Molecular systematics and population genomics of the tree-pathogenic fungus *Grosmannia clavigera*

by

Sepideh Massoumi Alamouti

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

Doctor of Philosophy

in

The Faculty of Graduate and Postdoctoral Studies

(Forestry)

The University of British Columbia
(Vancouver)

November 2013

Abstract

Ophiostomatoid fungi increasingly damage forests, but understanding their interactions with vectors and hosts is hampered by uncertainty over the validity of genera, relationships between genera, and species boundaries. To address some of these issues, I first generated a new multigene dataset from 67 taxa that represent the genus *Grosmannia* and other related genera. The multigene phylogeny resolved the *Grosmannia* fungi into a clade that was separated from previously intermixed species of the genera *Ambrosiella* and *Raffaelea*, and that corresponded to distinct ecological niches and vector associates, i.e. bark versus wood-boring beetles.

Second, I generated and used 15 gene genealogies to define species boundaries in *G. clavigera*. This destructive pine pathogen is vectored by two beetle species: mountain and Jeffrey pine beetles (MPB, JPB). MPB and its fungal associates have expanded into the largest epidemic in western North American history. I identified two phylogenetic species: Gs and Gc. Gc is present in the phylogenetically close *Pinus* species *ponderosa* and *jeffreyi*, which are infested by localized populations of their respective beetles. In contrast, Gs is an exclusive associate of MPB and its primary host *P. contorta*, although it is found in other pine species in current epidemic regions. These results suggest that host-tree species and beetle population dynamics are important factors in the genetic divergence and diversity of fungal associates in the beetle-tree ecosystems.

Finally, we generated new genomic sequences for eleven Gs and Gc strains to further assess evidence for divergence in these fungi as they adapt to different pine species, and to find genes that may be involved in species divergence. Aligning these genome sequences to the reference genome, we identified 103,430 SNPs that supported the Gs and Gc lineages and divided each lineage into two subclades. Genome-wide scans identified truncated genes and potential pseudogenes that differed between *Grosmannia* lineages, as well as seven genes that show evidence of positive selection. The seven genes are involved in secondary metabolism and in detoxifying host-tree defense chemicals (e.g. polyketide synthases, oxidoreductases), and their variants may reflect adaptation to the specific chemistries of *P. contorta*, *ponderosa*, and *jeffreyi*.

Preface

A version of chapter 2 has been published as “Massoumi Alamouti S, Tsui CKM, Breuil C. (2009). Multigene phylogeny of filamentous ambrosia fungi associated with ambrosia and bark beetles. Mycological Research. 113: 822–835”.

I conducted the experimental design, experiments and analyses with guidance from Colette Breuil. Clement Tsui provided some inputs for the Bayesian analysis. I wrote the manuscript with the guidance and assistance from Colette Breuil.

A version of chapter 3 has been published as “Massoumi Alamouti S, Wang V, DiGuistini S, Six DL, Bohlmann J, Hamelin RC, Feau N, Breuil C. 2011. Gene genealogies reveal cryptic species and host preferences for the pine fungal pathogen *Grosmannia clavigera*. Molecular Ecology. 20: 2581–2602”.

I conducted the experimental design, experiments and analyses with guidance from Colette Breuil. I received inputs from Nicolas Feau for the analyses and suggestions from Scott DiGuistini and Richard Hamelin for finding the polymorphisms. Vincent Wang helped me with DNA extraction and PCR for the fungal samples. Diana Six conducted the fungal samplings in California. I wrote the manuscript with the guidance and assistance of Colette Breuil and inputs from Nicolas Feau, Jöerg Bohlmann, Diana Six and Richard Hamelin.

A version of chapter 4 will be submitted to a peer-reviewed journal as “Massoumi Alamouti, et al. 2013. Comparative genomics of the pine pathogens and the beetle symbionts in the genus *Grosmannia*”.

I conducted the experimental design with guidance from Colette Breuil. Illumina sequencing was carried out in the Genome Science Center (GSC, Vancouver, BC, Canada). Sajeet Haridas generated the de novo genome assemblies. I proposed the improvements of the de novo assemblies (i.e. by mapping to the reference genome) that were generated by Sajeet Haridas. I performed the data analyses (SNP calling, orthologous gene finding, phylogenomics, gene genealogies and the evolutionary selection analyses). The selection analyses using PAML was ran by Nicolas Feau. Gordon Robertson showed me how to generate the graphics for figures 4.1b and 4.4a-b. I wrote the chapter with the guidance and assistance from Colette Breuil and Gordon Robertson and inputs from Nicolas Feau, Mary Berbee, Ye Wang and Jöerg Bohlmann.

Table of Contents

List of Tables

List of Figures

Acknowledgements

Without the support and patience of my advisors, colleagues, friends and family this thesis would not have been possible. It is to them that I owe my deepest gratitude. First I thank my supervisor Colette Breuil who patiently taught me about being critical, prepared and responsible. I appreciate her endless support and guidance throughout all these years. Her knowledge, enthusiasm and energy were motivating and contagious and made my PhD stimulating and productive. To Gordon Roberston for the excellent mentorship, I appreciate all his contributions of time and expertise that helped me to grow professionally and to move on. I'm grateful to my committee members Mary Berbee and Richard Hamelin who have been always encouraging and enthusiastic about my work. I am always thankful to Mary Berbee for being the source of inspiration to me from the very beginning and for teaching me so many things. To Jan Stenlid, Martin Adamson, Kermit Ritland, Loren Rieseberg and Les Lavkulich for accepting to be in my examining committee and for their time and insightful questions. It's a pleasure to acknowledge Joerg Bohlmann, who always kindly granted his time and provided insightful discussions and guidance to improve my study. An especial acknowledgement to Keith Seifert, Adnan Uzunovic and Diana Six, who have been always supportive and introduced me to their fascinating world of science. To Inanc Birol, Rod Docking, Shaun Jackman, Simon Chan and Greg Taylor who make the Genome Science Center a welcoming place for sharing ideas and a source of good advice and collaboration. To collaborators at the United States Barbara Bentz, Matthew Hansen and Jim Vandygriff

for their kind support with the fieldwork, and to Kathy Bleiker and Adrian Rice for sharing their fungal cultures. I also thank the great folks whom I worked with Vincent Wang, Lynette Lim, Vicente Hernandez, Huang-Ju Chen, Marie Josee, Thomas Wang, Ahmed Kenawy, Pablo Chung, Mohamed Ismail, Fariba Izadi, Kristin Tangen, Clement Tsui, Ljerka Lah, Lily Khadempour, Sajeet Haridas, Ye Wang and Scott DiGuistini. Their excellent work, encouragement and friendship made my PhD life more enjoyable and energizing. To my friends, especially Mahin Narimany and Farah Larki for helping me keep things in perspective and providing a much-needed time of occasional escapes. Special thanks to my amazing siblings, Siavash, Siamack and Sepehr for their love, support and encouragement. In particular, I would like to thank my brother Siavash. His determination and sense of joy have always kept me inspired, and I could not do this without him. My deepest respect to my partner Hedayat Alghassi, who endured this process with unselfishness and never-ending encouragement. His patience, love and support throughout all these years are something for which I will be forever grateful. Mom and Dad, Atiyeh Shaeri and Ahmad Massoumi Alamouti, I love you both and I wish you all the joy that you always ensured for me to experience. You raised me with love and supported me in all my pursuits. I cannot thank you enough.

To my loving mother and father

Atiyeh and Ahmad

Chapter 1 General introduction and research objectives

1.1 Introduction

Worldwide, beetles and their fungal associates remain among the most damaging forest pests (Wood 1982; Ayres and Lombardero 2000; Kirisits 2004). Pests that are at low and relatively stable endemic levels can quickly transition into substantial threats over regions far larger than the sites of origin by expanding into large-scale outbreaks or by invading new environments (Bentz et al. 2010). One such example is the mountain pine beetle (MPB; *Dendroctonus ponderosae*)-fungal outbreak, which in Canada alone has infested over 18.1 million hectares of *Pinus contorta* forests (http://www.for.gov.bc.ca/hfp/mountain_pine_beetle/facts.htm). Cataloging the diversity of species and populations is necessary for developing an understanding of how these organisms interact with their host trees, and assessing their potential threat to forest ecosystems. My thesis addresses some of these issues, focusing on the ophiostomatoid (Sordariomycetes, Ascomycota) *Grosmannia clavigera*. This fungus is an important component of the MPB-fungal outbreak. It is an obligate symbiont of the MPB and can be a destructive pathogen for *P. contorta* trees. We generated new molecular resources for *G. clavigera* and some related species, which I used to (1) establish the fungal phylogeny, species boundaries and population structures in both epidemic and localized populations infesting different species of pine, (2) assess the potential role of host trees and/or beetle vectors in fungal evolution and divergence, and (3) identify genes that may be involved in fungal species divergence and/or host adaptation. The outcome of

this work provides a foundation for future research on the comparative and functional genomics of this important group of fungi, and of related ophiostomatoid genera. In this introductory chapter I will give a brief overview of the following topics (1) symbiosis; (2) ophiostomatoid biology, ecology and systematics; (3) fungal species concepts and criteria; (4) fungal comparative and population genomics, and (5) the rationale and the objectives of my work.

1.2 Symbiosis

Fungi are heterotrophs and so they must interact with other organisms to obtain their food. This has led fungi to evolve saprophytic, mutualistic or parasitic lifestyles and nutritional modes. Fossils and molecular clock data suggest that fungi have repeatedly evolved partnerships (i.e. symbiosis) with plants, some of which are ancient and even ancestral to terrestrial plants (Selosse and Le Tacon 1998; Heckman et al. 2001). The establishment of plants and other eukaryotes on land was probably facilitated by such partnerships (Selosse and Le Tacon 1998). It is likely that initially some mycelium-forming fungi were saprophytic and, as they interacted with plants, continued to develop the ability to tolerate plant defenses, so that parasitic and, eventually, biotrophic interactions evolved (Taylor and Osborn 1996). While the earliest fungi have been reported from the Precambrian period, the first examples of plant responses to fungi have been traced back to early Devonian (Taylor et al. 1992; Heckman et al. 2001). While plants and fungi continued to form a close association from the early stages of terrestrial colonization, insects likely originated in the Silurian period (Taylor and Osborn

1996; Engel and Grimaldi 2004). Therefore, it is likely that fungi first adapted to plants and that interactions with insects developed later. Beetle-tree-fungal symbiosis has been dated as originating between 40 and 85 million years ago. Today, symbiotic associations are widespread in nature, are essential for the functions of ecosystems and the evolution of biodiversity, and involve complex interactions that are still poorly understood (Farrell et al. 2001; Six 2003; Harrington 2005; Jordal and Cognato 2012).

1.2.1 Bark beetle-tree interactions

Bark beetles belong to the Scolytinae subfamily, which includes over 6,000 described species within about 225 genera including many ambrosia and cone beetles that are specialized in exploiting woody plants (coniferous/hardwood trees and shrubs). Most bark beetles are harmless to healthy living trees, but some, in particular those in the genus *Dendroctonus*, are important pests in coniferous forests, especially in the boreal and temperate regions of the northern hemisphere. Each beetle species typically colonizes either a few (i.e. often closely related members of the same genus) or a single tree species (e.g. *D. rufipennis, D. pseudotsugae* and *D. jeffreyi*); the colonization occurs on specific parts of the host tree, and each tree can be inhabited by a large number of beetles (Wood 1982). However, some bark beetles (e.g. the MPB) can attack more distantly related host species; and can also switch to novel host species when beetle ranges expand into new habitats where 'naïve' host trees may have not developed sufficient defenses against large-scale attacks (Ayres and Lombardero 2000; Bentz et al. 2010; Safranyik et al. 2010; Cullingham et al. 2011). These potential new

interactions are a matter of concern, as they can result in extensive outbreaks and damage in forest ecosystems as the MPB-fungal complexes on *Pinus* forests in Canada (Kurz et al. 2008; Safranyik et al. 2010; Cullingham et al. 2011).

1.2.1.1 Mountain pine beetle: ecology, biology and systematics

Mountain pine beetles are native to North America and historically found in areas from northern Mexico to western Canada (BC and marginal regions in western Alberta where *P. contorta* trees are present), and from the Pacific Coast to South Dakota in the USA (Wood 1982; Mock et al. 2007). Although the MPBs' primary hosts are *P. contorta*, *P. ponderosa* and *P. monticola*, the beetles can successfully attack and reproduce in most pine species throughout their range. These pine species include: *P. albicaulis, P. strobus, P. flexilis*, *P. sylvestris* and *P. lambertiana* (Wood 1982; Safranyik and Wilson 2006). Host species have been shown to cause variation in some parameters of the beetles' life-cycle, like survival, phenology, development rate, and body size (Reid 1962; Reid 1963; Safranyik and Linton 1983; Langor 1989; Langor et al. 1990; Safranyik and Wilson 2006).

Mountain pine beetles have also been recorded attacking other coniferous genera such as *Picea engelmannii × glauca, P. mariana* and *Abies* species. Successful reproduction in non-pine hosts has only been reported for the MPB populations infesting *Picea* species. However, these observations are based on pheromone-baiting of the

beetles in epidemic regions and not on natural host-tree colonization and reproduction (Huber et al. 2009; Safranyik et al. 2010). The life cycle phase in which MPBs disperse and find new hosts often occurs during a short period between July and early September, depending on the geographic area (Safranyik and Wilson 2006). In large outbreaks, bark beetles are reported to locate a suitable host tree by randomly landing and testing the tree and its resistance capability (Hynum and Berryman 1980; Wood 1982). After the selection of host trees they use volatiles or pheromones to recruit more beetles, leading to a mass attack that defeats the chemical defenses of individual trees (Raffa and Berryman 1982; Raffa and Berryman 1983). During the last two decades of MPB-fungal outbreaks, the beetles have expanded their geographic ranges and have established in new host species (Cullingham et al. 2011). The beetle-fungal complexes have crossed the Rocky Mountain to the north-central Alberta where forest composition shifts to *P. banksiana* (jack pine), which is evolutionary close and form a hybrid zone with *P. contorta* (Bentz et al. 2010; Safranyik et al. 2010). Cullingham et al. (2011) used microsatellites to identify species and hybrids of pine and showed successful MPB-fungal attack and establishment in natural *P. banksiana.* Because *P. banksiana* is an important component of the North American boreal forest, it may permit the MPB to expand to the east across the north of Canada.

Outside the MPB epidemic populations, localized populations of the beetle-fungal complexes in the western and southern USA have been studied less extensively (Wood 1982; Mock et al. 2007). Generally, in response to host abundance and climate suitability, the beetle population size can progress through characteristic “endemic,

incipient epidemic, epidemic (i.e., outbreak) and post-epidemic (i.e., declining)" phases (Safranyik and Carroll 2006). During the endemic phase the "localized populations" are very small (less than 10 trees are attacked per hectare) and beetles typically infest damaged trees or those with compromised defense capacity (Carroll et al. 2006). Localized populations of beetles tend to be specific to some pine species (Wood 1982; Sturgeon and Mitton 1986a; Langor et al. 1990). However, given the right climate and suitable resources, they can erupt into large-scale epidemics causing significant losses of healthy pine stands of different species. Large-scale outbreaks of MPB-fungal complexes have caused mortality of hundreds of millions of trees including the primary host *P. contorta* and other pine species over large epidemic regions (Kurz et al. 2008; Safranyik et al. 2010; Cullingham et al. 2011). Two factors contribute to an ongoing MPB-fungal outbreak in western North America: 1) food supply (abundant mature *P. contorta* resulted from forest management practices such as extensive fire suppression over the past 50 years), 2) climate changes (mild winters lead to low brood mortality and dryer summers increase stress on trees).

1.2.1.2 Mountain and Jeffrey pine beetle taxonomy

Dendroctonus ponderosae was originally described from *P. ponderosa* in South Dakota (Hopkin 1902). Later, the mountain pine beetle (*D. monticolae*) was described from *P. contorta*, *P. ponderosa*, *P. lambertiana* and *P. monticola* while the Jeffrey pine beetle (*D. jeffreyi*, JPB) was described from *P. jeffreyi* (Hopkins 1909). The range of *D. ponderosae* extended south and east into Wyoming, South Dakota, Utah, Colorado,

Arizona and New Mexico, and *D. monticolae* was found in Idaho, Montana and British Columbia, as well as south into California along the west coast. Experimental mating of *D. ponderosae* and *D. monticolae* suggested that these two beetles might represent only one species that varied in body size with host and region (Hay 1956). The two species, *D. monticolae* and *D. ponderosae*, were synonymized as one single species of mountain pine beetle: "*D. ponderosae"* (Wood 1963). Additional evidence (i.e. phenology and karyology) supported the synonomy of *D. ponderosae* and *D. monticolae*, and confirmed the species status of *D. jeffreyi* (Lanier and Wood, 1968; Renwick and Pitman 1979; Zúñiga et al. 2002). Later, genetic studies showed differences in MPB colonizing *P. contorta* and *P. ponderosa* where they intermixed in localized populations in Colorado and California, and also between populations breeding in two varieties of *P. contorta* (i.e. var. *murrayana* and *var latifolia*) (Stock and Guenter 1979; Stock and Amman 1980; Sturgeon and Mitton 1986a; Kelley et al. 2000). Using mtDNA sequences and AFLP markers, Mock et al. (2007) have shown an isolation-by-distance pattern of separation, and some of the isolated populations (e.g. Arizona populations) were congruent with the original morphological subdivision between *D. ponderosae* and *D. monticolae.* While uncertainties around the host specificity and perhaps species boundaries in the MPB remain, this will soon change now that one genome sequence of the beetle is available (Keeling et al. 2013). A rapid and cost-effective re-sequencing of additional strains from distinct localized populations of the MPB would help clarify its population/species boundaries, which would support population genotyping analyses and identifying potential adaptive genomic differences.

While it seems that MPB can attack all species of pine across its range, there is no evidence of this beetle on *P. jeffreyi* – even in sympatric regions where MPBs and JPBs and their respective hosts are intermixed and the trees are attacked (Higby and Stock 1982; Wood 1982; Six and Paine 1999; Kelley et al. 2000). Instead *P. jeffreyi* is commonly infested by the JPB. This beetle species is a sibling of the MPB with an almost identical morphology, life cycle and gallery characteristics; however, JPB is a highly specialized beetle that only occur in *P. jeffreyi* trees in a limited geographic region in the western USA (Higby and Stock 1982; Wood 1982; Six and Paine 1999). The differences in host tree range and the physiological tolerances to various tree defense compounds, as well as a low genetic variation have distinguished these bark beetles as two distinct species (Kelley and Farrell 1998; Six et al. 1999; Kelley et al. 2000; Six 2003). The resistance of *P. jeffreyi* to the MPB may be related to heptane, which is a volatile chemical-defense unique to this pine species (Mirov and Hasbrouck 1976; Six and Paine 1998; Smith 2000). The JPB can periodically kill healthy trees, but because of its limited host range and restricted geographic distribution, it is economically less important than the MPB. The geographic range of JPBs follows roughly that of its host tree, extending from the northern border of California to the northern Baja peninsula, Mexico (Wood 1982). Based on molecular makers, Kelly et al. (2000) showed lower genetic diversity in JPB than in MPB populations, and they speculated that the specialist JPB has diverged from MPB, a more “generalist” beetle species.

1.2.2 Bark beetle-fungal interactions

Beetles and their associated fungi spend most of their life cycle protected either under the bark (e.g. bark beetles) or inside the wood (e.g. ambrosia beetles). The beetle life cycle usually requires one to three years; however, some beetles have several generations per year (Wood 1982). The MPB and its sister species JPB spend their whole one-year life cycle under the bark except for a short dispersal flight in summer (Whitney 1971; Six et al. 1999). The timing of MPB attacks is related to seasonal temperatures with attacks occurring earlier in warmer areas. If the beetles successfully overcome host tree defenses, the adult beetles construct vertical galleries in the phloem under the bark, where they lay eggs. During this process, spores of the fungal symbionts are transferred to the phloem and the outer surface of sapwood, where they start to grow and reproduce. Generally, the eggs hatch within a few days and eventually late larval instars overwinter underneath the bark. Before emerging, young adult beetles will come in contact with the fungi that have grown from the spores introduced by the parents. When new adult beetles emerge and fly to select and attack new host trees, they transport the slimy fungal spores on their exoskeletons, or in mycangia (organs specialized for transferring fungal spores to new hosts) and/or inside their gut. As beetles attack the trees, the cycle is repeated. Each generation generally takes one year, except at high elevations where it can take two years.

Beetle bodies and their galleries host a variety of microorganisms, mostly bacteria, yeasts and filamentous fungi. Among all fungal symbionts, ophiostomatoids are the

most common and relatively well-known associates of both bark and ambrosia beetles (Münch 1907; Wingfield et al. 1993; Wingfield and Seifert 1993; Jacobs and Wingfield 2001; Kirisits 2004; Harrington 2005; Massoumi Alamouti et al. 2007). The association of ophiostomatoid fungi with particular beetle species can be either specific or more casual (Kirisits 2004). In a specific association, different populations of the beetle consistently carry spores of certain fungal species that are generally not found on any other beetles. For example there has been a historical association of the MPB with two ophiostomatoid species: *G. clavigera* and *O. montium* (Rumbold 1941; Robinson 1962; Robinson-Jeffrey and Davidson 1968). Both species are present in the beetle's maxillary mycangia, but *O. montium* is isolated more frequently from the beetles' exoskeletons than from the mycangia (Whitney and Farris 1970; Six 2003). Fungal associates of JPBs have been sampled throughout its most geographic range but only from the mycangia. All isolates from JPBs have been identified as *G. clavigera* and no *O. montium* has been found (Six and Paine 1997). Although more work is required to assess the diversity of fungal species associated with the JPB, *G. clavigera* were shown to be the dominant and specific symbiont of this beetle and its sister species MPB (Six 2003).

Recent works by Lee et al. (2005; 2006a) on the epidemic populations of MPB infesting *P. contorta*, reported two distinguished additional fungal species closely associated with the beetle. When inoculated at high densities, the pathogenic *Leptographium longiclavatum* grows inside the *P. contorta* sapwood and can subsequently kill the tree. The non-pathogenic, slow growing *Ceratocystiopsis* sp. is

mainly found inside the beetle galleries. Heterogeneous assemblages of species from different ophiosotmatoid genera are rather common in bark beetle ecosystems (Six 2003; Kirisits 2004; Massoumi Alamouti et al. 2007), possibly because different fungal associates have different roles in the beetles' life cycle.

Beetle- fungal interactions are complex and poorly understood and thus can be controversial. Interactions are diverse, ranging from antagonistic to commensal or mutualistic relationships (Klepzig and Wilkens 1997; Six 2003). In many cases, beetle-fungal symbioses seem to be mutualistic, benefiting both the fungus and the beetle (Ayres and Lombardero 2000). The ophiostomatoid fungi almost completely rely on beetles for dispersal and clearly benefit by being transported to new hosts (Six 2003; Harrington 2005). Besides this protected transport to suitable habitats, beetle dispersal provides protection from desiccation and UV light (Klepzig and Six 2004). The fruiting structures of these fungi usually have long stalks bearing sticky and concave shaped spores that can easily adhere to beetle cuticles for dispersal. These morphological features are considered as adaptations to insect dispersal and to the bark beetle habitat (Francke-Grosmann 1967; Whitney 1971; Malloch and Blackwell 1993).

The evolution of mycangia also indicates that some beetles also benefit from the association with fungi (Paine et al. 1997; Harrington 2005). Based on a combination of evidence and speculation fungal symbionts may 1) increase the availability of nutrients (e.g. nitrogen and sterol) that are rather scarce in host trees so that beetles can complete their life cycle; 2) make the infested trees more favorable for the beetles by

detoxifying host defense metabolites and lowering the wood moisture content; and 3) reduce the effectiveness of tree defenses (Lieutier et al. 2009). Given this, the relationship between *G. clavigera* and its beetle vectors is mutualistic, partly due to the specificity of the association and the development of the beetles' mycangia for the maintenance and protection of fungal spores.

1.2.3 Fungal-beetle-tree interactions

Tree mortality results from complex interactions among the tree, beetles and their fungal associates (Robinson 1962; Yamaoka et al. 1990; Yamaoka et al. 1995; Paine et al. 1997; Lee et al. 2006b; Rice, et al. 2007a,b; DiGuistini et al. 2011; Wang et al. 2013). Pioneer fungal species, like *G. clavigera*, need to colonize tree tissues that are alive and to overcome tree defenses induced by the beetle attacks. While the suggestion that fungal pathogens help the beetle to colonize host trees is speculative (Six and Wingfield 2011), increasing evidence suggests that the beetle-tree-associated fungi have evolved specific mechanisms that allow them to colonize healthy trees (Hesse-Orce et al. 2010; DiGuistini et al. 2011; Lah et al. 2013; Wang et al. 2013). Fungi that do not display high levels of pathogenicity might be those invading tree tissues later and more slowly, following pathogenic fungal associates (e.g. *Ceratocystiopsis* sp. associate of the MPB). In the struggle to colonize a host tree, pathogen strategies may involve degradation or conversion of toxic host molecules, transport of host defense molecules out of the cell, modification of cell structures to avoid/exclude/sequester toxic host molecules (Katsir et al. 2008; DiGuistini et al. 2011; Wang et al. 2013). In *G. clavigera*, a number of genes

(e.g. cytochromes P450s, dehydrogenases and monooxygenases) potentially involved in a combination of these strategies for detoxifying defense chemicals from *P. contorta* have been identified or functionally characterized (i.e. ABC-transporters) (DiGuistini et al. 2011; Wang et al. 2013).

Conifers have evolved a complex chemical defense system that can be deployed against a wide range of pests and pathogens. The system produces both constitutive and inducible chemicals that are mainly composed of oleoresin and phenolics. Oleoresin is primarily composed of monoterpenes and diterpenes, but the chemical composition can vary significantly depending on the species of pine or different populations of the same species. For example, in contrast to other pine species, *Pinus jeffreyi* have very low amounts of terpene components in their resin. Instead, they contain aldehydes, which are diluted with heptane (Mirov and Hasbrouck 1976). Although tree phytochemistry plays a critical role in host selection by bark beetles (Raffa 2001), it also affects the rate at which fungal spores germinate and grow. In general, both terpenes and phenolics have shown deleterious effects on microorganisms (Delorme and Lieutier 1990; Savluchinske Feio et al. 1999; Hofstetter et al. 2005; Wang et al. 2013).

1.3 Ophiostomatoids ecology, biology and systematics

Fungi are the second largest group of eukaryotes with an estimate of over 1.5 million species, of which only 5 to 10% have been described (Hawksworth 2001; Spooner and Roberts 2010). Ophiostomatoid fungi are members of Ascomycota, which is the largest fungal phylum with over 64,000 described species (Kirk et al. 2008). In their sexual phase (teleomorph), ascomycetes produce sac-like structures called asci where the ascospores (sexual spores) are produced. The asci are enclosed in fruiting bodies called ascocarps. After ascospores are released, they germinate and produce a mycelium, from which the conidiophores and conidia (asexual spores) that characterize the asexual phase (anamorph) will develop. In their wide diversity of forms, conidiophores and conidia vary greatly from asci and ascospores. In filamentous ascomycetes, neither the teleomorph nor the anamorph is considered to be an organism by itself, as neither can exist without a mycelium. Therefore, the taxonomy of a fungus must include both the anamorph and teleomorph, but these phases do not always occur together at the same time or under the same conditions. Also, there are many species where only an asexual phase has been observed (e.g. *L. longiclavatum*). For many years, anamorphs had an independent taxonomy and a separate name (i.e. anamorph genera) — but when the teleomorph becomes known its name takes precedence over that of the anamorph (Seifert and Samuels 2000). Currently, while each fungus has only one name, researchers have proposed competing names for anamorphs and teleomorphs, following the “one fungus one name” proposal by the

2011 International Botanical Congress (Hawksworth 2011; Wingfield et al. 2012; Hawksworth et al. 2013).

Ophiostomatoids represent an artificial group of 397 accepted species (De Beer et al. 2013). These fungi are highly adapted for being dispersed by insects and colonizing host plants (Wingfield and Seifert 1993; Spatafora and Blackwell 1994; Six 2003; Harrington 2005; De Beer and Wingfield 2013; Wang et al. 2013). They are distributed in the Northern and Southern hemispheres on a wide variety of host plant substrates (Upadhyay 1993). These fungi are highly pleomorphic, and can grow as either mycelia or as unicellular yeast forms. The fungal sexual phase typically has long-necked, flask-shaped ascocarps called perithecia (Upadhyay 1981; De Beer and Wingfield 2013). The asexual structures come in a variety of forms, all of which produce similar sticky conidia (Upadhyay 1981; Six 2003; De Beer and Wingfield 2013). Some fungi like *G. clavigera* have yeast-like structures inside the mycangia, but form mycelia when grow in the tree phloem and sapwood (Tsuneda and Currah 2006).

