering, and in the absence of fertilization. Seedling location was randomized and periodically rotated throughout the experiment. After 158 to 180 days post-inoculation, all replicates with two living plants were harvested. Each replicate was removed from its pot, the root systems washed of soil and gently teased apart to separate the two plants. Each single root system was divided

into nine parts, randomized, and scored for % colonization with the aid of a 10X dissecting microscope. For *P. strobus* and *L. laricina* seedlings, 300 root tips were scored per plant unless less than 300 root tips were present, in which case all available root tips were scored. For *Q. rubra* seedlings, 1000 root tips were scored per plant due to the higher abundance of fine roots.

Mycelial Colonization Bioassay

Plants, growth media and fungal inoculum were prepared, grown, and harvested using the same methods and timeline (harvested 158 to 180 days after inoculation) as described above except that each pot was planted with combinations of two host species, with all host combinations represented (n = 9 pots / treatment = 9 plants / treatment) (Fig. 1.1c).

Primary Host Removal Bioassay

Cone-tainers (350 ml capacity) were co-planted with *P. strobus* and *Q. rubra* and inoculated as above with *S. subaureus* spores using the methods reported above. After six months (180 days) of growth, in half of the pots, *P. strobus* plants were hewn at the soil line, killing the seedling and removing the aboveground portion of the plant (n = 8 hewn and 8 unhewn). Plants were then grown for another 54 days before harvesting and scoring as above (Fig. 1.1d).

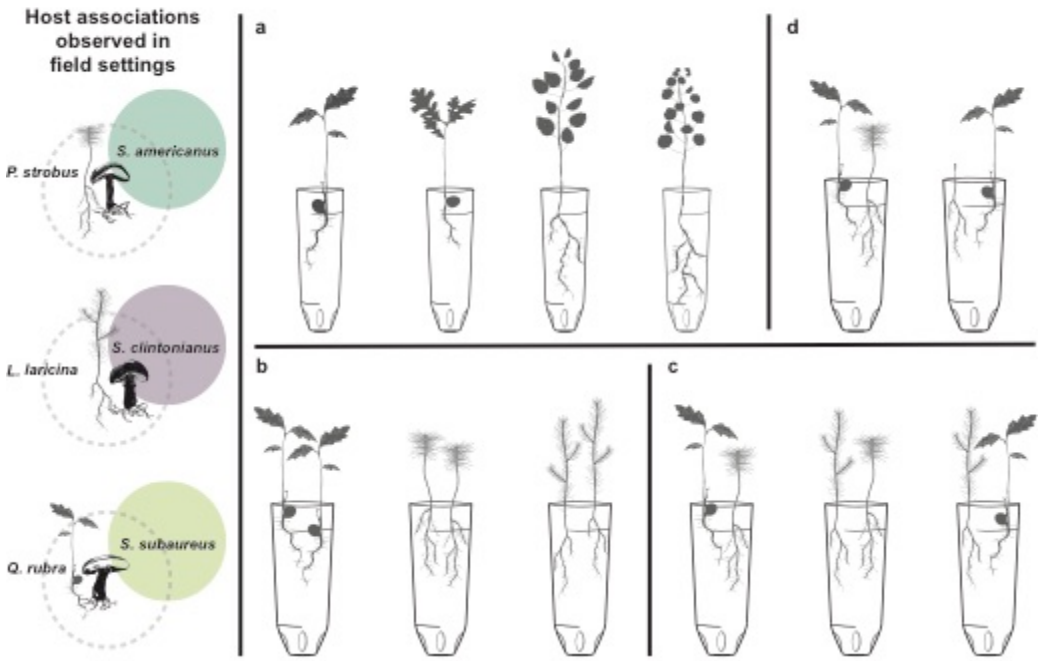

Fig. 1.1: Summary of experimental design for bioassays

a) *The* ***Angiosperm Spore Colonization Bioassay*** *tested angiosperms as spore germination triggers of Suillus subaureus: Inoculation of S. subaureus spores onto (left to right) Q. rubra, Q. macrocarpa, P. grandidentata and P. tremuloides by spore.* ***b)*** *The* ***Gymnosperm Spore Colonization Bioassay*** *tested single host species colonization: Inoculation of either S. americanus, S. clintonianus or S. subaureus spores onto a single host species, P. strobus, L. laricina or Q. rubra.* ***c)*** *The* ***Mycelial Colonization Bioassay*** *tested dual host species colonization: Inoculation of either S. americanus, S. clintonianus or S. subaureus spores into pots planted with two host species, with all pairwise-combinations of P. strobus, L. laricina and Q. rubra represented.* ***d)*** *The* ***Primary Host Removal Bioassay*** *examined disturbance as a mediator of host expansion: Inoculation of S. subaureus into pots planted with both P. strobus and Q. rubra. After sufficient growth (5 months), P. strobus seedlings were hewn at the soil line in half the replicates, removing P. strobus as a potential carbon source for the fungus.*

Morphological investigation of mycorrhizas

For all bioassays, representative and anomalous ectomycorrhizas were photographed using an Olympus Stylus TG4 and sequenced to confirm fungal identity. In all three bioassays, ITS sequencing identified that *Suillus* ectomycorrhizas were of the same species inoculated into the pots. Un-inoculated controls (n = 6 plants) remained uncolonized throughout the experiment. For analyses of Hartig net formation for *S. subaureus* on *Pinus* and *Quercus* hosts, a representative subset of ectomycorrhizas from the bioassays were reserved and stored in Formalin-Acetic-Alcohol fixative (ethanol:aceticacid:formalin:water at 50:5:10:35). To prepare for microcopy, ectomycorrhizas were rinsed in 0.1M sodium cacodylate buffer (10 min, 3x), post-fixed overnight at 4C in 1% osmium tetroxide in 0.1M sodium cacodylate buffer and dehydrated in an ethanol series. Ectomycorrhizas were then embedded in Embed 812 resin (Electron Microscopy Sciences, Hatfield, Pennsylvania) and semi-thin sections (0.5 μm thick) were cut on a Leica Ultracut UCT microtome (Leica Microsystems, Buffalo Grove, IL) using a diamond knife. Sections were then stained with 0.5% toluidine blue and observed using a Nikon Eclipse 90i light microscope (Nikon Instruments, Inc, Melville, NY) in bright field mode. Images were captured with a Nikon D2-Fi2 color camera using Nikon Elements software.

Statistical Analyses

To analyze differences in mycorrhizal colonization by treatment in the bioassays for which colonization was observed, we used a combination of statistical analyses. In the *Gymnosperm and Mycelial Colonization Bioassays*, we applied separate one-way non-parametric Kruskal-Wallis tests because assumptions of variance homogeneity could not

be met due to the lack of colonization in some treatments but not others. Based on the significance of both tests, post-hoc Wilcoxon tests were then used to determine specific differences among treatment means for each pair. For the *Primary Host Removal Bioassay*, we again observed high heterogeneity in colonization across treatments, so applied a one-tailed Mann-Whitney U test. In all cases, significance was determined at $P < 0.05$ using the R programing environment (R Core team, 2017) and JMP Pro 12 (Cary, NC, USA).

Results

Field analysis

From the six soil cores taken from underneath *S. subaureus* mushrooms at the Minnesota field site, four contained ectomycorrhizas exhibiting characters morphologically associated with *Suillus* species (white to off-white color with thick mantles and notable extramatrical mycelium). A total of ten root tips were identified as *S. subaureus* in three of the four cores for which suilloid tips were present. Plant DNA was successfully extracted from six of the ten root tips identified as *S. subaureus*. Of these, three yielded high-quality sequences, with the plant host identified as *Q. rubra* in all cases. The soil core from the Indiana field site also contained tips exhibiting characters morphologically associated with *Suillus* species. Fungal DNA was successfully extracted from six of the eight root tips taken for analysis and identified as *S. subaureus* in all cases. Plant DNA was successfully extracted from all six of those root tips and was identified as the genus *Quercus* in all cases (BLAST confidence was not high enough to identify the host DNA to species, but *Q. velutina* was the only *Quercus* species present at

the field site)

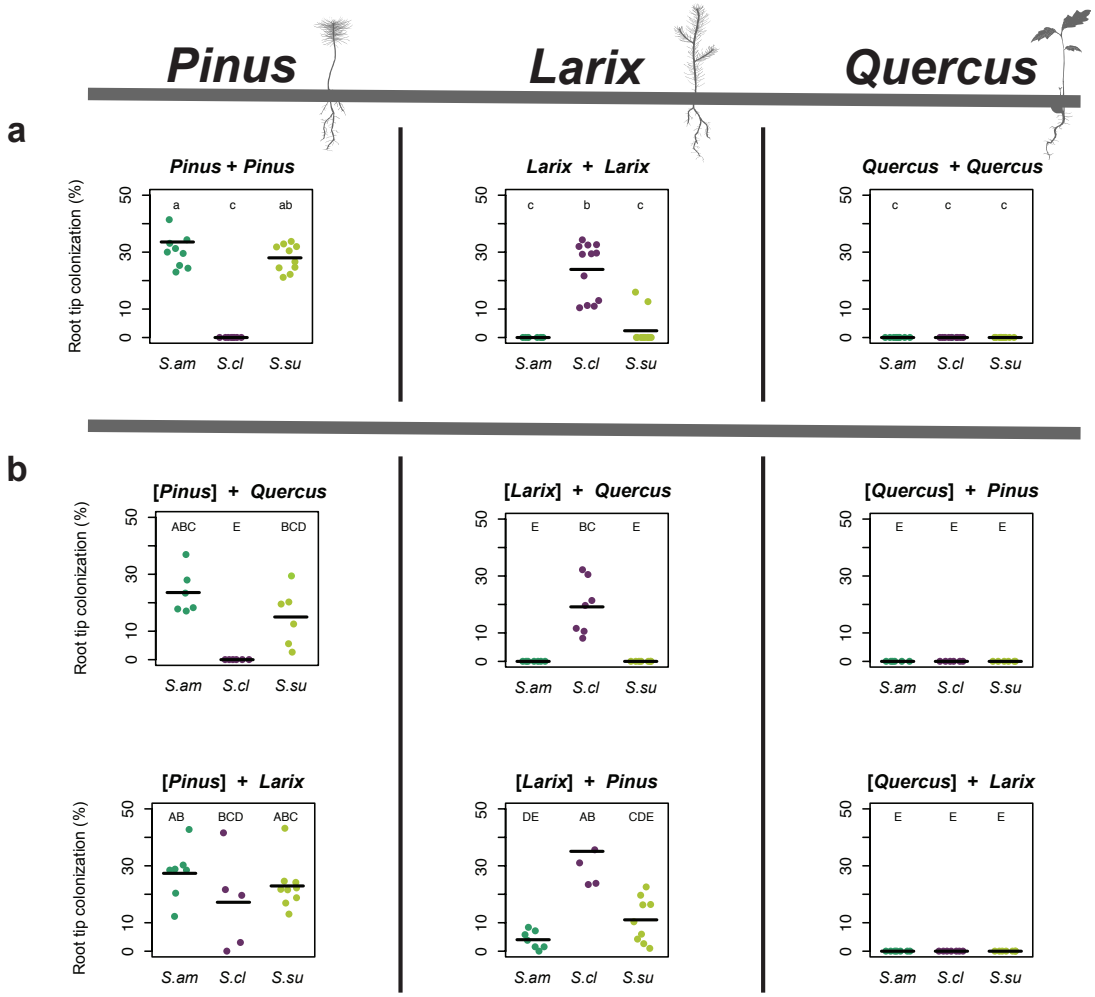

Fig. 1.2: Percent ECM root tip colonization

a) *the Gymnosperm Spore Colonization Bioassay and* ***b)*** *the Mycelial Colonization Bioassay. In the Gymnosperm Spore Colonization Bioassay, a single species of Suillus per pot was inoculated onto a single-species of host (two trees/pot) whereas in the Mycelial Colonization Bioassay a single species of Suillus per pot was inoculated onto pots containing two host species (two trees/pot). The first [bracketed] host genus indicates the host being quantified, and the second genus indicates the second tree species present in the pot. Different letters above treatments indicate significant differences in colonization among the nine treatments in the Gymnosperm Colonization Assay (lower case) or the 18 treatments in the Mycelial Colonization Assay (upper case)*

as determined by two separate one-way Kruskal-Wallis tests followed by Wilcoxon post-hoc tests for each comparison pair.

Angiosperm Spore Colonization Bioassay

For both the three and six month time points and for all angiosperm hosts tested, *Q. rubra*, *Q. macrocarpa, P. tremuloides* and *P. grandidentata*, spore inoculation failed to result in any colonization by *S. subaureus*.

Gymnosperm Spore Colonization Bioassay

On *Pinus*, *S. americanus* and *S. subaureus* colonized at statistically equivalent mean rates of 34% (n = 10, with all plants colonized) and 28% (n = 10, with all plants colonized), respectively, whereas *S. clintonianus* failed to form ectomycorrhizas on this host (Wilcoxon tests, $P < 0.05$) (Fig. 1.2). [Colonization rate is defined as the total number of root tips colonized by *Suillus* out of ~300 counted per plant.] On *Larix, S. clintonianus* formed ectomycorrhizas at a mean rate of 24% (n = 12, with all plants colonized), which was significantly higher than *S. subaureus* and *S. americanus*, which colonized at 2% (n=12, with 2 plants colonized at a mean of 14%) and 0%, respectively. Neither *S. americanus* (n = 10 plants), *S. clintonianus* (n = 12 plants), or *S. subaureus* (n = 12 plants) formed ectomycorrhizas with *Q. rubra* (hereafter referred to as *Quercus*).

Mycelial Colonization Bioassay

Suillus americanus formed ectomycorrhizas on *Pinus* at statistically equivalent mean rates of 27% (n = 7, with all plants colonized) when co-planted with *Larix* and 24%

(n = 6, with all plants colonized) when co-planted with *Quercus* (Fig. 1.2). On *Larix*, *S. americanus* formed ectomycorrhizas at a mean rate of 4% (n = 7, with 6 plants colonized averaging 5% colonization) when co-planted with *Pinus*, but did not form ectomycorrhizas (n = 7, with all plants uncolonized) when co-planted with *Quercus* (Wilcoxon test, $P > 0.05$). On *Quercus*, *S. americanus* failed to form ectomycorrhizas regardless of host species pairing. *Suillus clintonianus* formed ectomycorrhizas on *Larix* at statistically equivalent mean rates of 35% (n = 5, with all plants colonized) when co-planted with *Pinus* and 19% (n = 7, with all plants colonized) when co-planted with *Quercus*. On *Pinus*, *S. clintonianus* formed ectomycorrhizas at a mean rate of 17% (n = 5, with four plants colonized at a mean rate of 21%) when co-planted with *Larix*. This was significantly higher than the 0% colonization of *S. clintonianus* on *Pinus* when co-planted with *Quercus* or on any of the *Quercus* seedlings (Wilcoxon tests, $P < 0.05$). Finally, *S. subaureus* formed ectomycorrhizas on *Pinus* at the statistically equivalent mean rates of 23% when co-planted with *Larix* (n = 9, with all plants colonized), 15% (n = 6, with all plants colonized) when co-planted with *Quercus,* and 11% on *Larix* when co-planted with *Pinus* (Wilcoxon tests, $P < 0.05$). In contrast, *S. subaureus* did not form ectomycorrhizas on *Larix* when co-planted with *Quercus* (n = 7, with all plants uncolonized) and failed to form ectomycorrhizas on *Quercus* regardless of host species pairing.

Morphological description of S. americanus and S. clintonianus mycorrhizas

On *Larix*, *S. clintonianus* formed typical monopodial-pyramidal ectomycorrhizas with typical root swelling, an off-white mantle and prolific extramatrical mycelium (see Figure S1a in Supporting Information). On *Pinus*, *S. clintonianus* formed primarily

monopodial ectomycorrhizas with loosely attached mantle hypha and frequent dark patches (Fig. 1.S1b). On *Pinus*, *S. americanus* formed typical bifurcate ectomycorrhizas with a dense off-white mantle and prolific extramatrical mycelium (Fig. 1.S1c). On *Larix*, however, *S. americanus* formed primarily monopodial ectomycorrhizas, with a loose hyphal mantle and frequent dark patches (Fig. 1.S1b).

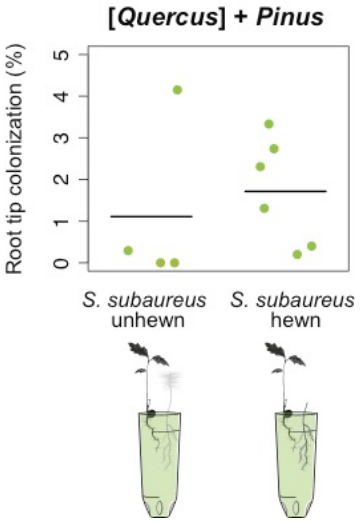

Fig. 1.3: Primary Host Removal Bioassay

In hewn treatments, Suillus subaureus colonized all six of Quercus rubra seedlings at a mean rate of 2%. In unhewn treatments (n = 4), two Q. rubra replicates were uncolonized, and two replicates formed ectomycorrhizas at rates of 4% and 0.29%. The first [bracketed] host genus indicates the host being quantified, and the second genus indicates the second species present in the pot. Letter sharing above treatments indicates no significant differences in colonization as determined by a Mann-Whitney U test.

Primary Host Removal Bioassay

In cone-tainers where *Pinus* was hewn after five months, *S. subaureus* successfully formed ectomycorrhizas on all six *Q. rubra* plants, at a mean colonization

rate of 2% (Fig. 1.3) [Note colonization rate on *Quercus* is defined as the total number of root tips colonized out of 1000 root tips counted per plant.] In cone-tainers where *Pinus* was unhewn, *S. subaureus* ectomycorrhizas were formed on two of the four *Q. rubra* plants. These two replicates were colonized at individual rates of 4% and 0.3% (mean rate = 2%). While more of the *Quercus* plants were colonized by *S. subaureus* when *Pinus* seedlings were hewn, there was no significant difference in the mean rates of colonization between the two treatments (N = 8, P = 0.225). To rule out contamination, fungal species identity was confirmed by sequencing the ITS region of individual mycorrhizas as described previously and were identified as *S. subaureus* in all cases.

Morphological description of S. subaureus mycorrhizas

Although *S. subaureus* formed ectomycorrhizas on *Pinus*, *Larix*, and *Quercus*, the morphology exhibited on each of these hosts differed (Fig. 1.4). Unlike the non-primary associations occasionally formed between *S. americanus* and *Larix* or between *S. clintonianus* and *Pinus* in the Gymnosperm Spore Colonization Bioassay, *S. subaureus* ectomycorrhizas were never monopodial and did not exhibit loosely attached mantles or dark discoloration on any of the host species tested. Rather, *S. subaureus* formed ectomycorrhizas that were white to orange in color (with larger, presumably older, ectomycorrhizas intensifying in color on all hosts), with thick mantles and prolific extramatrical mycelium. On *Pinus*, *S. subaureus* formed bifurcate ectomycorrhizas (like the ectomycorrhizas formed between *S. americanus* and *Pinus*). On *Larix*, *S. subaureus* ectomycorrhizas were monopodial-pyramidal (like the ectomycorrhizas formed between *S. clintonianus* and *Larix*) and on *Quercus*, *S. subaureus* ectomycorrhizas were notably coralloid (containing as many as 55 individual lobes per ectomycorrhiza) and, as a unit,

several times larger than those formed on either conifer host. Cross sections of *S. subaureus* mycorrhizas on both *Pinus* and *Quercus* revealed well-developed Hartig net structures on both hosts, with epidermal penetration on *Quercus* and outer cortical cell penetration on *Pinus*.

Discussion

Neighborhood effects as a function of time

Our results clearly demonstrate that the ECM fungus *S. subaureus* can associate with *Quercus* hosts, both in field and lab settings, making this species unique in an otherwise Pinaceae-specific clade. We have also shown that *S. subaureus* can colonize two Pinaceae host species, suggesting that this species is a host generalist rather than a *Quercus* specialist. Because the capacity to colonize alternative hosts can be controlled either at the point of spore germination or during downstream signaling processes, host identity may influence colonization differently depending on whether spores must be germinated in order to establish fungal presence, or whether extant mycorrhizas are already present on neighboring plants (Molina *et al.*, 1997; Kennedy *et al.*, 2012). Consistent with earlier studies i.e. (Massicotte *et al.*, 1994; Molina *et al.*, 1997) our bioassays indicated that the mode of colonization (i.e. spores *vs.* mycelium) strongly affects patterns of ECM host specificity. We observed that the spores of *S. americanus* and *S. clintonianus* germinated only in the presence of their primary hosts, and only formed a few (morphologically anomalous) mycorrhizas on alternate Pinaceae hosts when colonizing via mycelial networks. By contrast, *S. subaureus* germinated in the presence of both Pinaceae hosts and colonized all three hosts by mycelia (Fig. 1.5).

Importantly, the resulting ectomycorrhizas of *S. subauerus* were anatomically typical of functional host associations on all three hosts (Fig. 1.4). While the bioassay results indicated that only Pinaceae hosts could trigger germination of *S. subaureus* spores, Pineaceae trees were locally absent at both field sites. Because these hosts were historically present at both locations prior to anthropogenic disturbance events, it appears the current *S. subaureus*-*Quercus* associations represent carryover from past host presence. This pattern echoes other studies highlighting the role of neighborhood effects in structuring ectomycorrhizal fungal host specificity (Bogar & Kennedy, 2013; Bogar *et al.*, 2015) but, because *S. subaureus* mushrooms and mycorrhizas were found in angiosperm-only forests where Pinaceae hosts have long been locally extirpated, the spore germination triggers or mycelial inoculum originating from Pinaceae hosts cannot be considered a neighborhood effect in the traditional definition. Instead, the establishment of new *S. subaureus* colonies appears to depend on triggers provided by hosts long absent from the system, suggesting that neighborhood effects should not only be viewed in light of the current host community structure, but as a function of host community change over time.

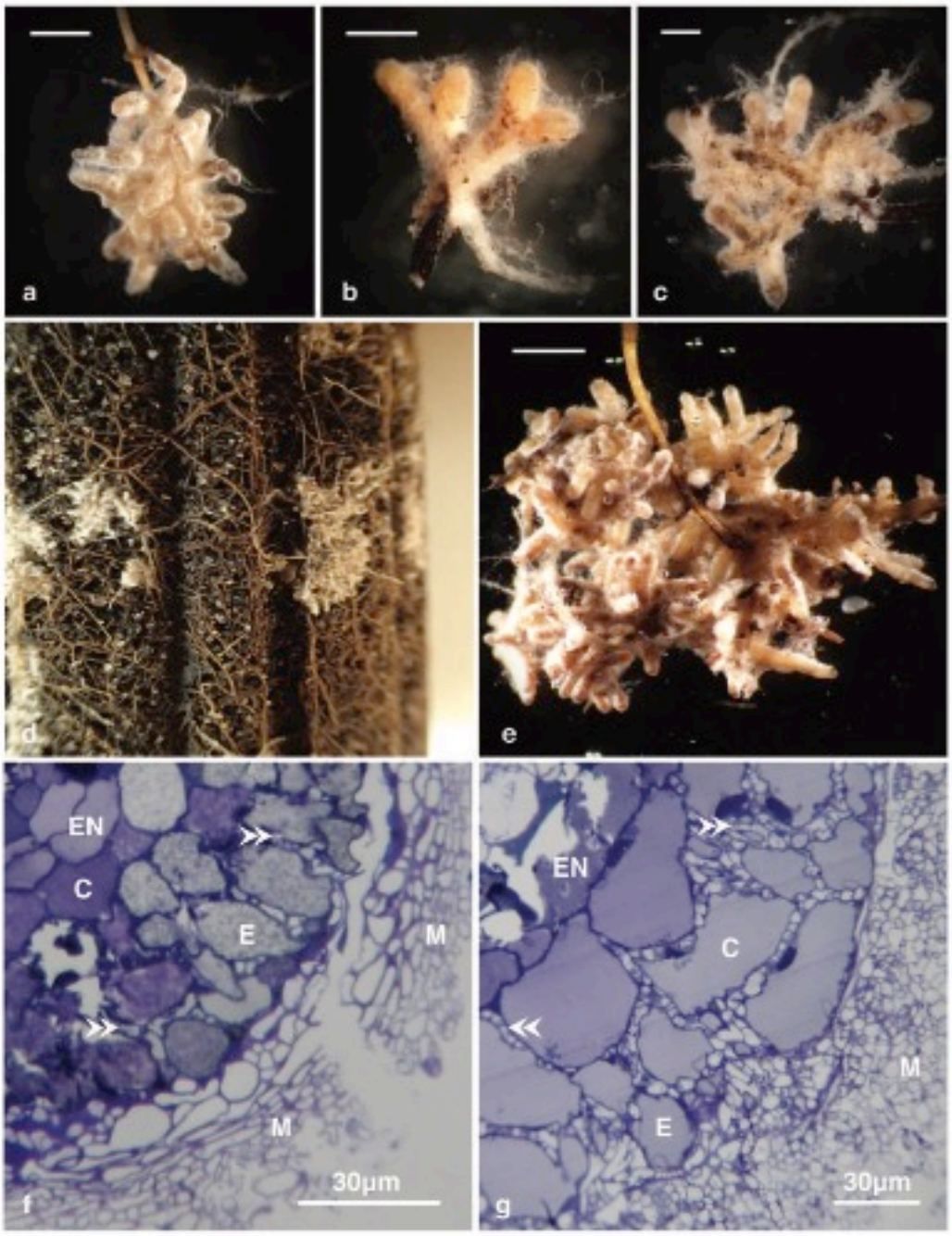

Fig. 1.4: Ectomycorrhizas formed by *Suillus subaureus* on three different host species, *Quercus rubra*, *Pinus strobus*, and *Larix laricina*

Ectomycorrhizas formed a thick white mantle that turned progressively orange with age. On Q. rubra, S. subaureus developed progressively multi-lobed coralloid structures with surface area per mycorrhiza much larger than that formed on P. strobus or L. laricina. Scale bars represent 1 mm unless otherwise noted. a) A young S. subaureus mycorrhiza on Q. rubra. b) A mature S. subaureus mycorrhiza on P. strobus. c) A mature S. subaureus mycorrhiza on L. laricina. d) Large coralloid mycorrhizas and extramatrical hyphae of S. subaureus on Q. rubra. e) A mature S. subaureus mycorrhiza on Q. rubra. f-g) cross sections of S. subaureus on Q. rubra (f) and S. subaureus on P. strobus (g) mycorrhizas, stained with Toluidine blue and visualized with light microscopy. M = mantle, >> = Hartig net, E = epidermis, C = cortical cells, EN = endodermis.

Evidence for host expansion rather than host switching

The deeply nested phylogenetic location of *S. subaureus* within the genus *Suillus* strongly suggests this species was ancestrally associated with Pinaceae hosts (Nguyen *et al.*, 2016a). If *S. subaureus* has lost the ability to colonize hosts in the Pinaceae, it would indicate this fungus has switched its association patterns to now associate exclusively with angiosperm hosts. Alternatively, the ability to colonize both angiosperm and gymnosperm hosts, would indicate that this fungus has simply expanded its host range to include angiosperms. In this study, the colonization of *Quercus* seedlings coupled with frequent colonization of *S. subaureus* on *Pinus* seedlings as well as the occasional colonization on *Larix* seedlings is consistent with a pattern of host expansion rather than host switching. These results add to the growing evidence that host specialization is not necessarily an evolutionary dead-end (Nosil, 2002; Desdevises *et al.*, 2002; Tripp & Manos, 2008; Ouvrard *et al.*, 2015)(Nosil, 2002; Desdevises *et al*., 2002; Tripp & Manos, 2008; Ouvrard *et al*., 2015), as famously suggested by Simpson (1953). In contrast to host-pathogen relationships, the evolutionary pressures structuring host range in fungal mutualists has been suggested to ultimately favor the maintenance of host generalism, where the capacity to colonize diverse hosts is assumed to have a positive net impact for

both plant and fungal partners (as discussed in Harley & Smith, 1983). However, the high host specificity observed in most *Suillus* species appears to be a derived trait which evolved from an ancestral habit of host generalism (Nguyen *et al.*, 2016a) bringing into question the assumption that expanded host range is an evolutionary driver that is beneficial to both partners. Experimental investigation regarding how the functional benefit to each partner might vary by species, and which partner (plant or fungus) controls the mutualism were not investigated in this study. However, examples such as *S. subaureus*, which appear to have the reverted capacity for host generalism, could provide an excellent experimental system for addressing these questions in ECM fungi.

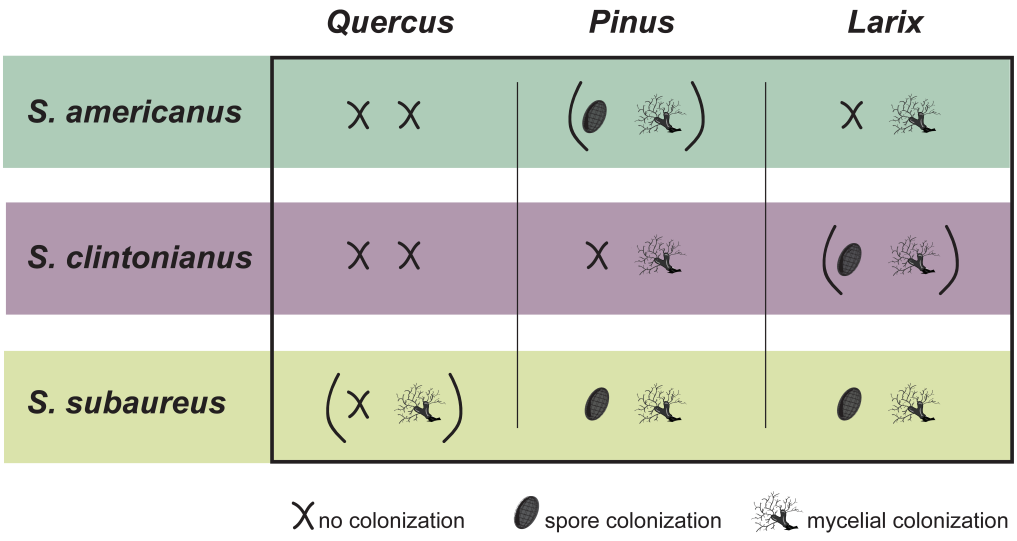

Fig. 1.5: Summary of differences in host colonization mode exhibited by the three *Suillus* species

Whereas Suillus americanus colonized its primary host, Pinus strobus, by both spore and mycelium, it only colonized Larix laricina via mycelial extension from extant mycorrhizas and did not form mycorrhizas with Quercus spp. Similarly, S. clintonianus colonized its primary host, L. laricina, by both spore and mycelium but colonized P. strobus only by

mycelium and did not colonize Quercus spp. Suillus subaureus formed mycorrhizas on both P. strobus and L. laricina from both spore and mycelium and was additionally able to form mycorrhizas on Q. rubra when colonized via mycelium. Brackets indicate the host associations observed in field settings.

Separating evolutionary pressure vs. environmental pressure

In a recent analysis of the ECM genus *Russula*, Looney *et al*., (2016) showed that changes in host association from Pinaceae to angiosperms occurred at a rate 15 times higher than the inverse, suggesting the transition may be a relatively common phenomenon. Long-term disturbance regimes resulting in selective host removal could act as a driver of ECM host expansion by placing pressure on specialist fungi to secure carbon from alternative hosts (given the obligate nature of the ECM symbiosis, it is very unlikely that ECM fungi can meet any significant part of their carbon needs by living saprotrophically (Baldrian, 2009; Kohler *et al.*, 2015). For example, repeated disturbances, such as fire, may favor alternative hosts such as *Quercus* spp. that are able to re-sprout from their existing tree bases (Crow, 1988). In the *S. subaureus* study system, we are unaware of any current populations of this fungus present in either young or mature angiosperm-only forests that have not at one point also contained hosts in the Pinaceae. However, our results suggest that Pinaceae host removal is not immediately necessary to induce angiosperm colonization by *S. subaureus* (and given the recent nature of the human disturbances (<200 years), it is not likely that anthropogenic influences are the selective agent directly responsible for inducing this broader host association). Rather, our results offer an example of the fitness advantage of an ectomycorrhizal fungus that is capable of acting as a generalist in the event of local extirpation of its primary host. Given the young age of the hosts used in glasshouse bioassays, future

research examining whether the timing of primary host removal (in regard to the age of the respective host trees and the time since mycorrhizal establishment) influences colonization rates on secondary hosts will also provide greater insight into the relative importance of evolutionary versus environmental pressure as drivers of observed host associations.

Mycorrhizal morphology and colonization patterns are influenced by host identity

Root tip colonization percentages of *S. subaureus* were notably lower on *Q. rubra* compared to *S. subaureus* colonization on *P. strobus* and *L. laricina*. This result is typical of *Quercus* ECM colonization due to the extensive production of fine roots generated by this host genus (He *et al.*, 2010; Chen *et al.*, 2016). Similarly, the difference in Hartig net development on *Quercus* (epidermal penetration) and *Pinus* (outer cortical cell penetration) is typical of angiosperm and gymnosperm ectomycorrhizal development, respectively (Smith & Read, 2008; Watkinson *et al.*, 2015). Less expected were the macro-morphological differences observed among *S. subaureus* on *Quercus* and the two gymnosperm hosts. On *Q. rubra*, *S. subaureus* produced prolific rhizomorphic mycelium and individual ectomycorrhizas exhibited greatly increased biomass and surface area over those produced on Pinaceae hosts (Fig. 1.4). Microscopic inspection (Fig. 1.4 f-g) coupled with the presence of *S. subaureus* ECM root tips directly under *S. subaureus* sporocarps in the field with no primary host (*Pinus*) in the vicinity suggests that *Quercus* - *S. subaureus* ectomycorrhizas are functional in terms of carbon acquisition by the fungus. Similar results were observed in experimental inoculations by Finlay *et al*., (1989), who found normal carbon allocation of *P. sylvestris* seedlings to *S. cavipes*, a species strictly associated with *Larix* hosts in field settings. Interestingly, the phosphorus

returned from that same experimental association was notably lower than when *P. sylvestris* seedlings were colonized by *Pinus*-specific *Suillus* species. Although we did not measure physiological traits in any of our experiments, and therefore cannot make any inferences about the efficacy of *Quercus* - *S. subaureus* symbioses, our combined results indicate that *S. subaureus* has the ability to both colonize and persist on both angiosperm and multiple gymnosperm hosts. The reason for the absence of *S. subaureus* on Pine in field conditions is not clear, but may reflect edaphic specialization or limited competitive ability by *S. subaureus*, as has been observed with *Suillus* species in other studies (Bidartondo *et al.*, 2001; Kennedy *et al.*, 2011). We are currently testing the competition hypothesis with seedling bioassay experiments, but additional field-based studies are needed to fully understand the ecological factors that make *S. subaureus* rare in both angiosperm-only and mixed host forests.

Resolving the long-standing question of angiosperm hosts for Suillus

Although there has been anecdotal evidence of some *Suillus* species being associated with angiosperm hosts under natural conditions (Miller & Miller, 2006), to date, no reliable confirmation of these associations has been established. Seedling inoculation trials claimed that ectomycorrhizas were formed between *S. luteus* and *S. granulatus* and four *Quercus* species (Dixon *et al.*, 1984; Dixon & Johnson, 1992), but in both of those studies, it was not accurately confirmed whether the ectomycorrhizas present belonged to *Suillus* or other ECM fungal species. In laboratory settings, by contrast, Molina and Trappe Molina & Trappe (1982) were able to successfully synthesize ectomycorrhizas between *S. brevipes*, *S. clintonianus*, *S. cavipes*, and *S. lakei* and a number of different host species, including the angiosperm host *Arbutus menziesii*.

It was later recognized, however, that the presence of glucose in the growth medium in that and other early ectomycorrhizal synthesis trials effectively reduced the host specificity barriers normally present among many ECM fungi (Duddridge, 1986; Theodorou & Reddell, 1991). Similarly, Murata *et al.* (2015), achieved superficial colonization between *S. luteus* and *Prunus speciosa* when grown in the presence of added glucose and sucrose. In this case, however, ECM colonization consisted of limited mantle development, no Hartig net, and frequent dark spotting.

Conclusions and future directions

Moving forward, we believe that assessing the effects of differences in mycorrhizal morphologies on nutrient trading dynamics, determining competitive ability, analyzing the genomic content and expression of *S. subaureus* will all aid in identifying the mechanisms that have facilitated host generalism in this species. Understanding the underlying ecological and evolutionary mechanisms driving host specificity in ECM symbioses is broadly important given the current rate of forest redistribution and changes to community composition caused by anthropogenic processes (Perry *et al.*, 1989; Dickie *et al.*, 2010; Pickles *et al.*, 2012; Bogar *et al.*, 2015; Hayward *et al.*, 2015). Specifically, as forest landscapes undergo host migration and current host species are displaced due to climate change, studying host expansion will help in understanding both how ECM hosts and fungi came to occupy their respective niches and how each will respond to future forest community dynamics.

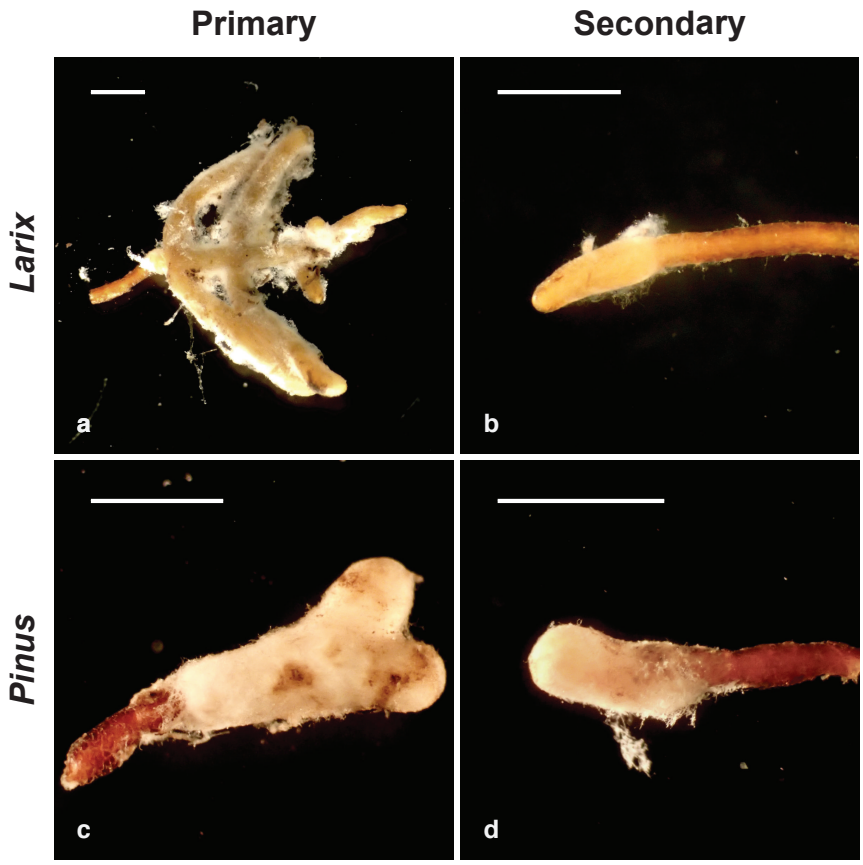

Fig. 1.S1: Morphology of *Suillus* ectomycorrhizas differ by host association

S1a) S. clintonianus ectomycorrhizas on L. laricina. S1b) S. americanus ectomycorrhizas on L. laricina. S1c) S. americanus ectomycorrhizas on P. strobus S1d) S. clintonianus ectomycorrhizas on P. strobus. Scale bars represent 1 mm.

Chapter 2: Genomic signatures of ectomycorrhizal host specificity encoded by the hyper-specialist genus *Suillus*

Synopsis

The genetic mechanisms and ecological drivers structuring host specificity in mutualisms are not well understood. Here, we use comparative genomics to investigate three potential genetic correlates of host specificity in the mutualistic ectomycorrhizal (ECM) genus *Suillus*, and relate our findings to the perceived ecological pressures structuring restricted host range. Based on previous studies of fungal host specificity, we target three suites of molecules which have been shown to be consistently upregulated during the process of ECM host colonization, including small secreted proteins, secondary metabolites, and G-protein coupled receptors. Our study contains two parts including 1) contrasting 19 genome-sequenced *Suillus* species (including 17 newly sequenced genomes) with nine non-*Suillus* ECM species and 2) a intrageneric comparison between White Pine, Red Pine and Larch associated *Suillus*. We then conduct phylogenomic analysis coupled with ancestral state reconstruction to identify the ancestral host of *Suillus* along with key host-jumping events in the lineage. We show that relative to other ECM species, *Suillus* have a marginally lower number of species specific small secreted proteins, a significant enrichment in terpene encoding secondary metabolite genes, and significant enrichment in G-protein coupled receptors. Intrageneric comparisons of *Suillus* by host association again support the role of species specific small secreted proteins, and G-protein coupled receptors in specific *Suillus* host specificity but do not support the role of terpene related secondary metabolites. Phylogenetic

reconstruction support multiple independent host jumps onto White Pine by Red Pine associated *Suillus* and point toward *Larix* as the ancestral host.

Introduction

Host specificity directly influences numerous ecological principles as varied as species distribution and composition, community dynamics and assembly, speciation, invasion, epidemiology, and biogeochemical cycling (Molina *et al.*, 1992a; Bruns *et al.*, 2002; Churchland & Grayston, 2014). Fungi display a multiplicity of host specificity relationships, ranging from a single host species to hundreds of host species spanning multiple kingdoms (Hawksworth, 2001; Gauthier & Keller, 2013). Our current understanding of the mechanisms that structure fungal host specificity has been heavily influenced by the field of plant pathology, where seminal work on host jumping, host range expansions/contractions, and context dependent compatibility scenarios have helped to elucidate both the genetic underpinnings and the ecological pressures selecting for the range of host specificity relationships observed across the fungal phylogeny (Gilbert & Webb, 2007; Schulze-Lefert & Panstruga, 2011; Lo Presti *et al.*, 2015). Despite these advances, the mechanisms facilitating host specificity in fungal mutualisms is not well understood.

Ectomycorrhizal fungi form wide spread mutualisms with ~ 10% of extant plant species, influencing carbon flux dynamics, the exchange of limiting nutrients, soil stabilization, and conferring increased water availability and chemical protection to their hosts plants (Smith & Read, 2008). The ability of fungi to form ectomycorrhizas arose

independently 78-82 times (Tedersoo & Smith, 2013), representing 7-10k ECM fungal species associating with approximately 8k ECM host species (Taylor & Alexander, 2005). Despite independent evolutionary trajectories, these lineages produce comparable structures, carryout comparable ecological functions, and in many cases, form these associations with the same host species. High host specificity is relatively rare in ECM fungi (Horton & Bruns, 1998; Bruns *et al.*, 2002). One of the best documented examples of high host specificity involves fungi in the genus *Suillus*, which primarily associate with plants in the family Pineaceae (particularly the genera *Pinus*, *Larix* and *Pseudotsuga*) (Kretzer *et al.*, 1996). These associations are tightly coupled, with a given *Suillus* clade tending to associated with a single host group. However, these host clade associations are not monophyletic, and represent the evolution of several independent host switching events (Nguyen *et al.*, 2016c). Ectomycorrhizal fungi that exhibit high host specificity like *Suillus* influence forest community dynamics at multiple levels of organization. For example, the common use of *Suillus* fungi in the establishment of successful tree nurseries, concurrently facilitates the invasion of naturalized host trees far beyond their native regions (Dickie *et al.*, 2010; Policelli *et al.*, 2019), while differences in fungal traits such as exploitation type and nutrient trading ratios, influence forest carbon and nitrogen dynamics when these traits co-vary with host distribution (Churchland & Grayston, 2014). *Suillus* fungi produce prolific, long-distance extramatrical mycelium, representing a major belowground carbon-sink and highlighting their importance in the biogeochemical dynamics in both native and introduced Pineaceae systems (Agerer, 2001; Bidartondo *et al.*, 2001).

The process of ECM colonization is divided into three stages including 1) an early pre-contact signaling stage where plant exudates stimulate spore germination, or are recognized by pre-established fungal hypha, 2) a late or post-contact stage where the process of mycorrhization occurs, involving effector-like suppression of host immune responses, and 3) the continued maintenance of the symbiosis over the life span of the ectomycorrhiza, thought to involve the perception of realized nutrient trading (Plett *et al.*, 2011; Garcia *et al.*, 2015). Currently, the molecular mechanisms facilitating host specialism is *Suillus* are poorly understood, but it is likely that compatibility is regulated at each of these stages independently. For example, the ability of a host to trigger germination (early stage) may be decoupled from the ability to form mycorrhizas (late stage), as seen in *S. subaureus*, which is able to form mycorrhizas with Oak in both laboratory and field conditions, even though there is no evidence that Oak triggers spore germination in this species (Lofgren *et al.*, 2018). In general, differences in host specificity could be related to differences in gene content, structural variation in the resultant metabolites or quantifiable differences in gene expression due to variation in transcription factors, promotors, or gene copy number. From the perspective of gene content, three molecular classes have garnered repeated attention in relation to fungal host specificity: small secreted proteins (SSPs), secondary metabolites (SMCs), and G-protein coupled receptors (GPCRs).

Small Secreted Proteins

The use of plant pathology principles to explain mutualistic fungal-host interactions led to the discovery and characterization of effector-like SSPs that play critical roles during the process of ECM mycorrhization (Plett *et al.*, 2011). Mycorrhizal-

induced small secreted proteins (MiSSPs) constitute 8-28% of the genes upregulated during symbiosis (Martin *et al.*, 2008). Many of these proteins are expressed only by the ectomycorrhizal species under investigation and termed species-specific small secreted proteins (SSSPs) (Martin *et al.*, 2008; Kohler *et al.*, 2015). Although the majority of these SSPs are uncharacterized and display little sequence conservation with known proteins, it has been hypothesized that the majority of fungal SSPs function as effectors (Kim *et al.*, 2016). In fungal pathogens, one of the canonically recognized functions of SSP effectors is conferral of virulence via the suppression of host defense pathways (Lo Presti *et al.*, 2015). Similar immunosuppression mechanisms take place during ECM colonization, as shown by MiSSP7 from *Laccaria bicolor*, which acts by suppressing host defense responses via the jasmonate pathway (Plett *et al.*, 2014). To date, MiSSP7 is the only MiSSP that is functionally characterized in ECM fungi. In pathogenic species, effectors act at multiple scales of molecular specificity, spanning kingdoms to individual tissue types (Skibbe *et al.*, 2010; Irieda *et al.*, 2018). This range of molecular specificity implies that not all effector targets are present in all host species, and specific suites of effectors are required to interact with specific host genotypes. Indeed, it has been suggested that fungal host specificity may be directly regulated by the specific complement of effectors produced, where mutation, loss, or gain of effectors, modulate colonization success and the resultant host range (Pritchard & Birch, 2011). This mechanism is supported in multiple pathogenic systems including the closely related oomycetes *Phytophthora infestans* and *Phytophthora mirabilis*, where changes to effector genes are connected to broad host jumps from *Mirabilis jalapa* on to Potato, and the subsequent speciation *P. infestans* (Dong *et al.*, 2014), in Smutt fungi, where the loss of effector genes is

associated with host expansion from monocots to dicots (Sharma *et al.*, 2014) and in *Fusarium oxysporum*, where the horizontal transfer of effector genes is the primary determinant of host range across a wide range of phylogenetically distant host species (vanDam *et al.*, 2016).

Secondary metabolites

The first molecular factors identified to play a role in fungal host specificity were host-specific toxins (HSTs) associated with fungal pathogenesis (Walton & Panaccione, 1993). In the 1950's the discovery of HSTs, generally small molecular weight compounds produced by secondary metabolism, increased the interest into fungal secondary metabolites already well established by the Pharmaceutical industry (Keller *et al.*, 2005). The genes responsible for secondary metabolism are generally clustered in the genome, allowing for the coordinated transcription of multi-step reactions leading to the biosynthesis of complex molecules (Keller & Hohn, 1997). These molecules represent a large number of bioactive compounds synthesized by a limited number of core biosynthetic enzymes, primarily non-ribosomal peptide synthases (NRPS), polyketide synthases (PKS), and terpene synthases or cyclases (TS/C). The products of SMCs function in a variety of ways including virulence (Collemare & Lebrun, 2011), antibacterial activity (de Weert *et al.*, 2007), communication (Brakhage, 2013), and host-metabolic changes, such as the induction of growth factors and genes related to nutrient acquisition (Contreras-Cornejo *et al.*, 2016). SMCs are also associated with fungal host specificity, although the mechanisms differ widely across fungal guild and phylogeny. In *Alternaria alternata*, host specificity is controlled by the presence of PKS genes located on accessory chromosomes that produce at least five different HSTs responsible for the

ability to switch between hosts as phylogenetically divergent as Tabaco and Tangerine (Tsuge *et al.*, 2016). The closely related *Cochliobolus carbonum*, uses a fundamentally different mechanism, encoding an NPRS producing HC-toxin, an HST that controls colonization outcomes between maize cultivars via cytostasis (Dunkle *et al.*, 1991; Walton, 2006). In entomopathogenic *Metarhizium* species, host specificity is associated with an increase in total SMCs as well as SMC diversity, coupled with the loss of a SMC encoding a broad acting (host-nondiscriminatory) toxin (Xu *et al.*, 2016). In *F. graminearum*, specific SMCs, including TS/C clusters are differentially regulated during colonization of Wheat, Barley and Maize hosts (Harris *et al.*, 2016). In *Suillus* fungi, host specificity is associated with the upregulation of cytochrome P450-related genes (involved in the production a vast number of secondary metabolites) as well as TS/C related genes (Liao *et al.*, 2016).

G-protein coupled receptors

G-protein coupled receptors are a large class of membrane proteins that function in the environmental sensing of a large number of external stimuli including other proteins and peptides, lipids, hormones, nutrients, water, and photons (Kochman, 2014). Although GPCRs exhibit low sequence similarity, they share a common architecture, including the presence of seven transmembrane domains, an extracellular N-terminus and intracellular C-terminus. The role of GPCRs in the transduction of environmental signals may extend to host and non-host recognition responses in fungi. PTH-11 like GPCRs are involved in host species recognition in entomopathogenic *Metarhizium* (Gao *et al.*, 2011), and are differentially expressed between fungal and insect hosts in *Tolypocladium* (Quandt *et al.*, 2016). In the ECM fungi *Laccaria bicolor* and *Tuber melanosporum*,

GPCR and G-protein related transcripts are the most highly upregulated signaling genes transcribed during ECM colonization (Voiblet *et al.*, 2001; Martin *et al.*, 2010; Plett *et al.*, 2012). In *Suillus*, the differential expression of GPCR transcripts are associated with host specificity and successful mycorrhizal establishment in compatible host interactions (Liao *et al.*, 2014).

Here, we use a comparative genomics approach to investigate these three putative genetic correlates of ECM fungal host specificity: SSPs, SMCs, and GPCRs. Our study contains two parts including 1) contrasting 19 genome-sequenced *Suillus* species with nine non-*Suillus* ECM species and 2) a intrageneric comparison between White Pine, Red Pine and Larch associated *Suillus*. We further conduct phylogenomic analysis coupled with ancestral state reconstruction to identify the ancestral host of *Suillus* along with key host-jumping events in the lineage, and link this information back to genomic diversification between *Suillus* specializing on different host groups.

Methods

Sequencing and assembly

Seventeen *Suillus* cultures were isolated from fungal fruitbodies from under *Pinus* s.g Pinus n = 10, *Pinus* s.g. Strobus n = 4, *Larix* n = 3, *Pseudotsuga* n = 1, or *Quercus* n = 1 (Table 2.S1). Isolates were grown in liquid MMN media on a shaker at room temperature. DNA and RNA was co-extracted using CTAB/chloroform and LiCl precipitation as described in (Liao *et al.*, 2014). Whole genome sequencing was carried

out at the Joint Genome Institute (JGI), on either Illumina or PacBio platforms (Table 2.1). Genomes were assembled using either AllPathsLG (Illimina) or Falcon (PacBio) and annotated using the JGI Annotation Pipeline. See Table 2.S2 for detailed genome assembly statistics.

Table 2.1: Species used in this study, identifying information, and sequencing platform used for each genome project

genus	specific epithet	JGI project code	host group	sequencing platform
Amanita	*muscaria*	Amamu1	Other	Illumina
Hebeloma	*cylindrosporum*	Hebcy2	Other	Solexa, PacBio, 454, Sanger
Laccaria	*bicolor*	Lacbi2	Other	Illumina, 454
Paxillus	*involutus*	Paxin1	Other	Sanger, Velvet, Solexa
Piloderm	*croceum*	Pilcr1	Other	Solexa
Pisolithus	*microcarpus*	Pismi1	Other	Solexa, 454, Sanger
Rhizopogon	*truncatus*	Rhitru1	Other	PacBio
Rhizopogon	*vulgaris*	Rhivul1	Other	PacBio
Scleroderma	*citrinum*	Sclci1	Other	Solexa
Suillus	*americanus*	Suiame1	White Pine	Illumina
Suillus	*ampliporus*	Suiamp1	Larch	PacBio
Suillus	*bovinus*	Suibov1	Red Pine	PacBio
Suillus	*brevipes*	Suibr2	Red Pine	Illumina and PacBio
Suillus	*clintonianus*	Suicli1	Larch	PacBio
Suillus	*cothurnatus*	Suicot1	Red Pine	PacBio
Suillus	*decipiens*	Suidec1	Red Pine	Illumina
Suillus	*granulatus*	Suigr1	Generalist	Illumina
Suillus	*hirtellus*	Suihi1	Red Pine	Illumina
Suillus	*lakei*	Suilak1	Pseudotsuga	PacBio
Suillus	*luteus*	Suilu4	Red Pine	PacBio
Suillus	*occidentalis*	Suiocc1	Red Pine	PacBio
Suillus	*paluster*	Suipal1	Larch	PacBio
Suillus	*pictus (spraguei)*	Suipic1	White Pine	Illumina and PacBio
Suillus	*placedus*	Suipla1	White Pine	PacBio
Suillus	*subalutaceus*	Suisu1	Red Pine	PacBio
Suillus	*subaureus*	Suisub1	Generalist	PacBio
Suillus	*tomentosus*	Suitom1	Red Pine	PacBio
Suillus	*variegatus*	Suivar1	Red Pine	PacBio

Genomics and Bioinformatics

Genome assemblies (repeat masked scaffolds) and gene annotations (Filtered Models) were transferred from JGI's MycoCosm database to Minnesota Supercomputing

Institute server space using Globus (Foster, 2006). In addition to the 19 *Suillus* species, the predicted proteomes of nine other ECM fungi were included for comparison (Table 2.1). These nine species represent 8 genera, with the two representatives from the genus *Rhizopogon*, a sister group to *Suillus*. Complete meta-data for the non-ECM species, and the two previously published *Suillus* species (*S. brevipes* and *S. luteus*) are publicly available on JGI's Mycocosm (Grigoriev *et al.*, 2014a). *Suillus* genomes were coded by host association as noted above and only groups representing more with n >= 3 were used for the within genus comparison, which resulted in the exclusion of the species *S. lakei*, which was the sole isolate known to associate with the genus *Pseudotsuga*. Two species, *S. subaureus*, isolated from under *Q. rubra*, but known to associate with *Pinus* s.g. pinus (Lofgren *et al.*, 2018) and *S. granulatus*, isolated under *Pinus* s.g. strobus but known to associate with both *Pinus* s.g. strobus and *Pinus* s.g. pinus (Jacobson & Miller Jr., 2007) were coded as generalist species, and excluded from the intrageneric comparison.