The current classification of the ophiostomatoid group is largely based on ribosomal gene DNA (rDNA) that places the 397 accepted species in 12 genera that are named according to the “one fungus: one name” proposal (Wingfield et al. 2012; De Beer and Wingfield 2013). Six of these genera are classified in the order Ophiostomatales: *Ophiostoma* sensu lato (s. l.) (including *Pesotum, Sporothrix* and *Ambrosiella*), *Ceratocystiopsis*, *Fragosphaeria*, *Graphilbum*, *Raffaelea* s. l., and *Leptographium* s. l. (including *Grosmannia,* and a few other unresolved groups). The other six are classified

in the order Microascales: *Ceratocystis* s. l. (including *Thielaviopsis* and *Ambrosiella*), *Graphium* sensu stricto, *Knoxdaviesia*, and *Sphaeronaemella*, *Cornuvesica* and *Custingophora*. Among the currently defined genera, *Ophiostoma, Leptographium* and *Ceratocystis* include the largest number of species (134, 94 and 72 spp., respectively). *Ceratocystis* species are characterized by endogenous conidia (Minter et al. 1983; Gebhardt et al. 2005). In contrast, species in the Ophiostomatales are characterized by a variety or a continuum of anamorphs (e.g. *Hyalorhinocladiella*, *Leptographium*, *Pesotum* and *Sporothrix*) that form exogenous conidia by building apical walls (Minter et al. 1982; Hausner et al. 1993; Jacobs and Wingfield 2001; Massoumi Alamouti et al. 2009). *Ceratocystis* species have less specific relationships with their beetle vectors than most genera in Ophiostomatales, particularly *Grosmannia* and *Leptographium* species, which are always associated with bark beetles and mainly colonize coniferous host trees (Jacobs and Wingfield 2001; Harrington 2005; Duong et al. 2012).

Traditional taxonomy in ophiostomatoid fungi is complicated by their limited range of morphological characteristics and by their convergent evolution for the insect dispersal (Spatafora and Blackwell 1994). Identification based on these structures has confused the taxonomy of these fungi for many years (reviewed by De Beer et al. 2013). For example, in the early 1970's *Ophiostoma* and *Ceratocystis* were considered to be synonymous, based on the features of their perithecia (Upadhyay and Kendrick 1975, Upadhyay 1981, De Hoog and Scheffer 1984, Wingfield et al. 1993). The genus *Ophiostoma* contains many species with a variety of ascospore shapes, including *Ceratocystiopsis* and *Grosmannia* species that until recently were synonymous with

Ophiostoma (Zipfel et al. 2006). Anamorph morphology has been preferred for defining fungal species for two reasons. First many species have no known sexual phase or do not produce fruiting body (ascomata) under artificial condition; for example, more than 20 species of *Leptographium* are known to have teleomorphs (Zipfel et al. 2006) but the sexual states are not known for at least an additional 70 species (e.g. *L. longiclavatum* associate of the MPB). Second anamorphs have diverse shapes that are easily observed in artificial media. But defining species using only the morphology of asexual phase can be problematic because many species develop a combination of anamorphs (e.g. synnematous and mononematous *Leptographium* anamorphs in *G. clavigera*) or reduced and non-distinctive asexual structures (Tsuneda and Currah 2006). Further, for many species, including *G. clavigera*, anamorphs can degenerate after repeated subculturing on artificial media (Robinson-Jeffrey and Davidson 1968; Tsuneda and Hiratsuka 1984; Okada et al. 1998; Six 2003). Therefore, for species identification and their classification at higher taxonomic level, morphological characterization needs to be complemented by molecular analysis or sufficient diagnostic markers. Analyses of DNA sequence data have redefined the status of several genera and species and have led to the discovery of many unrecognized species. This trend is likely to continue as more sequence, genomic and metagenomic data become available for this ecologically important fungal group (DiGuistini et al. 2011; Hintz et al. 2011; Massoumi Alamouti et al. 2011; Haridas et al. 2013; Khoshraftar et al. 2013).

Since rDNA was first used to estimate phylogenetic relationships, the classification of genera has evolved rapidly and names of most species have been changed

(Upadhyay 1981; Upadhyay 1993; De Beer et al. 2013). rDNA studies have shown that most of the genera in Ophiostomatales are polyphyletic, suggesting that similar morphological characteristics and an intimate association with beetles have originated more than once in *Ophiostomatales* genera (Spatafora and Blackwell 1994; Cassar and Blackwell 1996; Farrell et al. 2001; De Beer and Wingfield 2013). Phylogenetic analyses of rDNA sequences placed the genera *Ophiostoma* and *Ceratocystis* into two different ascomycetes orders, but kept the synonymy of the genera *Ophiostoma* and *Ceratocystiopsis*. More recent studies have shown that *Ophiostoma, Leptograhium, Raffaelea* and *Ambrosiella* species are polyphyletic; however, the partial rDNA data are not adequate to define monophyletic groups inside these genera (Zipfel et al. 2006; Massoumi Alamouti et al. 2009; Duong et al. 2012; De Beer and Wingfield 2013). Zipfel et al. (2006) used DNA sequence data from combined partial nuclear large subunit (nLSU) rDNA and β-tubulin genes, and suggested three well-supported, sexual-monophyletic clades in *Ophiostoma*. They re-introduced the teleomorph-genus *Ceratocystiopsis* to include *Hyalorhinocladiella* anamorphs and species with short perithecial necks. They proposed grouping together species with *Leptographium* anamorphs, including *G. clavigera* and close relatives, and to accommodate these species they re-instated the teleomorph-genus *Grosmannia*. However, when they tested the monophyly of *Leptographium*-forming species they did not consider some close relatives including *Ambrosiella*, *Raffaelea* and the monotypic genus *Dryodomyces* that are all associated with the ambrosia (fungal-feeding) beetles (Gebhardt et al. 2005). Previous nuclear small subunit (nSSU) rDNA phylogenies have shown that these fungi are dispersed within the *Leptographium*-forming species (Cassar and Blackwell 1996;

Jones and Blackwell 1998; Farrell et al. 2001; Rollins et al. 2001; Gebhardt et al. 2005). In chapter 2 we address some of these issues, and extend this work by generating the largest multigene phylogeny for a diverse set of genera in the ophiostomatales. We show that the *Leptographium*-forming species, which also represent the most common associates of bark beetles, form a clade separated from ambrosia-beetle associates. We also show that neither *Ambrosiella* nor *Raffaelea* are well-defined within ophiostomatales and both are likely to represent additional distinct genera (Massoumi Alamouti et al. 2009). Based on a large rDNA phylogeny of 216 Ophiostomatales taxa, De Beer and Wingfield 2013 suggested that there are at least 24 distinct groups (including *Ambrosiella* and *Raffaelea* clades resolved by our mutligene phylogeny in chapter 2) that might represent distinct genera. However, the rDNA data are not sufficient to define the unresolved groups, and thus for now they remain in the *Ophiostoma* s. l., *Leptographium* s. l. and *Raffaelea* s. l. until a multigene phylogeny becomes available for a larger number of taxa to define appropriate new combinations for these species.

1.4 Fungal species concept and recognition

Defining a “species” is fundamental for studying speciation and for understanding the biology and ecology of organisms. It is also essential for practical purposes such as pest controls and quarantine regulations. Despite disagreements about the origin and maintenance of species, there is a consensus view that species are “segments of separately evolving metapopulation lineages” (Mayden 1997; de Queiroz 1999; 2007).

During speciation processes, diverse events occur (e.g. initial separation, diagnosable characters, monophyly and reproductive lineages), and describing each of these events forms the bases for alternative species criteria (de Queiroz 1998). Properties used in each criterion can arise at different times and in different orders that can vary with factors like geography, demography, drift, selection and gene flow. For example, fixation of a nucleotide character in a lineage segment that originates from a large subdivision of the ancestral species may take longer than a lineage originating from a founder event or under strong positive selection. Different species criteria can lead to different conclusions regarding the boundaries and number of recently diverged species in different organisms (Avise 2004).

In fungi the criteria most often used to recognize and delimit species emphasize morphological divergence (morphological species criterion: MSC) and, less commonly, reproductive isolation (biological species criterion: BSC) (Brasier 1987; Hawksworth et al. 1995). Morphologically simple organisms like fungi may become genetically isolated due to selection and/or random genetic drift before morphological phenotypes or mating-behavioral differences have accumulated (Brasier 1987; Taylor et al. 2000). For a number of reasons, BSC cannot be a general method of choice for fungi. BSC focuses on the ability of a fungal species to interbreed in nature or in laboratory conditions. However, not all fungi can be cultured or mated in the laboratory, and no sexual stage is known for at least 20 percent of described fungi (Geiser et al. 1998). For example, *G. clavigera* can be cultured but not mated, and no sexual state is known for two of its close relatives, *L. longiclavatum* and *L. terebrantis*. However, failure to find the

sexual stage for a fungus does not mean that it does not exist; sexual reproduction has been demonstrated from crosses of fungi that were originally thought to be asexual (Hull et al. 2000; Magee and Magee 2000), and recombining population structures have been reported using multilocus nucleotide analyses (Burt et al. 1996; Geiser et al. 1998; Pringle et al. 2005; Matute et al. 2006). Some species that lack a sexual stage may have developed other strategies for overcoming the resulting shortfall of genetic recombination (Lynch et al. 1993; Butcher 1995), but it is difficult to conclusively rule out the presence of this stage (Pawlowska and Taylor 2004).

Phylogenetic and population genetic methods that emphasize nucleotide divergence (phylogenetic species criterion: PSC) define species as the "smallest monophyletic clade of organisms that share a derived character state" (Avise and Ball 1990). PSC became broadly applied in species recognition in the late 1980's, with the discovery of the Polymerase Chain Reaction (PCR) and the availability of many DNA characters through direct sequencing. These methods have resulted in an increasing number of reported cryptic species and species complexes in all taxonomic groups of living organisms (Bickford et al. 2005). Based on the outcomes of PSC in fungi, it is expected that most fungal species whose current description is based on morphology actually consist of more than one closely related cryptic or sibling species (Taylor et al. 1999; Taylor et al. 2000; Kasuga et al. 2003).

Accuracy in PSC requires information from multiple loci. Relying on the concordance of more than one gene genealogy (i.e. phylogenetic species recognition by

genealogical concordance: PSCGC) provides a higher resolution and can avoid subjectively determining the boundaries of species (Avise and Ball 1990; Slatkin and Maddison 1990; Baum and Shaw 1995; Taylor et al. 2000; Pringle et al. 2005; Matute et al. 2006). Methods that summarize population genetic and genealogical patterns across many loci are essential for diagnosing recent evolutionary lineages for which divergence time has not been long enough to detect complete reciprocal monophyly at many loci (Hudson and Coyne 2002; Maddison and Knowles 2006; Knowles and Carstens 2007).

1.4.1 Defining species boundaries and population structures in *G. clavigera*

During a MPB outbreak in the 1960's, Robinson described *G. clavigera* from the beetle and its primary host trees *P. contorta* and *P. ponderosa* (Robinson-Jeffrey and Davidson 1968). The fungus was first named *Europhium clavigerum* (Robinson and Davidson 1968), then *Ophiostoma clavigerum*, and was renamed *G. clavigera* [= *O. clavigerum* (Robinson and Davidson) Harrington] (Zipfel et al. 2006). The description of this species is based on an isolate recovered from the sapwood of *P. ponderosa* from Cache Creek, BC, Canada (Robinson and Davidson 1968). Later, *G. clavigera* was also reported from JPBs in western United States (Six and Paine 1997). *Grosmannia clavigera* is predominantly haploid except for a transient diploid phase occurring during sexual reproduction, which likely requires pairing of the two opposite mating types (heterothallic) (DiGuistini et al. 2011; Tsui et al. 2013). The sexual structures are rarely observed in nature and not produced under artificial conditions (Robinson-Jeffrey and Davidson 1968; Lee et al. 2003; Massoumi Alamouti et al. 2011).

Given the rarity of *G. clavigera* teleomorphs, mating incompatibility among strains cannot be determined, and the morphology of anamorphs is often unstable and unreliable for the identification of this species. The fungus is pleoanamorphic possessing several types of synnematous and mononematous anamorphs (Lee et al. 2003, Six et al. 2003, Tsuneda and Hiratsuka 1984, Upadhyay 1981). Furthermore, Tsuneda and Hiratsuka (1984) have observed holoblastic and annellidic-yeast states for this species. In addition to its pleoanamorphic nature, *G. clavigera* produces a broad range of conidiophores and conidia size and shape that overlap with those of *G. aurea*, *G. robusta, L. longiclavatum, L. pyrinum*, *L. terebrantis* and *L. wingfieldii* (Six et al. 2003, Lee et al. 2003). Molecular tools, including restriction fragment length polymorphism (RFLP), DNA fingerprinting and multigene phylogenies, have also shown that these seven *Leptographium*-forming fungi form a complex of closely related species that are very similar in morphology, found on the same host trees, and associated with bark beetles (Jacob and Wingfield 2001; Lim et al. 2004, Six et al. 2003). Therefore, in *G. clavigera* morphologically and genetically similar individuals can be recovered either from infected pine trees or from the beetles throughout the beetle distribution. Given the difficulty to identify the fungi in *G. clavigera*-species complex, it is not surprising to find misidentifications in the literature. The multigene phylogenies (i.e. rDNA and three house-keeping genes); however, have improved the molecular identification of *G. clavigera* distinguishing the pathogen from most close relatives, except for *L. terebrantis* (Six et al. 2003; Lim et al. 2004; Roe et al. 2010). Therefore, controversy had remained

over the species status of *L. terebrantis* and its phylogenetic relationship with *G. clavigera*.

According to morphological criteria and DNA fingerprinting analysis *G. clavigera* has a wide geographical range across Canada (British Columbia and western Alberta) and the Unites States (Washington, Oregon, and California inland to South Dakota, Colorado and New Mexico) where the major MPB and JPB hosts and other pines are found and attacked by the beetles (Upadhyay 1981; Zambino and Harrington 1992; Six and Paine 1999; Six et al. 2003; Lim et al. 2004). However, it seems reasonable to argue that the association of this fungal species with distinct host beetle and tree species might be an artifact of the morphological species recognition and shortcoming of the molecular tools that have been used in earlier systematic and population studies. Further, AFLP (amplified fragment length polymorphism) markers have suggested that two genetically distinct groups exist in the epidemic populations of *G. clavigera* associated with the MPBs (Lee et al. 2007). However, the AFLP groups have not been supported by our gene genealogies presented in chapter 3 or by microsatellite markers (Massoumi Alamouti et al. 2011; Tsui et al. 2012).

The beetle-pathogen population structures have been well studied in epidemic regions using different molecular tools (Lee et al. 2006a; Mock et al. 2007; James et al. 2011; Roe et al. 2011; Gayathri Samarasekera et al. 2012; Tsui et al. 2012). Based on the fungal surveys *G. clavigera* epidemic populations are divided into four groups according to four major geographic regions in northwestern BC, southern BC,

northeastern BC/Alberta, and Rockies. A north-south genetic structure is concordant among different studies for both the beetle and the pathogen pointing to the fungal pathogen's dependence on the beetle for the dispersal. On contrary to the epidemic regions, studies on the fungal localized populations in western and southern USA is limited to few allozyme markers for the California populations of *G. clavigara* associated with JPBs, showing low a genetic diversity within and among the populations.

1.5 Fungal genomics and their importance in *G. clavigera*

Sequencing technologies that enable the acquisition of whole genome sequences provide new approaches to address questions related to fungal systematics, evolution and speciation, as well as to identify genomic regions or particular genes involved in fungal divergence, host specialization and/or pathogenicity. Fungi are the eukaryotic group with the greatest number of completely, or nearly completely, sequenced genomes (http://www.ncbi.nlm.nih.gov/genomes/leuk s.cgi) (Stajich et al. 2009). This is not only due to their importance in ecology, medicine, agriculture and biotechnology, but also because their genomes are among the smallest and most compact eukaryotic genomes.

Saccharomyces cerevisiae was the first fungal genome to have been fully sequenced and annotated, making a major contribution to the basic understanding of eukaryote cell physiology, genetics and biochemistry (Goffeau et al. 1996). Today

sequencing advances have led to over a hundred fungal genome sequences being available for the phyla Ascomycota, Basidiomycota, Zygomycota and Chytridiomycota. Among the currently sequenced genomes, more than 30 represent important plant pathogens like *Cochliobolus, Fusarium*, *Mycosphaerella, Magnaporthe*, *Sclerotinia*, and *Ustilago* species, most of which have been sequenced by Sanger sequencing (Dean et al. 2005; Cuomo et al. 2007; Coleman et al. 2009; Ma, H Charlotte van der Does, et al. 2010; Raffaele et al. 2010; Schirawski et al. 2010; Spanu et al. 2010; Stukenbrock et al. 2010; Goodwin et al. 2011; Klosterman et al. 2011; Croll and McDonald 2012; Raffaele and Kamoun 2012; Stukenbrock et al. 2012). Genome analyses of these fungi have shown a remarkable diversity in genome size (i.e. 19–160 Mb), numbers of chromosomes, structural organization, repeat-induced point (RIP) mutations, transposable element activity and protein similarity (Galagan et al. 2005; Raffaele and Kamoun 2012). For examples, the genomes of rice blast fungus *Magnaporthe grisea* and its non-pathogenic relative *Neurospora crassa* have an average amino acid identity of only 47% and almost no conserved synteny (Berbee and Taylor 2001; Dean et al. 2005). These observations have been a powerful tool for inferring genome evolution and speciation process in plant pathogens, and to identify the molecular bases of host adaptation and pathogenicity.

Functional and genomic studies in ophiostomatoid fungi are at an early phase and at the time of writing are limited to three pathogens *G. clavigera*, *O. ulmi* and *O. novo ulmi* and one saprophyte *O. piceae* (DiGuistini et al. 2009; Forgetta et al. 2013; Haridas et al. 2013; Khoshraftar et al. 2013). *G. clavigera*, owing to its importance as the MPB

symbiont and a pine pathogen was the first beetle-tree-associated fungus for which the genome sequence was published. The genome was assembled from a combination of Sanger paired-end (PE) reads (0.3-fold coverage), 454 single reads (7.7-fold coverage), and Illumina paired-end reads (100-fold coverage), resulting in a high-quality draft genome sequence of 32.5 Mb (DiGuistini et al. 2009). Building on the draft sequences, the *G. clavigera* genome was manually finished, yielding 18 supercontigs with a total length of 29.8 Mb and 8,312 gene models that were supported by expressed sequence tag (EST) and RNA-seq data (DiGuistini et al. 2011). These genomic resources provided major insights into the biology of this fungus and the molecular mechanisms involved in its host-pathogen interactions (Hesse-Orce et al. 2010; DiGuistini et al. 2011; Lah et al. 2013; Wang et al. 2013).

Terpenoids, particularly monoterpenes, are among the most abundant defense chemicals that protect pine trees against pests and pathogens; however, beetle-tree-associated fungi, like *G. clavigera*, have evolved the ability to tolerate these rather toxic environments and to grow and become established inside the living trees (Keeling and Bohlmann 2006; Boone et al. 2011; Bohlmann 2012). Functional genomics and transcriptomic data have shown that monoterpenes and *P. contorta* extracts induce a stress response in *G. clavigera* and activate a ~100-kb cluster of genes. The cluster contains genes involved in β-oxidation pathway, as well as monooxygenases and alcohol/aldehyde dehydrogenases, that may be involved in detoxification and/or metabolism of host-defense chemicals (DiGuistini et al. 2011). A number of other genes that may be important for tree colonization by the pathogen include cytochromes P450

and ATP-binding cassette (ABC)-transporter families (Hesse-Orce et al. 2010; DiGuistini et al. 2011; Lah et al. 2013). A pleiotropic drug resistance transporter in *G. clavigera* has been functionally characterized and shown to be required to control monoterpene levels within the cells, enabling the fungus to grow on media containing these compounds (Wang et al. 2013).

1.5.1 Next generation sequencing and fungal population genomics

Despite (or indeed because of) much progress in the area of genome sequencing and functional genomics, even more information can be gained from sequencing not just one genome per species but rather from sequencing and comparing the genomes of different strains from the same or closely related species. Such work was made feasible by next-generation sequencing (NGS) technologies and supporting analysis methods that focus on addressing questions relevant to the biology/ecology of eukaryotic microorganisms (Nowrousian 2010). Different NGS technologies offer differences in read lengths and numbers of reads (e.g. ~450+ bp reads for Roche 454 and ~50-100-bp reads for Illumina), but all generate millions to hundreds of millions of reads per sequencing run at a substantially lower cost than Sanger sequencing (Mardis 2008; Metzker 2010). Despite the short sequence reads, NGS platforms have greatly facilitated genome sequencing, first for prokaryotes, and, within the last few years, for eukaryotic genomes (Reinhardt et al. 2009; Li et al. 2010; Nowrousian et al. 2010). NGS of different strains of the same species, when a reference genome is available (i.e. ‘resequencing’), can take full advantage of the high throughput, because the sequences

of the additional isolates can be mapped with high confidence to the reference sequence. This approach works well even for sequencing of large genomes. Sequence reads that are mapped to a reference can be used for detecting genomic variants that include single nucleotide polymorphisms (SNPs), insertions/deletions (indels) or structural variants (Wang et al. 2008; Wheeler et al. 2008; Eck et al. 2009). A number of novel algorithms for mapping NGS reads (i.e. aligning reads to a reference genome) have been developed to address issues like alignment accuracy, uniqueness and confidence, and identifying SNPs and small indels (Li et al. 2009; Li and Durbin 2009; Trapnell and Salzberg 2009). For *de novo* genome assembly of eukaryotes including the *G. clavigera* reference genome, NGS was once used in combination with Sanger sequencing (DiGuistini et al. 2009); but today, longer reads, the ability to sequence both ends of a DNA fragment (i.e. paired-end sequencing), and developments in assembly programs have allowed generating *de novo* draft genomes of eukaryotes like the giant panda and the filamentous fungus *Sordaria macrospora* (Korbel et al. 2007; Simpson et al. 2009; Li et al. 2010; Miller et al. 2010; Nowrousian et al. 2010; Nowrousian 2010).

Since the introduction of PCR thirty years ago, increasing amounts of data for RFLP, AFLP, microsatellites, and small-scale DNA sequencing have broadened the range of questions open to empirical analysis in evolutionary and population genetic studies. With the recent abundance of genome-wide SNP data, and now the advent of genomes sequenced across populations, evolutionary studies have become a data-driven discipline. Population genomics has enabled comprehensive views of genome-wide patterns of sequence variation within and between closely related species (i.e.

polymorphisms and divergence respectively), and of the evolutionary relationships between such species. Further, it has provided new insights into biological and ecological attributes of plant pathogens, and a rich resource for genome-wide assessment of adaptive evolution and functional variations. For filamentous plant pathogens, comparative genomics between different strains of the same species or between the closely related species have revealed genomic features that can have important roles in lifestyle and host ranges of these fungi. Examples include expansion/contraction of specific gene families (e.g. polyketide synthases, lytic enzymes and putative transporters), gene pseudogenizations or deletion, repetitive sequences, and distinct genomic regions (e.g. telomeres, gene clusters and rapidly-evolving genomic islands), or dispensable chromosomes (reviewed by Stukenbrock et al. 2011; Raffaele and Kamoun 2012; Stukenbrock and Bataillon 2012). Other mechanisms that drive genetic diversity and shape the genomes of plant pathogens include DNA point mutations (or SNPs when shared among different strains of the same species) and recombination.

1.6 Single nucleotide polymorphisms

As in many other organisms, SNPs represent a useful kind of genetic variation across the genomes of plant pathogens (Collins et al. 1998; Tyler et al. 2006; Ma et al. 2010; Neafsey et al. 2010; Amselem et al. 2011; Andersen et al. 2011; Desjardins et al. 2011; Klosterman et al. 2011; McCluskey et al. 2011; Stukenbrock et al. 2011; Xue et al. 2012; Condon et al. 2013). Generally, nucleotide changes that result in amino acid

replacements are called ‘nonsynonymous’ polymorphisms, while substitutions in coding regions that do not change an amino acid are called ‘synonymous’. Variations can also occur within non-coding regions of the genome such as intergenic regions, upstream (i.e. putative regulatory regions and/or promoters) and downstream of a gene, or within introns (Hartl and Clark 2007). SNPs in the coding regions of genes or in regulatory regions are more likely to cause functional differences than nucleotide variants elsewhere.

Genomic analyses have shown that SNP frequencies can vary among species and strains, and can also vary locally along chromosomes (Raffaele and Kamoun 2012). For example, the polymorphisms between two *Fusarium graminearum* strains were most frequently found near telomeres and within discrete chromosome regions that are likely involved in plant-fungus interactions (Cuomo et al. 2007). In other cases, SNP frequencies were more homogenous across the genomes, but local biases (e.g., in gene or genomic regions) in the ratio of synonymous to non-synonymous variants have been observed and interpreted as a signature of positive selection (Fedorova et al. 2008; Raffaele et al. 2010; Stukenbrock et al. 2010; Kemen et al. 2011; Stukenbrock et al. 2011; 2013). Therefore, population and comparative genomics approaches offer the possibility not only of capturing the evolutionary history of species, but also of identifying signatures of selection in genes involved in ecological divergence and host specificity.

1.7 Divergent adaptive evolution in fungi

In plant pathogens, speciation has been generally associated with ecological divergence (e.g. host shifting or specificity) or with large changes in genomic structure and composition due to hybridization (reviewed by Stukenbrock 2013). Genes or genomic regions important in fungal divergence and speciation can stand out because of an increased fixation of adaptive mutations — ecological divergence (e.g. host specificity) in particular will leave footprints of positive selection.

The relative abundance of non-synonymous and synonymous polymorphisms can reflect the effect of natural selection, which is generally expected to remove slightly deleterious non-synonymous variants in coding sequences (i.e. purifying selection). Therefore a significant increase in the ratio of non-synonymous to synonymous changes may reflect protein-coding genes that are favored by natural selection (i.e. positive selection) (Nielsen and Yang 1998; Yang and Swanson 2002; Yang 2007). Some analysis are designed to detect positive selection at individual coding sites within a gene or between lineages, which may be more sensitive than earlier methods that would average the ratio of non-synonymous and synonymous along a gene or between species (Yang and Nielsen 2000; Yang 2007). If genomic data from different strains and related species are available, rates of non-synonymous and synonymous substitutions between species can be assessed and compared with patterns of polymorphisms. Such tests allow the detection of ancient selection in homologous genes of related species (Fedorova et al. 2008; Li et al. 2008; Raffaele et al. 2010; Stukenbrock et al. 2010;

Kemen et al. 2011; Stukenbrock et al. 2011; 2013). For example, Stuckenbrock et al. (2010) using population genomic approaches within and between species have shown the role of beneficial mutations in host specialization of the wheat pathogen *Mycosphaerella graminicola*. Positive selection has been shown to have a significant impact on the evolution of filamentous plant pathogens, particularly on the effector genes that are responsible for modulating host physiology and enabling colonization of plant tissue (Hogenhout et al. 2009; Raffaele and Kamoun 2012).

Comparative and population genomic analyses are becoming more common in fungi. Such work assesses the genomic-wide pattern of variation within and between closely related species, and scans for the signature of positive selection across protein-coding genes. To date, however, most of these studies have analyzed crop pathogens and have not investigated beetle-tree-associated fungi. Due to their obligatory symbiotic association with the beetle vector and colonization of host-tree niches, ophiostomatoid species like *G. clavigera* represent an interesting group of organisms in which to study ecological divergence and assess genome-wide molecular adaptation. Despite its ecological importance, evolutionary and population studies of this fungal group are limited to a few protein-coding genes and neutral markers (i.e. AFLP and microsatellites), and little is known about the level and pattern of nucleotide polymorphisms and their functional importance in fungal fitness. This limitation is changing as more genome data becomes available (DiGuistini et al. 2011; Massoumi Alamouti et al. 2011; Forgetta et al. 2013; Haridas et al. 2013; Khoshraftar et al. 2013).