Secondary metabolite clusters were identified using antiSMASH Fungi 4.0 (Blin *et al.*, 2017) (ClusterFinder with default settings on, a default minimum probability of 60% and all optional features on). Investigation into terpene diversity was done using KEGG metabolic pathway analysis (Kanehisa *et al.*, 2019) on JGI server space. To predict SSPs, signalP5 (Almagro Armenteros *et al.*, 2019) was used to screen proteins containing a secretion signal peptide (eukaryote option with default settings). The resultant dataset was then filtered to include only proteins lacking predicted transmembrane helices using TMHMM (Krogh *et al.*, 2001). A custom R scrip was then used to filter proteins to those composed of < 300 aa. Putative effectors were identified

by processing the resultant dataset with EffectorP 2.0 (http://effectorp.csiro.au) which employs a machine learning approach on multiple criteria derived from characterized pathogenic effectors (Sperschneider *et al.*, 2018). Ortholog prediction of SSPs and SSSPs was carried out using OrthoFinder2 (Emms & Kelly, 2018a). GPCRs were identified using a custom pipeline consisting of the following: first, GPCRHMM with local scoring was used to identify putative GPCRs using an HMM model (Wistrand *et al.*, 2006). The resultant dataset was then filtered to contain only proteins with seven trans-membrane domains using Phobius (Krogh *et al.*, 2001). Finally, GPCR classification in relation to mammalian-orthologues was achieved using PCA-GPCR (Peng *et al.*, 2010). Phylogenomic analysis was conducted using OrthoFinder 2.0 running DIAMOND, with gene tree inference using DendroBLAST under default settings (Emms & Kelly, 2015). The species tree was inferred using STAG and rooted by STRIDE (Emms & Kelly, 2018b). Ancestral state reconstruction was achieved using the R packages phytools (Revell, 2012), and ape (Bolker *et al.*, 2014), employing a Bayesian stochastic mapping approach on a population of 100 character histories.

Statistics

To assess differences in in KEGG pathway enrichment, genome size, predicted proteome size, SSPs, and SSSPs diversity between *Suillus* and other ECM fungi, normality and variance assumptions were evaluated on each data set individually using Shapiro-Wilk and Cochran's C tests coupled with visual inspection, and log transformed when assumptions were not met. Significance was assessed by t-test or Welch's t-test if transformation did not improve equal variance. To assess differences in genome size, predicted proteome size, SSPs, and SSSPs diversity between *Suillus* associating with

different hosts normality and variance assumptions were evaluated using visual inspection, and data was transformed when assumptions of normality or variance were not met, using the transformation as recommended by boxCox testing. Significance values were assessed using a one-sided type 1 ANOVA. Significant differences were assessed using TukeyHSD at alpha <0.05. To account for unequal sample size, a second set of analyses was run using a series of randomization tests for genome size, proteome size, SSPs, SSSPs and effectors, for both *Suillus* compared to other ECM fungi, and *Suillus* analyzed by host association. For *Suillus* compared to other ECM fungi, and for each variable under evaluation, a two-factor randomization tests was run in base R over 10,000 permutations to generate random normal distribution of the mean difference between groups, and compared to the observed mean difference at alpha = 0.05 significane. For the *Suillus* by host association comparison, and for each variable under evaluation, multi-factor randomization tests were implemented using the *coin* package in R, at alpha = 0.05. Significant differences between groups were assessed using pairwise permutation tests, implemented with the package *rcompanion* with a Benjamini–Hochberg correction for multiple comparisons.

Differences in secondary metabolite clusters between *Suillus* and other ECM fungi, and between *Suillus* associating with different hosts (Red Pine, White Pine and Larch), were assessed using a two-way ANOVA with type two sum of squares to account for unbalanced sample design. For significant associations, differences between cluster type were assessed using post hock multiple t-tests at alpha = 0.05, with Holm adjustment for multiple comparisons. Differences in GPCR diversity for both comparisons of *Suillus*

to other ECM fungi, and for *Suillus* associating with different hosts was assessed using a two-way ANOVA with type two sum of squares, with GPCR class differences assessed using post hock multiple t-tests at alpha = 0.05, with Holm adjustment for multiple comparisons. All data analysis was carried out in the R programing environment (R Core team, 2017). All scripts and data associated with this project have been made open access and are available at: https://github.com/MycoPunk/Suillus_comp_genomics

Results

Genome size was significantly higher in *Suillus* than in Other ECM fungi (t = 2.2027, df = 22.048, p-value = 0.03836) assessed with T-test, but not significantly higher using randomization (p-value = 0.0628) (Table 2). This trend did not extend to total predicted proteome size, which was not significantly different between *Suillus* and other ECM fungi for either test. Neither genome size, or predicted proteome size were significantly different between *Suillus* associating with different hosts for either tests. KEGG metabolic pathway analysis showed no significant enrichment of major KEGG pathway categories between *Suillus* and other ECM fungi or between *Suillus* analyzed by host association. In total, KEGG mapped only a small percentage of predicted proteins to metabolic pathways, with an average of 19.7% of predicted proteins mapped for *Suillus* and 18.6% mapped for other ECM fungi (Table 2.S3).

Small Secreted Proteins

SSSPs as a percentage of SSPs were significantly more abundant in other ECM fungi than in *Suillus* (t = -4.9072, df = 9.4197, p-value = 0.0007 using T-test, and p-value = 0.0001 using randomization) (Fig. 2.1b). Overall SSSP abundance was lower in *Suillus*

than in other ECM fungi (t = -2.1905, df = 10.48, p-value = 0.0521 using T-test, and p-value = 0.0054 using Randomization) (Fig. 2.1e). No significant differences were found for SSPs as a percentage of total proteins (Fig. 2.1a), predicted effectors as a percentage of total proteins (Fig. 2.1c), total SSP abundance (Fig. 2.1d), or total abundance of predicted effectors (Fig. 2.1f) for either T-test or randomization.

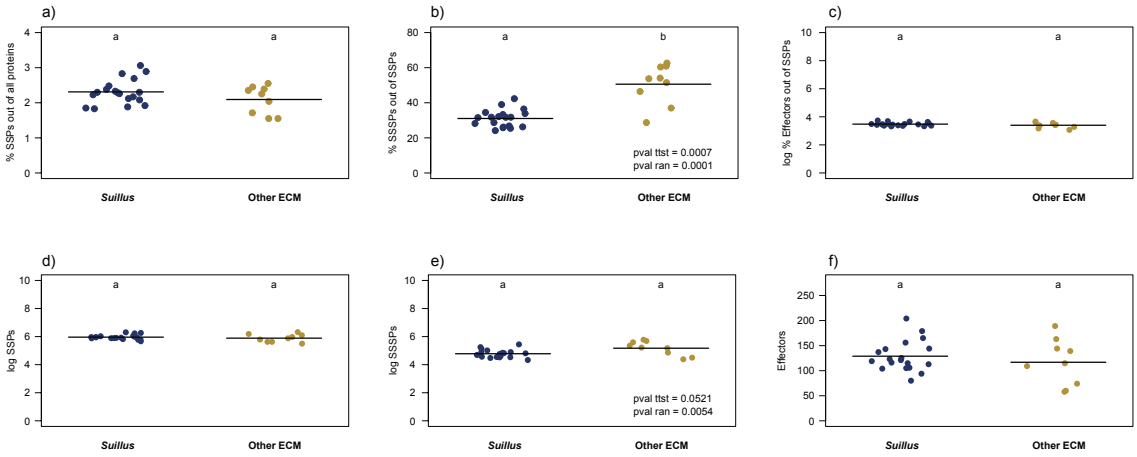

Fig. 2.1: Small molecule distribution in *Suillus* vs. other ECM fungi

SSP = small secreted proteins, SSSP= species-specific small secreted proteins. Normality evaluated by Shapiro-Wilk test, variance assumptions evaluated by Cochran's C test, and significance values assessed by t-test or Welch's t-test when variance assumptions were not met. Data was log transformed when assumptions of normality were not met. Different letters indicate significant differences between groups. ***a)*** *SSPs as a percentage of total proteins number of predicted proteins* ***b)*** *SSSPs as a percentage of SSPs was significant according to t-test (t = -4.9072, df = 9.4197, p-value = 0.0007) and randomization test (p-value = 0.0001)* ***c)*** *Effectors as a percentage of SSPs* ***d)*** *total abundance of SSPs* ***e)*** *total abundance of SSSPs was marginally significant according to t-test (t = -2.1905, df = 10.48, p-value = 0.0521) and significant according to randomization test (p-value = 0.0054)* ***f)*** *distribution of effectors.*

A similar trend was found for *Suillus* fungi associating with different hosts, where both SSSPs as a percentage of SSPs ($F = 9.239$, $df = 2$, p-value = 0.00319 for t-test and p-value = 0.00345 for Randomization) (Fig. 2.2b), as well as the total abundance of SSSPs ($F = 6.415$, $df = 2$, p-value = 0.012 for ANOVA and p-value = 0.011 for Randomization) (Fig. 2.2e), were significantly different between groups. For both SSSPs as a percentage of SSPs and total SSSPs abundance, TukeyHSD identified a difference between Red Pine associated *Suillus* and Larch Associated *Suillus* with White Pine associates appearing intermediate and not significantly different from the other groups (p-value = 0.0049 for SSSPs as a percentage of SSPs and p-value = 0.0242 for total SSSPs), however this effect was inverted between the two metrics, with Red Pine having the lowest number of SSSPs as a percentages of SSPs, but the highest number of total SSSPs. Randomization and pairwise comparisons differed slightly from t-test results, identifying Red Pine as significantly different than both Larch and White Pine, with no difference between Larch and White Pine, for both SSSPs as a percentage of SSPs (Red Pine vs. Larch with p-value = 0.0257 and Red Pine vs. White Pine with p-value = 0.0436), as well as for total abundance of SSSPs (Red Pine vs. Larch with p-value = 0.0252 and Red Pine vs. White Pine with p-value = 0.0466). No significant differences were found for SSPs as a percentage of total proteins (Fig. 2.2a), predicted effectors as a percentage of total proteins (Fig. 2.2c), total SSP abundance (Fig. 2.2d), or total abundance of predicted effectors (Fig. 2.2f) for either ANOVA or Randomization tests. See Table 2.2 for summary statistics.

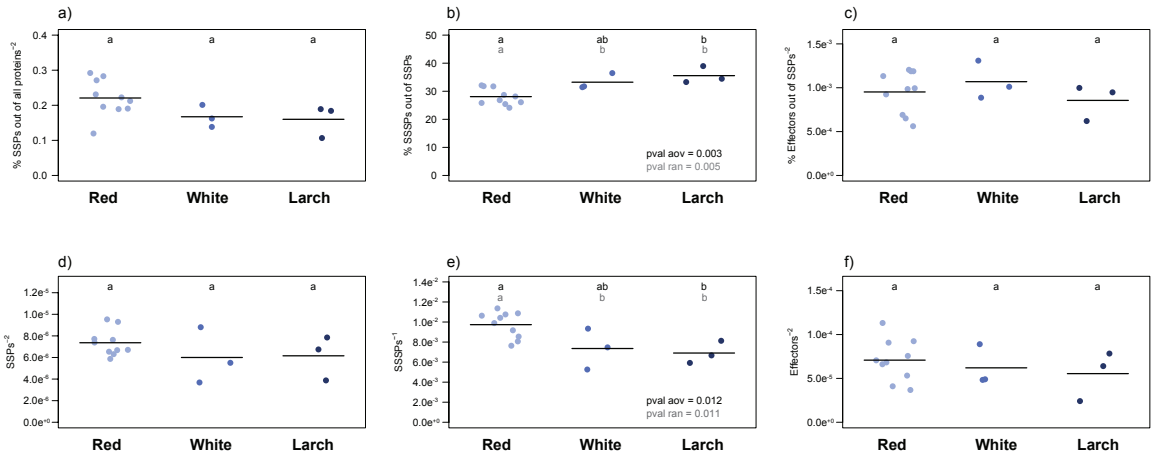

Fig. 2.2: Small molecule distribution in *Suillus* by host association

SSP = small secreted proteins, SSSP= species-specific small secreted proteins. Normality and variance assumptions evaluated by visual inspection. Data was transformed when assumptions of normality or variance were not met, with the transformation recommended by boxCox testing. Significance values were assessed using a one-sided type 1 ANOVA with type two sum of squares to account of unequal sample size, followed by TukeyHSD when significant at alpha < 0.05 (black) or by randomization test followed by pairwise multiple comparisons with Benjamini–Hochberg correction (Grey). Different letters indicate significant differences between groups. ***a)*** *SSPs as a percentage of total proteins number of predicted proteins* ***b)*** *SSSPs as a percentage of SSPs (F = 9.239, df = 2, p-value = 0.00319, TukeyHSD showed significant difference between Red and Larch hosts at p-value = 0.0049), randomization testing was significant at p-value = 0.005, with multiple comparisons showing significant differences for Red Pine vs. Larch (p-value = 0.025)7 and Red Pine vs. White Pine (p-value = 0.0436)* ***c)*** *Effectors as a percentage of SSPs* ***d)*** *abundance of SSPs* ***e)*** *abundance of SSSPs (F = 6.415, df = 2, p-value = 0.0115, TukeyHSD showed significant difference between Red and Larch hosts at p-value = 0.0242) Randomization testing was significant at p-value 0.011, with multiple comparisons showing significant differences for Red Pine vs. Larch (p-value = 0.0252) and Red Pine vs. White Pine (p-value = 0.0466)* ***f)*** *distribution of effectors.*

Table 2.2: Statistics table for *Suillus* vs. Other ECM, and *Suillus* by host association

First number in each field represent the mean, numbers proceeding +- represent standard error. Significance for Suillus vs. Other ECM assessed by t-test or permutation

test, and Suillus by host association assessed via ANOVA or permutation test, at alpha = 0.05. NS = not significant.

test used		genome size	proteins	SSPs	SSSPs	effectors	% SSP out of all prot.	% SSSPs out of SSPs	effectors out of SSPs
	Suillus	59.6 +- 3.24	17051 +- 423.60	393 +- 15.53	124 +- 8.92	129 +- 7.01	2.31 +- 0.08	31.04 +- 1.13	32.67 +- 0.98
	Other ECM	49.47 +- 3.26	18006 +- 1391.74	373 +- 34.83	195 +- 29.09	117 +- 15.44	2.09 +- 0.13	50.59 +- 3.82	30.43 +- 1.94
T-test	significant?	t = 2.2027, df = 22.048, p-value = 0.03836	NS	NS	marginal (p-value = 0.0521)	NS	NS	t = -4.9072, df = 9.4197, p-value = 7.3X10-4	NS
Randomization test	significant?	NS to marginal (p-value = 0.0628)	NS	NS	p-value = 0.0054	NS	NS	p-value = 1.0 X10-4	NS
	Red	59.22 +- 4.55	17276 +- 669.36	372 +- 9.21	105 +- 4.73	124 +- 7.19	2.18 +-0.10	28.1 +- 0.93	33.25 +- 1.52
	White	64.41 +- 11.42	17212 +- 1220.18	428 +- 53.13	144 +-24.44	131 +- 12.50	2.47 +-0.13	33.23 +- 1.62	30.89 +- 1.73
	Larch	55.72 +- 4.56	16214 +- 342.41	416 +- 46.38	147 +-13.35	147 +- 28.54	2.56 +- 0.25	35.56 +-1.73	34.76 +- 2.71
ANOVA	significant?	NS	NS	NS	p-value = 0.0115	NS	NS	p-value = 0.00319	NS
Randomization test	significant?	NS (p-value = 0.8497)	NS	NS	Z = 2.5452, p-value = 0.01092	NS	NS	Z = 2.9249, p-value = 0.00345	NS

Secondary metabolites

SMC analysis categorized core biosynthetic enzymes as belonging to either terpenes, t1pks, "other", nrps, indole, fatty acids, or "putative" (uncharacterized but meeting the criteria for cluster finder's prediction of SMCs). The overall abundance of SMCs was significantly higher in *Suillus* than in other ECM fungi (df = 1, F = 22.742, p-value = $3.783e^{-6}$), a result that was primarily driven by a diversity of terpene encoding genes (with an average of 23 in *Suillus* and 13 in other ECM fungi, p-value $1.7e^{-8}$) and SMCs falling into the "other" category (with an average of 12 in *Suillus* and 6 in other ECM fungi, p-value = $1.41e^{-2}$) (Fig. 2.3).

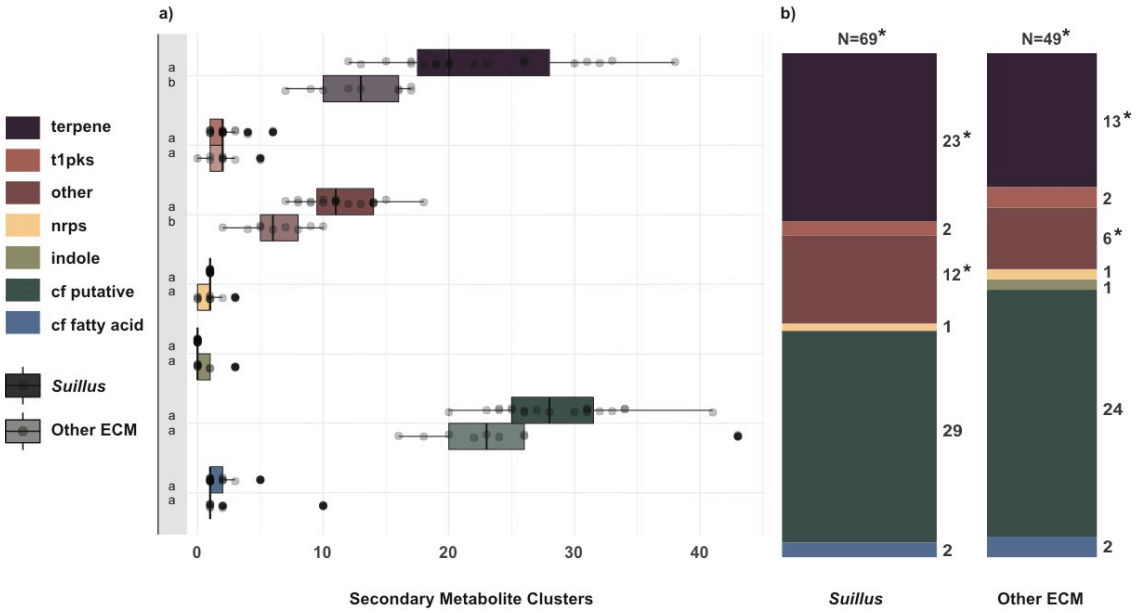

Fig. 2.3: Secondary metabolite distribution in *Suillus* vs. other ECM

"cf" = identified with the Cluster Finder algorithm, "putative" = uncharacterized, but identified by cluster finder as having the hallmarks of a SMC. Significance assessed with a two-way ANOVA using type two sum of squares to account for unbalanced sample design (df = 1, F = 22.742, p-value = $3.783e^{-6}$. Within group differences assessed using post hock multiple t-tests at alpha = 0.05, with Holm adjustment for multiple comparisons. Terpene difference significant at p-value = $1.7e^{-8}$, and "Other" (undefined) MSCs significantly different at p-value = $1.41e^{-2}$. ***a)*** *Box plots of SMC distribution representing the interquartile region intersected by the median. Different letters indicate significant differences between groups.* ***b)*** *Spine plots displaying the mean abundance of all ("N") and each (numbers to the right of each spine) SMC between* Suillus *and other ECM fungi. Significantly different groups are highlighted with an asterisk.*

No significant differences were found between SMCs in *Suillus* fungi associating with different hosts (Fig. 2.4). KEGG metabolic pathway analysis of terpene encoding genes classified terpenes primarily into di-terpene pathways, over tri/sesquiterpene pathways, regardless of the species in question (Table 2.S3).

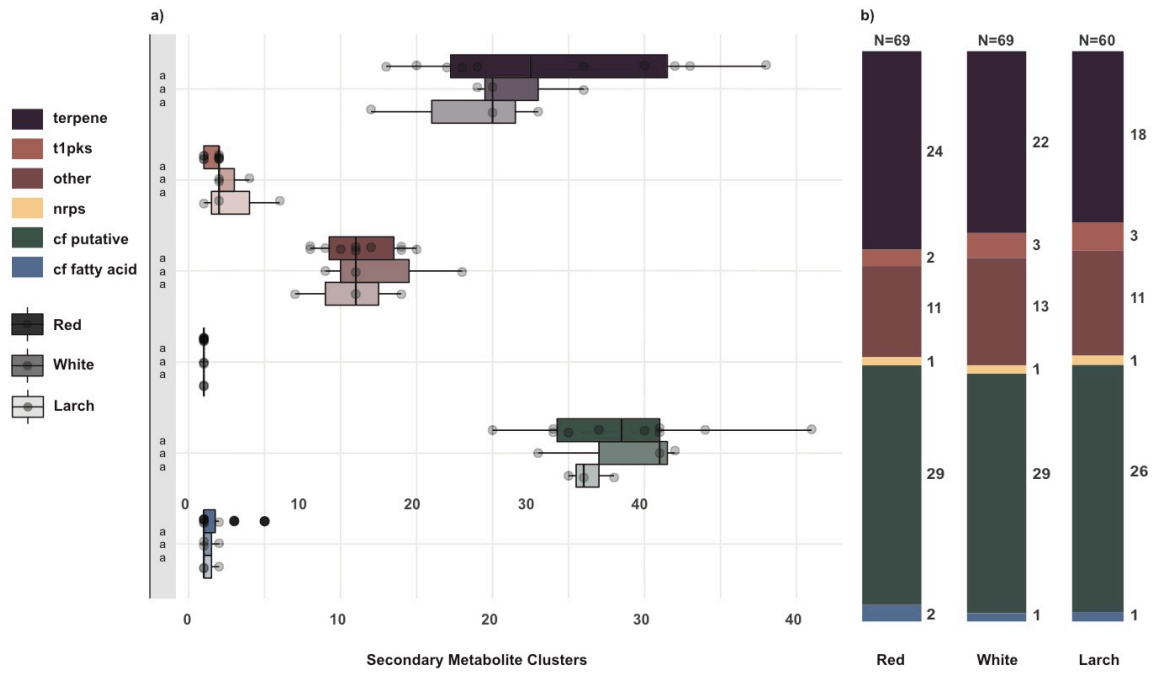

Fig. 2.4: Secondary metabolite distribution in *Suillus* by host association

"cf" = identified with the Cluster Finder algorithm, "putative" = uncharacterized, but identified by cluster finder as having the hallmarks of a SMC. Significance assessed with a two-way ANOVA using type two sum of squares to account for unbalanced sample design (df = 2, F = 0.6820, p-value = 0.5086). ***a)*** *Box plots of SMC distribution representing the interquartile region intersected by the median. Different letters indicate significant differences between groups. No differences were found.* ***b)*** *Spine plots displaying the mean abundance of all ("N") and each (numbers to the right of each spine) SMC between* Suillus *fungi associating with Red Pine, White Pine or Larch hosts.*

G-protein coupled receptors

GPCR abundance was significantly greater in *Suillus* than in other ECM fungi (df = 1, F = 5.363 P-value = 0.000189, with a mean GPCR count of 57 for *Suillus* and 38 for other ECM fungi). This difference was driven by class A GPCRs (p-value = $3.7e^{-7}$), whereas class B and D GPCRs were not significantly different between groups (Fig. 2.5).

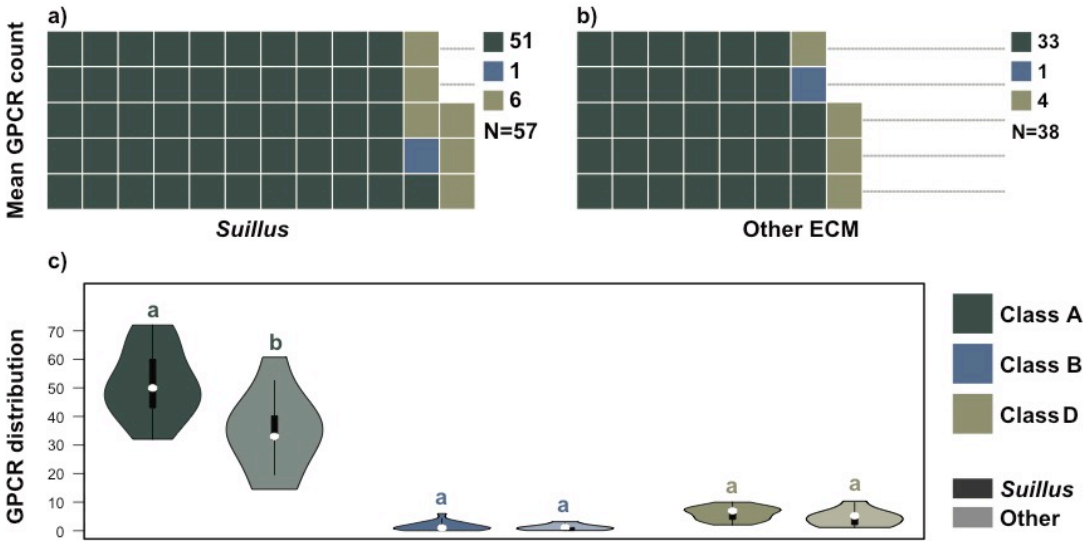

Fig. 2.5: GPCRs in *Suillus* vs. Other ECM fungi

Significance assessed with a two-way ANOVA using type two sum of squares to account for unbalanced sample design (df = 1, F = 5.363 P-value = 0.000189). Within group differences assessed using post hock multiple t-tests at alpha = 0.05, with Holm adjustment for multiple comparisons. Class A GPCRs were significantly different at p-value = $3.7e^{-7}$. ***a)*** *The distribution of GPCRs by class for* Suillus *or* ***b)*** *other ECM fungi, representing the mean number of GPCRs averaged over all* Suillus *(n = 19 species) or* ***b)*** *other ECM (n = 9 species). Numbers next to color keys represent the mean for each class, N = total mean. Numbers next to color keys represent the mean for each class, N = total mean.* ***c)*** *GPCR distribution by class, thin black lines represent 1.5x interquartile range, thick black lines represent interquartile range and white dots represent the median. Different letters indicate significant differences between groups.*

GPCRS were also significantly different between *Suillus* associating with different hosts (df = 2, F = 5.0086, p-value = 0.01159), where class A GPCRS were significantly higher for White Pine associated *Suillus* than for Red Pine or Larch associates (p-value = 0.0003 between White and Red Pine, and 0.0208 between White Pine and Larch associates) (Fig. 2.6).