1.8 Overview and purpose of this thesis

As an important pathogen of *P. contorta* forests in western North America, *G. clavigera* forms a tight symbiotic association with two sibling species of bark beetles: MPB and JPB. Despite their close phylogenic relationship, the two beetle vectors infest distinct host-tree species, where they spend most of their whole life under the bark, except during short dispersal flights. Thus, these organisms have a high degree of intimacy with their hosts that could reinforce local adaptive structures and consequent host differentiation of their accompanying fungi. However, important ecological and biological attributes of the symbiotic fungi may not have been detected because their species boundaries had not been clearly defined. *Grosmannia clavigera* shows morphological, ecological and evolutionary similarities with a number of pine-infesting species, leading to confusing taxonomic and phylogenetic relationships at the interspecific and generic levels. While *G. clavigera* and its close relatives have been re-instated into the genus *Grosmannia* inside the *Ophiostomatales*, the genus is intermixed with another ecologically distinct group of Ophiostomatales that are collectively called ambrosia fungi. Because the ambrosia fungi represent different genera, their inclusion in phylogenies made *Grosmannia* paraphyletic. Therefore the main goals of this thesis were to clarify the phyogenetic status of the genus *Grosmannia* within Ophiostomatales, to define species boundaries in *G. clavigera* and its evolutionary relationship with other closely related pine-infesting species, and finally to find genes or genomic regions that might be involved in species divergence and host specificity.

In an evolutionary sense, plant-fungal interactions are known to be older than interactions between fungi and insects. Studies of beetle-associated microflora have generally focused on reporting the fungal associates of different bark beetle species. However, the host tree that directly nourishes the fungus may have a more important role in the speciation and diversity of ophiostomatoid fungi than does the beetle that vectors the fungus. Given this likelihood, I hypothesize that *G. clavigera* is a species complex composed of more than one phylogenetic species, each being associated with a distinct host-tree species. A phylogenetic species may be associated with MPB or JPB beetles, depending on the tree on which the beetle feeds. I predict that adaptation to the specific chemistries of host tree will emerge as an important feature in the evolutionary divergence of these fungi, and finally that genes potentially involved in host-pathogen interactions have diverged in response to selection in different host environments.

Chapter 2 Multigene phylogeny of ophiostomatoid fungi associated with bark and ambrosia beetles

2.1 Introduction

Bark and ambrosia beetles are weevils (*Coleoptera*: *Curculionida*e) in the subfamilies *Scolytinae* and *Platypodinae* (Farrell et al. 2001; Marvaldi et al. 2002). They spend part of their life cycles in galleries that they mine under the bark (scolytid bark beetles) or in the wood of trees (scolytid and platypodid ambrosia beetles); and they vector diverse fungi that colonise the wood (Batra 1966). In coniferous forests, the most common fungi in beetle galleries are filamentous ascomycetes that are generally known as ophiostomatoids (Six 2003; Harrington 2005). In many countries, ophiostomatoid fungi include species that are involved in tree diseases, cause considerable value loss to the wood product industry, and are considered as quarantine pests (Wingfield et al. 1993; Alfaro et al. 2007; Fraedrich et al. 2008).

The current molecular classification of ophiostomatoids is largely based on nuclear rDNA (rRNA gene). It places the more than 140 species into morphologically similar, non-monophyletic teleomorph genera: *Ceratocystiopsis* Upadhyay and Kendrick, *Ceratocystis* Ellis and Halstead, *Grosmannia* Goidánich emend. de Beer, Zipfel and Wingfield and *Ophiostoma* H. and P. Sydow, as well as into a number of anamorph genera (Hausner et al. 2000; Zipfel et al. 2006). Species of genus *Ceratocystis* are

sensitive to the antibiotic cycloheximide; they are related to the *Microascales*, and are characterised by the *Thielaviopsis* Went anamorph and by phialidic conidia that are produced by the ring wall-building of the collarette (Minter et al. 1983). In contrast, species that are tolerant to cycloheximide are characterised by a variety of anamorphs that form conidia by building apical walls (e.g., *Hyalorhinocladiella*, *Leptographium*, *Pesotum* and *Sporothrix*); they are placed into the genera *Ceratocystiopsis, Grosmannia* and *Ophiostoma* of the *Ophiostomatales* (Hausner et al. 1993; Spatafora and Blackwell 1994; Zipfel et al. 2006). *Ceratocystis* species have less specific relationships with their beetle vectors than do members of *Ophiostomatales* teleomorph genera and related anamorphs, which are always associated with scolytid bark beetles (Kirisits 2004).

Ophiostomatoid fungi that were originally called ‘ambrosia’ are now classified in the anamorph genera *Ambrosiella* von Arx and Hennebert, *Raffaelea* von Arx and Hennebert and *Dryadomyces* Gebhardt (Batra 1967; Gebhardt et al. 2005). ‘Ambrosia’ represents heterogeneous groups of fungi that include yeasts and filamentous fungi (Batra 1963; Francke-Grosmann 1967). Except for two basidiomycetes (Batra 1972; Hsiau and Harrington 2003), all are ascomycetes, most of which belong to the genera *Ambrosiella* and *Raffaelea*. ‘Ambrosia’ fungi typically form dense mats of hyphae or clusters of small conidiophores (sporodochia) with conidia that germinate into mass of highly vacuolated sprout cells in beetle galleries; this is often referred to as the ‘ambrosia phase’ (Batra 1967). The majority of ‘ambrosia’ fungi have a symbiotic relationship with platypodid or scolytid ambrosia beetles (Batra 1967; Kubono and Ito

2002; Gebhardt et al. 2004; Gebhardt et al. 2005; Harrington et al. 2008). These beetles bore galleries into the wood of host trees, and dependent on their fungal symbionts to exploit the nutrient-poor xylem (Batra 1966; Roeper 1995). Many of the beetles have developed mycangia or similar structures that may support transferring fungi to new hosts (Batra 1966; Six 2003). In contrast to ambrosia beetles, scolytid bark beetles feed on the phloem of trees. While bark beetles seem to be less dependent on their fungal associates for nutrition, some may supplement their diets by consuming the fungal associates that they carry on their exoskeletons, and in their guts or mycangia, after the fungi have grown in the beetle galleries (Harrington 2005; Six 2003). *Ambrosiella* species are considered to be typical symbionts of ambrosia beetles; however, a number of *Ambrosiella* fungi are also reported from certain bark beetles (Batra 1967; Krokene and Solheim 1996; Rollins et al. 2001; Kirisits 2004). Confusingly, then, the term 'ambrosia' can refer to specific fungal associates and to a particular fungal morphological form (ambrosia phase) in beetle galleries that supports the beetle and its progeny developments (Hartig 1844; Batra 1967).

The taxonomy of 'ambrosia' fungi has been re-evaluated, because their classification was originally established using morphological characteristics that are poorly defined in artificial media, and because most are known only by their asexual state (Batra 1967; Gebhardt et al. 2005). For species that lack sexual structures, defining species using only the morphology of asexual phase can be problematic, because many species develop a combination of anamorphs or reduced and non-distinctive asexual structures (Tsuneda and Currah 2006). The genera *Ambrosiella* and

Raffaelea were differentiated based on the morphology of conidiogenous cells (von Arx and Hennebert 1965); *Raffaelea* have a series of cicatricial conidial scars, while *Ambrosiella* does not. However, applying electron microscopy to a number of *Raffaelea* and *Ambrosiella*, as well as to other asexual ophiostomatoids, has begun to reveal details of conidiogenesis that are not visible by light microscopy (Gebhardt et al. 2005; Gebhardt and Oberwinkler 2005).

Molecular phylogenies are clarifying the taxonomic status of most 'ambrosia' fungi among ascomycetes, and specifically in the ophiostomatoids (Cassar and Blackwell 1996; Jones and Blackwell 1998; Farrell et al. 2001; Rollins et al. 2001; Gebhardt et al. 2005). Nuclear small subunit (nSSU) rDNA phylogenies indicate that both *Ambrosiella* and *Raffaelea* are polyphyletic, suggesting that similar morphological characteristics and an intimate association with beetles have originated more than once in these genera (Cassar and Blackwell 1996; Farrell et al. 2001). Supporting the initial phylogenies, Gebhardt et al. (2005) showed that species of *Ambrosiella* in the order *Microascales* are also morphologically distinct from those in the order *Ophiostomatales*, suggesting that the taxonomic status of genus *Ambrosiella* should be re-evaluated. Despite the close relationship of genera *Ambrosiella* and *Raffaelea* with the ophiostomatoid fungi, they were not included in comprehensive ophiostomatoid fungi phylogenies (Hausner et al. 2000; Zipfel et al. 2006).

In the work described here, we address where the *Ambrosiella* and *Raffaelea* species should be placed within ophiostomatoids and their relationship to

Ophiostomatales genera. We report a multigene phylogenetic analysis of filamentous 'ambrosia' fungi that includes twenty-five species from the genera *Ambrosiella, Raffaelea* and *Dryadomyces* and thirty species from other ophiostomatoid clades. Our results indicate the limitations of using classical morphological traits and molecular analyses based on single genes to address the taxonomy of these fungi. Finally, we discuss whether the ecological characteristics of the beetle vectors (i.e., bark vs. wood), which originally contributed to the classification of 'ambrosia' fungi, appear to be phylogenetically significant.

2.2 Materials and methods

2.2.1 Taxon sampling

Twenty-two strains from the genera *Ambrosiella, Raffaelea* and *Dryadomyces* were requested from culture collections (Table 2.1). We also included three undescribed *Ambrosiella* species (sp. 1, sp. 2 and sp. 3) isolated from spruce-colonising bark beetles in Canada and Europe (Krokene and Solheim 1996; Massoumi Alamouti et al. 2007), as well as an undescribed species from the mycangia of ambrosia beetles *Trypodendron rufitarsus* and *T. lineatum* collected from lodgepole pines infested by the mountain pine beetle (MPB), *Dendroctonus ponderosae* in BC (Kuhnholz 2004). All the fungal strains are maintained at the Breuil Culture Collection (University of British Columbia, BC).

2.2.2 Morphological investigation

Fungal fruiting structures produced from one- to four-week-old cultures grown on Difco malt extract agar (MEA; 20g Difco malt extract, 10g Difco agar and 1L distilled water), potato dextrose agar (PDA) and MEA enriched with 1% Difco yeast extract (YEMEA), were mounted in water and observed using a Zeiss Axioplan compound light microscope. For scanning electron microscopy (SEM), small wood blocks (5x2x5 mm) bearing fungal structures were fixed using the method described by Lee et al. (2003). After fixation, samples were dried with a Blazers CPD 020 critical point drier. They were coated twice with gold palladium using a Nanotech Semprep II sputter coater and examined using a Hitachi S4700 scanning electron microscope.

2.2.3 DNA extraction, PCR amplification and sequencing

DNA was extracted from mycelia grown on Oxoid MEA [33 g malt extract agar (Oxoid CM59), 10 g agar 'tech. No.3' and 1L distilled water] plates overlaid with cellophane (gel dry grade, BioRad) following the method described by Möller et al. (1992). The nSSU was amplified and sequenced with primers NS1 and NS4 (White et al. 1990), and the nuclear large subunit (nLSU) region was amplified and sequenced with ITS3 or NL1/LR3 or LROR (Vilgalys and Hester 1990; O'Donnell 1992). The partial β-tubulin gene (β-tubulin) was amplified and sequenced using the primer set BT2E/BT12 (Kim et al. 2004). PCR amplification was performed as described by Kim et

al. (2004). PCR products were purified with a Qiaquick PCR Purification Kit (Qiagen, Ont, Canada). Sequencing was performed on an ABI 3700 automated sequencer (Perkin-Elmer, Foster City, CA) at the DNA synthesis and Sequencing Facility, Macrogen (Seoul, Korea). GenBank accession numbers of the new sequences obtained are shown in Table 2.1.

2.2.4 Phylogenetic analysis

Sequences from the representatives related to ophiostomatoid fungi in the *Ophiostomatale*s and *Microascales*, as well as those representing *Xylariales* and *Hypocreales* were included in the analysis (Table 2.1). Sequences were aligned using MAFFT (Katoh et al. 2002) and then manually adjusted with PHYDIT version 3.2 (http://plasza.snu.ac.kr/~jchun/phydit). The flanking regions were excluded from the analysis because sequence length varied with species. Phylogenetic analysis were conducted for the three loci (nSSU, nLSU and β-tubulin) under both maximum parsimony (MP) methods of PAUP*4.0b10 (Swofford 2003) and Bayesian inference of MrBayes v. 3.1.2 (Ronquist and Huelsenbeck 2003). Concordance of the three different gene datasets was evaluated with the partition homogeneity test (PHT) implemented with PAUP*4.0b10, using 1000 replicates and the heuristic general search option (William and Ballard 1996; Swofford 2003). *Taphrina populina* and *Penicillium expansum* were assigned as the outgroup taxa (Jacobs et al. 2003; Gebhardt et al. 2005).

For parsimony analysis, all characters were equally weighted and unordered. Separate analyses were conducted with gaps treated as missing data and as a fifth character state (Swofford 2003). The MP trees (MPTs) were identified by heuristic searches with 100 random stepwise addition replicates and tree-bisection-reconnection branch-swapping algorithms. Statistical support for the branches was assessed by bootstrap analysis (BS) using 1000 MP heuristic searches with ten random sequence addition replicates for each bootstrap replicate. Bayesian inference of phylogeny was calculated based on a Markov chain Monte Carlo analysis with the general time reversible (GTR+I+G) substitution model as determined by AIC criteria of Modeltest (Posada and Crandall 1998). The proportion of alignment sites was assumed to be invariable with gamma-distributed substitution rates of the remaining sites. Four simultaneous Markov chains were run from random starting trees for 1 000 000 generations and sampled every 100 generations (generating 10 001 trees). The first 5000 trees were discarded as burn-in, and inferences of posterior probability (PP) were calculated from 5001 trees.

2.3 Results

2.3.1 Sequence analysis

For multigene phylogenetic analysis, we generated 59 rDNA and 35 β-tubulin new sequences on the genera *Ambrosiella*, *Raffaelea*, *Dryadomyces*, *Ophiostoma* and *Ceratocystis* and retrieved 41 sequences of other ophiostomatoid taxa from GenBank (Table 2.1). We were able to amplify the target loci in all species in the analysis, except for the nLSU locus in *R. sulcati* and the β-tubulin locus in *A. gnathotrichi* and *R. arxii.* From ophiostomatoid taxa, no significant length variations were observed in nSSU and nLSU amplicons, whose lengths varied from 831–833 and 506–539 nucleotides, respectively. However, β-tubulin sequences varied from 550 to1095 nucleotides. This region contained four exons and three introns. Sequences of the four exons were of equal length for all ophiostomatoid taxa in the analysis, whereas sequences of the three introns varied highly in both nucleotide composition and length. Some taxa lacked either one or two introns, which accounted for the large difference in β-tubulin sequence lengths.

The aligned dataset consisted of 837 nucleotides from nSSU, 592 nucleotides from nLSU and 1258 nucleotides from β-tubulin loci. We excluded no nucleotides from nLSU and nSSU loci. However, 735 intron positions were excluded from the β-tubulin locus

because the large differences in length and composition of intron sequences across the ophiostomatoid orders and genera made the regions unalignable (Swofford et al. 1996).

We submitted the sequences to BLAST to assess potential misidentifications. The comparisons confirmed the species identity of all *Ambrosiella* and *Raffaelea* fungi in the analysis expect for *A. ips* and *A. sulcati. Ambrosiella ips* showed a high level of sequence identity (rDNA + β-tubulin: 99.7%) with that of *O. montium*, suggesting a potential misidentification and the possibility that *A. ips* and *O. montium* might represent a single species. *Ambrosiella sulcati* showed high sequence identity with that of *R. canadensis* (rDNA + β-tubulin: 99.6%), indicating that these two taxa may represent a single species. Since the closest match of *R. castellanii* was a *Dothideomycetes*, which is unrelated to ophiostomatoid fungi, this species was not included in the final phylogenetic analysis.

2.3.2 Phylogenetic analysis

MPTs from conserved individual loci (nSSU, nLSU and β-tubulin exons) showed weak resolution for the topology of deeper nodes and terminal branches. Although the partition homogeneity test (P-value< 0.01) did not indicate that the rDNA and β-tubulin datasets were concordant, MPT topologies from individual rDNA loci was not in conflict to the combined nSSU+nLSU+β-tubulin dataset, which had better resolution and higher support values. The concatenated matrix (nSSU, nLSU and β-tubulin) included sixty-

seven taxa from different ophiostomatoid genera (figure 2.1, Table 2.1) and 1952 aligned sites, of which 719 sites were variable and 534 sites were parsimony informative.

Under the first gap treatment (i.e., gaps as missing data), the parsimony analysis of the concatenated dataset resulted in nine MPTs with a length of 2703 steps (CI=0.39, RI=0.71). Gaps as a fifth character state resulted in eleven MPTs with the same length and topologies; thus, for the remainder of the analysis gaps were treated as missing data. The consensus phylogeny inferred from the Bayesian analysis revealed similar topology within and between groupings but with higher supporting relationships than those from the MP analysis (figure 2.1). The most visible difference was the placement of the genus *Ceratocystiopsis*, which appeared either as part of the *Ophiostoma* clade or as a basal group to the *Ophiostoma* and *Grosmannia* clades. However, neither method supported the placement of this genus strongly.

Analysis divided the ingroup taxa into four major clades, each receiving 79% or more bootstrap support and 100% posterior probabilities (figure 2.1A–D). These clades corresponded to the three teleomorph genera *Ophiostoma, Grosmannia,* and *Ceratocystiopsis*, recently re-instated by Zipfel et al. (2006) in the *Ophiostomatales*, as well as to the genus *Ceratocystis* of the *Microascales*. The twenty-five *Ambrosiella*, *Dryadomyces* and *Raffaelea* species in the analysis were divided into at least six well-resolved (>51% BS; 100% PP) groups which nested within the clades of *Ophiostoma* (A), *Grosmannia* (B) and *Ceratocystis* (D) (figure 2.1).

Within clade A, *Ambrosiella ips* grouped strongly (100% BS) with *O. montium* in the *Ophiostoma ips* complex (group 1) (Zhou et al. 2004). Other *Ambrosiella* taxa isolated from scolytid bark beetles formed a single, monophyletic group (group 2) with 99% BS and 100% PP (figure 2.1). This group contained *A. tingens*, *A. macrospora* and three undescribed *Ambrosiella* taxa, and was well separated from other *Ambrosiella* species that have been isolated from ambrosia beetles. The bark-beetle associated group showed a sister relationship to a group containing members of the *O. piceae* complex (Harrington et al. 2001) and both groups are also sibling of the *O. ips* complex (figure 2.1A).

Four *Ambrosiella* species clustered with the representatives of *Raffaelea* and *Dryadomyces* in clade B, which also included the genus *Grosmannia*. All 'ambrosia' fungi in this clade had a close association with platypodid and scolytid ambrosia beetles but their monopoly received a poor bootstrap support (62%) and a low posterior probability of 81%. Instead they were subdivided into two distinct well-resolved groups (groups 3, 4) that were supported (79% BS; 100% PP) as sibling of the genus *Grosmannia*. Group 3, which received strong supports (78% BS; 100% PP), encompassed all *Raffaelea* taxa, except *R. lauricola* and *R. montetyi*. The group included: *R. albimanens*, *R. ambrosiae, R. arxii, R. canadensis*, *R. santoroi*, *R. sulcati*, *R. tritirachium*, as well as two *Ambrosiella* species. *Ambrosiella sulcati* clustered with *R. canadensis*, with strong support (100% BS). While *A. gnathotrichi* was closely related to

R. arxii. These two taxa formed a monophyletic relationship with the *R. canadensis*-clade without bootstrap support and a posterior probability of 95%.

Representatives of *Raffaelea* concentrated in clade B, but the genus is not monophyletic. However, the relationships among various taxa (*R. albimanens*, *R. santoroi*, *R. sulcati*, *R. tritirachium*) were well resolved. The only exception to this was the unstable positioning of *R. ambrosiae*, the type species of *Raffaelea*, which depending on the locus tested, formed a monophyletic relationship with the *R. arxii* and *R. canadensis* group, or a basal taxon in the group 3.

Group 4, which was supported with high posterior probability (100%), contained representatives from *Ambrosiella, Raffaelea* and *Dryadomyces* but none of these constituted a monophyletic cluster. *Ambrosiella brunnea* and *A. sulphurea* mixed with *D. amasea* and with *R. lauricola* and *R. montetyi*. Group 4 also included one unidentified species (TR25) isolated from the mycangia of ambrosia beetle *T. rufitarsus* from lodgepole pines infested by the MPB in BC (Table 2.2). This fungus appeared as sister of *A. brunnea* and both species formed a well-supported (82% BS; 100% PP) monophyletic clade with the recently described species *R. lauricola* (Harrington et al. 2008).

Three species of *Ambrosiella* belonged to the *Microascales* (clade D) but they did not form a single, monophyletic group (figure 2.1). Instead they formed two distinct

groups (groups 5, 6) that form a close association with scolytid ambrosia beetles but interspersed with *Ceratocystis* fungi bearing a loose relationship with different insects. First group included the type species *A. xylebori* as well as *A. hartigii* (group 5) while the other included two strains of *A. ferruginea* (group 6).

2.3.3 Morphological investigation

Table 2.2 summarised morphological features that have been used in the literature to describe the genera *Ambrosiella, Raffaelea* and *Dryadomyces*: conidiomatal types (hyphal/singly, sporodochial, synnematous) and conidial proliferation (annellidic, phialidic, and sympodial).

Representative strains of the *Ophiostoma*-related *Ambrosiella* tended to sporulate better on the PDA and YEMEA media therefore morphological observations were made on these cultures. Similar to other *Ambrosiella*, the undescribed *Ambrosiella* spp. (*Ambrosiella* sp. 1 and *Ambrosiella* sp. 2) isolated from bark beetles in Canada (Massoumi Alamouti et al. 2007) and those from European bark beetles (*Ambrosiella* sp. 3) (Krokene and Solheim 1996) produced simple, mononematous conidiophores (figures 2.2B, 2.3); these were arranged in a discrete sporodochium-like structure (figure 2.2). Solitary conidiophores were also observed in the younger cultures of these species. Observations by light microscopy of these three species and the *A. tingens* type culture (CBS 366.53) revealed a non-phialidic conidiogenesis (e.g. figure 2.3D). In

contrast, SEM observations revealed both annellidic (figures 2.2B, 2.3A-B) and sympodial (figure 2.3C) conidiogenesis in *Ambrosiella* sp. 1 as well as in *Ambrosiella* sp. 2; however, conidial development seemed to occur more frequently through annellidic percurrent proliferation than sympodial proliferation. Because *A. macrospora, A. tingens* and the species from Europe produced few spores on artificial media and wood blocks, we were unable to determine with certainty whether their conidiogenesis was sympodial or annellidic. *Ambrosiella ips* (CBS 435.34) also failed to produce any fruiting structures on media and therefore its morphological characters could not be compared to those of the genetically identical species *O. montium* (CBS 151.78).

2.4 Discussion

We established a comprehensive phylogeny that clarifies the relationships between most filamentous 'ambrosia' fungi isolated from platypodid and scolytid beetles and their relationships with the ophiostomatoid fungi. Our results are consistent with studies that described the polyphyletic status of the genera *Ambrosiella* and *Raffaelea* (Cassar and Blackwell 1996; Farrell et al. 2001; Rollins et al. 2001; Gebhardt et al. 2005). These earlier phylogenies used mainly nSSU rDNA sequences to characterise members of *Ambrosiella* and/or *Raffaelea* at higher taxonomic levels, and sets of species that that did not adequately represent the morphological and ecological diversity of ophiostomatoid fungi. In the work described here, we addressed both limitations. We generated a new multigene dataset, and we characterised a diverse set of fungi that included representatives from the genera *Ambrosiella*, *Raffaelea* and *Dryadomyces*, as

well as from taxa that we selected from currently accepted ophiostomatoid teleomorph and anamorph genera. Our analysis indicated that these fungi evolved from three major teleomorph groups in two ascomycete orders: *Ophiostomatales* and *Microascales*. We will not discuss the *Ambrosiella* species within *Microascales* that include the type species *A. xylebori* because this aspect has been thoroughly studied by Paulin-Mahady et al. (2002) and their results agree with ours. In the subsequent discussion we will focus on the *Ambrosiella* species that belong to the *Ophiostomatale*s.

von Arx and Hennebert (1965) introduced the genus *Ambrosiella* to describe *A. xylebori*, the most frequent associate of the ambrosia beetle *Xylosandrus compactus*. Following the original description of the type species, nine additional species were assigned to the genus (Batra 1967). These species share certain morphological features: simple conidiophores, sporodochial arrangements of conidiophores and single terminal conidia (Batra 1967). Cassar and Blackwell (1996) showed that the genus *Ambrosiella* was not monophyletic within *Ophiostomatales* based on nSSU rDNA. Their phylogenies recognised two possible *Ambrosiella* groups that were closely related to either *Leptographium*-forming species or *Ophiostoma* species characterised by their *Pesotum* (e.g., *O. piceae*) and/or *Hyalorhinocladiella* (e.g., *O. bicolour*) anamorphs. Results from Farrell et al. (2001), Rollins et al. (2001) and Gebhardt et al. (2005) supported these groupings, but resolved neither the monophyly of different *Ambrosiella* groups nor their relationships with the closely related genera *Ophiostoma* and *Raffaelea.* Our multigene phylogeny clarified the relationships among these fungi and recognised at least four groups of *Ambrosiella* within the *Ophiostomatales*.

2.4.1 *Ambrosiella* associates of bark beetles are related to the teleomorph genus *Ophiostoma*

Our results supported a novel clade that consisted of five bark beetle associates: *A. macrospora*, *A. tingens* and the three undescribed *Ambrosiella* species from Canadian and European bark beetles (Krokene and Solheim 1996; Massoumi Alamouti et al. 2007). *Ambrosiella macrospora* and *A. tingens* were originally described in the genus *Trichosporium* Nannfeldt as *T. tingens* var. *macrosporum* and *T. tingens* (Lagerberg et al. 1927; Francke-Grosmann 1952). Batra (1967) reclassified these two bark beetle associates into *Ambrosiella*, while indicating that this genus should describe fungal associates of platypodid and scolytid ambrosia beetles. nSSU rDNA phylogenies subsequently showed that these two species were more closely related to *Ophiostoma* than to *Ambrosiella* (Cassar and Blackwell 1996; Rollins et al. 2001); however, limitations from nSSU rDNA sequences and taxon sampling prevented these studies from characterising the phylogenetic relationships in detail. In contrast, our multigene analysis resolved the *Ambrosiella* associates of bark beetles as a distinct group within *Ophiostoma.* We also showed that the bark-beetle associated group is a sibling to the *O. piceae* complex, members of which commonly inhabit sapwood and bark beetle tunnels in temperate forests (Harrington et al. 2001). Members of this complex are distinguished by their teleomorph fruiting bodies and *Pesotum* anamorph, which include both synnematous and *Sporothrix* conidiophore arrangements (Harrington et al. 2001, Seifert et al. 1993). *Ambrosiella* species produce conidiophores without denticles (Batra

1967; Gebhardt et al. 2005), which differentiate them from *Sporothrix*. *Ambrosiella* associates of bark beetles, including the undescribed species, do not produce synnemata; instead, their conidiophores are arranged in distinct sporodochial-like structures that resemble those of other *Ambrosiella* and *Raffaelea* species. Our analysis also strongly suggested that the bark beetle-associated *Ambrosiella* and the *O. piceae* complex form a monophyletic clade with members of the *O. ips* complex, which are distinguished by their pillow-shaped ascospores and continuum of anamorphs including *Hyalorhinocladiella* and *Pesotum* (Zipfel et al. 2006). This monophyletic clade is a sister to another *Ophiostoma* group that include a number of species (e.g., *O. abietinum* and *O. stenoceras*) foundo in a diverse range of ecological niches and identified by their *Sporothrix* anamorph.