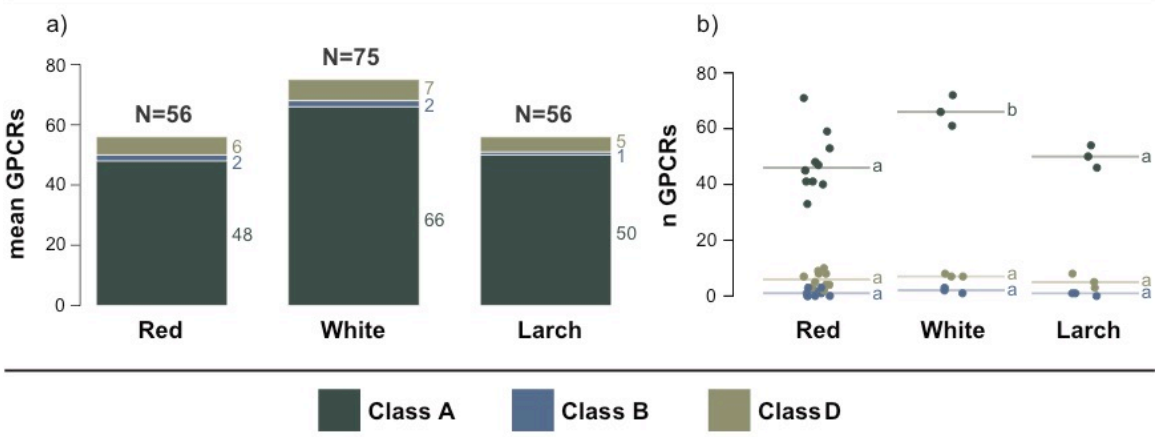

Fig. 2.6: *Suillus* GPCRs by host association

Significance assessed with a two-way ANOVA using type two sum of squares to account for unbalanced sample design. (df = 2, F = 5.0086, p-value = 0.01159). Within group differences assessed using post hock multiple t-tests at alpha = 0.05, with Holm adjustment for multiple comparisons. Class A GPCRS were significantly different between Red and White Pines at p-value = 0.0003 and Larch and White Pines at p-value = 0.0208. ***a)*** *Mean GPCRs by class. Numbers next to stacked bars indicate the mean abundance of all ("N") and each (numbers to the right of each spine).* ***b)*** *Distribution of GPCRs by class. For each class (color), different letters indicate significant differences between groups.*

Phylogenetic reconstruction

In total, 11,775 protein trees were constructed, of which 5406 contained representatives in all species and were used in consensus tree determination. Ancestral state reconstruction supported *Larix* as the ancestral host for *Suillus* (Fig. 2.7). The three Larix associated *Suillus* (*S. clintonianus*, *S. ampliporus*, and *S. paluster*) clustered on basal nodes of the tree, giving rise to a single independent origin for Red Pine associated *Suillus*. Conversely, the three White Pine associated *Suillus* species (*S. americanus*, *S.*

pictus and *S. placedus*) appeared to represent three independent host switching events, all from Red Pine associated ancestors.

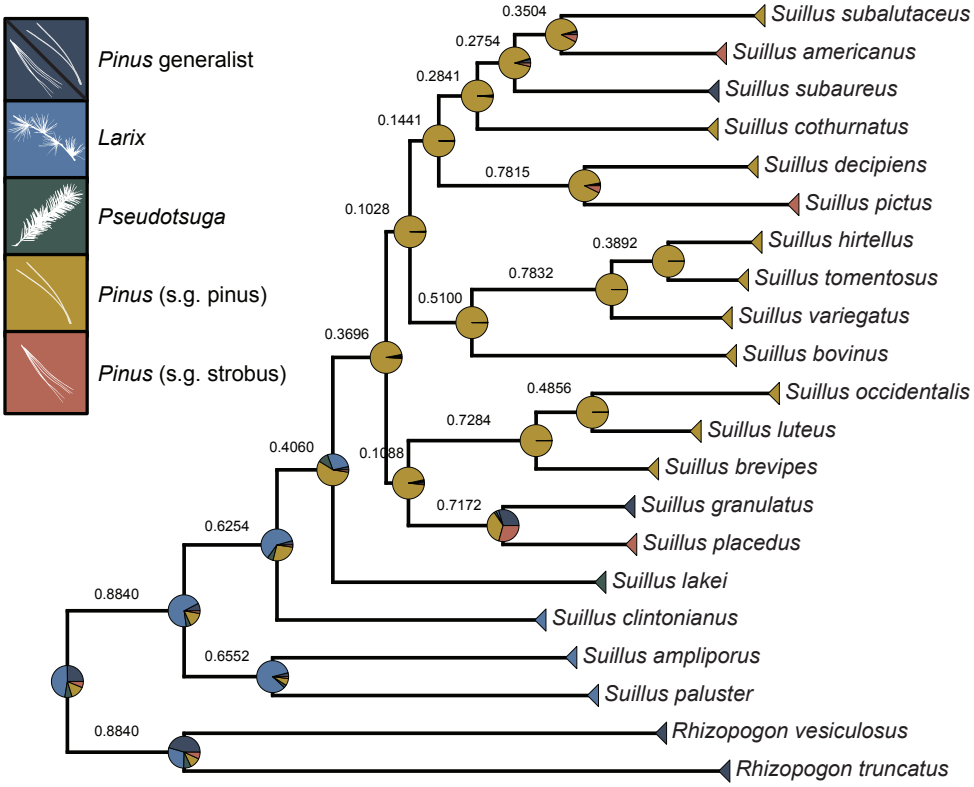

Fig. 2.7: Phylogenetic reconstruction of ancestral state and host switching

Orthologous gene groups shared between species were determined using OrthoFinder 2.0 running DIAMOND, with gene trees were inferred using DendroBLAST under default settings. The species tree was inferred using STAG and rooted by STRIDE. Ancestral state reconstruction employed a Bayesian stochastic mapping approach on a population of 100 character histories. Pie charts represent the posterior probabilities of ancestral host association at each internal node. Numbers above branches represent the proportion of bipartitions of individual gene trees where that bipartition occurs at the preceding node.

Discussion

Here, we show that relative to other ECM species, *Suillus* have a similar number of SSPs and canonically recognized effectors, but a lower number of SSSPs. *Suillus* also display significant enrichment in SMCs, particularly those encoding diverse terpene related enzymes, and significant enrichment in class A GPCRs. Intrageneric comparisons of *Suillus* by host association again support the role of SSSPs, and GPCRs in specific *Suillus*-host pairings, but do not support the role of SMCs. Phylogenetic reconstruction support multiple independent host jumps onto White Pine by Red Pine associated *Suillus* and point toward *Larix* as the ancestral host, a result which is in keeping with previous estimations (Nguyen *et al.*, 2016b).

In this study, we found no evidence to support the role of unique effector diversity in structuring *Suillus* specificity. Although it may be that the effector diversity is not driving host specificity in *Suillus* fungi, it is also possible that we were unable to detect a signal for effectors because the machine learning model used for effector identification was trained on pathogen data, and effector quality may intrinsically differ among ECM fungi. An abundance of SSSPs were identified across species in this study, and it is possible that many of these SSSPs do indeed act as effectors. The moderately lower number of SSSPs for *Suillus* relative to other ECM fungi is in keeping with the expectation that restricted host range is accompanied by gene losses, presumably corelated to the loss of traits needed to infect diverse hosts (Spanu *et al.*, 2010; Visser *et al.*, 2010; Baroncelli *et al.*, 2016). If SSSP diversity is associated with increased host generalism, we might expect that *Suillus* species which have undergone more recent host jumps (presumably those associated with White Pine) to have the highest number of

SSSPs relative to species that associate with more ancestral hosts such as Larch. However, although Larch associates had the lowest number of SSSPs overall, the number was not statistically different from the number of SSSPs encoded by White Pine associates. In contrast, Red Pine associates were significantly enriched in SSSPs over Larch associates and possibly White Pine associates (depending on the metric used). In this way, the diversity of SSSPs in Red Pine associates may actually be an indication of the genetic flexibility necessary for speciation and host switching onto White Pine. Permuted randomization tests were run on the SSP dataset and compared to results generated by t-test and ANOVA to assess possible bias introduced by unequal sample size. Because the conclusions drawn from t-test or ANOVA vs. randomization tests were not qualitatively different, randomization testing was dropped for the remainder of the analyses.

SMC enrichment in *Suillus* over other ECM fungi is driven primarily by genes encoding either TS/Cs or genes encoding unique enzymatic cores. TS/C genes are involved in signaling and communication across kingdoms in rhizosphere communities, playing critical roles in the process of recognition and response between fungi, bacteria, plants, and insects (Yoshida *et al.*, 2012; Schmidt *et al.*, 2017; Sharma *et al.*, 2017). Basidiomycete fungi produce primarily sesqui-, di-, and triterpenes (Quin *et al.*, 2014). Although many plant hosts, notably Pines, produce a large number of monoterpenes that inhibit fungal growth (Melin & Krupa, 1971; Huber & Bohlmann, 2006), to date, the only fungal monoterpene synthase genes described are from endophytic isolates in the ascomycete family *Xylariaceae* (Shaw *et al.*, 2015). In this study, KEGG analysis

classified most of the terpene encoding genes as di-terpene related, regardless of species, and only identified a few tri- or sesquiterpene encoding genes (tri- and sesquiterpene pathways are not separated in the KEGG database). Three of the genomes investigated here (*S. cothurnatus*, *S. subaureus*, and *S. luteus*) had a single positive KEGG hit for products assigned a mono-terpene encoding pathway. In all cases this was identified as (+)-neomenthol dehydrogenase. However, because KEGG mapping failed to identify any other enzymes or intermediary products associated with this pathway in any of the genomes in question, it is more likely that the identification of (+)-neomenthol dehydrogenase is an annotation error, than an indication of basidiomycete monoterpene production. Future work on the identification and classification of ECM TS/C genes would benefit from a high fidelity method specific to this this class of enzymes, such as that used by Quin *et al.*, 2013 to identify sesquiterpene encoding SMCs.

Several classification systems are recognized for GPCRs. The version used in the release of V 9.9.1 of GPCRDB and used to develop PCA-GPCR (used here), recognizes five classes of GPCRS, derived from across kingdoms: Class A Rhodopsin, Class B Secretin, Class C Metabotropic glutamate/pheromone, Vomeronasal receptors (here, equivalent to Class D fungal pheromones), and Taste receptors. The widely used GRAFS GPCR classification system also recognizes five classes, including Glutamate, Rhodopsin, Adhesion, Frizzled and Secretin, and was based on mammalian derived genes. Recent evidence shows that although some fungi possess canonical mammalian GPCR classes (except Secretin), fungi primarily encode a unique set of fungal-specific GPCRs (Krishnan *et al.*, 2012). Because of this, new classification schemes have been

suggested for fungal GPCRs, currently identifying 14 different classes, along with the recognition of low abundance orphan GPCR genes that do not resemble those currently characterized (Krishnan *et al.*, 2012; Brown *et al.*, 2018). Many of these GPCRs appear to be lineage-specific, such as the diversification of PTH-11 like genes in Pezizomycotina (Brown *et al.*, 2018). Currently, a systematic review of G-protein pathways in Basidiomycota is lacking, and may reveal yet undiscovered GPCR diversity, and necessitate further updates to fungal GPCR classification schemes. Lacking a bioinformatic tool capable of classifying these newly characterized fungal-specific GPCRs, alternative classification schemes using *de-novo* alignments of excised conserved 7-transmembrane domains, along with characterized references for each of the 14 putative classes will yield a better total picture of GPCR diversity in *Suillus*. Although the precise role of GPCRs in ECM mycorrhization is unclear, G-protein signaling is well established as a primary system for communication both between microbes, and between microbes and hosts. Although the canonical action of GPCRs is at the cell surface, it is now well established that GPCRs can also be internalized and act at the cell nucleus (Bhosle *et al.*, 2017). An additional role for fungal G-protein systems was suggested by Veneault-Fourrey & Martin (2011), who point out domain similarity between a viral capsid protein characterized by Meng & Li (2010) and a G-protein from *L. bicolor*, where the viral protein contains a characterized, functional, nuclear localization signal. It remains to be seen if this signal is similarly functional in ECM fungi, but if true, could suggest that GPCRs have the potential to interact with host machinery in a similar way to SSPs that require nuclear localization, such as MiSSP7 (Plett *et al.*, 2014).

In order for ECM colonization to occur, fungi must either suppress the host innate immune response, or prevent immunoactivation in the first place. Examples of immune avoidance can be evidenced by the collapse of gene families encoding cell wall degrading enzymes in ECM fungi, which could otherwise trigger host immune responses (Zamioudis & Pieterse, 2011; Veneault-Fourrey *et al.*, 2014; Pellegrin *et al.*, 2015). Active suppression of host immune responses by effectors, such as MiSSP7 undoubtedly play an important role in host colonization. However, effectors may be more famous for their ability to act as avirulence factors in fungal host-pathogen interactions. When recognized by the innate immune system of the host plant, SSP effectors (elicitors) betray their fungal origin, ultimately leading to disease resistance. Conversely, the ecological pressures and consequences of SSP effectors to act as avirulence factors in mutualistic associations is poorly understood. Mounting an initial defense to fungal invasion is necessary for host plants to exclude fungal pathogens, and the deployment of fungal immune suppression agents against these defenses would benefit invading hypha regardless of fungal lifestyle (Zamioudis & Pieterse, 2011). However, while it's assumed that there is positive selection for host plants to recognize effectors from fungal pathogens, rendering them avirulence factors, this is not the case for effectors from fungal mutualists.

It has long been proposed that high host specificity is driven by ecological trade-offs connected to resource specialization (MacArthur & Levins, 1964; Whittaker & Feeny, 1971). This argument assumes that maintaining access to diverse resources can only be accomplished at the sacrifice of performance, which in turn selects for an

optimized state of derived host specialization (Huey, 1984; Bruns *et al.*, 2002). Because gene loss is assumed to be permanent and host range contractions are often associated with genetic losses (Spanu *et al.*, 2010; Visser *et al.*, 2010; Baroncelli *et al.*, 2016), specialization was long considered to be an evolutionary irreversible state (Simpson, 1953; Moran, 2002). However, numerous examples have pointed to bidirectional transitions between generalist and specialist lineages, and it is now accepted that high host specificity is neither universally derived, or an innately optimized resource acquisition strategy (Appel & Martin, 1992; Desdevises *et al.*, 2002; Stireman, 2005; Hardy *et al.*, 2014; Ouvrard *et al.*, 2015). In ECM fungi, strict host specialists are rare (Horton & Bruns, 1998; Bruns *et al.*, 2002) and the high host specificity observed in *Suillus* does indeed appear to be a derived trait which evolved from an ancestral habit of host generalism (Nguyen *et al.*, 2016c). The prevalence of ECM host generalists is thought to be favored because, unlike fungal pathogens, the capacity to colonize diverse hosts is assumed to have a positive net impact for both plant and fungal partners (Harley & Smith, 1983). The ecological advantage driving host specialism in *Suillus* is unclear. Although the genus contains very few generalist species (host expansion, from one host on to many), host jumps (switching from one host onto another) appear to be common. The process of host jumping in *Suillus* cannot be assumed to be regulated by the same mechanistic or ecological processes that regulate host expansion from specialism into generalism. For example, whereas high specificity is associated with gene losses and generalism with gene family expansions, host switching (assuming both the gain-of-function ability of colonize a new host, and the loss-of-function ability to colonize a previous host) may be associated with both gene losses and gene gains.

The drivers of host jumping over host expansion in mutualisms are not clear. In general, host jumping could be either the result of partner switching (implying better resource opportunities with a new partner), or partner abandonment (implying a breakdown of the original mutualism, where the costs outweigh the benefits for one of the partners). Different ECM species display variability in nutrient trading dynamics with their hosts (Nara, 2006). This nutrient trading spectrum could result in host selection (filtering) scenarios that either punish poor performance or preferentially reward good performance, as has been suggested for legume-rhizobium, and arbuscular mycorrhizal mutualisms (Denison *et al.*, 2003; Kiers *et al.*, 2011). *Suillus* often constitute the most dominant fungal genus fruiting in Pineaceae dominant forests (Dahlberg, 1997), but, interestingly, a notably smaller fraction of colonized ECM root tips (Horton & Bruns, 1998), suggesting that *Suillus* fungi have high carbon-sink strength relative to other species. In order for the trading relationship between *Suillus* and host trees to be stable, *Suillus* may need trade nutrients more efficiently than ECM species with lower carbon-sink strength, or deploy mechanisms to keep their hosts even if more equitable symbionts are available. In fact, there is evidence that *Suillus* do not make up for their high sink-strength and trade nitrogen at lower ratios than other community members. In an *in-vitro* study using the ECM host generalist species *P. involutus* and several suilloid fungi (*Suillus* and *Rhizopogon*), *P. involutus* was found to transfer more than twice the amount of ammonium per mg of mycorrhizal tissue than suilloids (Bidartondo *et al.*, 2001). This scenario may be complicated if the net benefits of *Suillus* colonization are more complex than simple C:N trading efficiency, however, under certain environmental conditions,

even a marginally unequitable trading relationship could destabilize the mutualism toward occasional abandonment. In such cases, rather than employ active sanctions (such as controlling localized root mortality), ECM hosts could employ a strategy similar to the effector mediated resistance employed by plants against fungal pathogens, as suggested by Egger & Hibbett, (2004). In this scenario, host jumping over host expansion could be facilitated by selection for host recognition of effectors. In turn, this would drive selection for effector diversification on the part of the fungus, facilitating access to secondary host species which could become primary hosts if the ECM effector in question becomes an ECM elicitor and avirulence factor.

In this study, we did not find evidence for the diversification of canonical effectors in *Suillus* relative to other ECM fungi, but did find evidence for the diversification, GPCRs and SMCs. Although canonical effectors are undeniably involved in structuring fungal host compatibility, the definition of what an effector is or is not, is currently in revision as researchers expand the libraries of unique molecules known to interact with host nuclear machinery to facilitate colonization. Recent research into non-canonical effectors suggest that other molecular families, including SMCs, and perhaps even GPCRs, may have the capacity to contribute to fungal colonization in a similar way as canonical effectors (Veneault-Fourrey & Martin, 2011; Collemare *et al.*, 2019). Indeed, although MiSSPs are highly upregulated during symbiosis for model ECM basidiomycete *L. bicolor*, this is not case for the ECM ascomycete *Tuber melanosporum* (Martin *et al.*, 2010), suggesting that different lineages of ECM fungi may make use of different effectors, some of which may not fit the canonical definition. In addition to

expanding our characterization and understanding of mutualistic fungal effectors, the possibility exists for yet undiscovered regulatory pathways that facilitate host specific colonization via mechanisms that are entirely unique from effector biology.

Future directions and Concluding remarks

One potentially confounding factor in this study is the possibility of phylogenetic autocorrelation. Given the expected sequence similarity between closely related species, the diversity of a given group of molecules is difficult to assess across a wide phylogenetic scale without some measure of normalization. For example, in the all-*Suillus* group we would expect to find a smaller number of SSSPs than in the other ECM group, simply due to the decreased phylogenetic distance and assumed decrease in sequence diversity. In this way, looking at the diversity of SSSPs between *Suillus* and other ECM fungi, may result in an underestimate of relative SSSP diversity in *Suillus*. Accounting for these differences will necessitate the careful implementation of a correction factor for patristic distance over the entire species set. Although *in-silico* comparative genomic studies have greatly accelerated our ability to describe untapped genetic diversity in fungal genomes, as yet, comparative genomics alone cannot validate whether an enriched gene set is actually transcribed under a given context. Functional studies have largely been hindered in ECM fungi, due to difficulties transforming dikaryotic (n + n) organisms that do not reproduce sexually in culture. Directed by comparative genomics, future work using transcriptomics and functional characterization making use of new transformation platforms, will greatly advance our understanding of the mechanisms regulating ECM host specificity.

Much of what we know about the mechanisms facilitating host specificity in ECM fungi has been influenced by what we know about host-specificity in fungal pathogen systems. This perspective has led to the first mechanistic insights into the process of ECM colonization, and the identification of many molecular correlates of both compatible and incompatible ECM-host interactions. Just as the field of plant pathology has broadened its investigation of effector biology to include non-canonical and non-pretentious elicitors, broadening our investigation into the molecules responsible for mycorrhizal establishment on a given host is likely necessary to construct a full picture of how host compatibility and host range are structured in fungal mutualisms. In general, mutualists are more likely than pathogens to exhibit generalism over specialism (Borowicz & Juliano, 1991), but exceptions such as *Suillus* highlight blind spots in our understanding of the ecological benefits of specialization, the evolution and maintenance of stable trading relationships, and our knowledge of the genetic mechanisms defining compatibility and host range.

Table 2.S1: Culture origin and strain information for the 17 newly sequenced Suillus strains used in this study

specific_epithet	strain	strain_continent	strain origin	fruitbody_found_under	culture_holding	Notes
ampliporus	FC55	North America	Minnesota	Larix	Kennedy/Vilgalys	North American S. cavipes
subalutaceus	FC151	Eastern North America	Minnesota	Pinus (Pinus)	Kennedy/Vilgalys	
tomentosus	FC115	Western North America	Colorado	Pinus (Pinus)	Kennedy/Vilgalys	
occidentalis	FC124	Western North America	Colorado	Pinus (Pinus)	Kennedy	
paluster	FC165	Eastern North America	Minnesota	Larix	Kennedy/Vilgalys	
lakei	FC43	Western North America	Colorado	Pseudotsuga	Kennedy	
clintonianus	FC179	North America/Asia	Minnesota	Larix	Kennedy	North American S. grevillei
placidus	DOB743	Eastern North America	New York	Pinus (Strobus)	Vilgalys	
subaureus	MN1	Eastern North America	Minnesota	Pinus (Strobus)/Quercus	Kennedy	
variegatus	UH-Sva-Z1	Europe	Zolder, Belgium	Pinus (Pinus)	Colpaert	
bovinus	UH-Sbo-P2	Europe	Paal, Belgium	Pinus (Pinus)	Colpaert/Vilgalys	
spraguei	EM44	Eastern North America	Virginia	Pinus (Strobus)	Kennedy/Vilgalys	
weaverae	EM37	Eastern North America	Virginia	Pinus (Strobus)	Kennedy/Vilgalys	formerly North American S. granulatus
salmonicolor/cothurnatus	VC1858	Eastern North America	North Carolina	Pinus (Pinus)	Kennedy/Vilgalys	
americanus/sibiricus	EM31	Eastern North America	North Carolina	Pinus (Strobus)	Vilgalys	
decipiens	EM49	Eastern North America	North Carolina	Pinus (Pinus)	Vilgalys	
hirtellus	EM16	Eastern North America	North Carolina	Pinus (Pinus)	Vilgalys	

Table 2.S2: Sequencing and assembly statistics for all genomes used in this study

genus	se	JGI_project_code	isolate_number	genome_size	read_depth	contigs	scaffolds	scaffold_N50	scaffold_L50	gaps	% gaps	gene_models	sequencing_platform
Amanita	*muscaria*	Amamu1	Koide	40.7	125.7	3814	1101	54	0.17	2713	12	18153	Illumina
Hebeloma	*cylindrospori*	Hebcy2	h7	38.23	125.7	526	176	12	1.07	350	5.3	15382	Solexa, PacBio, 454, Sanger
Laccaria	*bicolor*	Lacbi2	v2	60.71	NA	584	55	5	4.31	529	2	23132	Illumina, 454
Paxillus	*involutus*	Paxin1	ATCC 200175	58.3	36.2	6222	2681	29	0.38	3541	15.7	17968	Sanger, Velvet, Solexa
Piloderm	*croceum*	Pilcr1	F1598	59.33	102.8	4469	715	33	0.53	3754	11.8	21583	Solexa
Pisolithus	*microcarpus*	Pismi1	441	53.03	87.1	5476	1064	89	0.15	4412	10.5	21064	Solexa, 454, Sanger
Rhizopogon	*truncatus*	Rhitru1	FC74	38.91	72.02	128	128	11	1	0	0	11852	PacBio
Rhizopogon	*vulgaris*	Rhivul1	FC72	39.9	281.22	1127	1127	71	0.14	0	0	11905	PacBio
Scleroderma	*citrinum*	Sclci1	Foug A	56.14	80.7	3919	938	63	0.24	2981	6.1	21012	Solexa
Suillus	*americanus*	Suiame1	EM31 v1.0	50.81	92	3604	1307	47	0.3	2297	7.1	17163	Illumina
Suillus	*ampliporus*	Suiamp1	FC55 v1.0	58.33	186.36	1601	1601	114	0.12	0	0	16527	PacBio
Suillus	*bovinus*	Suibov1	UH-Sbo-P2 v1.0	47.5	300.65	622	622	42	0.32	0	0	13537	PacBio
Suillus	*brevipes*	Suibr2	Sb2 v2.0	52.03	101.7	2205	1550	84	0.16	655	1.3	21458	Illumina and PacBio
Suillus	*clintonianus*	Suicli1	FC179 v1.0	46.84	76.3	288	288	41	0.31	0	0	15530	PacBio
Suillus	*cothurnatus*	Suicot1	FC179 v1.0	94.61	45.49	685	685	79	0.32	0	0	19836	PacBio
Suillus	*decipiens*	Suidec1	EM49 v1.0	62.78	91.3	3648	1391	48	0.34	2257	4.6	16894	Illumina
Suillus	*granulatus*	Suigr1	EM37 v1.0	42.34	80.3	1869	628	24	0.51	1241	4	15802	Illumina
Suillus	*hirtellus*	Suihi1	EM16 v1.0	49.94	109.5	1626	644	36	0.41	982	2.4	17067	Illumina
Suillus	*lakei*	Suilak1	FC43 v1.0	79.75	106.11	1154	1154	89	0.19	0	0	19384	PacBio
Suillus	*luteus*	Suilu4	UH-Slu-Lm8-n1 v3	44.49	80.09	67	67	10	1.39	0	0	16588	PacBio
Suillus	*occidentalis*	Suiocc1	FC124 v1.0	57.96	306.36	1584	1584	152	0.09	0	0	16030	PacBio
Suillus	*paluster*	Suipal1	FC165 v1.0	61.99	93.23	996	996	48	0.31	0	0	16585	PacBio
Suillus	*pictus (sprag*	Suipic1	EM44 v1.0	87.09	73.85	1400	1400	186	0.13	0	0	19349	Illumina and PacBio
Suillus	*placedus*	Suipla1	DOB743	55.32	284.53	753	753	36	0.42	0	0	15123	PacBio
Suillus	*subalutaceus*	Suisu1	FC151 v1.0	64.97	29.69	998	998	90	0.2	0	0	17080	PacBio
Suillus	*subaureus*	Suisub1	MN1 v1.0	57.66	64.5	668	668	24	0.68	0	0	15740	PacBio
Suillus	*tomentosus*	Suitom1	FC115 v1.0	53.06	88.2	203	203	33	0.52	0	0	17198	PacBio
Suillus	*variegatus*	Suivar1	UH-Sva-Z1 v1.0	64.86	240.13	1270	1270	99	0.17	0	0	17072	PacBio