Our analysis grouped *A. ips* with *O. montium* within the *O. ips* complex. Originally *A. ips* was described in the *Tuberculariella* von Höhnel (Leach et al. 1934), and then was transferred into the genus *Ambrosiella* (Batra 1967). However, our molecular results, as well as morphological and ecological descriptions from the literature suggested that these two species might represent a single taxon. We showed that the two species shared a high level of sequence identity (99.7%). We were unable to compare the morphology of *A. ips* and *O. montium* because, in our hands, the only *A. ips* strain available from the CBS culture collection and reported in the literature did not sporulate; however, the sporodochium-like structures illustrated in the description of *A. ips* (Leach et al. 1934) are similar to *Graphilbum* reported for *O. montium* and *O. ips* (Upadhyay 1981; Hutchison and Reid 1988). Because both *A. ips* and *O. montium* have

been isolated from bark beetles (*I. pini* and MPB, respectively) that infest the same pine-host trees in North America (Leach et al. 1934; Lee et al. 2006), it is possible that galleries of these beetles have overlapped, resulting in fungal associates being mixed and *A. ips* being misidentified.

2.4.2 *Ambrosiella* associates of ambrosia beetles are related to the teleomorph genus *Grosmannia*

In our analysis, the remaining *Ophiostomatales Ambrosiella* grouped with members of genera *Raffaelea* and *Dryadomyces* and formed a sister relationship with members of the genus *Grosmannia*. All members of this group (*Raffaelea*, *Dryadomyces* and *Ambrosiella*) are closely associated with the platypodid and scolytid ambrosia beetles (von Arx and Hennebert 1965; Guerrero 1966; Batra 1967; Funk 1970; Scott and Du Toit 1970). Our multigene phylogenies suggested that these ambrosia-beetle associates are monophyletic but with weak statistical support. Note that while we included most species described from ambrosia beetles, relatively few such associates have been fully characterised, and we were unable to further test the monophyly with sequence data for more isolates. The placement of ambrosia-beetle associates within the *Ophiostomatales* has been problematic because earlier phylogenetic studies consistently grouped them with species like *O. piceaperdum* and *O. serpens* of the well-defined genus *Grosmannia* (Rollins et al. 2001; Gebhardt et al. 2005; Hulcr et al. 2007). Zipfel et al. (2006) re-instated this genus to accommodate the most common fungal associates of bark beetles that are distinguished by their *Leptographium* anamorph (Jacobs and Wingfield

2001); however, when they tested the monophyly of *Grosmannia* they did not consider the close relatives of this genus, namely, members of genera *Ambrosiella* and *Raffaelea*. While our results confirmed that *Grosmannia* is monophyletic, we also included the *Grosmannia*-related associates of ambrosia beetles. Our analysis placed these into two distinct groups and provided the first robust indication that the ambrosia-beetle associates are close but independent relatives of *Leptographium*-forming species commonly isolated from bark beetles.

The first of the two groups included seven of the nine tested species of genus *Raffaelea*, as well as *A. sulcati* and *A. gnathotrichi.* The exceptions, *R. lauricola* and *R. montetyi*, were placed in the second group. The *Raffaelea* genus was established by von Arx and Hennebert (1965) to describe *R. ambrosiae*, which is frequently associated with the pinhole borer *Platypus cylindricus* in North America and Europe. Because the group members have similar ecological and morphological features and formed a strongly supported monophyletic clade, we suggested that they should be all recognised as species of genus *Raffaelea s. str.* (von Arx and Hennebert 1965; Guerrero 1966; Batra 1967; Funk 1970; Scott and du Toit 1970). We summarised evidence for this as follows. *Ambrosiella sulcati* and *R. canadensis* were respectively isolated from the ambrosia beetles *Gnathotrichus retusus* and *Platypus wilsonii* when these two beetles inhabited the same host, *Pseudotsuga menziesii* (Douglas fir) (Batra 1967; Funk 1970). Differentiating *Raffaelea* and *Ambrosiella* morphologically by assessing whether cicatricial conidial scars are present or absent using light microscopy is difficult, and often depends on subtle interpretations by researchers (von Arx and Hennebert 1965;

Batra 1967; Funk 1970). However, our results showed that these two species have high rDNA and β-tubulin sequence identify (99.6%) and formed a conspecific group within *Raffaelea* species, and so indicated that *R. canadensis* and *A. sulcati* represent a single taxon and that *A. sulcati* should be transferred into *Raffaelea. Ambrosiella gnathotrichi* is a frequent associate of the conifer-infesting species *G. retusus* in North America (Batra 1967). While *R. arxii* forms a close association with the *Xyleborus torquatus* on the *Cussonia umbellif* in South Africa (Scott and du Toit 1970). Gebhardt et al. (2005) suggested that *A. gnathotrichi*'s conidial ontogeny differs from that of the *Ambrosiella* type species and they showed that *A. gnathotrichi* form a sister taxon relationship with *R. arxii*. Our multigene analysis provided a higher resolution for the species-level phylogeny and showed that these two fungi are closely related species. Although, our analyses suggest that *A. gnathotrichi* be assigned to the genus *Raffaelea*, a more thorough morphological examination of the type material is needed, particularly of conidial ontogeny.

The second group included the remaining species from all three genera: *Ambrosiella*, *Raffaelea* and *Dryadomyces*. The group members were *A. brunnea*, *A. sulphurea*, *R. lauricola*, *R. montetyi* and *D. amasae*, and an unidentified species isolated from *T. rufitarsus* colonising MPB-attacked lodgepole pine in BC (Kuhnholz 2004). Our phylogenetic analysis consistently resolved these species as sister of other ambrosia-beetle associates in the *Raffaelea* clade, with weak statistical support. We will clarify briefly the group members. *Ambrosiella brunnea* and *A. sulphurea* have been isolated from North American and European hardwood species (*Quercus* and *Acer*) that

were infested with ambrosia beetles in the genus *Monarthrum* and with *Xyleborus saxesenii*, respectively (Verrall 1943; Batra 1967). Gebhardt et al. (2005) showed that these two *Ambrosiella* species produce a non-phialidic conidiogenesis, and included them in their *Raffaelea* phylogenetic clade. In contrast, our analysis segregated them from both *Ambrosiella* and *Raffaelea* genera. Our analysis showed that the *Ambrosiella* associate of *Monarthrum* spp. (*A. brunnea*), and the unidentified species isolated from *T. rufitarsus* in BC are sister taxa that form a distinct, well-supported monophyletic group with the new vascular wilt pathogen *R. lauricola* (Fraedrich et al. 2008). This pathogen is associated with the exotic ambrosia beetle *Xyleborus glabratus*, and causes substantial mortality of redbay and other *Lauraceae* in the USA (Fraedrich et al. 2008; Harrington et al. 2008). nSSU and nLSU sequences from Fraedrich et al. (2008) respectively suggested that the pathogen is related to *A. brunnea* and *Leptographium* spp. Our phylogenetic results were consistent with this pathogen being a distinct species; and included it with other ambrosia-beetle associates of genera *Ambrosiella*, *Raffaelea* and *Dryadomyces* that formed a sister group relationship with *Leptographium* -forming species. Harrington et al. (2008) suggested that this fungus most appropriately fit the genus *Raffaelea*; however, they did not provide a detailed morphology of the conidiogenous cells. Further, our analyses placed this pathogen in a clade that was clearly separated from other ambrosia-beetle associates in the genus *Raffaelea*. Therefore, the taxonomic status of the vascular wilt pathogen in the USA is not clear and needs further study. Our multigene phylogenies placed *A. sulphurea* into a distinct subclade that formed a highly supported sister relationship with *R. montetyi*. Previous nSSU phylogenies had also suggested that these two species were closely related

(Gebhardt et al. 2004, 2005; Hulcr et al. 2007). Both *A. sulphurea* and *R. montetyi* are frequently associated with ambrosia beetles in the genus *Xyleborus* that inhabit the oak trees (Gebhardt et al. 2004). Although, *R. montetyi* has been shown to produce conidia by annellidic percurrent proliferation, resolving the taxonomic placement of *R. montetyi* and that of *A. sulphurea* will require additional morphological studies. Finally, the second group comprises a monotypic genus, *Dryadomyces*, which accommodates the single species *D. amasae*. The genus was introduced by Gebhardt et al. (2005) to describe fungi frequently isolated from ambrosia beetle *Amasa concitatus* infesting hardwood timbers in Taiwan. nSSU results indicated that these fungi were phylogenetically related to species of genera *Raffaelea* and *Ambrosiella* but they were included into the new genus *Dryadomyces* based on their unique conidiogenous cells (Gebhardt et al. 2005). Later, Harrington et al. (2008) amended the genus indicating that until the taxonomy of the genus *Ophiostoma* was better resolved, *Raffaelea* should include all ambrosia beetle symbionts with affinities to *Ophiostoma*. Our multigene analysis provided a better resolution for the phylogenetic status of the genus *Ophiostoma* and consistent with the morphological observations of Gebhardt et al. (2005), recognised *D. amasae* as a distinct monotypic lineage that formed a highly supported monophyletic relationship with *R. montetyi* and *A. sulphurea.*

2.4.3 Morphological features

Morphological characters used to define *Ambrosiella* are less informative in phylogeny because the genus is polyphyletic. The morpho-taxonomy of *Ambrosiella* and

Raffaelea has been difficult and unstable. The shape (reduced conidiophores) and arrangement of conidiophores (sporodochia), as well as the mode of conidiogenesis are the key morphological characteristics that traditionally used to differentiate the genus *Ambrosiella* from the closely related genus *Raffaelea* (von Arx and Hennebert 1965; Batra 1967). However, complex or simple conidiophores did not correlate with the generic or sub-generic classification. Also the mode of conidiogenesis is difficult to observe under light microscopy (Tsuneda and Currah 2006). *Ambrosiella* and *Raffaelea* genera were reported as sympodial (von Arx and Hennebert 1965). But, Gebhardt et al. (2005) showed the presence of phialidic conidia for *A. xylebori*, and in other two *Ceratocystis*-related species: *A. hartigii* and *A. ferruginea*. Recently *Raffaelea* species having annellidic conidiogenesis and sympodial proliferations in *D. amasae* were also illustrated (Gebhardt et al. 2004; Gebhardt and Oberwinkler 2005).

Our phylogenetic results also indicated that the characters used to define anamorphs are convergent within the *Ophiostomatales*. The conidial proliferation in the unidentified *Ambrosiella* species, as well as *A. tingens* was non-phialidic, which clearly distinguished them from the type species of *Ambrosiella*. In addition, we observed that the conidiophores of these strains lacked denticles; this differentiated them from anamorph genera *Sporothrix* and *Dryadomyces,* both of which have conidia formed sympodially on denticles arising from undifferentiated hyphae (Gebhardt et al. 2005; Harrington et al. 2001). In our SEM micrographs, we found that the non-phialidic conidiogenesis observed for *Ambrosiella* sp. 1 (figure 2.3A) and *Ambrosiella* sp. 2 (figure 2.3B) was occurring through annellidic proliferation, and consequently was

identical to that found for *Ophiostomatales* anamorph genera *Raffaelea, Hyalorhinocladiella*, *Leptographium* and *Pesotum* (Seifert and Okada 1993; Benade et al. 1995; Tsuneda and Currah 2006). These two undescribed *Ambrosiella* also formed apical sympodial conidiogenesis but less frequently. Although, additional work is necessary to define the true mode of conidiogenesis for all *Ambrosiella* species related to the *Ophiostoma*, it is important to note that the conidium ontogeny, an early important taxonomic character, is now being challenged, because conidial fungi often develop more than one pattern of conidiogenesis and can be assigned to different anamorphic genera (Tsuneda and Currah 2006).

While *Ambrosiella* sp. 1 and *Ambrosiella* sp. 2 with their annellidic conidiogenesis and the morphology of their conidiophores are most similar to species of the genus *Raffaelea*, they were clearly distinguished from the *Raffaelea* group by our multigene phylogeny. Consistent with our phylogenetic classification, members of *Raffaelea* clade also colonise different ecological niches. The morphological characteristics of *Ambrosiella* sp. 1 and *Ambrosiella* sp. 2 also resemble those of *Hyalorhinocladiella* anamorph. This anamorphic state is not clearly delimited to a genus and is present in anamorphs of *Ceratocystiopsi*s and *Ophiostoma* (e.g., *O. ips* complex) (Upadhyay and Kendrick 1974; Benade et al. 1996). Currently, species of genera *Ambrosiella* and *Raffaelea* are differentiated from other ophiostomatoid genera including *Hyalorhinocladiella* by the formation of sporodochia. However, the production of sporodochia is variable and is often associated with the growth of the fungus in its

natural habitat (beetle gallery); as well, the importance of this structure for segregating anamorphic genera within the ophiostomatoid fungi has not been clarified.

In conclusion, we clarified the phylogenetic classification of *Ambrosiella* species isolated from ambrosia and bark beetles and that of the *Raffaelea* and *Dryadomyces* associates of ambrosia beetles, as well as the relationships between these species and ophiostomatoid relatives. We found that species of genus *Ambrosiella* are distributed in a number of distinct phylogenetic groups that each might be reassigned to different genera, and that the genus *Raffaelea* should be revised. While no morphological characteristics unambiguously supported the monophyletic groups that we report from our molecular data for the genera *Ambrosiella* and *Raffaelea*, these groups are clearly associated with the feeding behaviour of their beetle vectors. Specifically, *Ambrosiella* associates of scolytid bark beetles formed a monophyletic group in the genus *Ophiostoma*, while species associated with scolytid and platypodid ambrosia beetles form separate lineages that have a monophyletic relationship with the genus *Grosmannia*. Generating additional support for the monophyly presented will require characterising a range of morphological characters and/or ecological traits on an expanded collection of freshly isolated fungi from ambrosia and bark beetles.

2.5 Tables and figures

Table 2.1 Fungal species used in this study

Species	Source [a]	GenBank accession no. [b]		
		nSSU rDNA	nLSU rDNA	β-tubulin
Ambrosiella sp. 1	UAMH10632	EU984247	(DQ268582)	(DQ268618)
	UAMH10633	EU984248	(DQ268583)	(DQ268619)
Ambrosiella sp. 2	UAMH10634	EU984249	(DQ268584)	(DQ268620)
	UAMH10635	EU984250	(DQ268585)	(DQ268621)
Ambrosiella sp. 3	NISK-1994-166-39A	**EU984252**	**EU984282**	EU977458
	NISK-1994-176-B4	**EU984253**	**EU984283**	EU977459
A. brunnea Batra	CBS 378.68	(AY858654)	EU984284	EU977460
A. ferruginea (Mathiesen-Käärik) Batra	CBS 408.68	**EU984254**	EU984285	EU977461
	JB13 [CB]	EU984255	EU984286	EU977462
A. gnathotrichi Batra	CBS 379.68	(AY858655)	EU984287	N/A
A. hartigii Batra	CBS 404.82	**EU984256**	EU984288	EU977463
A. ips (Leach, Orr and Christensen) Batra	CBS 435.34	AY858657	EU984289	EU977464
A. macrospora (Francke-Grosmann) Batra	CBS 367.53	**EU984257**	EU984290	EU977465
A. sulcati Funk	CBS 805.70	(AY858658)	EU984291	EU977466
A. sulphurea Batra	CBS 380.68	(AY497509)	EU984292	EU977467
A. tingens (Lagerberg and Melin) Batra	CBS 366.53	**EU984258**	EU984293	EU977468
A. xylebori Brader ex von Arx and Hennebert	CBS 110.61	(AY858659)	EU984294	EU977469
Dryadomyces amasae Gebhardt	CBS 116694	(AY858661)	EU984295	EU977470
Raffaelea albimanens Scott and du Toit	CBS 271.70	**EU984259**	EU984296	EU977471
R. ambrosiae von Arx and Hennebert	CBS 185.64	(AY497518)	EU984297	EU977472
R. arxii Scott and Toit	CBS 273.70	(AY497519)	EU984298	N/A
R. canadensis Batra	CBS 168.66	(AY858665)	EU984299	EU977473

Species	Source [a]	GenBank accession no. [b]		
		nSSU rDNA	nLSU rDNA	β-tubulin
R. castellanii (Pinoy) de Hoog	MUCL 15755	EU984260	EU984300	EU977474
R. lauricola Harrington, Fraedrich and Aghayeva		(EU123076)	(EU123077)	N/A
R. montetyi Morelet	CBS 451.94	(AY497520)	EU984301	EU977475
R. santoroi Guerrero	CBS 399.67	**EU984261**	EU984302	EU977476
R. sulcati Funk	CBS 806.70	(AY858666)	N/A	EU977477
R. tritirachium Batra	CBS 726.69	**EU984262**	EU984303	EU977478
Unidentified species	TR25 [CB]	EU984251	EU984281	EU977457
Ceratocystiopsis manitobensis (Reid and Hausner) Zipfel, Beer and Wingfield	UM 237	EU984266	(DQ268607)	(DQ268638)
Cop. minuta (Siemaszko) Upadhyay and Kendrick	CBS 463.77	EU984267	(DQ268615)	EU977481
Cop. minuta-bicolor (Davidson) Upadhyay	CBS 635.66	EU984268	(DQ268616)	EU977482
Ceratocystis adiposa (Butler) Moreau,	CBS 600.74	EU984263	EU984304	EU977479
C. coerulescens (Münch) Bakshi	CL 13-12 [CB]	EU984264	(AY214000)	(AY140945)
*C. moniliformis (*Hedgc.) Moreau	CBS 155.62	EU984265	EU984305	EU977480
Grosmannia abiocarpa (Davidson) Zipfel, Beer and Wingfield	MUCL 18351	EU984269	(AJ538339)	(DQ097857)
*G. clavigera (*Rob.-Jeffr. and Davidson) Zipfel, Beer and Wingfield	ATCC 18086	EU984270	(AY544613)	(AY263194)
G. cucullata (Solheim) Zipfel, Beer and Wingfield	CBS 218.83	(AY497513)	(AJ538335)	EU977483
G. penicillata (Grosmann) Goid. Zipfel, Beer and Wingfield		(AY858662)	(DQ097851)	(DQ097861)
G. piceaperda (Rumbold) Goid. Zipfel, Beer and Wingfield		(AY497514)	(AY707209)	(AY707195)
G. serpens (Goid.) Zipfel, Beer and Wingfield		(AY497516)	(DQ294394)	(AY707188)
*Leptographium abietinum (*Peck) Wingfield	DAOM 60343	EU984271	(DQ097852)	(AY263182)
L. fruticetum M. Alamouti, Kim and Breuil	DAOM 234390	EU984272	(DQ097848)	(DQ097855)
L. longiclavatum Lee, Kim and Breuil	DAOM 23419	EU984273	(AY816686)	(AY288934)
L. lundbergii Lagerberg and Melin	UAMH 9584	EU984274	(AY544603)	(AY263184)
L. terebrantis Barras and Perry	UAMH 9722	EU984275	(AY544606)	(AY263192)
Ophiostoma abietinum Marm. and Butin	CMW 1468	EU984276	(DQ294356)	EU977484
O. bicolor Davidson and Wells		(AY497512)	(DQ268604)	(DQ268635)
O. canum (Münch) Sydow and P. Sydow	AU 30 [CB]	EU984277	(AJ538342)	EU977485
O. floccosum Mathiesen-Käärik		(AF139810)	(AJ538343)	(AY789142)
O. ips (Rumbold) Nannfeldt		(AY172021)	(AY172022)	(AY789146)

Species	Source [a]	GenBank accession no. [b]		
		nSSU rDNA	nLSU rDNA	β-tubulin
O. montium (Rumbold) Hunt	CBS 151.78	EU984278	(AY194947)	(AY194963)
O. novo-ulmi Brasier	NAN-MH75 [CB]	N/A	(DQ294375)	EU977486
O. piceae (Münch) Sydow and P. Sydow		(AB007663)	(AJ538341)	(AY305698)
O. pulvinisporum Zhou and Wingfield	CMW 9020	N/A	(DQ294380)	EU977487
O. setosum Uzunovic, Seifert, Kim and Breuil		N/A	(AF128929)	(AY305703)
O. stenoceras (Robak) Nannfeldt	C80	(M85054)	(DQ294350)	EU977488
O. quercus (Georgev.) Nannfeldt		(AF234835)	(DQ294376)	(AY789157)
O. ulmi (Buisman) Nannfeldt	W9 [CB]	(M83261)	(DQ368627)	EU977489
Claviceps sp.		(U32401)	(U17402)	(AF263569)
Daldinia sp.		(U32402)	(U47828)	(AY951701)
Epichloe typhina (Pers.) Tul. and C. Tul.		(AB105953)	(U17396)	(X52616)
Microascus cirrosus Curzi	CBS 217.31	EU984279	(AF275539)	EU977490
Penicillium expansum Link		(DQ912698)	(AF003359)	(AY674400)
Petriella setifera (Schmidt) Curzi	CBS 385.87	EU984280	(DQ470969)	EU977491
Taphrina populina (Fr.) Fr.		(D14165)	(AF492053)	(AF170968)
Xylaria sp.		(U32417)	(AY327481)	(AY951763)

[a] Source of isolates sequenced in this study: ATCC, American Type Culture Collection, Manassas, USA; C, Iowa State University, Dept. of Plant Pathology, USA; CBS, Centraalbureau voor Schimmelcultures, Utrecht, the Netherlands; CMW, Culture Collection Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, South Africa; DAOM, Canadian Collection National Fungus Herbarium and Culture Collection, Ottawa, Canada; NISK, Norwegian Forest Research Institute, Austria; MUCL, (Agro) Industrial Fungi and Yeasts Collection, Belgium; UAMH, University of Alberta Microfungus Collection and Herbarium, Edmonton, Canada; UM, University of Manitoba, Dept. of Botany, Winnipeg, Canada; CB, Colette Breuil's Culture Collection, University of British Columbia, Canada.

[b] Accession numbers of sequences newly produced, updated (bold) or downloaded from GenBank (parentheses); N/A, not available.

Table 2.2 Morphological and ecological characters reported for the genera *Ambrosiella*, *Raffaelea* and *Dryadomyces* in the literature or in this study

Fungal species	Morphological characteristic		Ecological characteristic	Reference
	Conidiophore arrangement	Conidiogenesis	Bark beetle ex host tree (geographic origin)	
Ambrosiella sp. 1 (UAMH10632-10633)	Distinct sporodochia-like and solitary [a]	Annellidic & sympodial [a]	*Ips* spp., *Dryocoetes affaber, Polygraphus rufipennis* ex *Picea* spp. (Canada)	Beaulieu [e], Harrison [e], Massoumi Alamouti et al. 2007
Ambrosiella sp. 2 (UAMH10634-10635)	Distinct sporodochia-like and solitary [a]	Annellidic & sympodial [a]	*Ips* spp., *Dryocoetes affaber, Polygraphus rufipennis* ex *Picea* spp. (western Canada)	Massoumi Alamouti et al. 2007
Ambrosiella sp. 3 (NISK-94-166/39A, 1994-176-B4)	Distinct sporodochia-like and solitary (data not shown) [a]	Non-phialidic (data not shown) [a]	*Hylurgops palliatus, Polygraphus poligraphus* ex *Picea abies* (Norway)	Krokene & Solheim 1996
A. brunnea	Distinct sporodochia	Sympodial [b]	*Monarthrum* spp. ex *Acer* spp., *Quercus* spp. (western Canada, USA)	Batra (1967) [d], Funk 1965, Verrall 1943
A. ferruginea	Effused sporodochia	Phialidic	*Trypodendron* spp., *Xyloterus signatus* ex *Betula* sp., *Fagus sylvatica*, *Larix* spp., *Picea* spp., *Pinus* spp., *Populus* sp., *Quercus* sp. (Canada, Europe, USA)	Batra (1967), Gebhardt et al. (2005), Kuhnholz 2004, Mathiesen-Käärik 1953
[*A. gnathotrichi*] [f]	Indistinct sporodochia (fascicles in younger part of colonies)	Non-phialidic	*Gnathotrichus retusus* ex *Picea engelmannii, Pinus ponderosa* (Colorado)	Batra (1967), Gebhardt et al. (2005)
A. hartigii	Distinct sporodochia	Phialidic	*Xyleborus dispar, Xylosandrus germanus* ex *Malus sylvestris* (Asia, Europe, USA)	Batra (1967), Gebhardt et al. (2005), Hartig 1844, Kajimura & Hijii 1992
[*A. ips*]	Solitary (sometimes indistinct sporodochia-like in old cultures)	Sympodial [b]	*Ips* spp., *I. sexdentatus* ex *Pinus* spp. (western USA, Europe)	Batra (1967), Leach et al. (1934)
A. macrospora	Effused sporodochia and solitary	Sympodial [b]	*Ips acuminatus* ex *Pinus sylvestris, Pinus* spp. (Europe)	Batra (1967), Francke-Grosmann 1952
[*A. sulcati*]	Distinct sporodochia	Sympodial [b]	*Gnathotrichus retusus* ex *Pseudotsuga menziesii* (Canada)	Funk (1970)
A. sulphurea	Distinct sporodochia	Non-phialidic	*Xyleborinus saxesenii* ex *Populus* spp., *Quercus* spp. (USA, Germany)	Batra (1967), Gebhardt et al. (2005)

Fungal species	Morphological characteristic		Ecological characteristic	Reference
	Conidiophore arrangement	Conidiogenesis	Bark beetle ex host tree (geographic origin)	
A. tingens	Solitary or distinct sporodochia	Non-phialidic [a]	*Tomicus minor, T. piniperda, Ips sexdentatus,* ex *Pinus* spp. (Europe)	Batra (1967), Francke-Grosmann 1952, Mathiesen-Käärik 1953
A. xylebori	Confluent sporodochia (fascicle in younger part of colonies)	Phialidic	*Xylosandrus compactus, X. crassiusculus, Corthylus columbianus,* ex *Coffea canephora, Acer rubrum* and *Ulmus* sp. (Africa, Ceylon, India, eastern USA, Taiwan)	Batra (1967), Brader 1964, Gebhardt et al. (2005), von Arx & Hennebert (1965)
Dryadomyces amasae	Confluent sporodochia	Sympodial	*Amasa* spp. (Taiwan)	Gebhardt et al. (2005)
Raffaelea albimanens	Distinct sporodochia, solitary and indeterminate synnemata	Annellidic	*Platypus externedentatus* ex *Ficus sycamorus* (South Africa)	Gebhardt & Oberwinkler (2005), Scott & du Toit (1970),
R. ambrosiae	Distinct sporodochia, solitary or loose fascicles	Annellidic	*Platypus* spp. (i.e. *P. wilsonii, P. cylindrus*) ex *Quercus* spp. (British Columbia, England, USA)	Batra (1967), Gebhardt & Oberwinkler (2005), von Arx & Hennebert (1965)
R. arxii	Confluent sporodochia, solitary and indeterminate synnemata	Annellidic	*Xyleborus torquatus* ex *Cussonia umbellif* (South Africa)	Gebhardt & Oberwinkler (2005), Scott & du Toit (1970)
R. canadensis	Solitary and sporodochia	Sympodial [b]	*Platypus wilsonii* ex *Pseudotsuga menziesii* (British Columbia, Oregon)	Batra (1967)
R. lauricola	Solitary and sporodochia	Sympodial [b]	*Xyleborus glabratus* ex *Persea borbonia*, other members of *Lauraceae* (USA)	Fraedrich et al. 2008, Harrington et al. (2008)
R. montetyi	Solitary or fascicles in the beetle galleries	Annellidic	*Platypus cylindrus, Xyleborus monographus, X. dryographusex Quercus* spp. (Europe, Portugal)	Gebhardt et al. (2004), Morelet (1998)
R. santori	N/A [c]	Sympodial [b]	*Platypus* spp. (i.e. *P. sulcatus*) ex N/A (Argentine)	Guerrero (1996)
R. sulcati	Effused sporodochia (dense fascicles)	Sympodial [b]	*Gnathotrichus sulcatus* ex *Pseudotsuga menziesii* (Canada)	Funk (1970)
R. quercivora	Distinct sporodochia or solitary	Sympodial [b]	*Platypus quercivorus* ex *Quercus* spp. (Japan)	Kubono & Ito (2002)
R. tritirachium	Confluent sporodochia (fascicles)	Sympodial [b]	*Monarthrum mali* ex *Quercus* spp. (Pennsylvania)	Batra (1967)
Unidentified species (TR25)	N/A	N/A	*Trypodendron rufitarsus* ex *Pinus contorta* (western Canada)	Kuhnholz 2004

[a] Reported in this paper.

[b] Sympodial proliferations have been reported by earlier studies through light microscopy, and the refore the true mode of conidiogenesis in these fungi needs further investigation.

[c] N/A, no description available.

[d] References in which the publication year is being shown in parentheses are dealing with the conidiomatal types or conidiogenesis.