Table 2.S3: Results of KEGG Metabolic pathway mapping analysis

Column names are JGI project identification codes, as identified in Table

	Sulpla	Sulgr	Sulvar	Sulbov	Sullak	Sulocc	Sulpal	Sulamp	Sulcli	Sulsu	Sulcot	Sulsub	Sultom	Sulpic	Sulame	Suldec	Sulbr	Sulhr	Sullu	Rhitru	Rhivul	Amamu	Hebcy	Lacbi	Paxin	Pilcr	Pismi	Sclci
#_1_ Amino Acid Metabolism	**386**	**465**	**422**	**403**	**474**	**402**	**416**	**421**	**394**	**429**	**490**	**420**	**431**	**534**	**720**	**534**	**490**	**530**	**388**	**355**	**385**	**425**	**445**	**446**	**482**	**488**	**366**	**398**
1_Alanine, aspartate and glutamate metabolism	25	30	35	26	31	30	25	33	31	33	35	28	31	64	47	33	29	38	30	27	26	26	31	35	28	28	25	33
1_Arginine and proline metabolism	63	54	71	66	69	63	78	72	57	63	94	62	70	93	75	65	53	63	59	54	61	44	43	45	42	47	40	47
1_Cysteine and methionine metabolism	46	48	51	44	50	45	39	37	40	52	54	44	40	48	61	46	51	45	45	40	40	43	37	41	47	41	32	43
1_Glycine, serine and threonine metabolism	59	70	71	67	72	68	65	83	62	75	66	65	72	75	97	78	75	85	66	60	60	56	65	55	52	87	45	49
1_Histidine metabolism	40	44	44	46	48	48	43	43	39	47	69	44	44	48	140	48	46	45	42	37	38	40	40	45	47	43	50	45
1_Lysine biosynthesis	19	23	21	20	21	21	19	19	22	19	24	20	19	32	38	26	28	29	18	19	26	24	26	26	21	24	23	24
1_Lysine degradation	58	58	59	63	73	63	60	58	52	57	74	58	69	66	108	65	64	65	57	50	58	51	55	57	61	60	49	45
1_Phenylalanine metabolism	38	30	44	42	46	42	51	46	40	48	56	41	40	65	94	32	30	33	39	37	39	29	26	26	32	28	29	30
1_Phenylalanine, tyrosine and tryptophan biosynthesis	22	18	25	19	25	20	17	20	20	25	25	21	21	25	30	16	17	19	25	20	21	18	17	14	18	16	17	16
1_Tryptophan metabolism	119	206	132	134	166	121	155	146	132	148	179	146	135	181	315	259	219	241	122	112	119	182	188	206	235	212	153	165
1_Tyrosine metabolism	50	50	53	54	64	60	43	45	51	53	79	62	62	69	138	52	50	44	60	51	42	47	43	44	44	53	41	43
1_Valine, leucine and isoleucine biosynthesis	24	23	24	24	30	21	25	24	24	29	33	24	25	32	29	25	22	23	22	22	21	21	19	17	26	18	19	18
1_Valine, leucine and isoleucine degradation	48	43	51	50	66	47	46	45	41	54	75	50	50	57	47	47	41	45	48	35	46	34	43	41	38	50	32	34
2_Biosynthesis of 12-, 14- and 16-membered macrolides	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2_Biosynthesis of ansamycins	2	2	2	2	2	4	2	2	2	2	2	2	3	2	2	2	3	2	2	3	2	1	3	2	2	2	2	3
2_Biosynthesis of siderophore group nonribosomal peptides	7	55	6	6	7	7	6	5	7	6	9	7	7	19	37	58	75	58	7	8	7	52	90	96	75	66	53	54
2_Biosynthesis of type II polyketide backbone	5	13	5	5	5	6	6	5	5	5	5	5	5	5	34	11	12	10	5	4	5	12	12	14	13	8	12	11
2_Biosynthesis of type II polyketide products	13	20	19	19	19	11	18	27	15	25	16	17	19	21	50	23	22	26	15	12	15	11	14	10	13	19	10	8
2_Biosynthesis of vancomycin group antibiotics	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	1	0	0	0
2_Nonribosomal peptide structures	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2_Polyketide sugar unit biosynthesis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	1	0	0	0
2_Type I polyketide structures	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#_3_ Biosynthesis of Secondary Metabolites	**210**	**285**	**237**	**238**	**291**	**220**	**255**	**264**	**251**	**272**	**312**	**269**	**252**	**306**	**492**	**349**	**288**	**334**	**227**	**225**	**215**	**238**	**221**	**221**	**296**	**303**	**190**	**219**
3_Acridone alkaloid biosynthesis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3_Anthocyanin biosynthesis	28	35	28	32	31	36	31	29	29	30	35	34	30	30	109	33	34	30	31	28	27	31	30	31	33	32	33	33
3_Benzoxazinoid biosynthesis	26	27	25	26	24	32	22	26	24	26	31	36	28	35	184	42	34	35	34	26	24	26	21	24	37	33	31	40
3_beta-Lactam resistance	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	4	2	0	2	0	0	
3_Betalain biosynthesis	19	26	22	26	25	24	22	20	21	27	22	24	24	51	23	25	23	20	17	21	22	22	21	25	22	21	24	
3_Brassinosteroid biosynthesis	10	19	8	8	10	8	8	11	8	12	10	12	9	15	136	34	28	30	12	10	8	19	16	17	33	26	29	38
3_Caffeine metabolism	36	70	46	42	68	34	57	52	54	57	53	55	49	59	16	101	57	95	35	35	37	61	41	44	77	58	32	33
3_Carotenoid biosynthesis	44	98	44	46	50	51	43	47	46	51	49	56	44	54	255	117	107	107	49	50	38	80	70	72	104	106	80	95
3_Clavulanic acid biosynthesis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3_Diterpenoid biosynthesis	25	31	31	28	25	31	27	33	27	30	32	34	31	38	154	44	37	37	32	31	28	27	22	26	45	29	31	42
3_Flavone and flavonol biosynthesis	11	16	11	15	14	14	14	12	12	12	13	14	13	12	16	14	18	17	12	11	11	11	13	8	13	14	11	16
3_Flavonoid biosynthesis	33	30	32	37	45	40	48	42	39	40	52	44	33	43	109	30	34	31	34	37	34	26	25	22	28	33	33	38
3_Glucosinolate biosynthesis	24	46	30	28	31	20	28	40	25	39	30	31	30	40	190	67	56	63	27	23	25	34	34	31	54	49	41	50
3_Indole alkaloid biosynthesis	2	2	3	4	3	2	1	1	2	2	3	2	2	5	2	2	2	3	2	1	1	2	2	1	2	1	0	1
3_Insect hormone biosynthesis	29	64	35	34	37	30	33	42	32	43	35	37	34	37	117	78	67	78	32	35	28	47	45	42	57	74	34	35
3_Isoflavonoid biosynthesis	4	2	5	4	6	5	4	7	4	8	7	4	4	5	28	2	4	5	4	5	4	3	3	2	4	5	9	7
3_Isoquinoline alkaloid biosynthesis	52	85	63	61	67	55	53	67	55	75	71	63	60	68	179	98	89	98	57	56	50	66	64	62	78	87	57	56
3_Limonene and pinene degradation	42	49	44	48	61	44	41	47	43	46	73	52	53	63	198	66	57	59	51	44	45	42	42	48	60	56	54	63
3_Monoterpenoid biosynthesis	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
3_Novobiocin biosynthesis	4	4	5	2	5	3	3	4	4	4	5	4	4	4	5	4	4	4	3	5	3	4	4	4	4	3	3	4
3_Penicillin and cephalosporin biosynthesis	3	0	2	4	3	2	5	4	3	3	3	4	5	3	6	0	0	0	4	3	3	0	4	2	0	2	0	1
3_Phenylpropanoid biosynthesis	40	45	35	40	50	44	55	46	50	52	61	50	34	50	208	64	57	56	42	42	35	49	40	38	56	65	56	62
3_Puromycin biosynthesis	4	35	4	4	6	3	3	3	5	5	4	4	3	3	5	45	36	45	3	10	2	27	23	24	35	45	14	19
3_Sesquiterpenoid biosynthesis	1	0	3	4	0	0	0	1	2	3	2	1	5	4	3	0	0	0	0	3	1	0	0	0	0	0	0	0
3_Stilbenoid, diarylheptanoid and gingerol biosynthesis	40	53	36	43	47	46	52	45	42	44	53	52	38	52	240	65	61	58	43	42	40	49	46	46	64	56	62	71
3_Streptomycin biosynthesis	7	7	6	7	7	7	7	9	7	6	8	6	6	6	10	6	7	5	6	8	7	5	7	8	6	11	7	7
3_Terpenoid backbone biosynthesis	9	25	8	8	10	8	7	8	10	8	10	8	8	9	24	21	25	20	8	8	7	8	16	18	22	13	14	15
3_Tetracycline biosynthesis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3_Tropane, piperidine and pyridine alkaloid biosynthesis	15	14	14	15	17	17	13	17	19	11	14	19	22	32	19	11	14	12	22	17	16	18	17	16	28	12	8	13
3_Zeatin biosynthesis	1	1	1	0	1	3	3	0	0	0	1	3	1	9	1	1	2	1	2	1	1	1	1	1	1	4	1	1
#_4_Carbohydrate Metabolism	**378**	**519**	**422**	**398**	**453**	**404**	**411**	**420**	**413**	**422**	**484**	**416**	**426**	**513**	**708**	**575**	**558**	**560**	**413**	**373**	**346**	**460**	**523**	**550**	**509**	**588**	**448**	**469**
4_Amino sugar and nucleotide sugar metabolism	91	121	103	106	115	100	101	112	106	114	118	103	102	145	179	139	124	138	103	103	92	104	112	105	110	132	76	86
4_Ascorbate and aldarate metabolism	63	60	72	68	78	61	66	84	61	74	89	67	71	74	225	79	67	76	69	49	56	50	49	48	57	69	49	55
4_Butanoate metabolism	55	67	60	64	71	55	60	69	57	70	80	59	63	65	121	72	69	73	56	43	51	51	59	59	57	64	51	47
4_C5-Branched dibasic acid metabolism	3	3	3	3	4	3	3	2	3	3	4	3	4	5	4	3	3	3	4	3	3	3	3	3	2	2	2	3
4_Citrate cycle (TCA cycle)	25	37	29	28	34	34	28	27	31	32	31	30	28	37	40	37	40	35	31	25	24	32	31	37	39	33	36	34
4_Fructose and mannose metabolism	60	68	71	69	78	59	66	79	67	77	74	71	72	91	110	75	78	77	66	59	58	56	56	44	59	65	50	52
4_Galactose metabolism	31	24	36	30	33	29	29	31	31	29	34	33	38	37	27	29	27	24	32	32	25	22	18	20	21	23	18	17
4_Glycolysis / Gluconeogenesis	41	46	43	46	50	46	44	51	48	53	72	44	48	61	54	48	51	48	42	35	38	47	46	49	45	65	43	46
4_Glyoxylate and dicarboxylate metabolism	15	20	17	17	16	16	28	16	14	15	21	16	16	22	23	21	21	22	17	17	23	16	20	20	16	18	24	25
4_Inositol phosphate metabolism	25	24	32	29	28	25	32	30	27	27	30	29	31	36	31	29	27	22	26	28	34	30	31	44	28	32	31	30
4_Pentose and glucuronate interconversions	27	53	34	29	34	25	24	26	27	28	36	28	34	31	19	56	55	55	30	29	20	36	36	37	45	56	31	45
4_Pentose phosphate pathway	28	59	30	24	30	30	25	34	32	29	29	29	32	46	91	60	66	58	30	20	21	54	62	55	61	70	75	70
4_Propanoate metabolism	29	28	28	34	41	30	27	28	30	28	52	32	37	41	31	29	28	30	35	28	31	23	25	24	24	35	22	22
4_Pyruvate metabolism	56	41	62	57	62	57	53	50	55	58	85	56	61	63	55	45	48	40	59	52	45	36	33	35	33	39	35	36
4_Starch and sucrose metabolism	53	95	59	56	61	58	62	58	59	60	69	56	65	72	57	111	104	110	56	55	50	82	131	148	97	117	85	73
#_5_ Energy Metabolism	**96**	**132**	**114**	**101**	**115**	**112**	**99**	**100**	**102**	**107**	**123**	**101**	**111**	**138**	**221**	**137**	**142**	**140**	**104**	**84**	**98**	**127**	**146**	**171**	**144**	**127**	**126**	**142**
5_Carbon fixation in photosynthetic organisms	19	18	23	17	20	21	18	18	21	19	22	18	21	33	22	20	21	19	23	17	16	19	21	19	20	16	17	22
5_Methane metabolism	20	11	32	24	29	25	30	27	21	30	31	23	34	26	38	12	13	13	22	15	28	14	11	8	16	13	16	14
5_Nitrogen metabolism	20	21	22	20	22	24	18	22	22	20	22	20	21	25	43	23	21	25	21	18	15	13	24	38	21	20	15	20
5_Oxidative phosphorylation	17	55	13	16	16	15	13	13	15	13	14	16	16	27	86	57	59	60	18	15	17	56	68	79	59	52	59	63
5_Photosynthesis	6	25	5	5	5	6	6	4	5	5	6	6	5	5	30	28	27	27	8	5	5	27	27	30	25	24	30	29
5_Photosynthesis - antenna proteins	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	1	1	1	1	1	1	1
5_Reductive carboxylate cycle (CO2 fixation)	9	14	11	9	11	13	8	6	10	11	17	11	8	15	14	13	14	12	11	9	8	11	10	14	11	13	10	12
5_Sulfur metabolism	14	15	16	19	20	17	15	17	16	18	23	17	14	17	24	14	16	13	14	13	16	15	13	15	20	14	11	14
#_6_ Glycan Biosynthesis and Metabolism	**105**	**219**	**109**	**111**	**128**	**113**	**112**	**107**	**107**	**114**	**124**	**108**	**118**	**137**	**314**	**212**	**237**	**224**	**113**	**104**	**97**	**238**	**239**	**264**	**244**	**259**	**222**	**221**
6_Chondroitin sulfate biosynthesis	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6_Glycosaminoglycan degradation	5	5	5	9	6	7	5	7	5	7	5	6	5	29	5	5	5	5	7	6	5	3	3	8	6	5	3	
6_Glycosphingolipid biosynthesis - ganglio series	19	27	20	24	25	24	23	19	21	21	24	22	23	21	48	25	26	24	20	20	19	23	22	22	28	23	24	25
6_Glycosphingolipid biosynthesis - globo series	16	16	18	21	26	21	20	16	18	19	19	19	20	18	22	15	16	16	18	23	17	13	12	9	17	19	14	16
6_Glycosphingolipid biosynthesis - lacto and neolacto series	12	13	12	16	15	16	15	13	13	15	14	14	14	14	15	12	13	12	13	11	11	10	10	7	11	14	10	13
6_Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	17	23	19	22	20	20	21	18	17	20	22	19	22	19	27	23	23	25	19	18	17	19	20	18	22	23	21	22
6_Heparan sulfate biosynthesis	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6_High-mannose type N-glycan biosynthesis	18	16	17	21	21	21	21	20	19	19	19	20	19	21	18	15	15	15	21	19	17	17	18	13	18	19	14	19
6_Keratan sulfate biosynthesis	12	13	12	16	15	16	15	13	13	15	14	14	14	14	15	12	13	12	13	11	11	10	10	7	11	14	10	13
6_Lipopolysaccharide biosynthesis	49	141	49	52	54	50	52	46	50	52	60	51	55	70	211	133	149	138	50	44	48	147	135	161	148	155	145	146
6_N-Glycan biosynthesis	28	39	27	30	34	28	31	32	29	29	31	29	29	31	36	36	36	36	27	27	26	38	36	36	38	42	35	40
6_O-Glycan biosynthesis	12	12	12	16	15	15	15	13	13	13	14	14	14	13	14	11	12	11	13	12	12	9	9	7	11	13	10	13
6_O-Mannosyl glycan biosynthesis	14	15	15	18	18	17	17	16	15	15	19	16	16	15	17	14	15	14	15	14	14	12	12	12	14	16	13	18
6_Other glycan degradation	12	25	14	12	17	16	12	10	12	13	16	12	13	14	13	29	37	33	16	15	9	35	50	49	47	46	28	25
6_Peptidoglycan biosynthesis	23	22	24	32	30	29	29	25	24	26	24	27	33	28	40	20	23	21	26	21	20	19	20	16	17	23	18	21

	Suipla	Suigr	Suivar	Suibov	Suilak	Suiocc	Suipal	Suiamp	Suicli	Suisu	Suicot	Suisub	Suitom	Suipic	Suiame	Suidec	Suibr	Suihr	Suilu	Rhitru	Rhivul	Amamu	Hebcy	Lacbi	Paxin	Pilcr	Pismi	Sclci	
#_7_Lipid Metabolism	**312**	**481**	**332**	**310**	**392**	**294**	**337**	**374**	**346**	**360**	**409**	**349**	**349**	**427**	**537**	**541**	**553**	**537**	**307**	**303**	**302**	**376**	**384**	**377**	**634**	**503**	**328**	**343**	
7_alpha-Linolenic acid metabolism	29	51	31	30	43	30	28	34	37	30	42	38	40	41	25	62	52	60	37	41	27	41	41	36	55	66	26	32	
7_Arachidonic acid metabolism	45	88	54	47	75	36	65	62	59	62	59	62	54	65	38	118	74	112	37	41	48	73	61	62	94	73	44	46	
7_Biosynthesis of unsaturated fatty acids	48	66	59	53	63	48	53	64	49	62	60	55	61	55	127	62	64	69	47	44	50	44	52	48	79	62	42	44	
7_C21-Steroid hormone metabolism	0	1	0	0	0	0	2	0	0	0	0	1	0	0	1	3	2	3	0	0	1	4	7	1	6	3	3	3	
7_Ether lipid metabolism	20	94	24	20	25	22	21	22	20	22	28	23	23	57	71	88	164	95	23	18	22	59	47	59	198	63	54	50	
7_Fatty acid biosynthesis	24	30	23	22	28	23	22	22	25	23	28	25	26	21	60	23	27	25	19	20	22	20	18	22	34	22	17	15	
7_Fatty acid elongation in mitochondria	9	12	9	10	14	9	8	8	8	8	10	8	9	8	15	12	11	13	9	7	8	11	14	14	14	22	10	11	
7_Fatty acid metabolism	70	115	88	80	113	65	89	86	87	96	112	87	95	101	57	151	99	140	64	65	68	91	84	79	114	118	64	66	
7_Glycerolipid metabolism	72	44	76	72	80	71	72	91	70	71	104	72	76	73	55	47	48	48	71	72	67	36	37	39	36	48	38	37	
7_Glycerophospholipid metabolism	39	49	44	38	41	38	39	42	39	40	40	38	39	75	76	52	48	47	40	35	35	44	44	41	52	50	43	47	
7_Linoleic acid metabolism	47	90	63	59	85	40	73	77	67	80	67	70	66	78	75	124	79	121	45	44	51	72	55	54	90	77	42	41	
7_Primary bile acid biosynthesis	16	16	14	13	21	16	19	21	19	28	25	21	16	18	24	16	15	14	18	17	19	8	9	13	9	17	8	8	
7_Secondary bile acid biosynthesis	4	33	4	3	6	3	3	3	5	5	5	2	4	3	3	6	43	34	42	3	10	2	25	21	21	34	43	12	17
7_Sphingolipid metabolism	54	77	49	54	64	56	60	60	60	57	67	59	54	74	200	98	94	91	61	56	48	87	98	100	98	109	85	93	
7_Steroid biosynthesis	23	50	22	23	23	22	19	21	24	25	27	28	27	44	134	66	58	61	26	24	23	36	35	40	82	61	48	53	
7_Synthesis and degradation of ketone bodies	9	7	7	7	8	8	8	9	8	7	8	8	7	7	8	7	7	7	7	3	6	4	6	7	4	4	4	5	
#_8_Metabolism of Cofactors and Vitamins	**265**	**407**	**307**	**284**	**341**	**273**	**312**	**315**	**308**	**324**	**343**	**311**	**300**	**374**	**638**	**474**	**423**	**467**	**276**	**279**	**263**	**381**	**338**	**368**	**422**	**444**	**337**	**365**	
8_Biotin metabolism	15	15	16	20	19	18	18	18	16	15	18	14	17	25	19	26	12	14	14	17	13	12	13	16	13	16	14	13	15
8_Folate biosynthesis	16	61	18	16	20	16	19	18	16	18	19	17	17	15	81	57	66	65	16	17	16	77	62	89	72	70	55	62	
8_Lipoic acid metabolism	4	0	4	3	3	3	3	3	3	3	3	3	3	4	1	0	0	0	3	3	3	0	0	0	0	0	0	0	
8_Nicotinate and nicotinamide metabolism	32	64	35	32	33	31	29	30	33	29	36	38	32	54	105	64	70	60	31	31	28	60	66	62	61	79	78	73	
8_One carbon pool by folate	12	14	14	13	12	11	11	12	11	11	12	12	13	13	15	12	12	13	11	10	11	14	12	13	14	13	13	15	
8_Pantothenate and CoA biosynthesis	23	22	35	27	29	25	28	26	27	27	30	27	30	39	24	26	23	24	27	22	21	18	18	14	20	17	16	17	
8_Porphyrin and chlorophyll metabolism	33	61	36	30	33	37	34	33	37	33	37	32	32	44	34	70	65	68	37	47	36	45	42	43	55	64	33	41	
8_Retinol metabolism	55	107	74	68	98	50	81	91	76	92	85	80	76	93	102	145	97	137	55	58	65	88	70	69	104	108	56	55	
8_Riboflavin metabolism	25	33	27	26	35	27	35	32	30	29	37	28	30	39	50	36	33	37	27	25	31	23	22	26	28	27	24	30	
8_Thiamine metabolism	9	10	11	9	13	9	11	11	8	9	9	10	11	13	24	10	9	10	10	10	10	9	11	10	11	10	10	9	
8_Ubiquinone and other terpenoid-quinone biosynthesis	43	50	41	43	53	50	50	51	54	59	60	52	38	49	216	66	62	56	48	46	39	48	38	42	57	54	55	62	
8_Vitamin B6 metabolism	12	9	13	11	14	10	10	9	10	10	15	11	12	10	4	9	9	12	10	10	7	11	8	12	9	14	11	12	
#_9_Metabolism of Other Amino Acids	**133**	**107**	**145**	**133**	**161**	**135**	**156**	**152**	**139**	**142**	**180**	**130**	**141**	**210**	**179**	**120**	**126**	**117**	**133**	**111**	**116**	**101**	**89**	**102**	**107**	**118**	**92**	**108**	
9_beta-Alanine metabolism	26	25	33	32	34	29	28	26	23	27	48	25	31	33	17	27	27	26	27	21	22	19	22	22	24	31	21	21	
9_Cyanoamino acid metabolism	26	21	28	26	32	27	41	34	29	30	37	26	28	46	47	26	25	23	23	24	25	16	13	11	18	22	17	18	
9_D-Alanine metabolism	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
9_D-Arginine and D-ornithine metabolism	3	0	2	3	3	2	5	4	3	3	3	3	3	2	2	0	0	0	3	3	3	0	0	0	0	0	0	1	
9_D-Glutamine and D-glutamate metabolism	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
9_Glutathione metabolism	57	40	57	48	65	52	60	67	60	58	65	50	56	74	63	47	50	46	55	41	43	48	37	49	43	46	39	49	
9_Phosphonate and phosphinate metabolism	7	8	6	7	7	7	6	7	6	5	7	6	6	34	10	8	8	9	8	6	6	4	7	5	6	7	3	5	
9_Selenoamino acid metabolism	16	17	19	17	20	19	16	14	18	19	20	20	17	21	38	16	18	17	17	16	17	18	15	22	20	17	16	19	
9_Taurine and hypotaurine metabolism	2	4	2	2	2	2	2	2	2	3	3	2	1	2	8	4	4	4	2	2	2	4	4	5	4	5	5	5	
#_10_Nucleotide Metabolism	**148**	**293**	**151**	**141**	**147**	**147**	**148**	**138**	**137**	**150**	**173**	**142**	**147**	**253**	**346**	**280**	**320**	**300**	**146**	**139**	**137**	**293**	**282**	**329**	**303**	**302**	**269**	**299**	
10_Purine metabolism	124	259	126	117	123	123	124	117	114	125	141	120	121	222	308	250	288	270	122	117	116	261	253	299	271	272	240	260	
10_Pyrimidine metabolism	66	117	65	61	62	61	59	52	58	64	76	62	65	85	117	114	129	119	61	58	56	114	114	119	120	121	110	117	
#_11_Overview	**292**	**352**	**327**	**306**	**375**	**300**	**345**	**336**	**338**	**359**	**392**	**345**	**332**	**502**	**555**	**409**	**359**	**396**	**302**	**296**	**291**	**309**	**289**	**305**	**388**	**351**	**293**	**293**	
11_Biosynthesis of alkaloids derived from histidine and purine	75	78	81	73	87	81	81	70	81	87	87	79	79	169	104	80	84	80	78	78	70	76	78	87	89	76	78	79	
11_Biosynthesis of alkaloids derived from ornithine, lysine and nicotinic acid	54	64	63	53	63	63	55	54	61	62	61	57	60	90	76	65	71	63	59	53	52	63	63	69	66	60	64	61	
11_Biosynthesis of alkaloids derived from shikimate pathway	134	174	163	147	190	143	162	164	168	191	187	165	154	195	228	213	173	213	138	128	131	161	143	145	186	166	133	129	
11_Biosynthesis of alkaloids derived from terpenoid and polyketide	54	84	63	57	67	65	56	56	67	68	71	61	61	93	85	82	90	86	61	55	55	73	76	88	108	86	78	72	
11_Biosynthesis of phenylpropanoids	83	98	89	80	107	93	98	96	102	105	112	98	80	120	220	113	116	115	90	88	78	105	95	103	118	117	114	124	
11_Biosynthesis of plant hormones	142	141	158	145	175	157	160	160	159	162	189	166	164	225	290	162	157	159	162	150	151	140	137	142	164	164	150	147	
11_Biosynthesis of plant secondary metabolites	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
11_Biosynthesis of terpenoids and steroids	73	112	85	77	89	83	77	86	86	96	93	90	81	119	258	127	124	118	83	73	73	105	99	114	123	115	116	119	
11_Overview of biosynthetic pathways	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
#12_Xenobiotics Biodegradation and Metabolism	**260**	**323**	**282**	**269**	**363**	**263**	**322**	**349**	**313**	**329**	**382**	**316**	**298**	**427**	**556**	**418**	**338**	**405**	**271**	**239**	**254**	**278**	**240**	**254**	**346**	**342**	**224**	**256**	
12_1,1,1-Trichloro-2,2-bis(4-chlorophenyl)ethane (DDT) degradation	3	0	4	3	4	4	4	6	4	4	3	7	3	5	17	0	0	3	4	3	4	0	0	0	0	6	1	2	
12_1,2-Dichloroethane degradation	12	8	14	16	18	14	15	13	12	14	39	14	16	15	8	11	8	9	12	9	12	5	5	6	4	8	6	5	
12_1,4-Dichlorobenzene degradation	28	27	26	25	31	24	30	28	32	29	32	31	24	41	147	41	34	37	28	25	22	31	25	28	40	34	37	51	
12_1- and 2-Methylnaphthalene degradation	30	43	31	27	42	31	32	39	33	48	50	35	32	37	107	45	43	39	33	33	35	29	35	38	36	47	35	34	
12_2,4-Dichlorobenzoate degradation	4	5	4	4	3	6	5	8	4	4	5	4	3	9	22	5	5	6	5	5	7	9	3	3	3	5	4	3	
12_3-Chloroacrylic acid degradation	13	17	16	17	20	15	15	15	12	16	34	13	18	17	12	21	17	16	14	11	13	11	12	12	9	24	12	12	
12_Atrazine degradation	3	2	3	3	3	3	3	2	3	3	3	3	3	4	6	2	2	2	3	4	3	2	3	2	2	3	2	2	
12_Benzoate degradation via CoA ligation	63	59	71	71	87	64	80	86	70	79	88	74	77	103	154	66	61	66	70	66	71	43	48	49	55	62	47	41	
12_Benzoate degradation via hydroxylation	12	7	13	11	14	13	14	18	13	13	10	15	12	17	24	7	7	8	13	13	13	7	8	7	8	11	9	8	
12_Biphenyl degradation	1	0	1	1	1	1	2	2	1	1	1	1	1	1	6	0	0	0	1	1	3	0	0	0	0	0	0	0	
12_Bisphenol A degradation	20	28	29	27	31	18	25	39	22	39	34	26	27	30	95	33	33	36	22	23	27	17	23	20	25	30	23	20	
12_Caprolactam degradation	32	24	25	30	39	32	26	43	37	31	43	35	34	46	38	26	22	30	38	27	24	22	29	29	22	37	15	21	
12_Carbazole degradation	2	0	2	1	3	2	2	4	2	2	0	1	2	1	0	0	0	0	2	2	2	0	0	0	0	0	0	0	
12_Drug metabolism - cytochrome P450	61	101	73	61	104	54	85	93	83	85	81	80	76	94	58	145	98	133	58	53	59	96	61	72	112	95	59	69	
12_Drug metabolism - other enzymes	48	89	57	53	82	43	71	63	68	72	67	65	59	112	32	117	76	112	42	44	46	81	63	72	114	76	44	51	
12_Ethylbenzene degradation	14	20	15	18	22	15	13	16	18	15	23	21	21	26	50	20	19	17	21	19	20	17	21	24	20	20	17	16	
12_Fluorene degradation	15	13	13	12	19	15	18	20	18	25	23	18	15	17	28	13	11	13	17	14	17	5	6	7	5	13	4	4	
12_Fluorobenzoate degradation	9	1	11	8	14	8	15	9	11	9	9	11	8	11	7	1	1	3	9	8	8	1	1	1	1	4	1	1	
12_gamma-Hexachlorocyclohexane degradation	74	107	96	88	139	70	120	113	115	111	127	114	102	131	195	172	113	166	78	84	90	94	68	77	134	109	86	99	
12_Geraniol degradation	13	11	11	15	23	13	11	11	15	10	14	17	19	25	18	12	10	13	20	15	13	10	14	14	14	21	9	10	
12_Metabolism of xenobiotics by cytochrome P450	63	96	76	65	106	57	88	97	85	91	84	83	80	99	56	132	93	124	61	58	58	85	59	64	103	91	51	60	
12_Naphthalene and anthracene degradation	50	53	59	55	76	55	57	74	70	75	82	76	65	76	251	81	68	74	62	61	59	45	39	42	63	65	65	69	
12_Styrene degradation	28	12	28	27	31	30	36	31	22	29	48	26	27	46	39	16	14	12	23	23	24	10	5	3	10	8	11	8	
12_Tetrachloroethene degradation	13	23	19	19	19	11	18	27	15	25	16	17	19	21	50	25	25	28	15	12	15	12	17	12	16	22	11	9	
12_Toluene and xylene degradation	14	7	25	22	35	17	21	33	31	26	27	25	30	28	29	18	17	22	19	23	20	9	6	4	15	9	25	19	
12_Trinitrotoluene degradation	3	33	3	2	5	2	2	2	4	4	1	3	2	2	0	43	34	42	2	9	1	25	21	21	34	43	12	17	

Chapter 3: Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles

Synopsis

Ribosomal DNA (rDNA) copy number variation (CNV) has major physiological implications for all organisms, but how it varies for fungi, an ecologically ubiquitous and important group of microorganisms, has yet to be systemically investigated. Here, we examine rDNA CNV using an *in silico* read depth approach for 91 fungal taxa with sequenced genomes and assess copy number conservation across phylogenetic scales and ecological lifestyles. rDNA copy number varied considerably across fungi, ranging from an estimated 14 to 1442 copies (mean = 113, median = 82), and copy number similarity was inversely correlated with phylogenetic distance. No correlations were found between rDNA CNV and fungal trophic mode, ecological guild, or genome size. Taken together, these results show that like other microorganisms, fungi exhibit substantial variation in rDNA copy number, which is linked to their phylogeny in a scale-dependent manner.

Introduction

Ribosomes are a central component of life on Earth and, to meet varying needs for protein production, the genomes of most eukaryotic organisms contain multiple copies of ribosomal DNA (rDNA). There is considerable rDNA copy number variation (CNV) both within and among taxonomic groups, typically totaling less than 15 copies in prokaryotes (Liao, 2000), 39-19,300 copies in higher animals (Prokopowich *et al.*, 2003),

150-26,048 copies in plants (Prokopowich *et al.*, 2003), and up to 315,786 copies in ciliates (Gong *et al.*, 2013). Copy number is a rapidly evolving trait and mechanisms for both rDNA copy number expansion and contraction have been described (Szostak & Wu, 1980). The consequences of rDNA CNV have received considerable attention in the context of DNA damage response (Ide *et al.*, 2010), DNA replication stress (Salim *et al.*, 2017), and the expression of non-ribosomal genes (Paredes *et al.*, 2011). Similarly, the ecological importance of rDNA CNV has also been well characterized, with rDNA copy number being linked to ecosystem stoichiometry (Elser *et al.*, 2000), growth rate and competitive ability (Klappenbach *et al.*, 2000; Nemergut *et al.*, 2016) as well as bias in estimates of organismal abundance in high throughput amplicon sequencing (Kembel et al. 2012, Perisin et al. 2016).

Relative to other microorganisms, estimates of rDNA CNV for fungi have been limited and consequently there has been no large-scale analysis of rDNA CNV with this ecologically important group of microorganisms. From the studies available, fungal rDNA CNV has been estimated to range between 28 and 511 (Maleszka & Clarkwalker, 1993; Liti *et al.*, 2009), which falls intermediate between prokaryotes and many larger eukaryotes. There have also been estimates of considerable rDNA CNV amongst strains of the same fungal species, with Liti et al. (2009) estimating that different strains of *Saccharomyces cerevasiae* had rDNA copies ranging from 54 to 511. The dikaryotic nature of many fungi suggests there may even be rDNA CNV amongst genetically distinct nuclei within a fungal individual (Zolan, 1995).

Despite a rapid increase in the sequencing of fungal genomes in recent years, estimates of rDNA CNV from annotated genomes has remained hindered by the collapsing of repetitive regions into a single representation. One solution to this problem is comparing the abundance of raw reads aligned to both single and multi-copy regions of DNA, an approach commonly known as relative read depth. Analysis of CNV using read depth was first developed to analyze repeat variation in tumor genomes (Chiang *et al.*, 2009), and later used to account for anomalies in 16S read abundance in bacteria (Perisin *et al.*, 2016). Here, we apply this approach to estimate rDNA copy number across a phylogenetically and ecologically diverse suite of fungi (Fig. 3.1).

Based on the significant positive relationships observed between rDNA copy number and phylogenetic relatedness among other microorganisms (Kembel *et al.*, 2012), we hypothesized that variation in rDNA copy number would exhibit significant phylogenetic signal in fungi. Additionally, due to the association of rDNA copy number and the physiological phenomena noted above, we hypothesized that rDNA CNV would also be linked with fungal ecological lifestyle. Specifically, because rDNA associated traits such as rapid growth or stress tolerance may be more crucial for some fungal lifestyles than others (e.g. pathogens vs. mutualist fungi), we predicted that there would be a significant association between fungal ecological lifestyle and rDNA copy number. Finally, because rDNA copy number has been reported to be significantly positively correlated with genome size in other eukaryotes (Prokopowich *et al.*, 2003), we investigated the relationship between rDNA copy number and genome size, both dependent and independent of size contributions from rDNA in each genome.

Materials and Methods

Copy number estimation pipeline

To assess rDNA CNV across a broad phylogenetic range of fungi, we selected 91 taxa with available genomic data, spanning phyla to interspecific populations. We also choose isolates to represent a wide variety of ecological lifestyles, including pathogens, saprotrophs, plant mutualists, and taxa capable of multiple lifestyles. Raw reads for each taxon were transferred from the Joint Genome Institute's MycoCosm site (Grigoriev *et al.*, 2014b) to server space at the Minnesota Supercomputing Institute (MSI) using Globus (Foster, 2006). Quality scores were converted to PHRED33 using Trimmomatic where necessary (Bolger *et al.*, 2014). The ITS and LSU gene regions as well as 10 single-copy reference genes were collected for each sequenced taxon. Single-copy genes were obtained as genomic .fasta files (with introns included) by keyword searching MycoCosm within the complete annotated assembly of each genome (Table 3.S1). Current sequencing technologies (including long-read platforms) do not produce reads long enough to span multiple copies of the full rDNA cassette. As such, reads from multi-copy regions, such as the internal transcribed spacer region (ITS) or the large subunit rRNA gene (LSU), are often unable to accrue the confidence values necessary to warrant placement and are therefore typically excluded from genome assemblies. To overcome this issue, we procured the ITS and LSU reference regions unique to each genome from the EST clusters associated with each sequencing project. This was accomplished by BLAST searching ITS and LSU ($E = 1.0x10^{-5}$, word size = 11) sequences from the same

genus (search sequence randomly chosen from NCBI) against the EST database associated with each genome on MycoCosm (Fig. 1.1). The nucleotide sequences of these EST clusters, internal to each genome, were then used in all downstream analyses. To confirm that EST clusters were high-fidelity sequence representatives, we compared EST derived ITS sequences with Sanger-sequenced ITS regions for a subset of the same strains (n = 7) that were used to generate the assemblies and found the average number of incongruences to be 1.2 bp. DNA for Sanger sequencing was extracted using the REDExtract-N-Amp plant kit (Sigma-Aldrich), followed by PCR amplification using the primer pair ITS1-F/ITS4 (White *et al.*, 1990; Gardes & Bruns, 1993) and sequenced at the University of Minnesota Genomics Center. Sequences were aligned using Sequencher v5.1 (Gene Codes Corporation, Ann Arbor, MI) using default parameters to count incongruencies between EST and Sanger derived sequences. For taxa where JGI annotations (single-copy genes), or EST clusters (multi-copy genes) were not available, reference sequences were procured from raw reads using -mpileup from bcftools in the SamTools package (Li *et al.*, 2009). ITS reference sequences were trimmed on either side of the priming regions for ITS1-F and ITS4 (White *et al.*, 1990; Gardes & Bruns, 1993) leaving ITS reference regions that were approximately 650 bp in length. LSU reference sequences were trimmed at the priming region for LROR (Rehner & Samuels, 1995) and again 750 bp downstream.

Reference sequences were indexed using Bowtie2 (Langmead & Salzberg, 2012). Demultiplexed paired-end reads for each genome were aligned to each reference gene individually [parameters: paired-end and –very-sensitive-local mode with a maximum

number of unknown base calls equal to 0.15 x read length, and alignment score benefits dependent on PHRED values] (Fig. 1.2). Sorting and depth calculations were carried out using SamTools v1.3 (Li *et al.*, 2009), with an increased max depth of 1 Mbp, and excluding reads with average quality scores < 20. To correct for GC bias, GC normalization was conducted using a custom R script (R Core team, 2017), employing a sliding window method as conceived by (Yoon *et al.*, 2009) (see script: gc_norm.R) (Fig. 1.3). Depth was then averaged over the length of each gene, minus the first and last 50 bp (which had misrepresentatively low depth due to alignment overhangs) (Fig. 1.4). For the 40% of taxa where such data was available, we analyzed sequences generated across two independent sequencing lanes to estimate stochastic variation introduced during the sequencing process. Single-copy genes with an average depth outside one standard deviation of the median value for each independently sequenced lane were excluded from the analysis. The copy number for multi-copy regions was estimated by dividing the GC normalized depth of the average depth of ITS and LSU by an average of the GC normalized depth across all single copy regions (Fig. 1.5), and averaged against the two independently sequenced lanes (where possible). All analyses were carried out using batch submission to the MSI computing cluster (see cnv_pipeline.pbs for pipeline bash script).

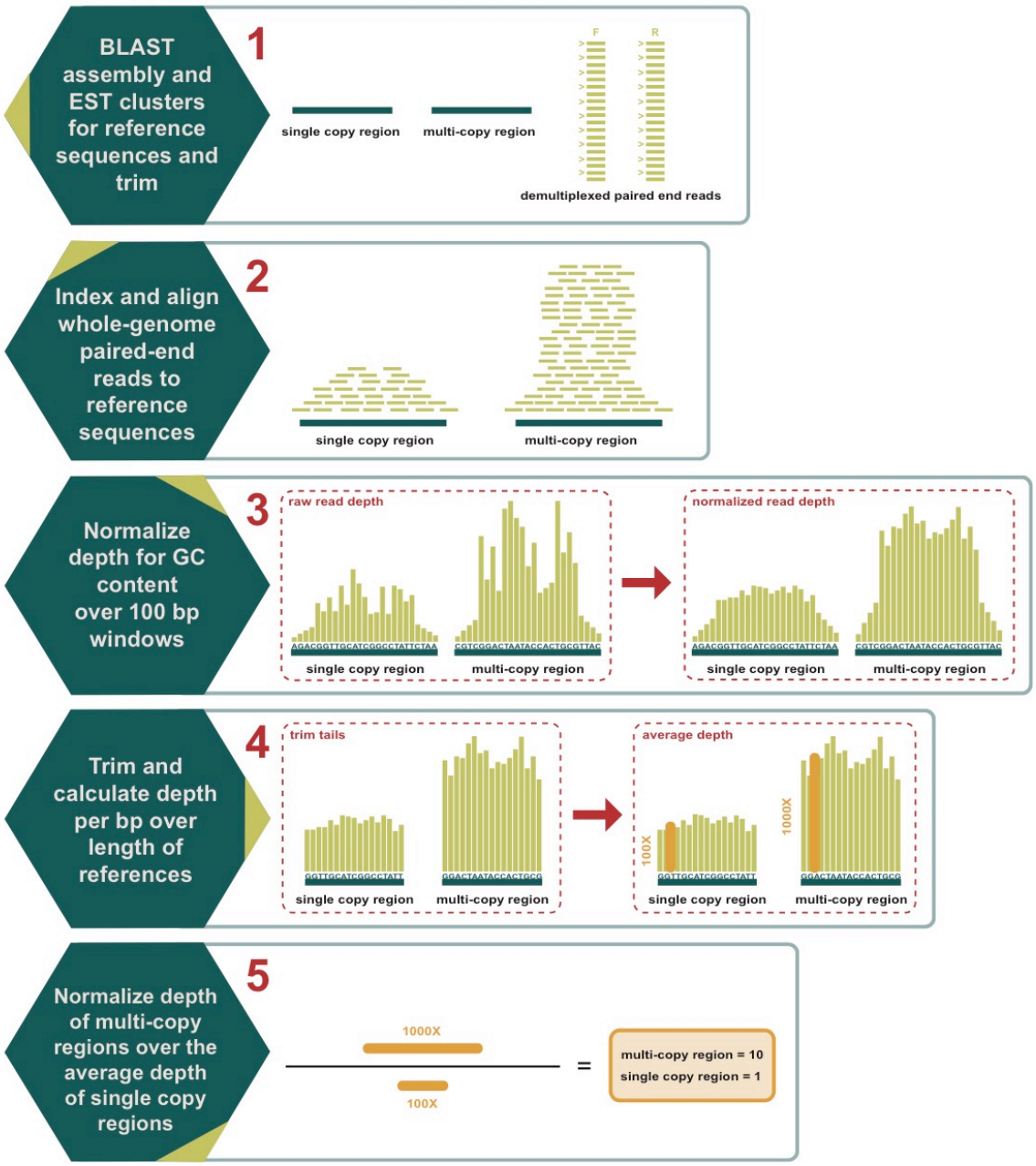

Fig. 3.1: Analysis pipeline for estimating rDNA copy number

1) *Demultiplexed paired-end reads from whole genome sequencing projects, along with 10 single-copy reference genes, are collected for each species.* ***2)*** *paired-end reads aligned to indexed references using Bowtie2.* ***3)*** *Alignment depth over each reference bp is normalized for variable GC content using a 100 bp sliding window.* ***4)*** *Overhangs are trimmed from alignments and average depth calculated over the length of each reference.*

5) *Number of rDNA copies is calculated by dividing the average depth of single copy alignments by the average depth of multi-copy alignments (ITS and LSU).*

In-silico verification of copy number estimation pipeline

A mock genome was generated consisting of 52 million randomly drawn base pairs (which falls within the genome size range of the fungal taxa included) in R. Using the reference regions for *Suillus brevipes* (a randomly chosen reference species), 60 concatenated multi-copy cassettes consisting of tandem ITS and LSU repeats, along with the 10 single-copy reference genes for *S. brevipes* were inserted into known, non-overlapping, locations in the mock genome (see script: generate_mock_genome.R). Twenty seven independently drawn sets of paired-end reads were then generated, varying in size from 1 to 50 million reads, formatted as .fastq files with idealized quality scores of ~ (representing the highest possible PHRED value in ASCII code), and run through the ITS CNV pipeline (see script: generate_mock_reads.R) (Fig. 3.2).

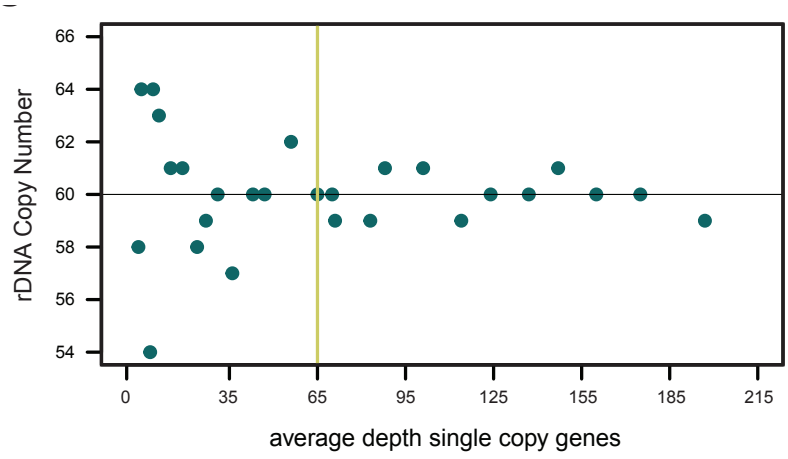

Fig. 3.2: Mock genome pipeline validation

Observed rDNA copy number estimates for a mock genome containing 60 rDNA copies. Black horizontal line at 60 represents expected number of copies. Green vertical line indicates where copy number estimates are +/- 1 copy from expected after a depth of ~65x/bp.

Phylogenetic analysis

A phylogeny containing the 91 fungal taxa was constructed using DNA sequences from three single-copy genes: TOP2, GH63 and MCM7. Alignments for each gene were carried out using MUSCLE (Edgar, 2004) on the CIPRES portal (Miller *et al*. 2010), and trimmed using trimAl (Capella-Gutiérrez *et al.*, 2009) to remove gaps and non-informative positions. Sequences from the three genes were then manually concatenated, realigned, and re-trimmed resulting in 8096 informative positions. Phylogenetic analysis was conducted using RAxML HPC2 (Stamatakis, 2006) on XSEDE (Towns *et al.*, 2014) run with default parameters, which utilized a 16 state GTR model and calculated bootstrap support based on 1000 iterations. Results were visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree).

Statistical analyses

To determine whether fungal rDNA CNV displayed phylogenetic signal (i.e. conservation of rDNA copy number among more closely related taxa), we used the R package 'phylosignal' (Keck et al., 2016) on the non-ultrametric tree described above. This package calculates multiple evolutionary- and correlation-based metrics and allows for tests within internal nodes to identify significant 'local hotspots' of trait conservation. Based on the recommendations of Münkemüller et al. (2012), Bloomberg's K and Pagel's λ were selected for the evolution-based metrics, while Abouheif's C_{mean} and Moran's I

were selected for the spatial correlation metrics. The assessment of phylogenetic signal at internal nodes was conducted using the 'lipaMoran' function, which calculates local Moran's I (I_i). To determine if ecological lifestyle and rDNA CNV are related, we first assessed fungal taxa grouped by trophic mode – saprotrophic, pathotrophic, symbiotrophic, as well as those belonging to multiple trophic modes. We also tested differences in rDNA CNV among specific guilds containing sufficient taxon replication ($N \geq 5$): soil/litter/organic matter saprotroph vs. pathogen within the Ascomycota and wood rot saprotroph vs. ectomycorrhizal with the Basidiomycota. By delineating these latter two analyses by phyla, we sought to minimize the effect of phylogenetic relatedness (see results below). To determine significance in these ecological analyses, we used either parametric (ANOVA) or non-parametric (Kruskal-Wallis) tests depending on variance heterogeneity. Given the highly divergent rDNA copy number estimate for *Basidiobolus meristosporus* relative to all other taxa (see results below), we took a conservative approach and performed all the phylogenetic and ecological analyses with this taxon excluded.

For all genomes where full assembly sizes were published (n = 79), we analyzed the correlation between rDNA CNV and genome size. Because repeat regions are not included in genome size estimates based on assembly size, we analyzed the relationship between rDNA copy number and genome size both including and independent of the length contribution of the rDNA cassettes themselves. Length contribution from rDNA for each genome was estimated by taking the number of rDNA copies estimated for each genome and multiplying that by an assumed average rDNA cassette length of 9.1 kb

(Miyazaki & Kobayashi, 2011) and then adding that additional size to each assembly size. Because phylogenetic signal analysis showed that rDNA CNV differed significantly by phyla (see below), we conducted correlational tests on both the whole data set and when subset by phylum. We used both parametric (Person's r) and non-parametric tests (Kendall's tau and Spearman's rho). For Pearson's r, the data was log transformed when appropriate to normalize the distribution, according to visual inspection (Plotting) and numeric evidence (Shapiro-Wilk test of normality). To account for phylogenetic autocorrelation, we also constructed a Phylogenetic Generalized Least Squares Model (PGLS) with a sub-set tree constructed as above, and implemented in the R package Caper (Orme *et al.*, 2014).

Results

From the 27 independent mock genome read libraries simulating variable sequencing depths, we found that our CNV estimation pipeline consistently returned the number of copies expected (+/-1 copy) after a read depth of 65X (Fig. 2) (see scripts: generate_mock_genome.R, and generate_mock_reads.pbs). As such, we used 65X as the minimum read depth necessary to confidently estimate rDNA CNV. The estimates of rDNA copy number among the 91 taxa analyzed exhibited some variation between sequencing lanes, with an average between lane difference of 14.9% (±2.4% S.E.) (Table 3.S2). The upper (251) and lower (11) limits of rDNA copy number estimates fell within the range of previous estimates for fungi, with the exception of *Basidiobolus meristosporus*, which had an estimated 1442 rDNA copies (across fungi mean = 113 copies (98 with no outlier), median = 82 copies with or without outlier), Fig. 3.3a). Both

the evolutionary (Bloomberg's K and Pagel's λ) and spatial correlation (Abouheif's C_{mean} and Moran's I) metrics indicated significant phylogenetic signal in rDNA CNV (Fig. 3.3b). Across the entire fungal phylogeny, there was a significant positive correlation between rDNA copy number and taxa at closer phylogenetic distances, but a significant negative correlation at greater distances (Fig. 3.3c). The negative correlation was particularly notable at the level of phylum, where, on average, Ascomycota taxa had only half as many copies as those belonging to the Basidiomycota or early diverging lineages (Fig. 3.3d). A similar trend in phylogenetic signal was also observed in the local Moran's index analyses, where all taxa with significant negative I_i values (rDNA copy number lower than expected) were in the Ascomycota and all those with significant positive I_i values (rDNA copy number higher than expected) belonged to the Basidiomycota or early diverging lineages (Fig. 3.3a).

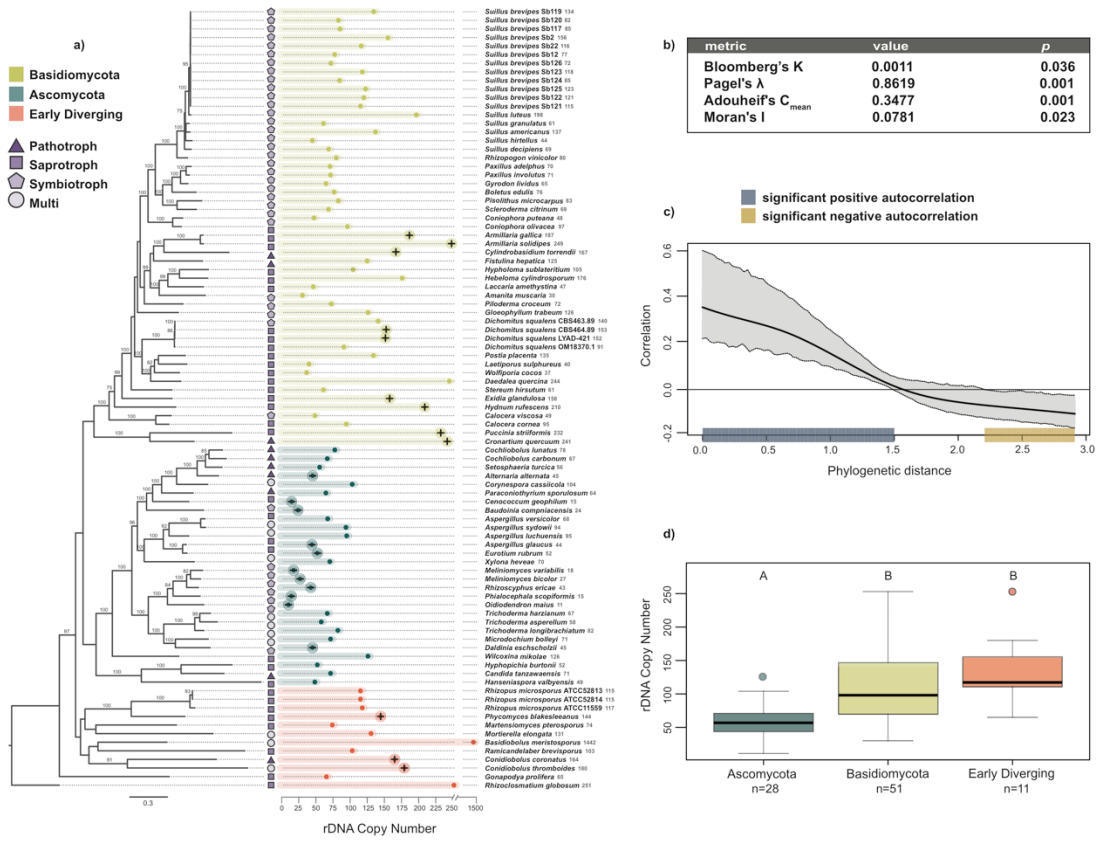

Fig. 3.3: rDNA copy number variation across multiple phylogenetic scales

a) a maximum likelihood phylogenetic reconstruction of the 91 fungal taxa included in this study based on concatenation of three single-copy genes (TOP2, GH63, MCM7). Branch values represent % bootstrap support from 1000 iterations. Grey numbers next to taxa names indicate rDNA copy number. Ending targets on the copy number scale indicate values that are significantly positive (+) or significantly negative (-) according to local Moran's I, and highlight local hotspots of autocorrelation. b) Significance tests of phylogenetic signal in rDNA copy number using both evolutionary (Bloomberg and Pagel) and autocorrelation (Adouheif and Moran) metrics. c) Phylogenetic correlogram of autocorrelation based on Moran's I. The x-axis represents the patristic distance (unitless) of all pairwise comparisons for all taxa under investigation. Shaded area indicates the 95% confidence interval of autocorrelation values. Significance based on comparison to the null hypothesis of zero phylogenetic autocorrelation (horizontal black line at 0). d) Distribution of rDNA copy number by fungal phylum. Different letters above groups indicate significant differences. Variance assumptions evaluated by a Cochran's C test, and significance values assessed by ANOVA and Tukey HSD. See Figure S1 for a validation that the observed differences in average copy number at the phylum level are not caused by overrepresentation of specific taxa.