[e] Unpublished data from spruce-beetle-fungal survey in eastern Canada by Dr. Ken Harrison (Natural Resources Canada, Canadian Forest Service-Atlantic) and Marie-Eve Beaulieu in Centre d'étude de la forêt, Université Laval Research.

[f] Fungi of doubtful genus identity based on the multigene dataset are shown in brackets.

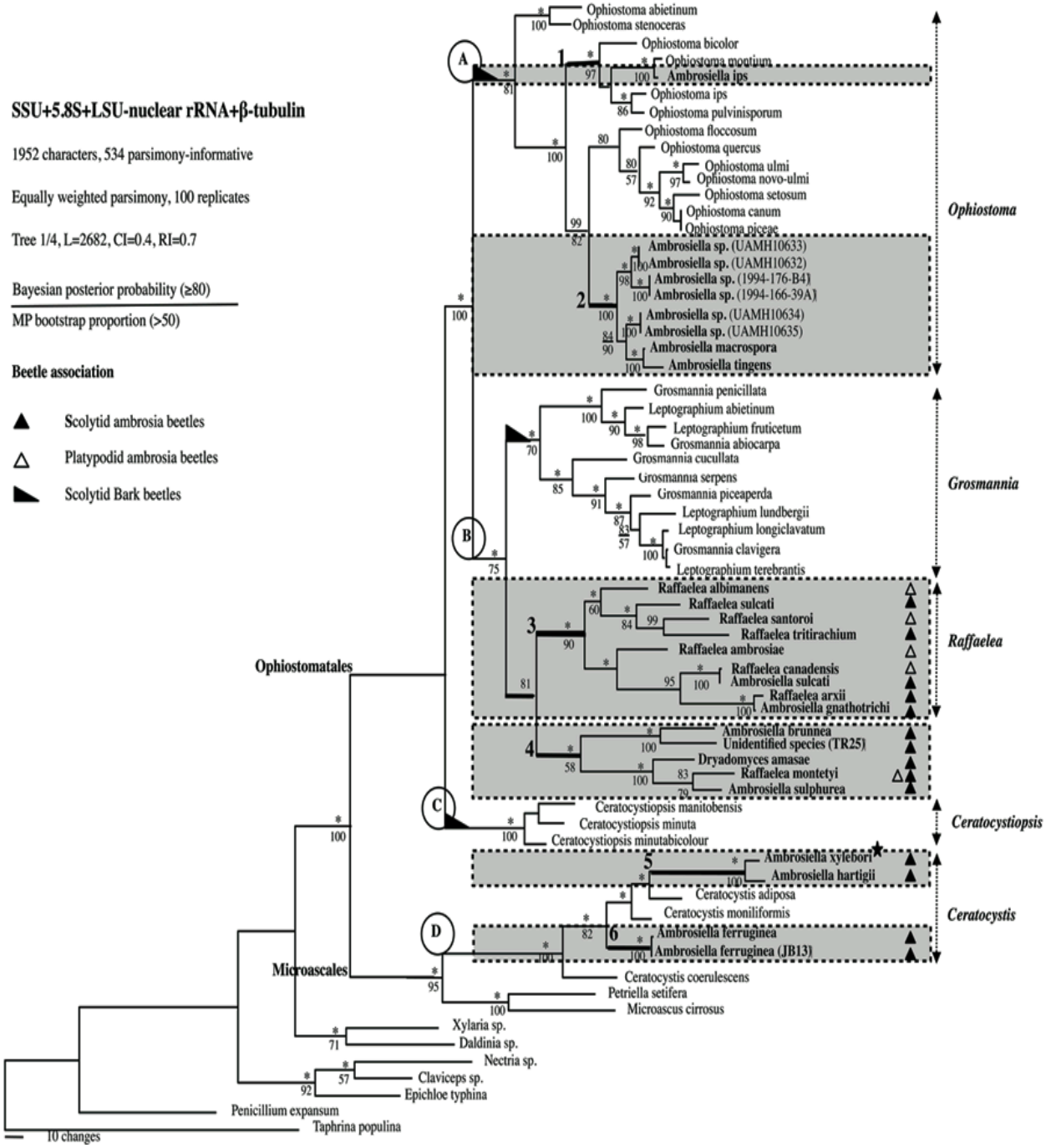

SSU+5.8S+LSU-nuclear rRNA+β-tubulin
1952 characters, 534 parsimony-informative
Equally weighted parsimony, 100 replicates
Tree 1/4, L=2682, CI=0.4, RI=0.7
Bayesian posterior probability (≥80)
MP bootstrap proportion (>50)
Beetle association
Scolytid ambrosia beetles
Platypodid ambrosia beetles
Scolytid Bark beetles
Ophiostoma abietinum
Ophiostoma stenoceras
Ophiostoma bicolor
Ophiostoma montium
Ambrosiella ips
Ophiostoma ips
Ophiostoma pulvinisporum
Ophiostoma floccosum
Ophiostoma quercus
Ophiostoma ulmi
Ophiostoma novo-ulmi
Ophiostoma setosum
Ophiostoma canum
Ophiostoma piceae
Ambrosiella sp. (UAMH10633)
Ambrosiella sp. (UAMH10632)
Ambrosiella sp. (1994-176-B4)
Ambrosiella sp. (1994-166-39A)
Ambrosiella sp. (UAMH10634)
Ambrosiella sp. (UAMH10635)
Ambrosiella macrospora
Ambrosiella tingens
Grosmannia penicillata
Leptographium abietinum
Leptographium fruticetum
Grosmannia abiocarpa
Grosmannia cucullata
Grosmannia serpens
Grosmannia piceaperda
Leptographium lundbergii
Leptographium longiclavatum
Grosmannia clavigera
Leptographium terebrantis
Raffaelea albimanens
Raffaelea sulcati
Raffaelea santoroi
Raffaelea tritirachium
Raffaelea ambrosiae
Raffaelea canadensis
Ambrosiella sulcati
Raffaelea arxii
Ambrosiella gnathotrichi
Ambrosiella brunnea
Unidentified species (TR25)
Dryadomyces amasae
Raffaelea montetyi
Ambrosiella sulphurea
Ceratocystiopsis manitobensis
Ceratocystiopsis minuta
Ceratocystiopsis minutabicolour
Ambrosiella xylebori
Ambrosiella hartigii
Ceratocystis adiposa
Ceratocystis moniliformis
Ambrosiella ferruginea
Ambrosiella ferruginea (JB13)
Ceratocystis coerulescens
Petriella setifera
Microascus cirrosus
Xylaria sp.
Daldinia sp.
Nectria sp.
Claviceps sp.
Epichloe typhina
Penicillium expansum
Taphrina populina
10 changes
A
B
C
D
1
2
3
4
5
6
Ophiostoma
Grosmannia
Raffaelea
Ceratocystiopsis
Ceratocystis
Ophiostomatales
Microascales

Figure 2.1 Multigene phylogeny of ophiostomatoid fungi

The tree is one of the nine MPTs of the combined nSSU+5.8S+nLSU-rDNA+β-tubulin dataset from *Ambrosiella*, *Raffaelea* and *Dryadomyces* taxa with the selected members of different ophiostomatoid teleomorph and anamorph genera including the *Grosmannia*. Bootstrap percentage values (≥ 50%) generated from 1000 replicates from maximum parsimony and posterior probabilities (≥ 80%) from Bayesian analysis are shown on the branches. Posterior probabilities of 100% are shown by *. Thickened black branches represent the major ambrosia clades of genera *Ambrosiella*, *Raffaelea* and *Dryadomyces* produced from the combined dataset. The taxa given in bold represent all the ambrosia fungi included in the analysis. The beetle association (ambrosia and bark beetles) is also mapped onto the phylogenetic tree. *Taphrina populina* and *Penicillium expansum* from orders *Taphrinales* and *Eurotiales*, respectively, were used as outgroup taxa to root the phylogenetic tree.

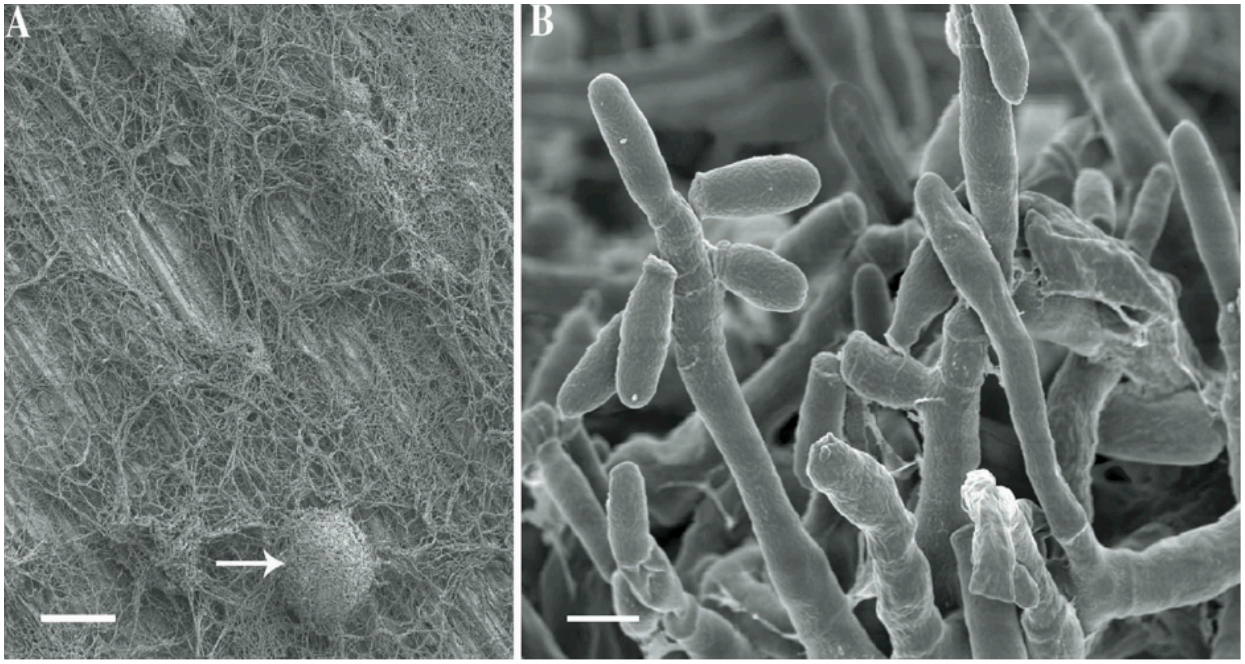


Figure 2.2 Sporodochium-like formation

Scanning electron micrograph of *Ophiostoma*-related *Ambrosiella* (UAMH 10635). (A) Low magnification of sporodochium-like structures developed from an interwoven mat of hyphae. (B) Close-up of the conidiophores and annellidic conidiogenous cells comprising a portion of the sporodochium-like structures indicated in figure 2.2A (arrow). Bar = (A) 300 µm, (B) 2.5 µm.

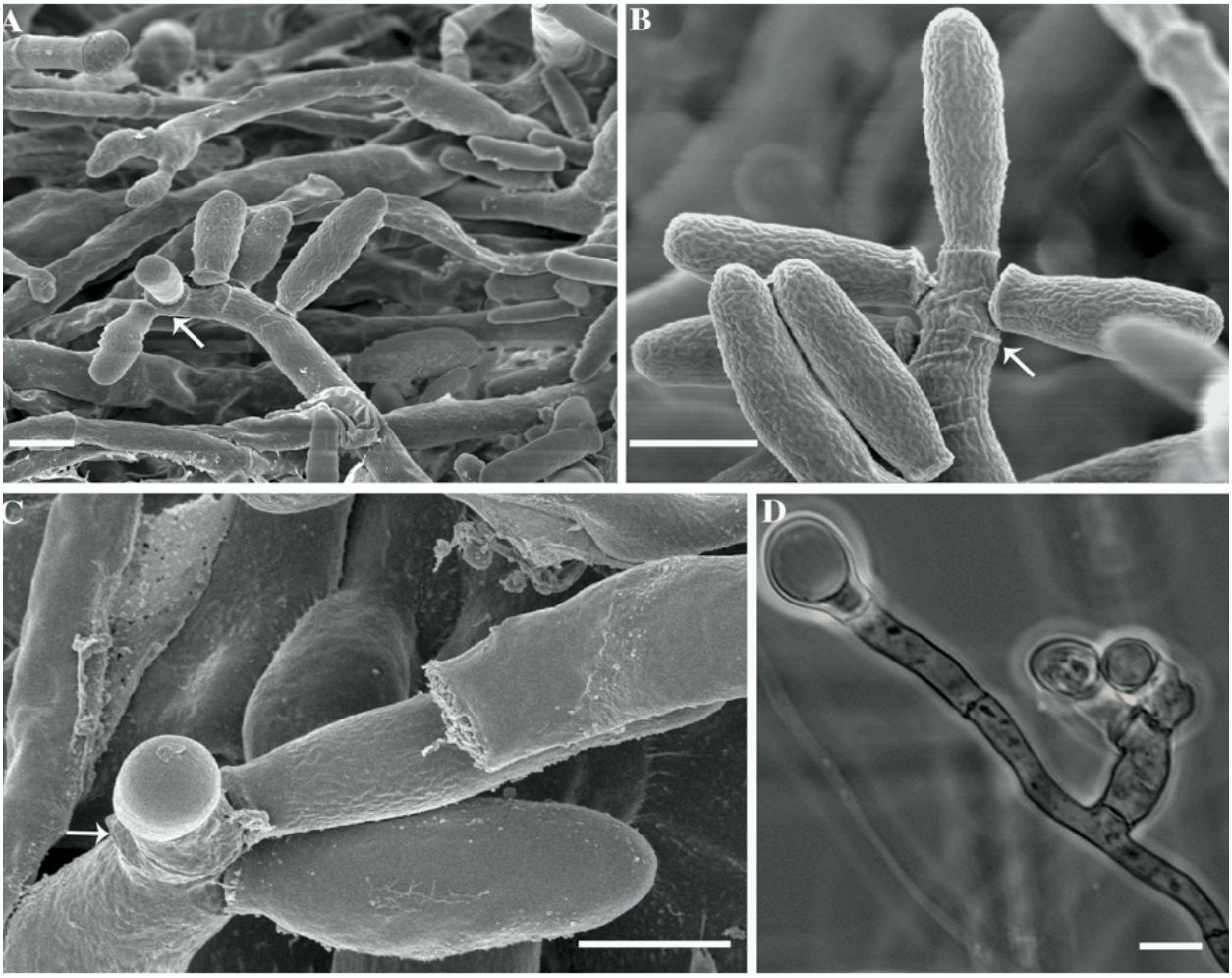


Figure 2.3 Annellidic conidiogenous cells

Scanning electron (A–C) and light micrographs (D) of *Ophiostoma*-related *Ambrosiella* spp. (A, B) Annellidic conidiogenous cells (arrowed) of *Ambrosiella* sp. 1 UAMH 10632 and *Ambrosiella* sp. 2 UAMH 10635, developing from mononematous conidiophores tapering toward the apex. (C) Conidiogenous cell of *Ambrosiella* sp. 2 UAMH 10635 showing apical conidia produced through sympodial proliferation (arrowed). (D) *A. tingens* CBS 366.53, producing conidia through the non-phialidic conidiogenesis. Bars = 2.5 µm.

Chapter 3 Gene genealogies reveal cryptic species and host preferences for the pine fungal pathogen *Grosmannia clavigera*

3.1 Introduction

Because of global trade, and environmental and climate changes, phytophagous insects and insect-vectored fungi that are pathogenic to trees have the potential to undergo rapid population expansion and cause substantial ecological changes (Anderson et al. 2004). A key aspect of estimating risks to ecosystems due to the spread of native or introduced pathogenic species involves defining species boundaries and genetic diversity. A growing number of fungal pathogens that were originally reported as dispersed generalists are now described as collections of populations or sister species adapted to new hosts or environments (Burnett 2003; Giraud et al. 2008). However, like many other organisms, it is difficult to define species boundaries in fungi. Molecular approaches, such as phylogenetic species recognition by genealogical concordance (Taylor et al. 2000; Dettman et al. 2003), can be more effective than traditional concepts. While it can be challenging to identify a genetic threshold that defines a species boundary, it is becoming increasingly practical to generate genomic sequence data for delimiting species with many independent gene genealogies (Knowles and Carstens 2007).

Native bark beetles and their fungal associates, which evolve within coniferous trees, are among the most damaging forest pests in North America (Harrington 2005).

The current *Dendroctonus ponderosae* (mountain pine beetle: MPB) outbreak is the largest epidemic in recorded history. It has affected more than 18 million hectares of *Pinus contorta* forest in western Canada (www.for.gov.bc.ca/hfp), leading to major impacts on ecosystem dynamics and associated economic (Kurz et al. 2008). MPB normally remains at low population levels within pine forests for many decades, but can rapidly erupt into large-scaled outbreaks, killing large areas of susceptible host trees (Raffa 1988). Climate change and large areas of susceptible host trees likely contribute to the epidemic expanding northward and into high-elevation pine forests, beyond the MPB's recorded historical range (Safranyik et al. 2010). Population studies of both beetles and fungal associates (i.e. *Grosmannia clavigera*) confirmed population expansion in the northern part of the beetle/fungal species range, where outbreak activity is currently increasing (Mock et al. 2007; Lee et al. 2007; Roe et al. 2011). Further, if conditions continue to be suitable for MPB in its current geographic range, there is a risk that the outbreak will expand eastward into the boreal forests via *P. banksiana* (Logan and Powell 2001; Bentz et al. 2010; Safranyik et al. 2010).

One of the most common characteristics of bark beetles is their association with the wood-colonizing filamentous ascomycetes grouped as ophiostomatoid fungi (Six and Wingfield 2011). *Grosmannia clavigera* is an ophiostomatoid tree pathogen that forms a symbiotic association with MPB (Robinson and Davidson 1968) and its sister species *D. jeffreyi* (jeffrey pine beetle: JPB; Six and Paine 1997). While MPB and JPB have only subtle phenotypic and genetic differences, they inhabit different host trees. JPB is highly specialized, infests only *P. jeffreyi*, and has no history of large-scaled epidemics,

whereas MPB inhabits its primary host *P. contorta* and 22 other pine species, but not *P. jeffreyi* (Wood 1982; Safranyik et al. 2010). These bark beetles carry similar mycoflora and their geographic distributions overlap in some regions of the USA (Wood 1982; Six and Paine 1997; Kelley and Farrell 1998). *G. clavigera* is one of the most important fungal associates of MPB and JPB (Robinson-Jeffrey and Davidson 1968; Six and Paine 1997; Lee et al. 2006a; Rice, Markus N. Thormann, et al. 2007), and a central component of current MPB epidemics. Vectored fungi benefit from the association because the beetles carry them through the tree bark into a new host's tissues (Six and Wingfield 2011). The benefits reported for the beetle and its progeny include the fungi providing a suitable substrate for brood development, participating in weakening tree defenses, and serving as a source of nutrients (Raffa and Berryman 1983; Harrington 2005; Bleiker and Six 2009; Lieutier et al. 2009; DiGuistini et al. 2011). While both beetle and fungi contribute to tree death, *G. clavigera* can kill trees without the beetle when manually inoculated into the host at a certain density (Solheim and Krokene 1998; Lee et al. 2006b). Mountain pine beetles have also been reported to attack other coniferous genera in epidemic regions; however, the beetles were only colonizing and reproducing inside *Picea* trees that have been baited with pheromone (Huber et al. 2009; Safranyik et al. 2010).

While species identification is important to understand the ecology and biology of organisms, boundaries between closely related species often lack clear limits and diagnostic characteristics. In *G. clavigera,* the sexual state (teleomorph) has been rarely found in nature and is not produced in the laboratory (Lee et al. 2003). Therefore, the

fungus is predominantly haploid through its life cycle and is known to mainly reproduce asexually (Lee et al. 2007; Six and Paine 1999). Because teleomorphs are rare, morphological identification relies on asexual structures, which occur in a variety of forms including the anamorph *Leptographium* (Jacobs and Wingfield 2001; Six et al. 2003). Conidiophores and conidia in *G. clavigera* show a great variation of shapes and sizes that can become confusingly indistinguishable from those of other *Leptographium*-forming species (Tsuneda and Hiratsuka 1984; Six et al. 2003). Early molecular studies have not been successful in separating *G. clavigera* from morphologically similar species, due to the lack of diagnostic DNA substitutions in the loci that have been commonly used for systematics (Zambino and Harrington 1992; Six et al. 2003). Multigene phylogenies, using ribosomal DNA and housekeeping genes (e.g. actin, elongation factor 1, alpha and beta tubulin), have provided better resolution for separating *G. clavigera* from its most closely related morphological species (e.g. MPB-associate *L. longiclavatum*), with the exception of a generalist fungus *L. terebrantis* (Six et al. 2003; Lim et al. 2004; Roe et al. 2010).

Based on morphological and phylogenetic species criteria, *G. clavigera* has been found colonizing different species of pine infested by MPBs and JPBs across western North America (Six et al. 2003; Lim et al. 2004). Using AFLP and microsatelitte makers, as well as multilocus sequencing, *G. clavigera*'s population structure has been mainly studied for those associated with the MPB epidemics in BC, Alberta and the USA (Lee et al. 2007; Roe et al. 2010; Tsui et al. 2012). All these studies have shown low genetic divergence and low nucleotide diversity between and within different epidemic

populations with a northern to southern pattern of differentiation. AFLP markers have also suggested the presence of two genetically distinct groups within *G. clavigera* associated with MPB epidemics infesting *P. contorta* in the Rocky Mountain of Alberta (i.e. Banff) and the northern USA (Lee et al. 2007); however, these groups have not been supported by microsatellite markers (Tsui et al. 2012). Similarly, using multilocus sequencing, Roe et al. (2011) reported no evidence of the two AFLP groups; but in this work the sampling regions focused on BC and Alberta epidemics and did not include the region where the second AFLP group had been reported.

Beetle-associated plant pathogens like *G. clavigera* depend on beetle vectors and host trees to complete their life cycles (Harrington 2005). Therefore, detecting genetic isolation in relation to the degree of host specialization or evolution of symbiosis is relevant to this group of fungi. As they grow, *G. clavigera* develop tight ecological and/or biological association with different species of pines, as well as with the two distinct sibling beetle species MPB and JPB. Six and Paine (1998) suggested that *G. clavigera* isolated from MPB or JPB had different tolerance to host defense metabolites. These beetles segregate in different ecological niches; as such we hypothesize that this segregation may have resulted in genetic divergence of their fungal associate *G. clavigera* producing a complex of distinct phylogenetic species adapted to different host tree species.

Defining species boundaries is essential for understanding the potential adaptive variations and the ecological and/or biological traits that may impact the evolution of

beetle-associated fungi. Theoretical models that incorporate adaptation and divergence among pathogens are applicable to risk assessment and to developing control measures, and detailed genetic information on evolving species should improve such models (Giraud et al. 2010). Currently, information on genetic variation in *G. clavigera* is limited to few protein coding genes and non-coding markers (Six and Paine 1999; Lee et al. 2007; Tsui et al. 2009; Roe et al. 2011). Here, we screened nucleotide polymorphisms in 67 loci and applied "phylogenetic species recognition by genealogical concordance" (Taylor et al. 2000; Dettman et al. 2003) using a subset of 15 protein-coding genes to assess whether genetically isolated lineages occur within *G. clavigera*, and whether host beetle and/or tree specialization may influence the evolution of these fungi. We combined the sequence data from the 15 loci to clarify how the phylogenetic species were related to each other. We show evidence of recombination in these apparently asexual fungi. Finally, we conclude that *G. clavigera* consists of Gc and Gs lineages, and discuss the ecology and biology of these fungi.

3.2 Materials and methods

3.2.1 Samples

We examined 166 isolates of *G. clavigera* and eight additional isolates of its four closely related species *G. aurea, L. longiclavatum*, *L. terebrantis* and *L. wingfieldii* (Table 3.1 and Appendix A.1). The *G. clavigera* isolates were collected from the two

beetle associates MPB and JPB and their host trees at different sites in Canada and the USA. In California, where both MPB and JPB are present (i.e. sympatric regions), we collected 30 and 25 isolates from *P. jeffreyi* and *P. contorta*, respectively, as well as a few isolates from *P. ponderosa*. We also included *G. clavigera* from locations where only MPB is present (i.e. allopatric regions). Sixty-seven isolates were from *P. contorta* in Canada and the USA, 29 isolates from *P. ponderosae* in South Dakota and British Columbia (BC), and a limited number of isolates from other MPB-host species, including *P. albicaulis* in BC and *P. strobiformis* in Arizona. Our fungal collection provides a comprehensive coverage of the beetles' geographic distribution. This included samples from current MPB epidemics in Canada, Idaho and Montana as well as from previous outbreaks in the 1960s and 1980s. It also included samples from small, geographically isolated outbreak populations in South Dakota, California and Arizona. In the work described here, we refer to such populations as 'localized'. Figure 3.2a shows the sampling locations.

3.2.2 Isolation

Fungal isolations from beetle exoskeletons or mycangia, or from galleries in infested trees were carried out following the methods described by Six and Paine (1997) and Massoumi Alamouti et al. (2007). Identification and molecular analyses were done from single-spore isolates. All cultures are maintained at the Breuil culture collection (University of British Columbia, Canada). Morphological features were determined from colonies grown on 2% MEA (20 g Difco malt extract, 10 g Difco agar and 1 L distilled

water) or from fungi taken from beetle galleries. The *G. clavigera* reproductive structures were examined and compared to those described by Robinson and Davidson (1968), using light microscopy.

3.2.3 Polymorphism detection

We identified polymorphic loci using two approaches. The first approach involved sequencing 28 candidate genes that were available from the *G. clavigera* genome sequence and EST-supported gene predictions (DiGuistini et al. 2007; 2009; 2011), followed by polymorphism discovery and verification. The second approach involved sequencing 39 putative polymorphic loci identified from an expressed sequence tag (EST) library obtained by pooling mRNA from eight *G. clavigera* isolates (DiGuistini et al. 2009) that were characterized as distinct haplotypes (i.e. unique sequence type) using the first approach. The target loci were identified from the genomic resource using CLCbio Genomics Workbench (CLC) 3.7.1 (Aarhus, Denmark). To discover polymorphisms, we sequenced the 67 loci across nine *G. clavigera* isolates (Table 3.1) chosen from distinct ecological and geographical sources. The sequences were aligned and analyzed for polymorphisms using CLC. We validated the novel polymorphisms in 15 genes (Table 3.2) selected for further characterization in an additional 53 *G. clavigera* isolates and eight isolates of four closely related species (Table 3.1). Genomic locations and gene descriptions of the 67 *G. clavigera* loci screened for polymorphisms are listed in Appendix A.2 and concatenated alignment of these datasets are deposited in TreeBASE (TB2: www.treebase.org).

3.2.4 DNA extraction, primer design and sequencing

We followed DNA extraction method by Möller (et al. 1992) for mycelia grown on 2% MEA (33 g Oxoid malt extract agar, 10 g Technical agar No.3, and 1 L distilled water) plates overlaid with cellophane (gel dry grade, BioRad). Primer pairs were designed with optimal melting temperatures of 58–62°C, using CLC (Appendix A.2). PCR amplifications were performed following standard methods (Lim et al. 2004). Amplicons were purified and sequenced at the Sequencing and Genotyping Platform, CHUL Research Center (Québec, Canada). Sequence data were collected from one strand, except for new haplotypes, which were all confirmed by sequencing both strands. All sequences are available at GenBank (accession nos. HQ633073–HQ634118).

3.2.5 Sequence alignments and analyses

Sequences were edited and aligned using Geneious 5.1 (Biomatters Ltd, New Zealand). Coding, intronic and untranslated (UTR) regions were determined based on alignment of DNA sequences to the *G. clavigera* genome sequence and gene prediction models. Genetic diversity indices and divergence analyses were assessed using DnaSP 5.10 (Librado and Rozas 2009). Net nucleotide divergence (Dxy) (Nei 1987) was

calculated with the Tamura-Nei gamma correction model using Mega 4.0 (Tamura et al. 2007).