With respect to ecological lifestyle, there were no significant differences in rDNA CNV across the three trophic modes or for taxa capable of belonging to multiple trophic modes (Fig 3.4a, Table 3.S3). When comparing amongst specific guilds, rDNA CNV was also not significantly different between pathogens and soil/litter/organic matter saprotrophs in the Ascomycota (Fig 3.4b) or between wood saprotrophs and ectomycorrhizal fungi in the Basidiomycota (Fig 3.4c). All tests examining the relationship between rDNA CNV and genome size failed to produce evidence that these metrics were correlated, regardless of the statistic used or the contribution of rDNA cassette length to total genome size (Fig. 3.5).

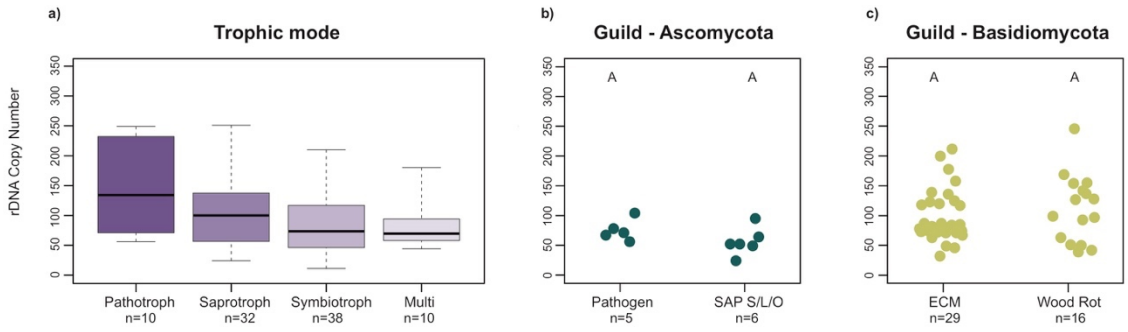

Fig. 3.4: rDNA copy number variation by fungal ecological lifestyle

a) Boxplot summaries of rDNA copy number variation by fungal trophic mode across the 91 taxa included. b) rDNA copy number variation of Ascomycota taxa assigned to a specific fungal guild (sensu Nguyen et al, 2016); SAP S/L/O = soil, litter and organic matter saprophyte. c) rDNA copy number variation of Basidiomycota taxa assigned to a specific fungal guild; ECM = ectomycorrhizal. For both b and c, only guilds with $n \geq 5$ replicate taxa were assessed. Variance assumptions evaluated by a Cochran's C test and significance values assessed by Kruskal-Wallis (a) or ANOVA (b and c) tests; no significant differences were found.

Discussion

Our results indicate that rDNA CNV and phylogenetic relatedness are linked in fungi, but that this relationship is scale-dependent. At close phylogenetic scales (i.e. within species and genera), there was an overall trend of greater similarity in rDNA copy number, while at more distant scales (i.e. phyla) rDNA copy number became more divergent. Despite this general trend, we observed multiple examples that warrant caution when considering how rRNA copy number varies among fungi. For example, among the 12 different isolates of *Suillus brevipes*, estimates ranged from 72 to 156 copies, while across the genus (five additional species of *Suillus*) the estimated range was only slightly greater (44 to 198 copies). Interestingly, even at the very closest phylogenetic scale, CBS464.89 and CBS463.89 of *Dichomitus squalens*, which represent independent monokaryons from the same dikaryotic individual, had an estimated copy number difference of 13 (140 vs 153). Although our analyses do confirm that total rDNA copy number is generally an order of magnitude greater for fungi than for bacteria or archaea, all three of these microbial groups display similar levels of variance in rDNA copy number (Větrovský & Baldrian 2013, (Stoddard *et al.*, 2015). One notable exception to this trend was *B. meristosporus*. Because this species represents only a single isolate and a single sequencing library, this estimate should also be interpreted cautiously. However, *Basidiobolus spp.* have several properties that are unique, including a non-canonical nucleus associated organelle, markedly large nuclei, and a genome that appears to be highly prone to duplication events (McKerracher & Heath, 1985; Henk & Fisher, 2012). Although there is evidence for the regulation of rDNA copy number, and some of the genetic mechanisms for rDNA copy number maintenance have been identified, (Szostak & Wu, 1980; Russell & Rodland, 1986; Kobayashi *et al.*, 1998), high rDNA variants

have been reported in plants, animals, and yeast (Rogers & Bendich, 1987; Liti *et al.*, 2009; Long *et al.*, 2013) and it is currently unknown whether high rDNA CN strains represent a conserved or a temporary state (Pukkila & Skrzynia, 1993; Simon *et al.*, 2018). Looking forward, research focused at multiple phylogenetic scales (e.g., is the amount of rDNA CNV observed within the genus *Suillus* common or an exception? Why are taxa in the Ascomycota consistently lower in rDNA copy number than other phyla?) represent important directions of study.

a) Genome size independent of rDNA contribution

metric	value	p
Pearson's pm	1.4797	0.147
Kendall's tau	0.6759	0.499
Spearman's rho	9529.9	0.515
Pearson's pm	-0.4199	0.678
Kendall's tau	-0.7713	0.441
Spearman's rho	4401.5	0.296
Pearson's pm	1.4071	0.193
Kendall's tau	1.2494	0.212
Spearman's rho	133.80	0.233
PGLS R^2	0.0021	0.7641

b) Genome size with rDNA contribution

metric	value	p
Pearson's pm	1.7421	0.090
Kendall's tau	0.9089	0.363
Spearman's rho	9083.9	0.363
Pearson's pm	-0.3778	0.709
Kendall's tau	-0.6526	0.514
Spearman's rho	4346.5	0.334
Pearson's pm	1.5055	0.167
Kendall's tau	1.4056	0.160
Spearman's rho	122.78	0.174
PGLS R^2	0.0047	0.5492

c)

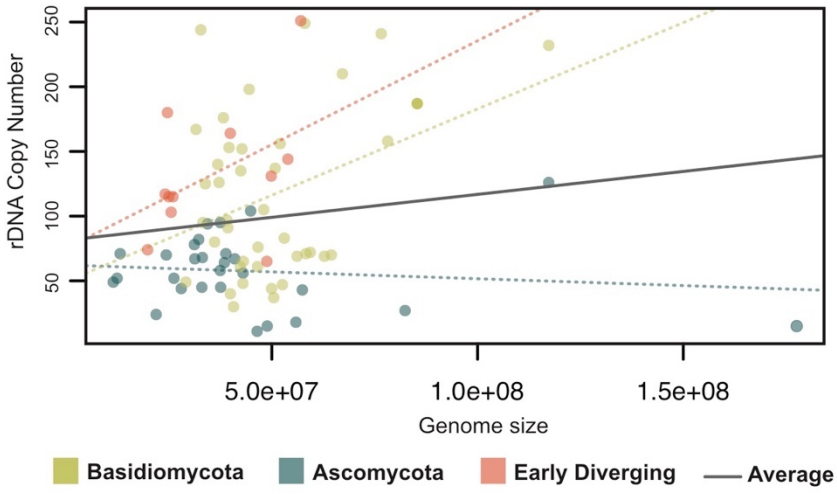

Fig. 3.5: Fungal genome size and rDNA copy number variation

a) results based on genome assembly size, without including the length added by the rDNA cassettes b) results after accounting for length added by rDNA cassettes c) Grey solid line represents the average across all taxa included, while dotted lines correspond to the rRNA CNV-genome size relationship for specific phyla. Relationships displayed are based on genome size without length contribution of rDNA cassettes included.

Although rDNA copy number is thought to have important physiological implications, such as allowing for more rapid growth (Stevenson & Schmidt, 2004) and increased DNA damage response (Ide *et al.*, 2010), our results did not find that rDNA copy number is coupled with fungal ecological lifestyle. Pathogenic fungi had a non-significant trend toward higher rDNA copy number in comparisons across trophic modes and between guilds, but there was considerable variation within this ecological lifestyle. Further, while genomic studies of fungi capable of using multiple trophic modes (e.g. saprotroph and symbiotroph) indicate that gene content and expression differs from taxa using a single trophic mode (Martino *et al.*, 2018), we did not find evidence that this increased metabolic flexibility was correlated with rDNA copy number. Finally, within the Basidiomycota, ectomycorrhizal fungal representatives had rDNA copy number estimates that were very comparable to their saprotrophic wood rot relatives (Kohler *et al.*, 2015). Given that previous studies have shown positive associations between rDNA copy number and traits relevant to lifestyle (Stevenson & Schmidt, 2004; Ide *et al.*, 2010), we suspect that the relatively coarse ecological scale of our analyses was not sufficient to capture clear links to fungal lifestyle. These results do, however, have notable ecological implications for estimates of fungal species abundances in high throughput amplicon sequencing datasets (Baldrian *et al.*, 2012). Systematic bias may be introduced as a consequence of CNV-associated differences in template DNA concentrations of barcoding regions (such as ITS) that fall within the rDNA cassette. Given our demonstration that rDNA copy number can vary widely among closely related taxa, comparisons based on ITS sequence read abundance even among members of the same species may strongly over or underestimate actual individual fungal abundance.

Efforts to account for these effects, as has been applied in other microorganisms (Kembel *et al.*, 2012; Stoddard *et al.*, 2015), remain a significant research priority.

In other eukaryotic organisms (e.g. plants and animals) rDNA CNV has been shown to have a strong positive correlation with genome size, independent of size contributions from rDNA cassettes (Prokopowich *et al.*, 2003; Wencai *et al.*, 2018). Conversely, investigation into this correlation for bacteria has shown no such relationship (Fogel *et al.*, 1999). Contrary to other eukaryotes, we found no indication that rDNA CNV is related to genome size in fungi (and regardless of whether or not rRNA cassette size was also considered). The finding that fungi do not conform to the pattern recognized between rDNA CNV and genome size may offer an interesting counterpoint for future analyses into the mechanisms structuring this relationship in plants and animals.

Using an *in silico* approach coupled with computational benchmarking, we have demonstrated that fungi exhibit substantial rDNA CNV that is inversely correlated with phylogenetic relatedness. While we did not observe strong links between rDNA CNV and ecological lifestyle, the continued use of this analysis pipeline on the rapidly increasing number of fungal genomes being generated will enable greater consideration of this trait in future studies. Similarly, using this pipeline in conjunction with studies characterizing rDNA gene expression will further enhance our understanding of fungal responses to a broad range of environmental conditions. Importantly, the range of rDNA copy numbers estimated for fungi, which have often been thought to bridge the macro- and

microbiological worlds, falls between lower rDNA copy numbers in prokaryotes and higher rDNA copy numbers in many other eukaryotes. As such, identifying the evolutionary and ecological mechanisms constraining CNV for fungi will help facilitate a broader understanding of the influence of rDNA CNV across all domains of life.

Data accessibility statement

All unpublished data used in this project was used with permission of the project PIs. Meta-data for each genome sequencing project can be found at the JGI genome pages for each sequencing project. All code and data associated with this project has been made open access and can be found on GitHub at:

https://github.com/MycoPunk/rDNA_CNV

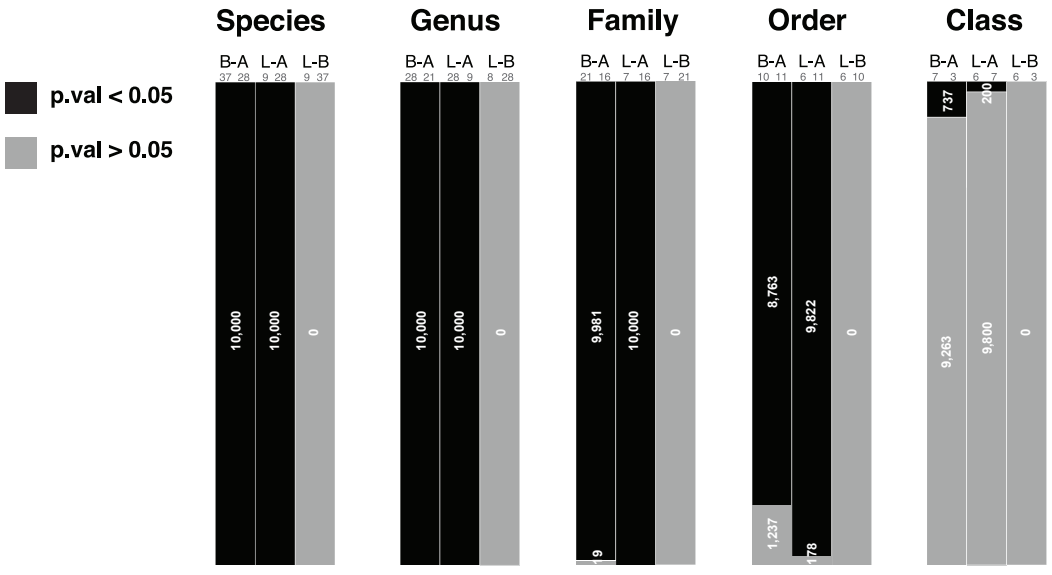

Fig. 3.S1: Statistical check for the overrepresentation of specific taxa

To assess the possibility that phylum-level differences in average copy number (CN) are caused by overrepresentation of specific taxa, we conducted a series of additional comparisons that increasingly down-weighted the representation of any specific taxa. We used a resampling approach with 10,000 iterations of each pair-wise comparison (i.e. Ascomycota v. Basidiomycota, Ascomycota v. early diverging lineages, Basidiomycota v. early diverging lineages). We compared differences in average CN when each group (i.e. Ascomycota, Basidiomycota, early diverging lineages) was represented by randomly selecting 1) a single representation of each species in each group, 2) a species from each genus within group, 3) a species from each family within each group, and 4) a species from each order within each group, and 5) finally a species from each class within each group. For the species to order level analyses, the difference in average CN for Ascos and Basids as well as early diverging lineages is not affected by overrepresentation of particular taxa. At the class level, the pattern becomes more random, which is expected given the much smaller sample size (e.g. N = 3).

Bibliography

Agerer R. **2001**. Exploration types of ectomycorrhizae A proposal to classify ectomycorrhizal mycelial systems according to their patterns of differentiation and putative ecological importance. *Mycorrhiza* **11**: 107–114.

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. **2019**. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*.

Appel HM, Martin MM. **1992**. Significance of metabolic load in the evolution of host specificity of Manduca sexta. *Ecology* **73**: 216–228.

Baldrian P. **2009**. Ectomycorrhizal fungi and their enzymes in soils: is there enough evidence for their role as facultative soil saprotrophs? *Oecologia* **161**: 657–660.

Baldrian P, Větrovský T, Cajthaml T, Dobiášová P, Petránková M, Šnajdr J, Eichlerová I. **2012**. Estimation of fungal biomass in forest litter and soil. *Fungal Ecology* **6**: 1–11.

Baroncelli R, Amby DB, Zapparata A, Sarrocco S, Vannacci G, Le Floch G, Harrison RJ, Holub E, Sukno SA, Sreenivasaprasad S, *et al.* 2016. Gene family expansions and contractions are associated with host range in plant pathogens of the genus Colletotrichum. *BMC Genomics* **17**.

Bhosle VK, Rivera JC, Chemtob S. **2017**. New insights into mechanisms of nuclear translocation of G-protein coupled receptors. *Small GTPases*: 1–10.

Bidartondo MI, Ek H, Wallander H, Soderstrom B. **2001**. Do nutrient additions alter

carbon sink strength of ectomycorrhizal fungi? *New Phytologist* **151**: 543–550.

Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, De Los Santos ELC, Kim HU, Nave M, *et al.* 2017. AntiSMASH 4.0 - improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Research* **45**: W36–W41.

Bogar LM, Dickie IA, Kennedy PG. **2015**. Testing the co-invasion hypothesis: ectomycorrhizal fungal communities on Alnus glutinosa and Salix fragilis in New Zealand (DM Richardson, Ed.). *Diversity and Distributions* **21**: 268–278.

Bogar LM, Kennedy PG. **2013**. New wrinkles in an old paradigm: neighborhood effects can modify the structure and specificity of Alnus-associated ectomycorrhizal fungal communities. *FEMS Microbiology Ecology* **83**: 767–77.

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

Bolker B, Claude J, Sien H, Des- R, Durand B, Dutheil J, Gas- O, Heibl C, Lawson D, Lefort V, *et al.* 2014. Package 'ape'. *Analyses of Phylogenetics and Evolution*: 1,6.

Borowicz VA, Juliano SA. **1991**. Specificity in host-fungus associations: Do mutualists differ from antagonists? *Evolutionary Ecology* **5**: 385–392.

Brakhage AA. **2013**. Regulation of fungal secondary metabolism. *Nature Reviews Microbiology* **11**: 21–32.

Brown NA, Schrevens S, Van Dijck P, Goldman GH. **2018**. Fungal G-protein-coupled receptors: Mediators of pathogenesis and targets for disease control. *Nature Microbiology*

3: 402–414.

Bruns TD, Bidartondo MI, Taylor DL. **2002**. Host specificity in ectomycorrhizal communities: what do the exceptions tell us? *Integrative and Comparative Biology* **42**: 352–359.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. **2009**. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

Chen W, Koide RT, Adams TS, DeForest JL, Cheng L, Eissenstat DM. **2016**. Root morphology and mycorrhizal symbioses together shape nutrient foraging strategies of temperate trees. *Proceedings of the National Academy of Sciences of the United States of America* **113**: 8741–6.

Chiang DY, Getz G, Jaffe DB, O'Kelly MJT, Zhao X, Carter SL, Russ C, Nusbaum C, Meyerson M, Lander ES. **2009**. High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nature Methods* **6**: 99–103.

Churchland C, Grayston SJ. **2014**. Specificity of plant-microbe interactions in the tree mycorrhizosphere biome and consequences for soil C cycling. *Frontiers in Microbiology* **5**.

Collemare J, Lebrun M-H. **2011**. Fungal Secondary Metabolites: Ancient Toxins and Novel Effectors in Plant-Microbe Interactions. Effectors in Plant-Microbe Interactions.377–400.

Collemare J, O'Connell R, Lebrun M. **2019**. Nonproteinaceous effectors: the terra

incognita of plant–fungal interactions. *New Phytologist*.

Contreras-Cornejo HA, Macías-Rodríguez L, del-Val E, Larsen J. **2016**. Ecological functions of Trichoderma spp. and their secondary metabolites in the rhizosphere: interactions with plants. *FEMS microbiology ecology* **92**: fiw036.

Covey K, Carroll CJW, Duguid MC, Dorji K, Dorji T, Tashi S, Wangdi T, Ashton M. **2015**. Developmental dynamics following selective logging of an evergreen oak forest in the Eastern Himalaya, Bhutan: Structure, composition, and spatial pattern. *Forest Ecology and Management* **336**: 163–173.

Crow TR. **1988**. Reproductive Mode and Mechanisms for Self-Replacement of Northern Red Oak (Quercus rubra) - A Review. *Forest Science* **34**: 19–40.

Cuddington K. **2013**. Legacy Effects: The Persistent Impact of Ecological Interactions. *Biological Theory* **6**: 203–210.

Dahlberg A. **1997**. Population ecology of Suillus variegatus in old Swedish Scots pine forests. *Mycological Research* **101**: 47–54.

Dahlberg A, Finlay RD. **1999**. Suillus. Ectomycorrhizal fungi key genera in profile. Berlin, Heidelberg: Springer Berlin Heidelberg, 33–64.

Denison RF, Bledsoe C, Kahn M, O'Gara F, Simms EL, Thomashow LS. **2003**. Cooperation in the rhizosphere and the 'free rider' problem. *Ecology* **84**: 838–845.

Desdevises Y, Morand S, Legendre P. **2002**. Evolution and determinants of host specificity in the genus Lamellodiscus (Monogenea). *Biological Journal of the Linnean Society* **77**: 431–443.

Dickie IA, Bolstridge N, Cooper JA, Peltzer DA. **2010**. Co-invasion by Pinus and its mycorrhizal fungi. *New Phytologist* **187**: 475–484.

Ditengou FA, Müller A, Rosenkranz M, Felten J, Lasok H, van Doorn MM, Legué V, Palme K, Schnitzler J-P, Polle A. **2015**. Volatile signalling by sesquiterpenes from ectomycorrhizal fungi reprogrammes root architecture. *Nature communications* **6**: 6279.

Dixon RK, Garrett HE, Cox GS, Marx DH, Sander IL. **1984**. Inoculation of Three Quercus Species with Eleven Isolates of Ectomycorrhizal Fungi. I. Inoculation Success and Seedling Growth Relationships. *Forest Science* **30**: 364–372.

Dixon RK, Johnson PS. **1992**. Synthesis of ectomycorrhizae on northern red oak seedlings in a Michigan nursery. *Journal of Arboriculture* **17**: 166–272.

Dong S, Stam R, Cano LM, Song J, Sklenar J, Yoshida K, Bozkurt TO, Oliva R, Liu Z, Tian M, *et al.* 2014. Effector specialization in a lineage of the Irish potato famine pathogen. *Science* **343**: 552–555.

Duddridge A. **1986**. The Development and Ultrastructure of Ectomycorrhizas. III. Compatible and Incompatible Interactions between Suillus grevillei (Klotzsch) Sing. And 11 Species of Ectomycorrhizal Hosts In vitro in the Absence of Exogenous Carbohydrate. *The New Phytologist* **103**: 457–464.

Dunkle LD, Cantone FA, Ciuffetti LM. **1991**. Accumulation of host-specific toxin produced by Cochliobolus carbonum during pathogenesis of maize. *Physiological and Molecular Plant Pathology* **38**: 265–273.

Egger KN, Hibbett DS. **2004**. The evolutionary implications of exploitation in

mycorrhizas. *Canadian Journal of Botany* **82**: 1110–1121.

Elser JJ, Sterner RW, Gorokhova E, Fagan WF, Markow TA, Cotner JB, Harrison JF, Hobbie SE, Odell GM, Weider LW. **2000**. Biological stoichiometry from genes to ecosystems. *Ecology Letters* **3**: 540–550.

Emms DM, Kelly S. **2015**. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**.

Emms DM, Kelly S. **2018a**. OrthoFinder2: fast and accurate phylogenomic orthology analysis from gene sequences. *bioRxiv*: 466201.

Emms E., Kelly S. **2018b**. STAG: Species Tree Inference from All Genes. *bioRxiv preprint*.

Finlay RD. **1989**. Functional aspects of phosphorus uptake and carbon translocation in incompatible ectomycorrhizal associations between Pinus sylvestris and Suillus grevillei and Boletinus cauipes. *New Phytologist* **112**: 185–192.

Finlay RD, Ek H, Odham G, Soderstrom B. **1988**. Mycelial uptake, translocation and assimilation of nitrogen from 15N-labelled ammonium by Pinus sylvestris plants infected with four different ectomycorrhizal fungi. *New Phytologist* **110**: 59–66.

Fogel GB, Collins CR, Li J, Brunk CF. **1999**. Prokaryotic genome size and SSU rDNA copy number: Estimation of microbial relative abundance from a mixed population. *Microbial Ecology* **38**: 93–113.

Foster I. **2006**. Globus toolkit version 4: Software for service-oriented systems. *Journal of Computer Science and Technology* **21**: 513–520.

Fraterrigo JM, Balser TC, Turner MG. **2006**. Microbial community variation and its relationship with nitrogen mineralization in historically altered forests. *Ecology* **87**: 570–579.

Fries N. **1978**. Basidiospore germination in some mycorrhiza-forming hymenomycetes. *Transactions of the British Mycological Society* **70**: 319–324.

Fries N. **1984**. Spore germination in the higher Basidiomycetes. *Proceedings: Plant Sciences* **93**: 205–222.

Gao Q, Jin K, Ying SH, Zhang Y, Xiao G, Shang Y, Duan Z, Hu X, Xie XQ, Zhou G, *et al.* 2011. Genome sequencing and comparative transcriptomics of the model entomopathogenic fungi Metarhizium anisopliae and M. acridum. *PLoS Genetics* **7**.

Garcia K, Delaux P-M, Cope KR, Ané J-M. **2015**. Molecular signals required for the establishment and maintenance of ectomycorrhizal symbioses. *The New phytologist* **208**: 79–87.

Gardes M, Bruns TD. **1993**. ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts. *Molecular Ecology* **2**: 113–118.

Gauthier GM, Keller NP. **2013**. Crossover fungal pathogens: The biology and pathogenesis of fungi capable of crossing kingdoms to infect plants and humans. *Fungal Genetics and Biology* **61**: 146–157.

Gilbert GS, Webb CO. **2007**. Phylogenetic signal in plant pathogen-host range. *Proceedings of the National Academy of Sciences* **104**: 4979–4983.

Gong J, Dong J, Liu X, Massana R. **2013**. Extremely High Copy Numbers and

Polymorphisms of the rDNA Operon Estimated from Single Cell Analysis of Oligotrich and Peritrich Ciliates. *Protist* **164**: 369–379.

Goodale CL, Aber JD. **2001**. The long-term effects of land-use history on nitrogen cycling in northern hardwood forests. *Ecological Applications* **11**: 253–267.

Grigoriev I V., Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, *et al.* 2014a. MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Research* **42**.

Grigoriev I V, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, *et al.* 2014b. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic acids research* **42**: D699-704.

Hardy NB, Otto SP, Gilbert L, Ehrlich P, Raven P, Yoder J, Rueffler C, Dooren T Van, Leimar O, Abrams P, *et al.* 2014. Specialization and generalization in the diversification of phytophagous insects: tests of the musical chairs and oscillation hypotheses. *Proceedings. Biological sciences / The Royal Society* **281**: 117.

Harley JL, Smith SE. **1983**. *Mycorrhizal symbiosis.*

Harris LJ, Balcerzak M, Johnston A, Schneiderman D, Ouellet T. **2016**. Host-preferential Fusarium graminearum gene expression during infection of wheat, barley, and maize. *Fungal Biology* **120**: 111–123.

Hawksworth DL. **2001**. The magnitude of fungal diversity: the 1.5 million. *Mycol Research* **105**: 1422–1432.

Hayward JA, Horton TR. **2012**. Edaphic factors do not govern the ectomycorrhizal

specificity of Pisonia grandis (Nyctaginaceae). *Mycorrhiza* **22**: 647–652.

Hayward J, Horton TR, Pauchard A, Nuñez MA. **2015**. A single ectomycorrhizal fungal species can enable a *Pinus* invasion. *Ecology* **96**: 1438–1444.

He X-H, Critchley C, Bledsoe C. **2010**. Nitrogen Transfer Within and Between Plants Through Common Mycorrhizal Networks (CMNs). *Critical Reviews in Plant Sciences* **22**: 531–567.

Van Der Heijden MGA, Bardgett RD, Van Straalen NM. **2008**. The unseen majority: Soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters* **11**: 296–310.

Henk DA, Fisher MC. **2012**. The gut fungus basidiobolus ranarum has a large genome and different copy numbers of putatively functionally redundant elongation factor genes. *PLoS ONE* **7**.

Hollingsworth TN, Johnstone JF, Bernhardt EL, Chapin FS, Noble I, Slatyer R, Diaz S, Cabido M, Casanoves F, Ozinga W, *et al.* 2013. Fire Severity Filters Regeneration Traits to Shape Community Assembly in Alaska's Boreal Forest (KO Reinhart, Ed.). *PLoS ONE* **8**: e56033.

Homola R, Mistretta P. **1977**. *Ectomycorrhizae of Maine 1: A Listing of Boletaceae with the Associated Hosts*. Life Sciences and Agricultural Experiment Station, University of Maine at Orono.

Hortal S, Plett KL, Plett JM, Cresswell T, Johansen M, Pendall E, Anderson IC. **2017**. Role of plant-fungal nutrient trading and host control in determining the

competitive success of ectomycorrhizal fungi. *ISME J*.

Horton TR, Bruns TD. **1998**. Multiple-host fungi are the most frequent and abundant ectomycorrhizal types in a mixed stand of Douglas fir (Pseudotsuga menziesii) and bishop pine (Pinus muricata). *New Phytologist* **139**: 331–339.

Horton TR, Molina R, Hood K. **2005**. Douglas-fir ectomycorrhizae in 40- and 400-year-old stands: Mycobiont availability to late successional western hemlock. *Mycorrhiza* **15**: 393–403.

Huber DPW, Bohlmann J. **2006**. The role of terpene synthases in the direct and indirect defense of conifers against insect herbivory and fungal pathogens. Multigenic and Induced Systemic Resistance in Plants.296–313.

Hubert NA, Gehring CA. **2008**. Neighboring trees affect ectomycorrhizal fungal community composition in a woodland-forest ecotone. *Mycorrhiza* **18**: 363–374.

Huey RB. **1984**. Is a jack-of-all temperatures a master of none? *Evolution* **38**: 441–444.

Ide S, Miyazaki T, Maki H, Kobayashi T. **2010**. Abundance of ribosomal RNA gene copies maintains genome integrity. *Science* **327**: 693–696.

Irieda H, Inoue Y, Mori M, Yamada K, Oshikawa Y, Saitoh H, Uemura A, Terauchi R, Kitakura S, Kosaka A, *et al.* 2018. Conserved fungal effector suppresses PAMP-triggered immunity by targeting plant immune kinases. *Proceedings of the National Academy of Sciences* **116**: 496–505.

Ishida TA, Nara K, Tanaka M, Kinoshita A, Hogetsu T. **2008**. Germination and infectivity of ectomycorrhizal fungal spores in relation to their ecological traits during

primary succession. *New Phytologist* **180**: 491–500.

Jacobson KM, Miller Jr. OK. **2007**. Physiological variation between tree-associated populations of Suillus granulatus as determined by in vitro mycorrhizal synthesis experiments . *Canadian Journal of Botany* **70**: 26–31.

Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. **2019**. New approach for understanding genome variations in KEGG. *Nucleic Acids Research* **47**: D590–D595.

Keller NP, Hohn TM. **1997**. Metabolic pathway gene clusters in filamentous fungi. *Fungal Genetics and Biology* **21**: 17–29.

Keller NP, Turner G, Bennett JW. **2005**. Fungal secondary metabolism - from biochemistry to genomics. *Nature reviews. Microbiology* **3**: 937–47.

Kembel SW, Wu M, Eisen JA, Green JL. **2012**. Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance. *PLoS Computational Biology* **8**.

Kennedy PG, Higgins LM, Rogers RH, Weber MG. **2011**. Colonization-Competition Tradeoffs as a Mechanism Driving Successional Dynamics in Ectomycorrhizal Fungal Communities (M Moora, Ed.). *PLoS ONE* **6**: e25126.

Kennedy PG, Smith DP, Horton TR, Molina RJ. **2012**. Arbutus menziesii (Ericaceae) facilitates regeneration dynamics in mixed evergreen forests by promoting mycorrhizal fungal diversity and host connectivity. *American Journal of Botany* **99**: 1691–1701.

Kiers ET, Duhamel M, Beesetty Y, Mensah JA, Franken O, Verbruggen E, Fellbaum CR, Kowalchuk GA, Hart MM, Bago A, *et al.* 2011. Reciprocal rewards

stabilize cooperation in the mycorrhizal symbiosis. *Science* **333**: 880–882.

Kikuchi K, Matsushita N, Suzuki K, Hogetsu T. **2007**. Flavonoids induce germination of basidiospores of the ectomycorrhizal fungus Suillus bovinus. *Mycorrhiza* **17**: 563–570.

Kim K-T, Jeon J, Choi J, Cheong K, Song H, Choi G, Kang S, Lee Y-H. **2016**. Kingdom-Wide Analysis of Fungal Small Secreted Proteins (SSPs) Reveals their Potential Role in Host Association. *Frontiers in Plant Science* **7**.

Klappenbach JA, Dunbar JM, Schmidt TM. **2000**. rRNA operon copy number reflects ecological strategies of bacteria. *Applied and Environmental Microbiology* **66**: 1328–1333.

Kobayashi T, Heck DJ, Nomura M, Horiuchi T. **1998**. Expansion and contraction of ribosomal DNA repeats in Saccharomyces cerevisiae: Requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. *Genes and Development* **12**: 3821–3830.

Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canbäck B, Choi C, Cichocki N, Clum A, *et al.* 2015. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nature Genetics* **47**: 410–415.

Kretzer A, Li Y, Szaro T, Bruns TD. **1996**. Internal transcribed spacer sequences from 38 recognized species of Suillus sensu lato: Phylogenetic and taxonomic implications. *Mycologia* **88**: 776–785.

Krishnan A, Almén MS, Fredriksson R, Schiöth HB. **2012**. The origin of GPCRs: Identification of mammalian like rhodopsin, adhesion, glutamate and frizzled GPCRs in

fungi. *PLoS ONE* **7**.

Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. **2001**. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology* **305**: 567–580.

Kropp BR, Trappe JM. **1982**. Ectomycorrhizal fungi of Tsuga heterophylla. *Mycologia* **74**: 479–488.

Kuo M, Methven A. **2010**. *100 Cool Mushrooms*. University of Michigan Press.

Langmead B, Salzberg SL. **2012**. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. **2009**. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Liao D. **2000**. Gene conversion drives within genic sequences: Concerted evolution of ribosomal RNA genes in bacteria and archaea. *Journal of Molecular Evolution* **51**: 305–317.

Liao H-L, Chen Y, Bruns TD, Peay KG, Taylor JW, Branco S, Talbot JM, Vilgalys R. **2014**. Metatranscriptomic analysis of ectomycorrhizal roots reveals genes associated with Piloderma-Pinus symbiosis: improved methodologies for assessing gene expression in situ. *Environmental microbiology* **16**: 3730–42.

Liao H-L, Chen Y, Vilgalys R, Dodds P, Rathjen J, Plett J, Kemppainen M, Kale S, Kohler A, Legué V, *et al.* 2016. Metatranscriptomic Study of Common and Host-

Specific Patterns of Gene Expression between Pines and Their Symbiotic Ectomycorrhizal Fungi in the Genus Suillus (F Martin, Ed.). *PLOS Genetics* **12**: e1006348.

Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, *et al.* 2009. Population genomics of domestic and wild yeasts. *Nature* **458**: 337–341.

Lofgren L, Nguyen NH, Kennedy PG. **2018**. Ectomycorrhizal host specificity in a changing world: can legacy effects explain anomalous current associations? *New Phytologist*.

Long Q, Rabanal FA, Meng D, Huber CD, Farlow A, Platzer A, Zhang Q, Vilhjálmsson BJ, Korte A, Nizhynska V, *et al.* 2013. Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. *Nature Genetics* **45**: 884–890.

MacArthur R, Levins R. **1964**. Competition, habitat selection, and character displacement in a patchy environment. *Proceedings of the National Academy of Sciences* **51**: 1207–1210.

Malajczuk N, Molina R, Trappe JM. **1982**. Ectomycorrhiza Formation In Eucalyptus. I. Pure Culture Synthesis, Host Specificity And Mycorrhizal Compatibility With Pinus Radiata. *New Phytologist* **91**: 467–482.

Maleszka R, Clarkwalker GD. **1993**. Yeasts Have a 4-Fold Variation in Ribosomal DNA Copy Number. *Yeast* **9**: 53–58.

Martin F, Aerts A, Ahrén D, Brun A, Danchin EGJ, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V, *et al.* 2008. The genome of Laccaria bicolor provides insights into mycorrhizal symbiosis. *Nature* **452**: 88–92.

Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O, Montanini B, Morin E, Noel B, Percudani R, *et al.* 2010. Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* **464**: 1033–1038.

Martino E, Morin E, Grelet GA, Kuo A, Kohler A, Daghino S, Barry KW, Cichocki N, Clum A, Dockter RB, *et al.* 2018. Comparative genomics and transcriptomics depict ericoid mycorrhizal fungi as versatile saprotrophs and plant mutualists. *New Phytologist* **217**: 1213–1229.

Massicotte HB, Molina R, Luoma DL, Smith JE. **1994**. Biology of the ectomycorrhizal genus, Rhizopogon. II.Patterns of host-fungus specificity following spore inoculation of diverse hosts grown in monoculture and dual culture. *New Phytologist* **126**: 677–690.

McKerracher L, Heath I. **1985**. The Structure and Cycle of the Nucleus-Associated Organelle in Two Species of Basidiobolus. *Mycologia* **77**: 412–417.

Melin E, Krupa S. **1971**. Studies on Ectomycorrhizae of Pine II. Growth Inhibition of Mycorrhizal Fungi by Volatile Organic Constituents of Pinus silvestris (Scots Pine) Roots. *Physiologia Plantarum* **25**: 337–340.

Meng B, Li C. **2010**. The capsid protein of Grapevine rupestris stem pitting-associated virus contains a typical nuclear localization signal and targets to the nucleus. *Virus Research* **153**: 212–217.

Metz MR, Frangioso KM, Wickland AC, Meentemeyer RK, Rizzo DM. **2012**. An emergent disease causes directional changes in forest species composition in coastal California. *Ecosphere* **3**: art86.

Miller OK, Miller HM. **2006**. *North American Mushrooms: A Field Guide To Edible And Inedible Fungi*. Blue Ridge Summit, PA: Flacon Guides.

Miyazaki T, Kobayashi T. **2011**. Visualization of the dynamic behavior of ribosomal RNA gene repeats in living yeast cells. *Genes to Cells* **16**: 491–502.

Molina R. **1979**. Pure culture synthesis and host specificity of red alder mycorrhizae. *Canadian Journal of Botany*.

Molina R, Massicotte H, Trappe JM. **1992a**. Specificity phenomena in mycorrhizal symbioses: community-ecological consequences and practical implications. Mycorrhizal functioning an integrative plantfungal process.357–423.

Molina R, Massicotte H, Trappe J. **1992b**. *Specificity phenomena in mycorrhizal symbioses: community-ecological consequences and practical implications*. London, UK: Chapman and Hall.

Molina R, Smith JE, Mckay D, Melville LH. **1997**. Biology of the ectomycorrhizal genus, Rhizopogon: III. Influence of co-cultured conifer species on mycorrhizal specificity with the arbutoid hosts Arctostaphylos uva-ursi and Arbutus menziesii. *New Phytologist* **137**: 519–528.

Molina R, Trappe JM. **1982**. Lack of Mycorrhizal Specificity by the Ericaceous Hosts Arbutus menziesii and Archtostaphylos uva-ursi. *New Phytologist* **90**: 495–509.

Moran NA. **2002**. The Evolution of Host-Plant Alternation in Aphids: Evidence for Specialization as a Dead End. *The American Naturalist* **132**: 681–706.

Mujic AB, Durall DM, Spatafora JW, Kennedy PG. **2015**. Competitive avoidance not edaphic specialization drives vertical niche partitioning among sister species of ectomycorrhizal fungi. *New phytologist* **209**: 1174–83.

Münkemüller T, Lavergne S, Bzeznik B, Dray S, Jombart T, Schiffers K, Thuiller W. **2012**. How to measure and test phylogenetic signal. *Methods in Ecology and Evolution* **3**: 743–756.

Murata H, Yamada A, Yokota S, Maruyama T, Shimokawa T, Neda H. **2015**. Innate traits of Pinaceae-specific ectomycorrhizal symbiont Suillus luteus that differentially associates with arbuscular mycorrhizal broad-leaved trees in vitro. *Mycoscience* **56**: 606–611.

Nara K. **2006**. Ectomycorrhizal networks and seedling establishment during early primary succession. *New Phytologist* **169**: 169–178.

Nemergut DR, Knelman JE, Ferrenberg S, Bilinski T, Melbourne B, Jiang L, Violle C, Darcy JL, Prest T, Schmidt SK, *et al.* 2016. Decreases in average bacterial community rRNA operon copy number during succession. *ISME Journal* **10**: 1147–1156.

Nguyen N, Vellinga E, Bruns T, Kennedy P. **2016a**. Phylogenetic assessment of global Suillus ITS sequences supports morphologically defined species and reveals synonymous and undescribed taxa. *Accepted to Mycologia*.

Nguyen NH, Vellinga EC, Bruns TD, Kennedy PG. **2016b**. Phylogenetic assessment

of global Suillus ITS sequences supports morphologically defined species and reveals synonymous and undescribed taxa. *Mycologia* **5514**: 16–106.

Nguyen N, Vellinga EC, Bruns TD, Kennedy P. **2016c**. Phylogenetic assessment of global Suillus ITS sequences supports morphologically defined species and reveals synonymous and undescribed taxa. *Mycologia*: 16–106.

Nosil P. **2002**. Transition Rates Between Specialization And Generalization In Phytophagous Insects. *Evolution* **56**: 1701–1706.

Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N, Pearse W. **2014**. Caper: Comparative analyses of phylogenetics and evolution in R. *R package version 0.5.2/ r121*: 1–36.

Ouvrard D, Chalise P, Percy DM. **2015**. Host-plant leaps versus host-plant shuffle: a global survey reveals contrasting patterns in an oligophagous insect group (Hemiptera, Psylloidea). *Systematics and Biodiversity* **13**: 434–454.

Palm M, Stewart E. **1984**. In vitro Synthesis of Mycorrhizae between Presumed Specific and Nonspecific Pinus + Suillus Combinations. *Mycologia* **76**: 579–600.

Paredes S, Branco AT, Hartl DL, Maggert KA, Lemos B. **2011**. Ribosomal DNA deletions modulate genome-wide gene expression: rDNA-sensitive genes and natural variation. *PLoS genetics* **7**: e1001376.

Pellegrin C, Morin E, Martin FM, Veneault-Fourrey C. **2015**. Comparative analysis of secretomes from ectomycorrhizal fungi with an emphasis on small-secreted proteins. *Frontiers in Microbiology* **6**.

Peng ZL, Yang JY, Chen X. **2010**. An improved classification of G-protein-coupled receptors using sequence-derived features. *BMC Bioinformatics* **11**.

Perisin M, Vetter M, Gilbert JA, Bergelson J. **2016**. 16Stimator: statistical estimation of ribosomal gene copy numbers from draft genome assemblies. *The ISME Journal* **10**: 1020–1024.

Perry DA, Margolis H, Choquette C, Molina R, Trappe JM. **1989**. Ectomycorrhizal mediation of competition between coniferous tree species. *New Phytologist* **112**: 501–511.

Pickles BJ, Egger KN, Massicotte HB, Green DS. **2012**. Ectomycorrhizas and climate change. *Fungal Ecology* **5**: 73–84.

Plett JM, Daguerre Y, Wittulsky S, Vayssieres A, Deveau A, Melton SJ, Kohler A, Morrell-Falvey JL, Brun A, Veneault-Fourrey C, *et al.* 2014. Effector MiSSP7 of the mutualistic fungus Laccaria bicolor stabilizes the Populus JAZ6 protein and represses jasmonic acid (JA) responsive genes. *Proceedings of the National Academy of Sciences* **111**: 8299–8304.

Plett JM, Kemppainen M, Kale SD, Kohler A, Legué V, Brun A, Tyler BM, Pardo AG, Martin F. **2011**. A secreted effector protein of Laccaria bicolor is required for symbiosis development. *Current biology : CB* **21**: 1197–203.

Plett JM, Kohler A, Martin F. **2012**. De-constructing a mutualist: How the molecular blueprints of model symbiotic fungi are changing our understanding of mutualism. Fungal Associations, 2nd Edition.93–117.

Policelli N, Bruns TD, Vilgalys R, Nuñez MA. **2019**. Suilloid fungi as global drivers of pine invasions. *New Phytologist*.

Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S, Kahmann R. **2015**. Fungal Effectors and Plant Susceptibility. *Annual Review of Plant Biology* **66**: 513–545.

Pritchard L, Birch P. **2011**. A systems biology perspective on plant-microbe interactions: Biochemical and structural targets of pathogen effectors. *Plant Science* **180**: 584–603.

Prokopowich CD, Gregory TR, Crease TJ. **2003**. The correlation between rDNA copy number and genome size in eukaryotes. *Genome* **46**: 48–50.

Pukkila PJ, Skrzynia C. **1993**. Frequent changes in the number of reiterated ribosomal RNA genes throughout the life cycle of the basidiomycete Coprinus cinereus. *Genetics* **133**: 203–211.

Quandt CA, Di Y, Jaiswal P, Spatafora JW, Elser J. **2016**. Differential Expression of Genes Involved in Host Recognition, Attachment, and Degradation in the Mycoparasite Tolypocladium ophioglossoides . *G3: Genes|Genomes|Genetics* **6**: 731–741.

Quin MB, Flynn CM, Schmidt-Dannert C. **2014**. Traversing the fungal terpenome. *Natural Product Reports* **31**: 1449–1473.

Quin MB, Flynn CM, Wawrzyn GT, Choudhary S, Schmidt-Dannert C. **2013**. Mushroom hunting by using bioinformatics: application of a predictive framework facilitates the selective identification of sesquiterpene synthases in basidiomycota.

Chembiochem : a European journal of chemical biology **14**: 2480–91.

R Core team. **2017**. R Core Team. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing , Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.* **55**: 275–286.

Rehner S, Samuels G. **1995**. Molecular systematics of the Hypocreales: a teleomorph gene phylogeny and the status of their anamorphs. *Canadian Journal of Botany* **73**: 816–832.

Revell LJ. **2012**. phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**: 217–223.

Rogers SO, Bendich a J. **1987**. Heritability and Variability in Ribosomal RNA Genes of Vicia faba. *Genetics* **117**: 285–95.

Russell PJ, Rodland KD. **1986**. Magnification of rRNA gene number in a Neurospora crassa strain with a partial deletion of the nucleolus organizer. *Chromosoma* **93**: 337–340.

Salim D, Bradford WD, Freeland A, Cady G, Wang J, Pruitt SC, Gerton JL. **2017**. DNA replication stress restricts ribosomal DNA copy number. *PLoS Genetics* **13**.

Schmidt R, De Jager V, Zühlke D, Wolff C, Bernhardt J, Cankar K, Beekwilder J, Van Ijcken W, Sleutels F, De Boer W, *et al.* 2017. Fungal volatile compounds induce production of the secondary metabolite Sodorifen in Serratia plymuthica PRI-2C. *Scientific Reports* **7**.

Schulze-Lefert P, Panstruga R. **2011**. A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. *Trends in Plant*

Science **16**: 117–125.

Sharma E, Anand G, Kapoor R. **2017**. Terpenoids in plant and arbuscular mycorrhiza-reinforced defence against herbivorous insects. *Annals of Botany* **119**: 791–801.

Sharma R, Mishra B, Runge F, Thines M. **2014**. Gene Loss rather than gene gain is associated with a host jump from monocots to dicots in the smut fungus melanopsichium pennsylvanicum. *Genome Biology and Evolution* **6**: 2034–2049.

Shaw JJ, Berbasova T, Sasaki T, Jefferson-George K, Spakowicz DJ, Dunican BF, Portero CE, Narváez-Trujillo A, Strobel SA. **2015**. Identification of a fungal 1,8-cineole synthase from Hypoxylon sp. with specificity determinants in common with the plant synthases. *Journal of Biological Chemistry* **290**: 8511–8526.

Simon L, Rabanal FA, Dubos T, Oliver C, Lauber D, Poulet A, Vogt A, Mandlbauer A, Le Goff S, Sommer A, *et al.* 2018. Genetic and epigenetic variation in 5S ribosomal RNA genes reveals genome dynamics in Arabidopsis thaliana. *Nucleic Acids Research* **46**: 3019–3033.

Simpson GG. **1953**. *The Major Features of Evolution*. New York: Columbia Univ. Press.

Skibbe SD, Doehlemann G, Fernandes J, Walbot V. **2010**. Maize Tumors Caused by Ustilago Genes in Host and Pathogen. *Science* **67**: 89–93.

Smith, A.H., Thiers HD. **1964**. *A Contribution Toward a Monograph of North American Species of Suillus*. Ann Arbor, Michigan: Lubrecht & Cramer Ltd.

Smith SE, Read DJ. **2008**. *Mycorrhizal Symbiosis*.

Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, Van Themaat

EVL, Brown JKM, Butcher SA, Gurr SJ, *et al.* 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **330**: 1543–1546.

Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. **2018**. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Molecular Plant Pathology* **19**: 2094–2110.

Stamatakis A. **2006**. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.

Stevenson BS, Schmidt TM. **2004**. Life History Implications of rRNA Gene Copy Number in Escherichia coli. *Society* **70**: 6670–6677.

Stireman JO. **2005**. The evolution of generalization? Parasitoid flies and the perils of inferring host range evolution from phylogenies. *Journal of Evolutionary Biology* **18**: 325–336.

Stoddard SF, Smith BJ, Hein R, Roller BRK, Schmidt TM. **2015**. rrnDB: Improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Research* **43**: D593–D598.

Szostak JW, Wu R. **1980**. Unequal crossing over in the ribosomal DNA of Saccharomyces cerevisiae. *Nature* **284**: 426–430.

Taylor AFS, Alexander I. **2005**. The ectomycorrhizal symbiosis: Life in the real world. *Mycologist* **19**: 102–112.

Tedersoo L, Põlme S. **2012**. Infrageneric variation in partner specificity: multiple

ectomycorrhizal symbionts associate with Gnetum gnemon (Gnetophyta) in Papua New Guinea. *Mycorrhiza* **22**: 663–668.

Tedersoo L, Smith ME. **2013**. Lineages of ectomycorrhizal fungi revisited: Foraging strategies and novel lineages revealed by sequences from belowground. *Fungal Biology Reviews* **27**: 83–99.

Theodorou C, Reddell P. **1991**. In vitro synthesis of ectomycorrhizas on Casuarinaceae with a range of mycorrhizal fungi. *New Phytologist* **118**: 279–288.

Towns J, Cockerill T, Dahan M, Foster I, Gaither K, Grimshaw A, Hazlewood V, Lathrop S, Lifka D, Peterson GD, *et al.* 2014. XSEDE: Accelerating scientific discovery. *Computing in Science and Engineering* **16**: 62–74.

Tripp EA, Manos PS. **2008**. Is Floral Specialization An Evolutionary Dead-End Pollination System Transitions In Ruellia (Acanthaceae). *Evolution* **62**: 1712–1737.

Tsuge T, Harimoto Y, Hanada K, Akagi Y, Kodama M, Akimitsu K, Yamamoto M. **2016**. Evolution of pathogenicity controlled by small, dispensable chromosomes in Alternaria alternata pathogens. *Physiological and Molecular Plant Pathology* **95**: 27–31.

vanDam P, Fokkens L, Schmidt SM, Linmans JHJ, Corby Kistler H, Ma LJ, Rep M. **2016**. Effector profiles distinguish formae speciales of fusarium oxysporum. *Environmental Microbiology* **18**: 4087–4102.

Veneault-Fourrey C, Commun C, Kohler A, Morin E, Balestrini R, Plett J, Danchin E, Coutinho P, Wiebenga A, de Vries RP, *et al.* 2014. Genomic and transcriptomic analysis of Laccaria bicolor CAZome reveals insights into polysaccharides remodelling

during symbiosis establishment. *Fungal genetics and biology : FG & B* **72**: 168–81.

Veneault-Fourrey C, Martin F. **2011**. Mutualistic interactions on a knife-edge between saprotrophy and pathogenesis. *Current Opinion in Plant Biology* **14**: 444–450.

Visser B, Le Lann C, den Blanken FJ, Harvey JA, van Alphen JJM, Ellers J. **2010**. Loss of lipid synthesis as an evolutionary consequence of a parasitic lifestyle. *Proceedings of the National Academy of Sciences* **107**: 8677–8682.

Voiblet C, Duplessis S, Encelot N, Martin F. **2001**. Identification of symbiosis-regulated genes in Eucalyptus globulus - Pisolithus tinctorius ectomycorrhiza by differential hybridization of arrayed cDNAs. *Plant Journal* **25**: 181–191.

Walker JKM, Cohen H, Higgins LM, Kennedy PG. **2014**. Testing the link between community structure and function for ectomycorrhizal fungi involved in a global tripartite symbiosis. *New Phytologist* **202**: 287–296.

Walton JD. **2006**. HC-toxin. *Phytochemistry* **67**: 1406–1413.

Walton JD, Panaccione DG. **1993**. Host-Selective Toxins and Disease Specificity: Perspectives and Progress. *Annual Review of Phytopathology* **31**: 275–303.

Watkinson SC, Boddy L, Money NP. **2015**. *The Fungi: Third Edition*.

de Weert S, Kuiper I, Lagendijk EL, Lamers GEM, Lugtenberg BJJ. **2007**. Role of Chemotaxis Toward Fusaric Acid in Colonization of Hyphae of Fusarium oxysporum f. sp. radicis-lycopersici by Pseudomonas fluorescens WCS365 . *Molecular Plant-Microbe Interactions* **17**: 1185–1191.

Wencai W, Tao W, Aleš K, Becher H, Kuderova A, Leitch I, Garcia S, Leitch A.

2018. Remarkable variation of ribosomal DNA organization and copy number in gnetophytes, a distinct lineage of gymnosperms. *Annals of Botany* **myc172**.

White TJ, Bruns S, Lee S, Taylor J. **1990**. Amplification and direct sequencing of fungal ribosomal RNA genes for phyologenetics.pdf. *PCR Protocols: A Guide to Methods and Applications*: 315–322.

Whittaker RH, Feeny PP. **1971**. Allelochemics: Chemical interactions between species. *Science* **171**: 757–770.

Wistrand M, Käll L, Sonnhammer ELL. **2006**. A general model of G protein-coupled receptor sequences and its application to detect remote homologs. *Protein Science* **15**: 509–521.

Xu YJ, Luo F, Li B, Shang Y, Wang C. **2016**. Metabolic conservation and diversification of metarhizium species correlate with fungal host-specificity. *Frontiers in Microbiology* **7**.

Yoon S, Xuan Z, Makarov V, Ye K, Sebat J. **2009**. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome research* **19**: 1586–92.

Yoshida S, Kameoka H, Tempo M, Akiyama K, Umehara M, Yamaguchi S, Hayashi H, Kyozuka J, Shirasu K. **2012**. The D3 F-box protein is a key component in host strigolactone responses essential for arbuscular mycorrhizal symbiosis. *New Phytologist* **196**: 1208–1216.

Zamioudis C, Pieterse CMJ. **2011**. Modulation of Host Immunity by Beneficial Microbes. *Molecular Plant-Microbe Interactions* **25**: 139–150.

Zolan ME. **1995**. Chromosome-length polymorphism in fungi. *Microbiological Reviews* **59**: 686–698.