3.2.6 Gene trees and concatenated data phylogeny

Phylogenetic analyses were conducted using maximum parsimony (MP) and Bayesian inference of each of the 15 genes, as well as the combined dataset of these genes. The best-fit model of sequence evolution for each gene was determined using the Akaike information criterion (AIC) implemented in JModelTest 0.1.1 (Posada 2008). MP trees were identified using PAUP* 4.0b10 (Swofford 2003) by heuristic searches and 100 random sequence additions. Gaps were treated as missing data and no weighting was introduced in single gene analysis. Confidence was examined using bootstrapping (BS) with 1000 replicates and the heuristic option (Felsenstein 1985). Bayesian analyses were run using MrBayes 3.2 (Ronquist and Huelsenbeck 2003), under the best-fit substitution model. Each run consisted of four incrementally heated Markov chains, with default heating values. The chains were initiated from a random tree, and were run for 2 million generations with sampling every 1000 generations. Posterior probabilities (PP) were inferred with a 50% majority-rule consensus tree sampled after the likelihood scores had converged. The 15 nuclear genes were concatenated to conduct partitioned maximum likelihood (ML) analysis (with 1000 nonparametric replicates bootstrap) using RAxML-VI-HPC 7.0.4 (Stamatakis 2006) and partitioned Bayesian analysis. The partitioned-ML and -Bayesian analysis utilized the substitution models selected by the AIC in JModelTest for each gene locus. The

combined dataset was also analyzed with weighted parsimony, with the weighting inversely proportional to the number of parsimony informative characters at each locus. Weighting allowed each locus to contribute equally to the combined data tree. All phylogenetic trees were rooted using *G. aurea* as outgroup (Massoumi Alamouti et al. 2007). Monophylies supported by both BS ≥ 70% and PP ≥ 95% were considered as significant.

Constraints on topologies were applied in PAUP* and the Wilcoxon signed-rank (WSR; Templeton 1983) test was employed to assess significant differences among topologies. For this test, up to 100 MPTs recovered were used as constraint topologies. When testing the constraint of lineage-specific monophyly, the lack of significance in the WSR tests indicates that nonmonophyly could be the result of insufficient phylogenetic signal.

3.2.7 Network approaches and evidence for recombination in G. clavigera

For each of the 15 gene datasets, we generated parsimony networks of *G. clavigera* haplotypes, which is described in the supporting information Method S1. We applied three approaches to detect the presence or absence of recombination in *G. clavigera*. First, we applied the index of association (I_A) to estimate the extent of clonality in *G. clavigera*, using the program Multilocus 1.3b (Agapow and Burt 2001). I_A determines to what extent individuals that are the same at one locus are more likely

than random to be the same at other loci. We used 10,000 randomizations on the subset of polymorphic sites that showed the most balanced distribution of alleles in each gene (i.e. excluding the uninformative sites). The test assumes an infinite amount of recombination so significant departure ($p<0.05$) from simulated recombined datasets suggest the presence of clonality (Smith et al. 1993). Second, we used the NeighborNet algorithm (Bryant and Moulton 2004) for decomposition analysis with SplitsTree 4.10 (Huson and Bryant 2006) to visualize the incongruence generated by recombination from the pairwise distance matrix of the *G. clavigera* concatenated sequence dataset estimated under the GTR model. Third, we estimated the pairwise homoplasy index (PHI; Bruen et al. 2006) in SplitsTree. Using a 100 bp window, compatibility among sites was calculated and, assuming no recombination, significance was determined with a permutation test.

3.3 Results

3.3.1 Polymorphism discovery

Sixty-seven loci, which represented 50 nuclear protein-coding genes with predicted functions, were sequenced and screened for polymorphisms (Appendix A.2). Some genes were constitutively expressed, e.g. housekeeping genes; others were differentially expressed in specific growth conditions, and were potentially involved in growth, metabolic processes or host tree pathogenicity (DiGuistini et al. 2007; 2009;

2011). A number of genes lacked significant homology with proteins or domains of known functions. We obtained ~50 kb of high quality sequence data for these genes in nine *G. clavigera* isolates (TB2:S11053) and identified 128 polymorphic sites (i.e. substitutions) across 33 genes. The majority of variations (63%), i.e. 81 single nucleotide polymorphisms (SNP) in 31 genes, separated the seven isolates representing the MPB associates at the epidemic sites from the two other isolates: *G. clavigera* holotype (ATCC 18086) and JPB associate (DLS1575). A subset of 18 informative (i.e. shared by two or more isolates) SNP (14%) in 12 genes, were exclusive polymorphisms that segregated only within the seven epidemic isolates. The rest of polymorphisms were substitutions that were unique to one isolate (i.e. singletons).

3.3.2 Polymorphism validation

For further analysis, we selected 15 genes (Table 3.2) that showed different levels and patterns of variation in the polymorphism-discovery panel and sequenced them in an additional 53 *G. clavigera* isolates (Table 3.1). These isolates were selected to represent the beetle associates MPB and JPB, and their respective primary host trees *P. contorta* and *P. jeffreyi,* as well as a few other MPB pine host species. Within *G. clavigera* isolates, we identified a total of 86/13,198 (0.65%) base substitutions and three indels in the concatenated 15-gene dataset. No site had more than two alleles (biallelic). The most polymorphic gene region was CFEM-II (π= 0.0039) and the least polymorphic was alpha-tubulin (π= 0.00073). Of the 86 polymorphic sites, 68 were informative and 18 were singletons. Eighteen of the changes were predicted in

noncoding locations (i.e. intronic and UTR), and, for the 68 that were in coding regions, 33 were synonymous and 35 were non-synonymous. The vast majority of variants were identified either as fixed SNPs (n=33) or as exclusive polymorphisms (n=49) that segregate only within one of the two potential *G. clavigera* lineages. The remaining four SNPs and one indels were the only shared polymorphisms found. The number of sites in the different classes of polymorphisms for each gene as well as for the concatenated dataset is shown in Tables 3.3 and Appendix A.3.

3.3.3 Single-gene phylogenies, phylogenetic species recognition and concatenated phylogeny

Using 15 gene phylogenies, we assessed *G. clavigera* species boundaries and phylogenetic relationships with related pine-infesting fungi: *G. aurea, L. longiclavatum*, *L. terebrantis* and *L. wingfieldii*. A summary of the phylogenetic data and model parameters inferred for each locus and the combined dataset are presented in Table 3.4. The target genes were amplified in all species, except for TRPG and MPEP in the outgroup taxon *G. aurea* and anonymous-I in *L. wingfieldii*. MP and Bayesian consensus trees inferred similar topologies that are only shown for MP trees (figure 3.1). MP analyses yielded one to five trees for each locus, which mainly differed in the branching orders of two close relatives *L. terebrantis* and *L. wingfieldii*. The majority of gene trees (10/15) resolved the pathogen *G. clavigera* into two distinct clades. We referred to these clades as Gs with 40 isolates and Gc with 22 isolates (figure 3.1). The Gs clade contained all isolates from epidemic MPBs, as well as those from localized

populations except for those collected from *P. ponderosa* trees. This clade was significantly (BS ≥ 70% and PP ≥ 0.95) supported by five loci (TRPG, MPEP, P450-I, LAH, anonymous-II). The Gc clade encompassed all JPB associates, as well as isolates from MPB that were infesting *P. ponderosa* trees in sympatric regions (California) including *P. jeffreyi* and *P. contorta,* and in allopatric regions of South Dakota. The *G. clavigera* holotype (ATCC18086; Robinson and Davidson, 1968) was also placed within Gc. This clade was significantly supported by the same subset of loci that supported the Gs clade. Clades in gene trees that did not agree with this partitioning were either not fully resolved (figure 3.1 40SRP, P450-II) and/or not significantly supported (figure 3.1 CFEM-II: PP ≤ 0.95 and/or BS ≤ 70%). While one additional group showed a high level of support (figure 3.1: BS=100% and PP=1.0) in the TRPG and another in the CFEM-I (figure 3.1: BS=75% and PP=0.95) phylogenies, we considered neither clade to be an independent lineage, since their partitions contradicted each other and neither was supported in the concatenated phylogeny (figure 3.2b). In the constraint analysis forcing the monophyly of Gs and Gc, WSR results were significant (p=0.04) for only CFEM-II, indicating that incongruence from the constraint phylogeny is only significant in 1/15 of the loci.

Of the 15 genes, TRPG, MPEP and anonymous-II showed the highest resolving power for species boundaries, supporting five monophyletic groups: Gs, Gc, *L. terebrantis*, *L. wingfieldii* and *L. longiclavatum*. While species-level clades were strongly supported by a number of single-gene phylogenies, relationships between species were difficult to resolve. For example, *L. terebrantis* showed a non-robust phylogenetic

placement among trees, and it was collapsed into polytomy in at least ten single-gene phylogenies. Although positioning of some ingroup taxa varied among gene trees, TRPG, MPEP, LAH and anonymous-II significantly supported a sister group relationship between Gs and Gc.

The concatenated matrix of 15 gene sequences resulted in 13,239 bp of aligned nucleotide positions, 402 variable sites, and 226 informative characters (Table 3.4). MP, partitioned ML and partitioned Bayesian analyses resulted in similar topologies that had only minor differences in the placement of terminal taxa (figure 3.2b, ML tree). The topology of concatenated phylogeny was consistent with the single-gene tree partitions resolving the *G. clavigera* isolates into two monophyletic clades, and with the sister-group relationship between Gs and Gc (MP and ML BS=100%, PP=1.0).

Finally, we challenged our phylogenetic results by testing whether the polymorphism distribution of *G. clavigera* into two groups was due to independent evolutionary histories or to random sorting of genetic variations. The probability of observing different groups that, by chance, do not share polymorphisms, were tested by random shuffling the 15 dataset across (nonpartitioned dataset) and within (partitioned dataset) the two phylogenetic species. For the randomization, the association of polymorphic sites within each gene was left intact (i.e. each gene was randomized as blocks). In 1000 such randomizations, we found no partition that would create groups with no shared polymorphisms. Among 62 *G. clavigera* isolates, the shortest trees acquired from the nonpartitioned, randomized dataset were significantly ($p < 0.001$)

longer (510–570 steps) than trees obtained from the randomized dataset considering the Gs and Gc partitions (61–185 steps).

Within Gs (n=40), we identified 36 distinct haplotypes that were characterized by 33 base substitutions across 12 polymorphic genes (See appendices A.3 and A.5). Of the 33/13,198 (0.25%) polymorphic sites, 23 were informative. The number of haplotypes ranged from 1 to 4 among the genes. Gene/haplotype diversity (H) ranged from 0.0 to 0.73 in CFEM-II. The diversity over the combined dataset showed a high value of 0.99; however, genetic differentiation within the isolates was low, resulting in minor nucleotide diversity (π=0.00068). Similar haplotypes did not cluster based on either geographic locations or the pine host species (See appendix A.5). Within Gs, we found seven isolates representing three identical haplotypes. Two isolates with the same haplotype were from the same California locality and two isolates with the same haplotype were from Alberta. Identical haplotypes were also isolated from different localities, two from BC and one from Arizona. The probability of identical haplotypes (i.e. isolates sharing the same sequence type at all 12 polymorphic loci) resulting from random mating and recombination was small ($4.2 \times 10^{-3} - 1.8 \times 10^{-6}$), suggesting that they represent epidemic clones from the asexual reproductions. In comparison, the Gc isolates (n=22) showed a similar pattern but with a slightly lower level of nucleotide diversity (Appendix A.3). They represented a collection of 22 unique haplotypes (i.e. H=1.0) that could be distinguished by a total of 24 base substitutions across 12 polymorphic genes. Of the 24/13,198 (0.18%) polymorphic sites, 14 were informative. As was the case for the Gs group, CFEM-II showed the highest level of both haplotype and nucleotide diversity.

However, some genes that showed a high level of variation within Gs isolates (e.g. TRPG, LPL, PLT) showed almost no polymorphisms among Gc isolates. Based on these genes dataset, haplotypes did not correlate with the host beetle/tree species, except for one allele in MPEP that was only found for MPB*/P. ponderosa* associates; however, this partition was not statistically supported.

3.3.4 Evidence of recombination

For Gs, we evaluated I_A for all isolates, as well as for the reduced-by-haplotype dataset in which we excluded identical haplotypes. When all isolates were included (n=40), the I_A=1.6 was significantly different (p=0.007) from the values obtained for the simulated recombined dataset, leading us to reject the null hypothesis of recombination. However, the I_A=1.3 for the unique haplotypes (n=36) was indistinguishable (p=0.09) from that expected for a recombinant population (figure 3.3b–Gs). Within Gc, the I_A=0.5 (p=0.3) also suggested recombination, both when all 22 isolates were included or only those from JPB (figure 3.3b–Gc). Split decomposition analysis also provided evidence for network relationships, giving a graphical support for the evidence of recombination and/or lineage sorting within both Gs and Gc (figure 3.3a). Finally, PHI provided another significant evidence (p= 0.00006) of recombination history.

3.3.5 Ecological and morphological characteristics

To assess the host and distribution ranges of the *G. clavigera* lineages in more detail, we sequenced a single informative locus, PCAS (Table 3.2), in an additional 104 isolates (Appendix A.1, figure 3.2a). Locus PCAS contains two fixed SNPs (2/685=0.29%) that differentiate the two *G. clavigera* lineages and possesses exclusive polymorphisms that are not shared between Gs and Gc. This locus has been tested against a large number of other closely related species, and has been used as target-specific PCR-primers to detect and differentiate microbial communities associated with the MPB (Khadempour et al. 2010). We generated the data for *P. contorta*-associated isolates (n=67) from BC, Alberta, Montana and Idaho, as well as isolates from *P. ponderosa* trees in BC (n=13). Consistent with results from 15-gene phylogenies, the deeper single-locus sampling showed that the Gc group was not present in the epidemic populations of MPB; instead, the Gc lineage largely represented isolates from *P. ponderosa* (n=18) and *P. jeffreyi* (n=30) trees attacked by the localized populations of respective beetle associates MPB and JPB in South Dakota and California. In contrast, Gs (n=117) occurred on MPB in epidemic populations of the beetle and its pine-host species in western Canada and the USA, as well as in localized populations infesting *P. contorta* in California and *P. strobiformis* in Arizona. Both Gs and Gc were found in MPB and JPB localized populations in California, where the two beetle associates live in sympatry on *P. contorta, P. jeffreyi* and *P. ponderosa*. In South Dakota, where the localized population of MPB infests *P. ponderosa* trees, we found Gc (n=15) but no evidence of Gs.

We compared the reproductive structures of five isolates representing the Gs group to those of the *G. clavigera* holotype, which was included in our analysis and represented the Gc group. The anamorph (conidia and conidiophore) and teleomorph (i.e. ascocarp and ascospores) morphologies of Gs (figure 3.4) representatives agreed with the formal *G. clavigera* holotype description by Robinson and Davidson (1968). The conidiophores and conidia sizes varied among isolates, but all measurements (Appendix A.4) agreed with the *G. clavigera* original descriptions, as well as with descriptions of *P. contorta* associates (Robinson and Davidson 1968; Six and Paine 1997; Lee et al. 2003).

3.4 Discussion

We generated the first comprehensive dataset of protein-coding gene variability in the bark-beetle symbiont and pine pathogen *G. clavigera*. We used this dataset to characterize patterns of DNA polymorphism and divergence within the pathogen and among four close relatives that also inhabit pine trees. In contrast to the current taxonomy, our results show species diversity and ecological complexity with respect to host species. Paine and Hanlon (1994) and Six and Paine (1998) showed that the *G. clavigera* isolates of JPB were more tolerant to host oleoresin than those of MPB, suggesting some potential physiological differences between these two types of isolates. Here we suggest that the genetic divergence and diversity in *G. clavigera*

isolates may have resulted from the fungus adapting to particular pine species and to extensive expansion of the epidemic.

Our phylogenetic analyses identified two distinct lineages in *G. clavigera*. While the combined dataset of nuclear ribosomal DNA and the protein-coding genes have improved the phylogenetic positioning of *G. clavigera* (Lim et al. 2004; Roe et al. 2010), these loci failed to distinguish the two lineages identified in this study. These results indicate that the sequences currently available for ophiostomatoid systematics are inadequate for phylogenetic species recognition and inferring evolutionary relationships in the genus *Grosmannia.* We demonstrated that sequencing more genomic regions is more effective for inferring species boundaries. Given this, care should be taken when interpreting ecological characteristics of this group of fungi. The literature suggests speculative evolutionary processes (Six et al. 2003; Lim et al. 2004; Roe et al. 2011) that rely on data that are insufficient for identifying species and on an imperfectly known phylogeny. Six et al. (2003) and Lim et al. (2004) suggested that *G. clavigera* is a recently diverged morphological variant of the generalist fungus *L. terebrantis*. Our results show that *L. terebrantis* is a distinct species separated from both *G. clavigera* lineages; we also found that some isolates assigned as *L. terebrantis* were genetically different from the *L. terebrantis* holotype isolated from *D. terebrans* (Six and Massoumi Alamouti unpublished data), suggesting that this fungus represents a complex of closely related species that need to be taxonomically and ecologically re-assessed. Ecological descriptions of this species and other close relatives will be more valuable if based on solid taxonomic foundations.

Below, we provide two main lines of evidence to show that *G. clavigera* lineages represent two distinct species: a) they are evolutionary independent, and b) they are ecologically distinguishable. Because a lineage can represent a species, a clone, or a divergent group within a population, we will discuss these two concepts and discuss the evidence of recombination and ecological significance in each species. Current concepts agree that species correspond to “segments of separately evolving lineages” (de Queiroz 2007); however, different characteristics (e.g. morphological, reproductive and nucleotide divergence) are used to infer boundaries for species, clones, and divergent groups. Such characteristics do not arise at the same time during the process of speciation, and so each type of evidence can lead to different conclusions regarding species boundaries (Avise 2004).

3.4.1 Evolutionarily independent lineages

Phylogenetic species recognition by genealogical concordance (Taylor et al. 2000; Dettman et al. 2003) stipulates that when lineages are separated for long periods of time relative to population size, genealogies from the majority of loci should be congruent. This criterion considers a clade to be an independent evolutionary lineage and a phylogenetic species if it is present in the majority of single locus phylogenies (Dettman et al. 2003). Here, the concordance of ten genealogies define *G. clavigera*

lineages as two sibling phylogenetic species, and suggest genetic isolation—even when the lineages occur in the same geographic region, as in California.

Enforcing topological constraints for the monophyly of Gs and Gc did not result in significantly worse fit of the data to the tree (compared with the fit to an unconstrained tree) for 14 of the genomic regions. This means that for three regions, lack of reciprocal monophyly and/or lack of nodal support were the result of insufficient phylogenetic signal.

In general, the pattern of gene genealogies and the level of polymorphism depend on the timing of the speciation event, historical population sizes, modes of reproduction, extent of hybridization and natural selection (Avise 2004). For only one (CFEM-II) of the 15 regions, the fit of the data to the constrained tree was significantly worse (at the α = 0.05 level). For CFEM-II, we found no evidence for significant departure from neutrality (data not shown), intragenic recombination, or paralogs in the *G. clavigera* genome's predicted gene models (DiGuistini et al. 2009; 2011). Therefore, none of these mechanisms can explain the incongruent pattern. Introgression can occur when interspecific hybridization results in the transfer of genetic material from one species into another, which leads to paraphyly of recipient species; alternatively, incomplete lineage sorting or recombination before species divergence can result in incongruent genealogies if species divergence occurred too recently for ancestral polymorphisms to have sorted into reciprocal monophyly (Avise 2004).

Distinguishing between interspecific hybridization and lineages sorting is difficult, because both result in the same pattern of incongruence (Hey and Nielsen 2004). However, when population genomic datasets are available, one could expect to find an alternative topology of a set of independent gene trees more frequently for hybridization event than other equally possible topologies under stochastic nature of lineage sorting (Huson et al. 2005). While we could not estimate the divergence time of *G. clavigera* lineages with certainty, due to the lack of fossils and the great variance in fungal nucleotide substitution rates (Kasuga et al. 2002), two observations suggest that these fungi diverged recently. First, the low interspecific nucleotide divergence (0.0037± 5.7 x 10–4) and the unresolved species phylogeny suggest that not all loci have reached reciprocal monophyly. Second, when we compared ingroup and outgroup taxa of two or four species, a large number of ancestral polymorphisms appeared to predate divergence, consistent with the speciation event being so recent that ancestral polymorphisms were retained.

3.4.2 Evidence of recombination

Because classical phylogenetic trees can give only a snapshot of the actual complex relationships that can be encountered when intraspecific details are considered, we describe *G. clavigera* population structure with modified phylogenies using split decomposition analysis. In this, network relationships account for

recombination within both Gs and Gc that, in agreement with our gene phylogenies, are separated into two phylogenetic groups. Within each group, I_A values were not significantly different from artificially recombined datasets, and the occurrence of unique sequence types suggested a history of recombination within each phylogenetic species that had created different combinations of alleles. While these results can also be explained by convergent or parallel mutations, the very low sequence divergence and lack of multiple alleles observed for each polymorphic site, even when compared against other close relatives, indicate that the most likely explanations are recombination (i.e. current and/or historical) and incomplete lineage sorting.

We also observed direct evidence for clonal propagation in Gs with the occurrence of the same haplotype over a wide geographic area. In this species, applying the I_A test for all isolates and for the reduced-by-haplotype dataset suggested the existence of epidemic clonality (Smith et al. 1993). Overall, the recombination component appears greater in Gc (100% unique haplotypes and lower I_A); however, concordant with the fungal asexual reproductions in natural environments (Six and Paine 1997), I_A greater than zero still suggests some deviation from complete panmixia.

These results agree with the genomic analysis of *G. clavigera sensu lato*; both suggested that this fungus is a heterothallic sexual species (Tsui et al. 2009; DiGuistini et al. 2011). Consistent with this, *G. clavigera* ascocarps have been occasionally reported at epidemic sites in one-year old MPB galleries (Robinson and Davidson 1968; Lee et al. 2003). Although a sexual state has yet to be reported for *G. clavigera*

associated with JPB under either field or experimental conditions, molecular results suggest history of recombination in this fungus. However, sexual reproduction seems to occur in older galleries when competition and predation increases and when environmental variables change. The asexual state is abundant in the galleries and pupal chambers during the active life cycle of the two beetles; as well it is abundant on artificial media used for fungal isolations. Systematic investigations with more isolates from different phases of the JPB life cycle may also allow the discovery of the sexual reproductive mode in Gc.

3.4.3 Ecologically distinguishable

Evidence for host-specific differentiation between the two *G. clavigera* lineages is as follows. While we expected that the Gc and Gs would be specific to beetle vectors, our ecological data indicate that one lineage (Gc) occurs on MPBs colonizing *P. ponderosa* as well as on JPBs infesting *P. jeffreyi*, whereas the other (Gs) is exclusively associated with MPBs. Gc was only isolated from two geographically distinct and localized US populations, one of which was populated with *P. contorta, P. jeffreyi* and *P. ponderosa* and the other only with *P. ponderosa*. In contrast, Gs was associated with epidemic and localized populations of MPB inhabiting *P. contorta,* as well as other pine species in the epidemic regions, but not with *P. jeffreyi* and the localized *P. ponderosa* supporting the Gc clade. Further, our phylogenetic data showed that *G. clavigera* from the same host species in different geographic areas are genetically closer than those collected from different host species occurring in the same geographic region (e.g.

California). While our data in some geographic areas were limited, preventing us from assessing the role of geographical isolation in speciation, overall, the data indicate that both a beetle vector's preference for a host tree species, and the geographic isolation of the host species, can contribute to progressive differentiation of the vectored fungal species.

G. clavigera lineages develop all phases of their life cycles on host trees and are dispersed by their respective beetle vectors via a specific association (Harrington 2005). Between beetle generations, these fungi are protected and maintained inside the specialized beetle structures called mycangia. Given this, the fates of the mutualistic fungus and beetle partners are linked, and mating is more likely to occur between fungi within the specific host tree. Such a degree of inherent isolation has been suggested to facilitate adaptive differentiation in a large number of fungal plant pathogens recognized as complexes of specialized sibling species (Giraud et al. 2006). The frequent asexual reproduction and sexual recombination in fungi can also promote ecological divergence by creating new combinations of alleles and rapid reproduction of those combinations that favor host adaptation (Giraud et al. 2010).

During the early phases of a massive attack by a beetle-fungal complex, healthy standing pine trees release constitutive or induce defense chemicals such as oleoresin. To survive in such hostile and toxic environments, beetle-fungal complexes must have mechanisms for modifying or metabolizing tree defense compounds (DiGuistini et al. 2009; 2011). While pine species have similar chemical defense systems, there are

quantitative and specific chemical differences among pine species and even between populations of the same host species (Forrest 1980; Gerson et al. 2009). For example, β-phellandrene is the most abundant monoterpene in *Pinus contorta* while heptane is the major volatile chemicals in *Pinus Jeffreyi* (Mirov and Hasbrouck 1976; Smith 2000). Heptane has been found at moderate concentration in the hybrid between *P. Jeffreyi* and *P. ponderosae* but has not been reported in *P. ponderosae*. However, tree chemical data are limited, especially for *P. ponderosae,* which needs to be systematically characterized across its range in western North America. Given this, specific association of the fungal pathogen with a host tree may also be maintained by the ability of pathogen to overcome and adapt to a tree's chemical defense systems.

Concordant with our results showing a distinct phylogenetic separation between *P. jeffreyi* (Gc) and *P. contorta* (Gs) associates, Six and Paine (1998) showed that *G. clavigera* from *P. contorta* exhibit a poor growth in *P. jeffreyi*. They also indicated that JPB associates were tolerant to a wider range of host chemicals. These differences might be due to the pathogen adapting to tree's chemical defense compounds, e.g. β phellandrene being at higher concentration in *P. contorta* than either *P. jeffreyi* and *P. ponderosae.* Further, molecular phylogeny of *Pinus* species is concordant with the monophyly of *G. clavigera* from localized populations of *P. ponderosae* and *P. jeffreyi*, and with the separation of the Gc from these two pine species from the Gs of *P. contorta*. *P. jeffreyi* and *P. ponderosae* are genetically and morphologically close relatives; they can hybridize and are classified in the *Pinus* subsection *Ponderosa* (Gernandt et al. 2009), while *P. contorta* is phylogenetically distinct and is classified in

the subsection *Contorta* (Krupkin et al. 1996). Similarly, MPB genetic divergence related to host trees have been also reported between localized beetle populations in mixed *P. contorta*/*P. ponderosa* forests in California, Colorado, Utah and Alberta (Stock and Amman 1980; Stock et al. 1984; Sturgeon and Mitton 1986b; Langor et al. 1990; Kelley et al. 2000). Hopkins (1909) described MPB as two species, *D. ponderosae* and *D. monticola*. Although these species were synonymized by Wood (1982), they attack and breed in different pine species (Stock et al. 1984). Genetic studies using allozyme and AFLP markers have reported contradictory results: host-dependent (e.g. *P. contorta* versus *P. ponderosa*) differentiation between localized beetle populations for allozymes and no host-dependent differences between MPB populations for AFLP (Mock et al. 2007). However, tree species and geographic areas vary between these studies, and it will be necessary to sample additional populations in the eastern and southern portion of the MPB range, and from different host trees including *P. contorta*, *P. ponderosa* and *P. flexilis* to resolve these contradictory results.

While MPB can attack and breed in different pine species, it is important to note that localized populations of MPB prefer one host pine species, even when that species is intermixed with other species that MPB could colonize (Wood 1982; Langor et al. 1990). A combination of events may contribute to the accumulation of host-adapted genes in MPB localized populations; for example, selective pressures on developing broods imposed by different tree species, host preferences by the beetle, differences among trees and allochronic separation of beetles' emergence from different hosts (Sturgeon and Mitton 1982; Borden 1984; Langor et al. 1990). Localized populations are

also characterized by temporary small outbreaks that are often initiated by secondary bark beetles attacking stressed trees (Smith et al. 2010); beetle populations in such regions may maintain a stable diversity of fungal species for extended periods of time. In contrast, epidemic populations of MPB often contain a high number of beetles relative to the preferred pine species in a given geographic range, and so attack other pine species (Wood 1963; Logan and Powell 2001; Bentz et al. 2010; Safranyik et al. 2010). Furthermore, during extensive outbreaks, MPBs have been reported occasionally as attacking and reproducing in non-host pine such as *Picea* when faced with a shortage of host trees (Huber et al. 2009). Consequently the spread of epidemics, which is affected by host tree's susceptibility, availability and continuity on large geographic regions (Safranyik et al. 2010), may dilute or replace older fungal populations that have become host adapted during the non-epidemic phases (Sturgeon and Mitton 1982; Langor et al. 1990). Such a population change was suggested by AFLP analysis of both MPB and *G. clavigera* populations (Lee et al. 2007; Mock et al. 2007; Roe et al. 2011). Lee et al. (2007) reported two genetically distinct groups of *G. clavigera* associated with *P. contorta* in the epidemic regions; the major group contains 166 individuals from BC and the Rocky Mountains and the second group include nine individuals from the Rocky Mountains. They suggested that the latter might represent the original population of the Rocky Mountains that was mixed with the larger group that was introduced into the region by the eastward expansion of MPB epidemic. Although representative isolates were included in our dataset, we found no evidence of these two MPB-associated *G. clavigera* groups. While support of distinct lineages based on independent gene

genealogies would indicate more ancient divergence among these fungi, microsatellite makers have also not supported such a distinction (Tsui et al. 2012).

Although the data from localized populations (i.e. California and South Dakota) suggested that *P. ponderosa* might not be a preferred host of the Gs lineage, this tree species was found hosting Gs in the epidemic regions (BC and Rocky mountains). This might be the result of the current rapid expansion of MPB and the pathogen (Gs) from their primary preferred host *P. contorta* to other pine species, including *P. ponderosa*. The holotype (ATCC 18086; Robinson and Davidson 1968) is the only remaining isolate from *P. ponderosa*-infested trees before the current epidemic in BC. It clusters genetically with other current localized *P. ponderosa*-associates, and not with Gs isolates from epidemic regions; this is consistent with the MPB rapidly expanding its population and geographic range in the epidemic. While no other historical isolates of *G. clavigera* are available, we would expect to find additional evidence for host tree preferences among *G. clavigera* lineages by sampling populations from different infested-tree species in the eastern and southern portion of the MPB range, i.e. in areas that have not been reached by the current epidemics. If fungal lineages are adapted to host species, then lineages should correlate with host species locations; however, this assumes, simplistically (Thompson 1994), that ecological constraints or genetic structure of host beetles/trees and pathogen are the same in different geographic regions. But they are not; both the beetle and host trees vary genetically and phenotypically between different geographic regions (Krupkin et al. 1996; Richardson 2000; Mock et al. 2007; Gernandt et al. 2009). And there are significant chemical

differences between trees at different geographic locations and with environmental conditions that need to be further characterized (Mirov 1948; Latta et al. 2003).

While the nomenclatural name *G. clavigera* is tied to the species that is genetically and ecologically represented by the holotype (Robinson and Davidson 1968), we showed that the fungus consists of Gs and Gc lineages. These are distinct sibling species that should be recognized taxonomically. Gc should retain the name *G. clavigera*, while Gs should be described as a new species. In the future we can anticipate that Gc genetic variation will evolve slowly while Gs might go through further genetic variation, and we outline two scenarios. In BC we already observed a post-epidemic phase in which the MBP population is decreasing and we anticipate that this population will collapse in the near future due to the lack of mature *P. contorta*. In the first scenario, only a small number of Gs haplotypes survive the MPB collapse and are maintained through the endemic cycle of the beetle until young pine trees reach maturity. At that point, in a future outbreak, the population and the fungal genetic diversity will increase, as it did in the current epidemic, leading to an array of closely related new haplotypes. In a second scenario that is potentially a shorter term concern a subset of the current large population in Alberta succeeds in becoming established in a new host tree species, and, with its fungal symbionts, adapts to the new physical and chemical environment presented by this host. There is evidence that this may already be occurring, as it has recently been shown that the beetle can successfully reproduce in the wild, in hybrids between *P. contorta* and *P. banksiana.* Significantly, *P. banksiana* occurs across the northern Canadian boreal forest. While *P. banksiana* is more closely

related to *P. contorta* than to *P. ponderosa* or *P. jeffreyi*, landscape and environmental conditions prevailing in the boreal forest would lead the symbiotic partners to evolve as the MPB spread across the boreal forest. Extending the work described above could characterize how Gc and Gs populations are evolving, and so help to assess threats related to the above scenarios. Even if MPB does not become established on *P. banksiana*, in the near future climate change will affect geographic distributions of trees and beetles, and populations of fungal associates will evolve with vectors and hosts. Similar work on other MPB host trees or other beetle systems could establish accurate species diversity and provide a foundation for understanding ecological interactions of the ophiostomatoid group that includes the most common fungal symbionts associated with bark beetles.

3.5 Tables and figures

Table 3.1 Fungal isolates used in this study

Fungal species	Beetle associate	Host tree	Collection site (Map no. [a])	No. isolates [b]	Source [c]	ID [d]	Collector (Date sampled)
***Grosmannia* sp. (Gs clade)**	*Dendroctonus ponderosae*	*Pinus contorta*	Canada, BC, Riske Creek	1	UAMH 4585	B01	Whitney (1982)
.	.	.	Canada, BC, Terry Fox Creek	1	NOF 1280	B02	Hiratsuka & Maruyama (1987)
.	.	.	Canada, BC, Houston (1)	10	UAMH (11153)	B03	Lee (2003)
.	.	.	Canada, BC, Tweedsmuir park (2)	1	CB SLA11	B04	.
.	.	.	Canada, BC, Williams Lake (3)	2	CB W14	B05	.
.	.	.	Canada, BC, Kamloops (4)	2	UAMH (11150)	B06	Lee (2003)
.	.	.	.		UAMH (11151)	B07	Lee (2004)
.	.	.	Canada, BC, Kelowna (5)	**2**	CB KDW4	B08	M. Alamouti (2007)
			.		UAMH (11152)	B09	.
.	.	.	Canada, BC, Manning Park (6)	5	CB M6	B10	Lee (2003)
.	.	.	Canada, Alberta, Westcastle	1	UAMH 4818	A01	Tsuneda (1983)
.	.	.	Canada, Alberta, Carbondale	1	NOF 842	A02	.
.	.	.	Canada, Alberta, Blairmore	1	NOF 2893	A03	Unknown (1983)
.	.	.	Canada, Alberta, Banff (7)	15	CB B20	A04	Lee (2003)
.	.	.	.		UAMH (11154)	A05	.
.	.	.	.		CB B6	A06	.
.	.	.	.		CB BW26	A07	.
.	.	.	.		CB B14	A08	.
.	.	.	.		UAMH (11155)	A09	.
.	.	.	Canada, Alberta, Cypress Hills (8)	**5 (2 trees)**	CB (CHMC3)	A10	M. Alamouti (2007)
.	.	.	.		CB CHDSC7	A11	
.	.	.	.		CB CHEBC10	A12	
.	.	.	USA, Montana, Hidden Valley (9)	10	CB HV14	M01	Six (2003)
.	.	.	.		CB HV30	M02	.

Fungal species	Beetle associate	Host tree	Collection site (Map no. [a])	No. isolates [b]	Source [c]	ID [d]	Collector (Date sampled)
	.	.	USA, Idaho, Hell Roaring (10)	10	CB D1128	I01	Six (2002)
	.	.	.		CB D1151	I02	.
	.	.	USA, California, Sierra Nevada (11)	**2**	DLS 1061	C01	Six (1995)
	.	.	.		DLS 1037	C02	.
	.	.	.	**23 (5 trees)**	CB 12G13	C03	M. Alamouti (2009)
	.	.	.		CB 23G23	C04	.
	.	.	.		CB 55B11	C05	.
	.	.	.		CB 68B21	C06	.
	.	.	.		CB 710G16	C07	.
	D. ponderosae	*P. albicaulis*	Canada, BC, Nelson (12)	**5**	CB Pa–9	B11	Blaiker (2007)
	.	.	.		CB Pa–6	B12	.
	D. ponderosae	*P. strobiformis*	USA, Arizona, Pinaleño mountains (13)	**7**	CB GCA02	Z01	Six (2009)
	.	.	.		CB GCA04	Z02	.
	D. ponderosae	*P. ponderosae*	Canada, BC, Kamloops (4)	**8 (4 trees)**	CB PY2–3b	B13	M. Alamouti (2007)
	.	.	.		CB PY8–8	B14	.
	.	.	Canada, BC, Kelowna (5)	**5 (5 trees)**	CB KGW5	B15	.
***G. clavigera* (Gc clade)**	.	.	Canada, BC, Cache Creek	1	(ATCC 18086)	B16	Robinson-J. (1968)
	.	.	USA, South Dakota, Black Hills (14)	**15 (5 trees)**	CB 15B29C1	D01	Blaiker (2009)
	.	.	.		CB 34B94C6	D02	.
	.	.	.		CB 24B166C8	D03	.
	.	.	USA, California, Sierra Nevada	**1**	DLS 15	C08	Six (1993)
	.	.	USA, California, Lassen (15)	**1**	DLS 24	C09	.
	D. ponderosae	.	.	**1**	DLS 56	C10	.
	D. jeffreyi	*P. jeffreyi*	USA, California, Sierra Nevada	10	C 843	C11	Harrington (1993)
	.	.	.		DLS 554	C12	Six (1999)
	.	.	.		DLS 776	C13	.
	.	.	.		DLS 833	C14	.
	.	.	.		DLS 681	C15	.
	.	.	.		DLS 771	C16	.
	.	.	USA, California, Lassen	10	DLS 173	C17	.
	.	.	.		DLS 210	C18	.

Fungal species	Beetle associate	Host tree	Collection site (Map no. [a])	No. isolates [b]	Source [c]	ID [d]	Collector (Date sampled)
	.	.	.		DLS 237	C19	.
	.	.	USA, California, San Bernardino (16)	1	DLS 52	C20	Six (1993)
	.	.	.	**9**	DLS 1560	C21	A. Hansen (2006)
	.	.	.		DLS 1565	C22	.
	.	.	.		UAMH (11156)	C23	.
	.	.	.		DLS 1588	C24	.
	.	.	.		DLS 1595	C25	.
Total number of isolates				166	62		
G. aurea	*Dendroctonus* sp.	*P. contorta*	Canada, BC, Invermere	1	CBS 438.69	UB	Davidson (1963)
Leptographium longiclavatum	*D. ponderosae*	*P. contorta*	Canada, BC, Kamloops	2	CB SLKW1436	LB	Lee (2003)
	D. jeffreyi	*P. jeffreyi*	USA, California, Sierra Nevada		**DLS 845**	LC	Six (1999)
L. terebrantis	*D. ponderosae*	*P. contorta*	Canada, BC, Kamloops	3	**CB 878AW1-2**	TB1	Kim (2004)
	.	**.**	**.**		**CB LPKRLT-3**	TB2	Kim (2003)
	D. brevicomis	*P. ponderosae*	USA, California, Sierra Nevada		C 418	TC	Harrington (2003)
L. wingfieldii	*Tomicus piniperda*	*P. sylvestris*	France, Orléans	2	CBS 645.89	WF	Morelet (1984)
	NA	*P. brutia*	Greece, Thessaloniki		CBS 648.89	WG	Skarmoutsos (1987)

[a] Generalized map location of collection sites corresponding to figure 3.2a

[b] Number of isolates analyzed for the ecological assessment using single-locus sequencing; Samples isolated and identified in this study are bolded; Isolates from the same locality are originated from different sources (i.e. from beetles and/or galleries collected from different tree individuals); otherwise number of isolation sources are shown in parentheses

[c] Isolates selected for 15 single-gene phylogenies; Source of isolates: UAMH, University of Alberta Microfungus Collection and Herbarium, Canada; NOF, Culture Collection of Northern Forestry Centre, Canada; ATCC, American Type Culture Collection, USA; CBS, Centraalbureau voor Schimmelcultures, Netherlands; Isolates beginning with CB, DLS and C are from culture collections of C. Breuil, University of British Columbia, Canada; D.L. Six, University of Montana, USA; and T.C. Harrington, Iowa State University, USA; respectively; Nine isolates chosen for the polymorphism discovery are shown in parentheses

[d] Letters indicate the location and colors indicate the host trees corresponding to figure 3.2a; Numbers indicate the number of isolates from each location

Table 3.2 Primer sequence and gene description for loci used in phylogenetic and population genetic analyses

Primer sequence 5'→3'		Gene description (abbreviation)	*G. clavigera* sequence length in bp*				GenBank accession no.
Forward	Reverse		Total	Exon	Intron	UTR	
TCACGCCCACCGTTACCGACA	TGGAAATGGTCGGTGCCGAGGT	40S ribosomal protein S3 (40SRP)	742	585	0	157	HQ633911 - HQ633980
TCCAGACGAACCTGGTGCCGT	CAGGCGTCATCGAGCAAGCGA	alpha-tubulin	640	489	59	92	HQ633073 - HQ633142
ATGTGCAGGGTGGCGAGCGAA	GAATACCGCTCCGCTCGCACA	ATP-binding-cassette multidrug transporter (ABC)	549	549	0	0	HQ633143 - HQ633212
TGATTCGACTTTCCCCCT	CGTCGAACACAAACTCCT	Anthranilate synthase (TRPG)	1,925	1,925	0	0	HQ633981 - HQ634049
GGAGTTTGTGTTCGACGAG	GAATGACAAGGCTATGAAGGGA						
TAAGGAAAGGGAGGGCGGT	TGGGTGCGTGATGAGCGA	Metallo-peptidase (MPEP)	1,672	1,386	215	71	HQ634050 - HQ634118
ATTCCCCTCCCCTACTCC	CTTCCATGTCCTCCTTCC						
GACATTGTAGAGGGCAGC	AGATGGGAGGTTGGAGAG	Cytochrome P450 (P450 I)	1,596	1,440	113	43	HQ633213 - HQ633282
AGTAGAACACCGCCGACAG	CCGACCAAACACACCGCA						
TGCAGCAATGGGACCGGATGA	TCGTCACGTTCTCCCAGCGCT	P450 II	710	710	0	0	HQ633283 - HQ633352
CACACGGACCAACGACGA	CTCTCCTGCCCCTCTTCTC	Lipid acyl hydrolase (LAH)	1,123	1,123	0	0	HQ633353 - HQ633422
CTCTTCTTTGCCGGCCTTGCTGT	CGCAACGCAAACGCCAGAAGA	Fungal extracellular membrane protein (CFEM I)	667	510	58	99	HQ633423 - HQ633492
GCGTCCATTGATCGGCGTGATGT	AACCGCCAACATGGCAACGG	CFEM II	491	427	64	0	HQ633493 - HQ633562
TGCTGTCGAGAACTGGAGGCGT	CGGCAGGACCTGGAACAGGAA	Lysophospholipase (LPL)	568	443	125	0	HQ633563 - HQ633632
CGGTCGCCCGCTCTACATTGA	CTCAGCCTCTAAGCCGTTGCCT	Phosphatidylinositol transferase (PLT)	570	570	0	0	HQ633633 - HQ633702
TGCCGACAAGGTGGCCAAGTTC	GCGCAGCGCAACATTGACGACT	Peroxisomal-coenzyme A synthetase (PCAS)	685	117	24	544	HQ633703 - HQ633772
CACGACGACGAACTCCTCTCCCA	CAGGATGCCCTCGGCCTCTAAC	Anonymous I	455	296	3	156	HQ633773 - HQ633840
ACGCCGGCAAGACCTACACCA	TGCCAGACTGGTCCACATCTGCA	Anonymous II	805	240	61	504	HQ633841 - HQ633910
Concatenated dataset			13,198	10,810	722	1,666	

*Base pair

Table 3.3 Fixed and shared polymorphisms between the two monophyletic clades in *G. clavigera*

Locus	Fixed polymorphisms				Shared polymorphisms			Exclusive polymorphisms to		Genetic differentiation [a]	
	Total	Noncoding	Synonymous	Replacement	Total	SNPs	Indels (bp)	Gs	Gc	Dxy (10^{-3})	SD of Dxy (10^{-4})
40SRP	0	0	0	0	0	0	0	2	1	1.27	3.4
alpha-tubulin	1	0	1	0	0	0	0	0	0	1.56	4.1
ABC	2	0	0	2	0	0	0	1	0	3.69	9.4
TRPG	7	0	3	4	1	0	1 (1) *	6	1	4.79	9.5
MPEP	4	2	0	2	0	0	0	2	2	3.44	7.8
P450 I	6	3	2	1	0	0	0	3	3	4.48	9.3
P450 II	0	0	0	0	0	0	0	2	3	1.36	3.6
LAH	4	0	0	4	0	0	0	0	1	3.33	8.7
CFEM I	0	0	0	0	0	0	0	5	1	2.81	5.6
CFEM II	0	0	0	0	3	3	0	1	2	4.18	6.9
LPL	1	0	0	1	0	0	0	2	0	2.43	5.5
PLT	1	0	1	0	0	0	0	2	1	2.32	5.1
PCAS	2	2	0	0	0	0	0	1	2	3.10	7.7
Anonymous I	2	0	2	0	0	0	0	0	1	4.51	11.6
Anonymous II	3	3	0	0	1	1	0	2	2	7.00	14.2
Concatenated dataset	33	10	9	14	5	4	1	29	20	3.71	5.7

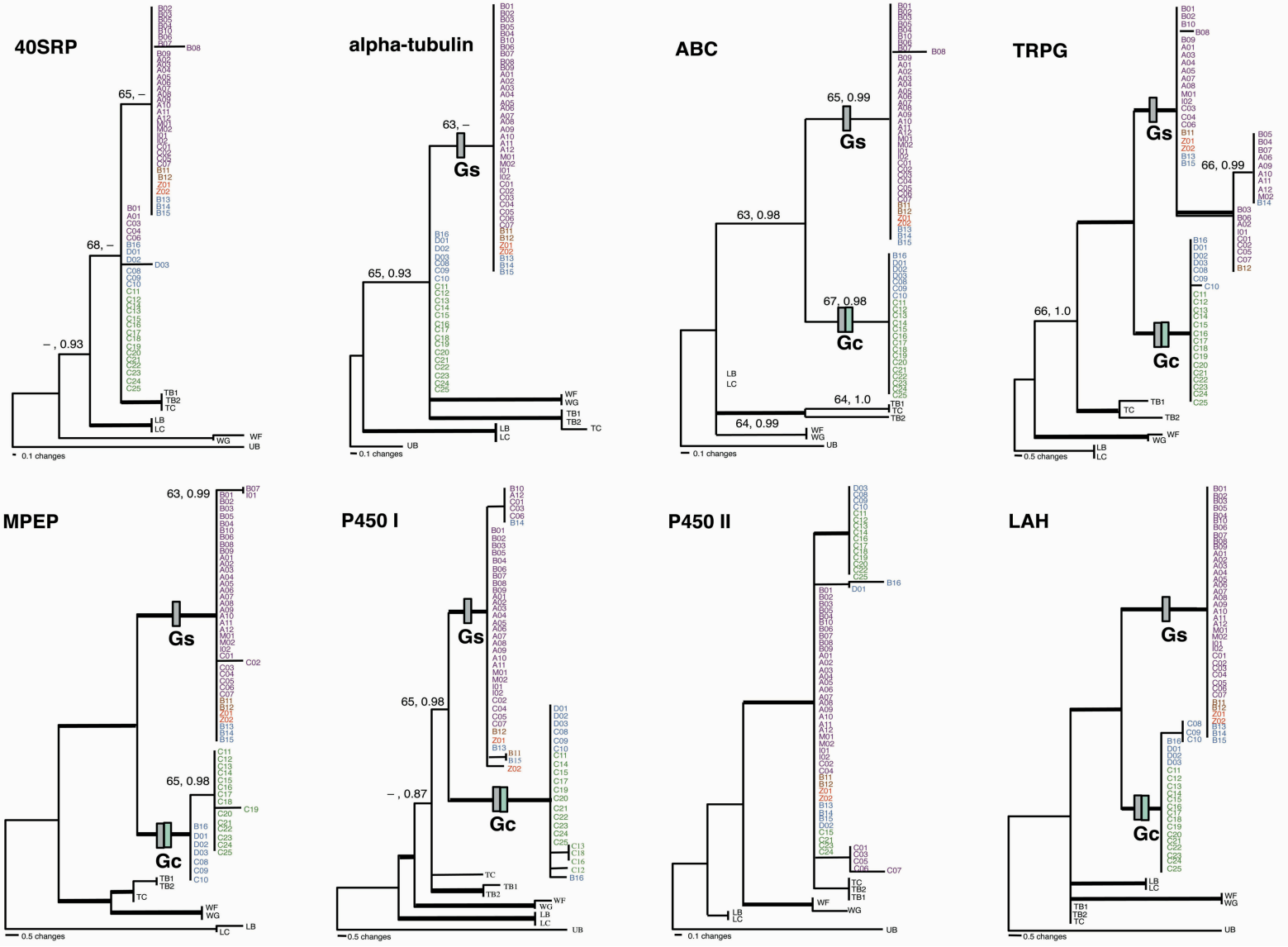
*Coding regions

[a] Dxy, net nucleotide divergence for the pairwise comparison of the two monophyletic clades in *G. clavigera* ; SD, standard deviation

Table 3.4 Information on phylogenetic dataset sequenced from *G. clavigera* and its close relatives

Locus	Sample size	Total number of characters	Variable sites	Parsimony informative characters (PI)	Number of tree steps (TS)	Number of MP trees	Homoplasy level		Nucleotide substitution model
							PI/TS	CI	
40SRP	70	742	23	11	23	1	2.09	1.00	TrN
alpha-tubulin	70	640	10	8	10	1	1.25	1.00	HKY
ABC	70	549	10	6	10	1	1.67	1.00	HKY
TRPG	69	1,925	37	33	45	3	1.36	0.89	TPM1uf+I
MPEP	69	1,672	30	28	31	2	1.11	0.96	HKY+I
P450 I	70	1,597	53	29	55	4	1.90	0.98	TrN
P450 II	70	710	17	8	18	3	2.25	0.94	TRN
LAH	70	1,123	25	15	26	2	1.73	1.00	TPM1uf+I
CFEM I	70	673	31	12	33	4	2.75	0.96	TIM1+I
CFEM II	70	491	27	12	29	5	2.42	0.93	TIM1+I
LPL	70	569	28	13	30	2	2.31	0.97	HKY+I
PLT	70	570	10	6	10	1	1.67	1.00	HKY
PCAS	70	706	27	9	27	1	3.00	1.00	HKY
Anonymous I	68	467	22	9	22	1	2.44	1.00	HKY
Anonymous II	70	805	46	27	47	2	1.74	0.98	GTR
Concatenated dataset	70	13,239	402	226	503	100	2.23	0.81	GTR

Information with outgroup taxon.


40SRP
65, –
68, –
– , 0.93
0.1 changes
alpha-tubulin
63, –
Gs
65, 0.93
0.1 changes
ABC
65, 0.99
Gs
63, 0.98
67, 0.98
Gc
64, 1.0
64, 0.99
0.1 changes
TRPG
Gs
66, 0.99
66, 1.0
Gc
0.5 changes
MPEP
63, 0.99
Gs
65, 0.98
Gc
0.5 changes
P450 I
Gs
65, 0.98
– , 0.87
Gc
0.5 changes
P450 II
0.1 changes
LAH
Gs
Gc
0.5 changes

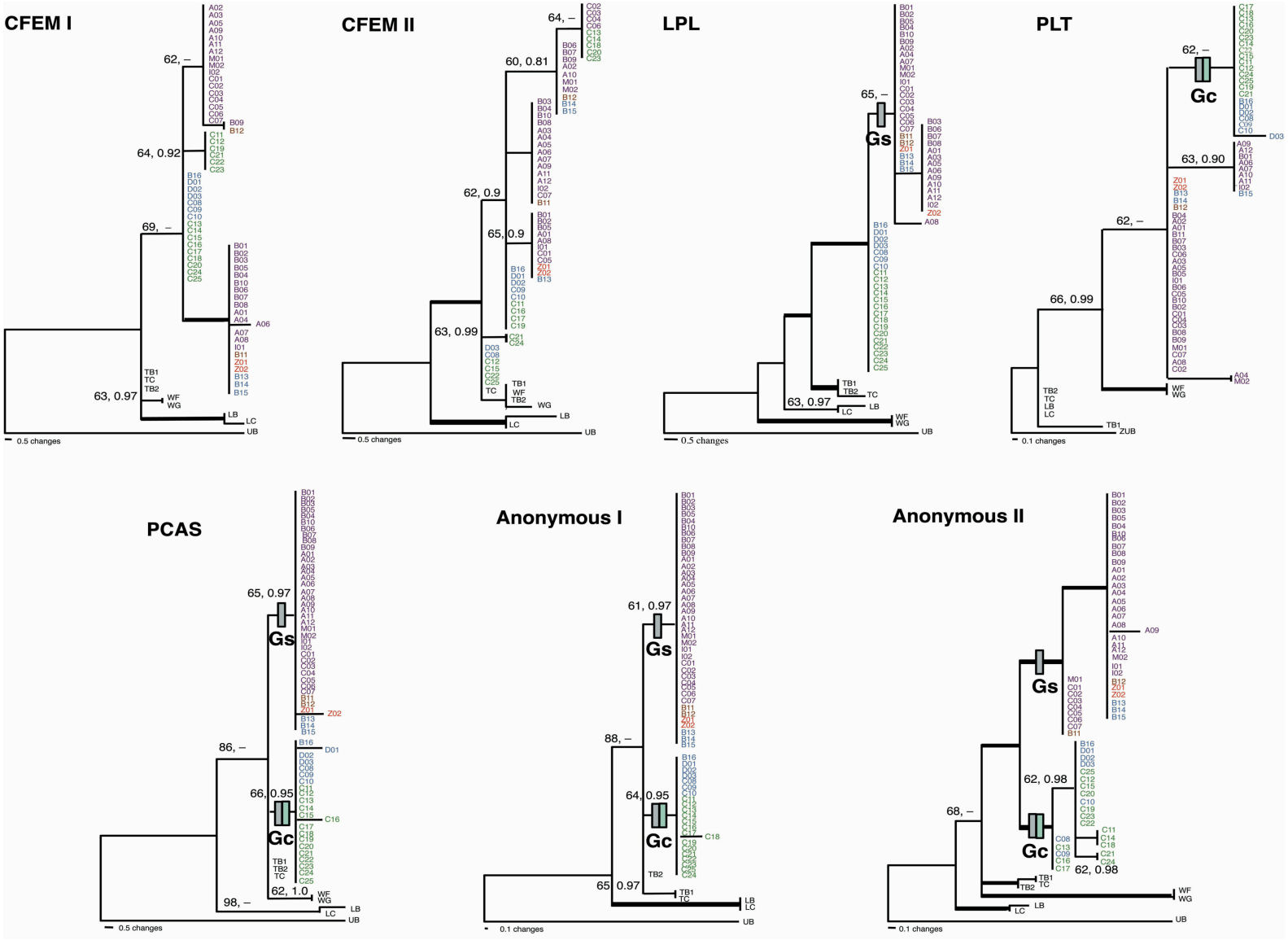

CFEM I
62, –
64, 0.92
69, –
63, 0.97
TB1
TC
TB2
WF
WG
LB
LC
UB
A06
0.5 changes
CFEM II
64, –
60, 0.81
62, 0.9
65, 0.9
63, 0.99
TB1
WF
TB2
WG
LB
LC
UB
0.5 changes
LPL
65, –
Gs
63, 0.97
TB1
TB2
TC
LC
LB
WF
WG
UB
0.5 changes
PLT
62, –
Gc
63, 0.90
62, –
66, 0.99
TB2
TC
LB
LC
TB1
ZUB
WF
WG
A04
M02
0.1 changes
PCAS
65, 0.97
Gs
86, –
66, 0.95
Gc
TB1
TB2
TC
62, 1.0
98, –
WF
WG
LC
LB
UB
Z02
D01
C16
0.5 changes
Anonymous I
61, 0.97
Gs
88, –
64, 0.95
Gc
TB2
65, 0.97
TB1
TC
LB
LC
UB
C18
0.1 changes
Anonymous II
Gs
68, –
62, 0.98
Gc
62, 0.98
TB2
TB1
TC
LC
LB
WF
WG
UB
A09
C11
C14
C18
C21
C24
0.1 changes

Figure 3.1 Single-locus phylogenies of 15 genes in *G. clavigera* and its four close relatives.

Bootstrap (BS>50) and posterior probabilities (PP>0.8) from MP and Bayesian analyses are shown along the branches. Thick branches indicate nodes with PP ≥ .95 and BS ≥ 70. The two bars indicate the *G. clavigera* monophyletic clades color-coded according to their beetle associates: MPB (gray) and JPB (green). Trees are rooted with *G. aurea*, except for TRPG and MPEP that miss the outgroup taxon and therefore are midpoint rooted. Refer to figure 3.2a and table 3.1 for color codes and lables.

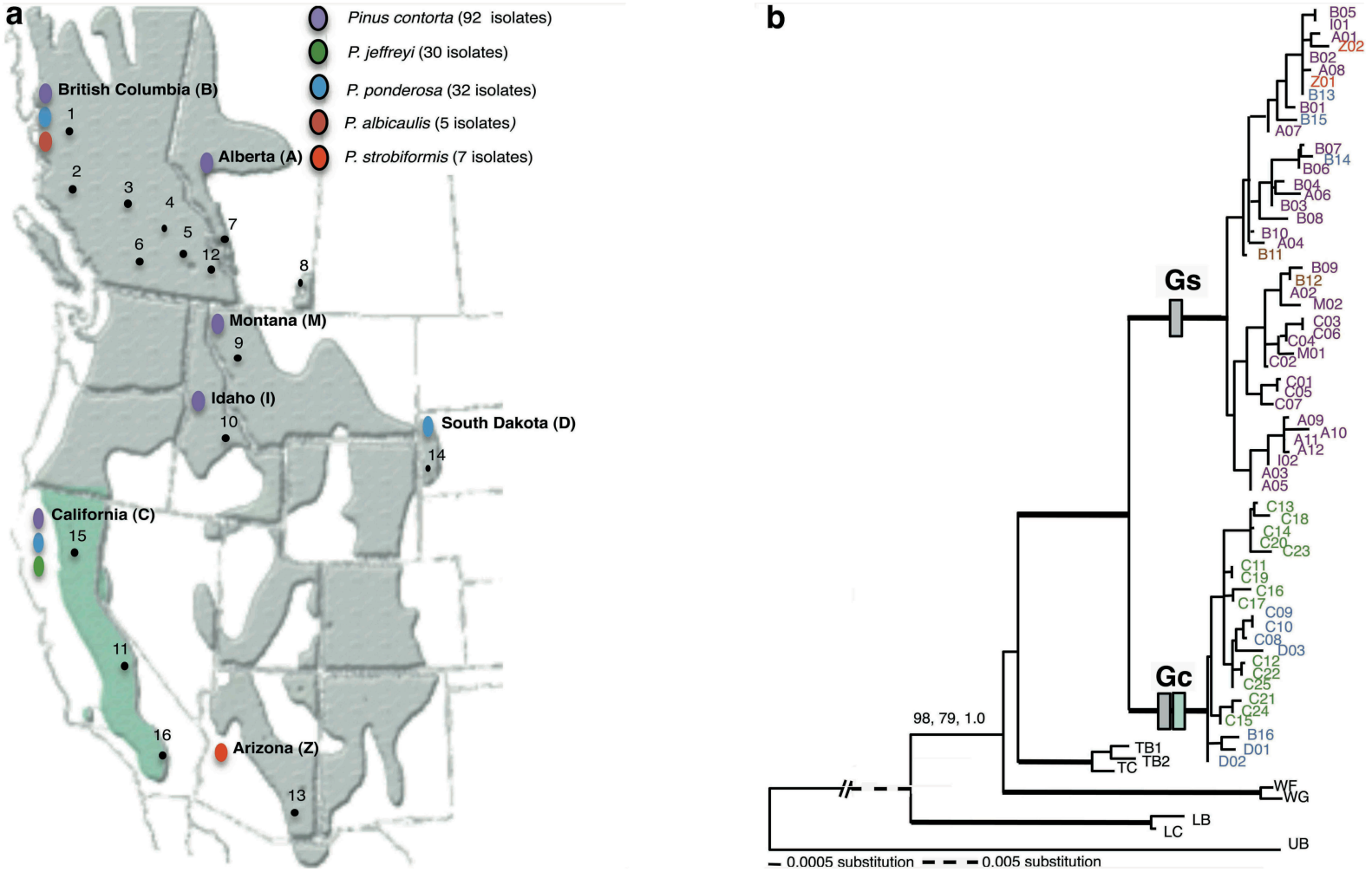
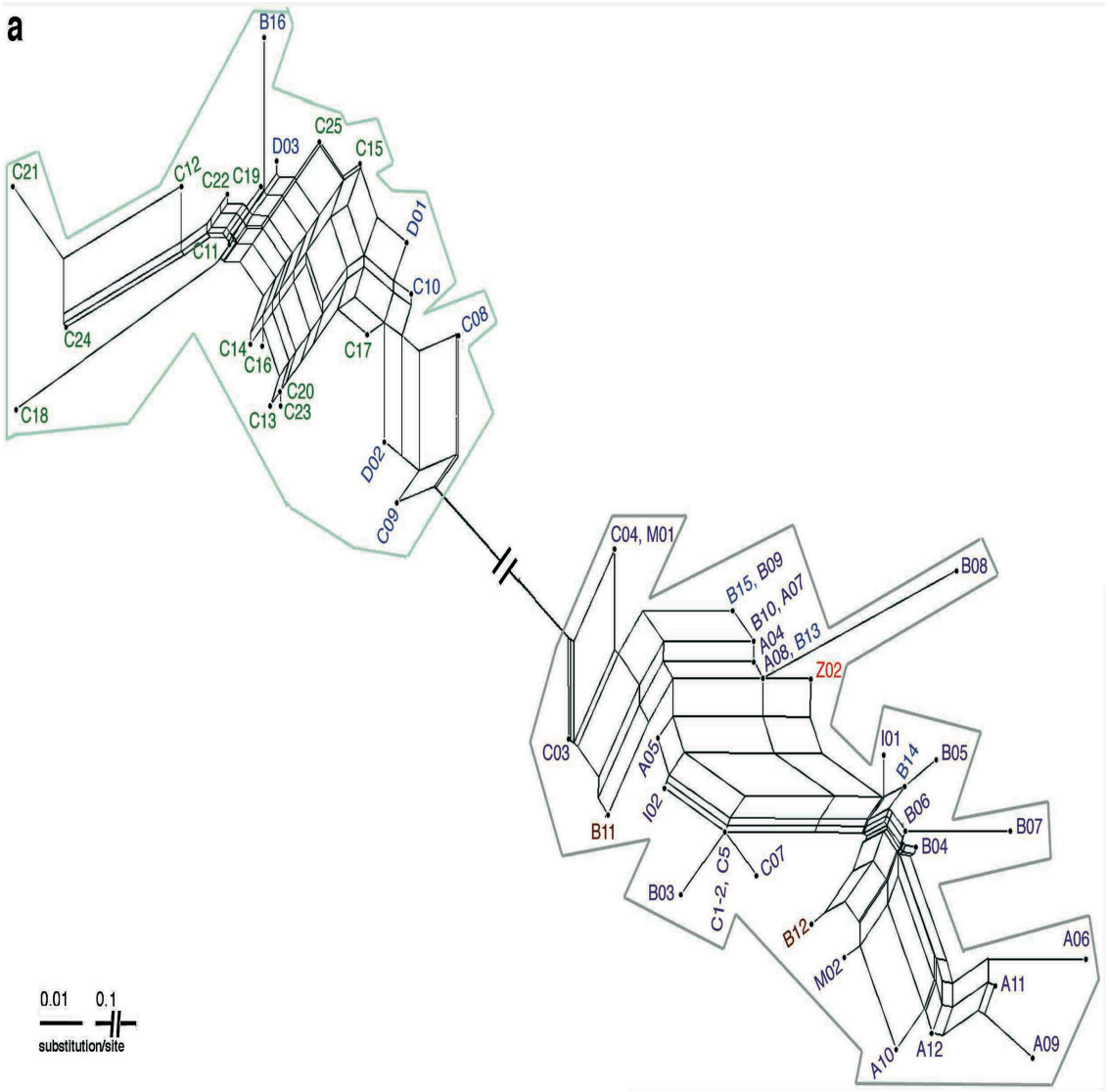
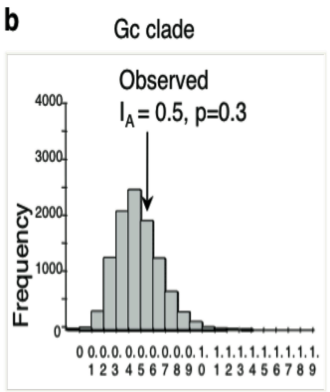

a
Pinus contorta (92 isolates)
P. jeffreyi (30 isolates)
P. ponderosa (32 isolates)
P. albicaulis (5 isolates)
P. strobiformis (7 isolates)
British Columbia (B)
Alberta (A)
Montana (M)
Idaho (I)
South Dakota (D)
California (C)
Arizona (Z)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
b
Gs
Gc
98, 79, 1.0
B05
I01
A01
Z02
B02
A08
Z01
B13
B01
B15
A07
B07
B14
B06
B04
A06
B03
B08
B10
A04
B11
B09
B12
A02
M02
C03
C06
C04
M01
C02
C01
C05
C07
A09
A10
A11
A12
I02
A03
A05
C13
C18
C14
C20
C23
C11
C19
C16
C17
C09
C10
C08
D03
C12
C22
C25
C21
C24
C15
B16
D01
D02
TB1
TB2
TC
WF
WG
LB
LC
UB
0.0005 substitution
0.005 substitution

Figure 3.2 Fungal collection sites and 15–gene phylogeny of *Grosmannia clavigera* complex

a) Map of western North America showing fungal collection sites where only one (MPB: gray) or two (MPB & JPB: green) *G. clavigera* beetle associates are present. Host tree species are color-coded and number of fungal isolates from each tree species is shown in parentheses. b) ML analysis of 15-gene combined dataset showing how the species recognized by PSR are related to each other and to other closely related species. Thick branches indicate nodes with ML and MP BS values of 100 and the Bayesian PP of 1.0. Gs and Gc monophyletic clades are labeled with bars color-coded according to beetle associates: MPB (gray) and JPB (green). Letters indicate the collection localities and colors indicate host tree species corresponding to the map and table 3.1. Dashed line indicates an adjustment of scale.


a
B16
C25
C15
D03
C21
C12
C22
C19
D01
C11
C10
C08
C24
C14
C16
C17
C18
C20
C13
C23
D02
C09
C04, M01
B15, B09
B10, A07
B08
A04
A08, B13
Z02
C03
A05
I01
B14
B05
I02
B06
B07
B11
B04
B03
C1-2; C5
C07
B12
A06
M02
A11
A10
A12
A09
0.01
0.1
substitution/site


b
Gc clade
Observed
I_A = 0.5, p=0.3
Frequency
4000
3000
2000
1000
0
0 0.0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9

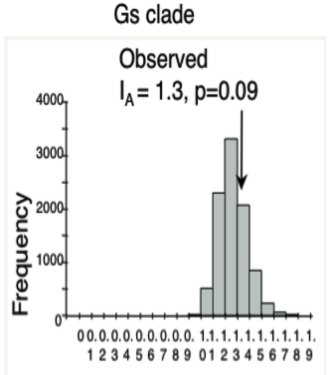

Gs clade
Observed
I_A = 1.3, p=0.09
Frequency
4000
3000
2000
1000
0
0 0.0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9

Figure 3.3 Recombination analysis

a) Split decomposition analysis of the 15-gene combined dataset. The scale represents phylogenetic distances between the haplotypes using a GTR+I+G substitution model with parameters estimated using JModelTest. Colored boxes represent the two *G. clavigera* monophyletic clades: Gc (green box) and Gs (gray box). The interconnected networks are suggestive of recombination and/or lineage sorting within both Gc and Gs clades. The center of network where it branches to the Gs and Gc groups was slightly netted (implying that the data support conflicting deep splits as expected from the among-gene incongruences, e.g. CFEM-II, as well as generally for deep branches) but it has been simplified and shortened to fit the graph. The labels refer to *G. clavigera* isolates listed in table 3.1). The I_A values for Gs and Gc clades are shown and compared against histograms of I_A values for 10,000 simulated recombined dataset.

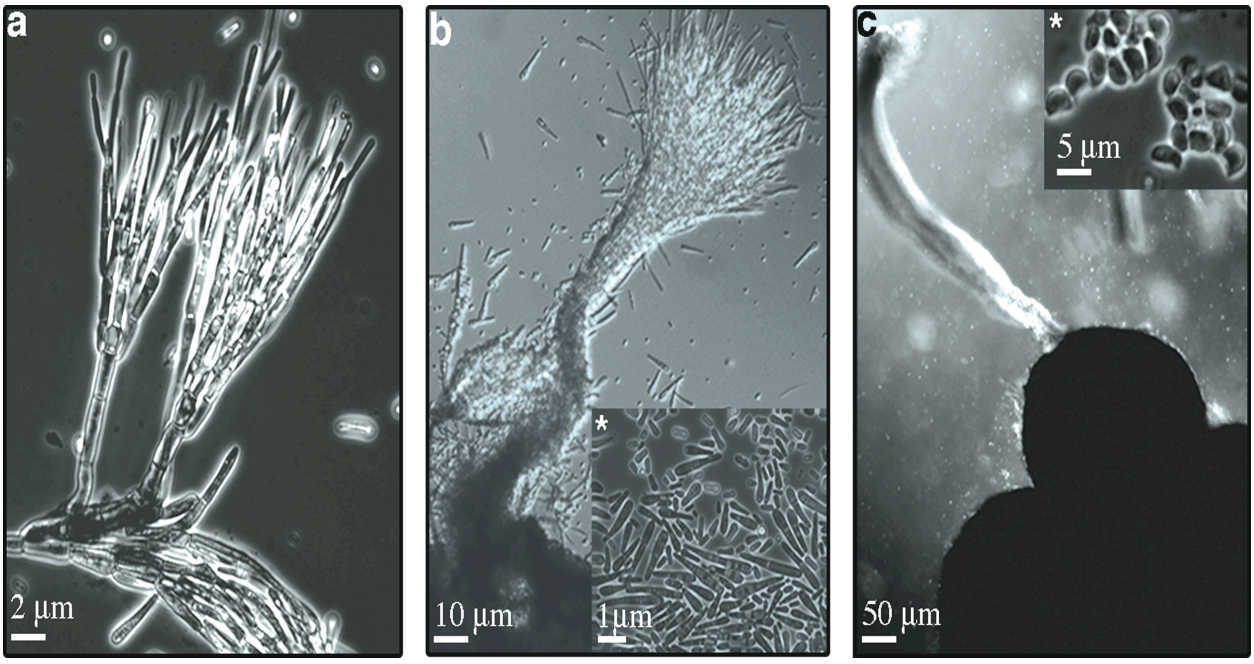


Figure 3.4 Asexual and sexual stage in Gs

Reproductive structures in Gs. Light micrographs of asexual stage characterized with mononematous (a) and synnematous (b) conidiophores reproducing conidia (*). Light micrograph of sexual structure (c) characterized by a spherical ascocarp oozing ascospores (*).

Chapter 4 Comparative genomics of the pine pathogens and the beetle symbionts in the genus *Grosmannia*

4.1 Introduction

Over tens of millions of years conifer forests around the world have provided unique ecological niches for native bark beetles and their fungal symbionts. Interactions between conifer hosts, bark beetle vectors and their fungal associates have influenced the evolution of tree chemical defenses (e.g. terpenoids), beetles and fungal symbionts (Seybold et al. 2000; Farrell et al. 2001; Jordal 2013). Although beetle-tree-associated fungi have significant effects on forest ecosystems, knowledge has improved only recently about the specificity for host trees or beetle vectors in this group of fungi (Wingfield and Seifert 1993; Kurz et al. 2008). Currently, little is known about the genetic differences that are associated with fungal speciation and adaptation, and that may have been shaped by evolutionary processes. Fungal diversification and specialization for hosts may depend on genetic differences that include genomic rearrangements, gene losses/duplications, and coding and non-coding sequence variants that may be under selective pressure in particular genes (Aguileta et al. 2009; Stukenbrock et al. 2010; Manning et al. 2013). The extent of adaptive processes at the genome level can be quantified by identifying genomic differences within and between fungal lineages that have recently diverged and specialized onto different host trees (Stukenbrock et al. 2010).

In North America, tree-inhabiting beetles and their fungal symbionts are among the most diverse and damaging forest pests (Harrington 2005; Jordal and Cognato 2012). For example, in western Canada alone, the mountain pine beetle (MPB; *Dendroctonus ponderosae*) and its fungal associates have killed over 18 million hectares of *Pinus contorta* forests (http://www.for.gov.bc.ca/hfp/mountain_pine_beetle/facts.htm; http://cfs.nrcan.gc.ca/pages/276), dramatically altering forest ecosystem dynamics and forest-dependent economic activities (Kurz et al. 2008). Further, the recent spread of the MPB-fungal complex into Alberta and Saskatchewan, and into *P. banksiana* raises the risk that the epidemic will spread eastward into and potentially across Canada's boreal forests (Cullingham et al. 2011). Of the fungal associates, the ophiostomatoid (Sordariomycetes, Ascomycota) *Grosmannia clavigera sensu lato* is crucial to the epidemic as an obligate symbiont of MPB and a pathogen of *P. contorta* that can kill living trees through beetle mass colonization (Lee et al. 2006). This fungus forms a symbiotic relationship with MPB and its sister species the jeffrey pine beetle (JPB; *D. jeffreyi*). While the two beetles are morphologically and genetically similar, they are adapted to different host trees (Six and Paine 1997). JPB is highly specialized and colonizes only *P. jeffreyi*, while MPB primarily inhabits *P. contorta* but can also successfully colonizes more than 20 pine species, but not *P. jeffreyi* (Wood 1982).

Complexes consisting of beetles, trees, and fungi provide unique systems for understanding ecological divergence or speciation (Thompson 1994; DiGuistini et al. 2011; Massoumi Alamouti et al. 2011). Theoretical studies suggest that dispersal of the

plant pathogen between hosts, and aspects of the pathogen life cycle can promote ecological divergence; e.g. reproduction is frequently asexual, and sexual recombination is constrained because it occurs within a host's tissues (Giraud et al. 2006; Giraud et al. 2008). Concordant with this theoretical framework, protein-coding genealogies have identified two cryptic species within G. clavigera (Massoumi Alamouti et al. 2011). One species (Gs) is an exclusive associate of MPB and its primary host tree P. contorta, while the other (Gc) is found on localized populations of MPB and JPB where these beetles colonize the closely-related P. jeffreyi and P. ponderosa. Although the two Grosmannia lineages can occur in the same geographic region (e.g. California), no evidence of gene flow between Gs and Gc was detected based on sequence analysis of 15 nuclear coding loci, suggesting that host tree species and beetle population dynamics are important factors in the evolution and divergence of these fungi (Massoumi Alamouti et al. 2011).

Defining species boundaries in *G. clavigera* has been difficult, because this fungus belongs to a complex that consists of many related species that have little morphological and genetic variation (Tsuneda and Hiratsuka 1984; Zambino and Harrington 1992; Six et al. 2003; Lim et al. 2004; Roe et al. 2010; Massoumi Alamouti et al. 2011; Roe et al. 2011). Genetic variation has been characterized using allozymes in ten populations of Gc associated with JPB in California, and using multilocus sequencing, AFLP and microsatellite markers in the epidemic populations of Gs in British Columbia and Alberta (Canada), Idaho and Montana (USA) (Six and Paine 1999; Lee et al. 2007; Tsui et al. 2012). Based on these surveys the Gs epidemic populations

were divided into four distinct groups with some gene flow and admixture between the groups. Molecular evidence of random mating and linkage equilibrium suggests that both Gs and Gc have life cycles with sexual stages (Massoumi Alamoutiet al. 2011; Tsui et al. 2012). However, these fungi are predominatly found reproducing asexually inside the beetles' galleries under the bark. Sexual fruiting bodies have been found occasionally in nature for Gs and Gc from *P. ponderosae* but not for Gc from *P. jeffreyi*, and fruiting bodies that demonstrate successful crosses have not been produced in the laboratory (Lee et al. 2007; Alamouti et al. 2011; Tsui et al. 2012). While our recent genealogical study identified cryptic species, this differentiation was not evident using other methods (e.g. multigene phylogenies or DNA finger printing). Fungal genome sequence and structure can be highly variable across populations and species (Galagan et al. 2005; Raffaele and Kamoun 2012), and single nucleotide polymorphisms (SNP) are now widely used for characterizing population genetics because they can also provide evolutionary information on ecological adaptation and speciation (Brumfield et al. 2003; Morin et al. 2004; Raffaele and Kamoun 2012).

Recently we reported the genome sequence of a Gs strain (slkw1407) isolated from *P. contorta* trees in the epidemic region (DiGuistini et al. 2011). The ~30 Mb genome assembly consisted of 18 supercontigs and 8,312 protein-coding gene models. We characterized some aspects of the functional genomics of the fungus, including its interaction with host-defense chemicals (Hesse-Orce et al. 2010; Wang et al. 2013). This work suggested that *Grosmannia* can tolerate, detoxify and utilize host defense chemicals. Given that host defense chemicals vary among pine species (Keeling and

Bohlmann 2006; Gerson et al. 2009; Boone et al. 2011; Hall, et al. 2013a;b), here we hypothesize that genes involved in host-pathogen interactions, secondary metabolite production and fungal interactions and differentiation, like cytochrome P450s, monooxygenases, membrane proteins such as ATP-binding cassette (ABC) and major facilitator superfamily (MFS) transporters, polyketide synthases genes (PKS), and vegetative incompatibility genes may have diverged to a greater extent in response to selection in different host environments.

In the current work, we use the reference Gs genome to enable comparative analysis of evolutionary divergence in distinct populations of Gc and Gs. We sequenced eleven strains, assembled their draft genome sequences, and reported a comprehensive assessment of intra and interspecies genomic variations relative to the Gs reference sequence. We applied genome-wide SNP phylogenies of twelve *Grosmannia* strains and gene genealogies of additional strains to test whether the genome dataset confirm our recent genealogical study that Gs and Gc are distinct lineages and whether it provides further evidence of ecological and/or geographic divergence in these fungi. Focusing on SNPs that are predicted to alter proteins encoded by genes noted above that are likely to be crucial for fungal colonization of host trees, we assess evidence for fungi diverging as they evolve to adapt to different pine species (*P. contorta*, *P. jeffreyi*, *P. ponderosa*). We identify genes that show evidence of adaptive selection, and relate these variations to differences in fungal ecology and biology.

4.2 Material and methods

4.2.1 Fungal samples

We sequenced eight Gs genomes from two distinct populations of MPB-infested *P. contorta* trees: a) epidemic regions in Canada and the USA, and b) localized populations in small geographically isolated outbreaks in California. We also sequenced three genomes of the sibling species Gc. The sibling group included two strains from JPB-infested *P. jeffreyi* trees in California, as well as the Gc holotype described by Robinson-Jeffrey and Davidson (1968) from MPB-infested *P. ponderosa* trees in British Columbia (BC). We deposited these fungi at the University of Alberta Microfungus Collection and Herbarium (UAMH) along with additional strains used for SNP validations, phylogenies and physiological studies (Table 4.1, Appendix B.12).

4.2.2 Illumina paired-end library construction, sequencing and assembly

Fungal mycelia were grown on 2% malt extract (MEA; 33 g Oxoid malt extract agar, 10 g Technical agar No.3, and 1 L distilled water) overlaid with cellophane. DNA from the mycelia was extracted using the method of Möller et al. (1992). DNA samples were processed at the Genome Science Center (GSC, Vancouver, BC, Canada) for paired-end sequencing following Illumina protocols (Illumina, Hayward, CA, USA). The library for each strain was amplified in a single flow-cell and sequenced to either 50 or 76

nucleotide base (nt) reads on the Illumina Genome Analyzer (GA) II or IIx following the manufacturer specifications.

Genomes were assembled from paired-end reads of 200 base DNA fragments generated by Illumina sequencing using the ABySS assembler v1.2.7 (Simpson et al. 2009). Reads that passed the chastity filter were assembled with a relative short kmer (25–31nt) for higher sensitivity. The resulting contigs were used as single end reads along with the original paired end data and reassembled with a higher kmer (35–61nt), which has a higher specificity (Appendix B.1). The assembly was cleaned and gaps closed using Anchor (www.bcgsc.ca/platform/bioinfo/software). Ambiguous base calls were resolved by mapping the reads back to the assembly using BWA v0.5.9 (Li and Durbin 2009) and calling consensus bases using SAMtools 'mpileup' v0.1.18 (Li et al. 2009). Assembled contigs and scaffolds larger than 200 nucleotides were ordered and oriented using MUMmer (Kurtz et al. 2004) based on the *Grosmannia* published genome (slkw1407; NCBI Genome PID: 39837; DiGuistini et al. 2009). Contigs that did not align with slkw1407 were ignored for our analysis. The assembly statistics after ordering and orientation are shown in Appendix B.2.

4.2.3 Gene predictions and ortholog determination

To detect *Grosmannia* orthologs, we first generated gene annotations for each draft genome using the homology-based gene predictor genBlastG (She et al. 2011). We

used protein sequences from the slkw1407 reference genome as the query (n=8,312) and the genome of another strain as the target database. Gene annotations and pairwise homology between the slkw1407's gene models and those from each genome were assigned based on genBlastG hits with an E–value cutoff of ≤ 1e-10 and a query coverage of >50%. The genBlastG output can result in redundant gene predictions when the query gene belongs to a multigene family, paralogous genes, or tandem gene duplications. Given this, for downstream analyses we applied a filtering procedure so that each genomic region would contain only one gene prediction with the highest global sequence percent identity (PID) to the query. The filtering procedure was carried out as follows: (i) all gene predictions were sorted by PID, (ii) for each two overlapping gene model, if the overlapping region was > 5% of the length for either gene, then only the prediction with higher PID was kept, (iii) all gene models were required to have PID ≥ 70% to the query, and (iv) to avoid assigning paralogs to query-target pairs, the best match had to have a PID 10% higher than the next best match. Non-overlapping gene models with high similarity to the same query were reported as putative paralogs and were removed from analysis. For genes with alternative splicing variants, the longest transcript was selected to represent the gene. After filtering incomplete genes and discarding genes with frame-shifts, which could have been caused by the draft quality of the genomes, only high quality 1:1 orthologous genes were retained for analysis. Gene models for all *Grosmannia* genomes are provided in annotation files.gff (https://tria website). Appendix B.3 summarizes genBlastG output used to find the pairwise homology between reference gene models and those of each draft genome.

4.2.4 Mapping and variant calling

We performed variant calling among the *Grosmannia* genomes by mapping reads from each strain to the slkw1407 reference genome sequence. Before read mapping, we filtered raw reads to remove low-quality and duplicate sequences using PRINSEQ lite v0.17.1 (Schmieder and Edwards 2011). We discarded reads that failed the Illumina chastity filter (Haridas et al. 2011), contained uncalled bases, and had an average *Phred*-scaled quality of less than 10 in the last 20 base calls. For the retained reads, the initial (5') five nucleotides showed GC-content bias (data not shown), and were trimmed, leaving 45 and 71 nt reads for mapping. We also filtered potential duplicate reads resulting from amplification of the identical DNA fragments during library preparation and sequencing. The numbers of reads used for SNV calling and processing steps are shown in Appendix B.4.

For each strain, filtered reads were mapped to the slkw1407 reference genome sequence using BWA v0.5.9 (Li and Durbin 2009). Initial mapping results were converted into the indexed and sorted Binary Alignment/Map (BAM) format using SAMtools v0.1.18 (Li et al. 2009) and Picard v1.54 (http://picard. sourceforge.net). To enhance the quality of the alignments for more accurate variant detection, we used Genome Analysis Toolkit (GATK) (McKenna et al. 2010) to locally realign the BAM files in complex regions, e.g. containing insertions/deletions (indels). For each alignment, BWA assigned a mapping quality score (MAPQ). We used reads with MAPQ greater

than zero to estimate the coverage and average read depth of final BWA alignments, using BEDtools v2.13.4 (Quinlan and Hall 2010). The individual BAM datasets are available at https://. Once reads from individual strains were mapped to the slkw1407 genome, we used SAMtools 'mpileup' to assess variant sites, applying Base Alignment Quality computation and a –C50 argument to minimize alignment artifacts and base-calling errors. Single nucleotide variants (SNVs) were identified using the Bayesian variant calling models implemented in 'bcftool' (Li et al. 2008). After consensus base calling, we filtered the initial variants for strand and distance biases (*p*-value<0.0001) using SAMtools 'vcfutils.pl'. The final set of high-quality calls also required a candidate site to be biallelic and to meet the following criteria: minimum *Phred*-scaled base calling score of 20, MAPQs of at least 30, read depths of more than four and less than 500, and a minimum 10 nt distance from indels. Variant calls that failed to meet these criteria were likely to be false positives. Because SAMtools 'mpileup' assumes a diploid model and our samples represent haploid genomes, we also removed heterozygote calls.

4.2.5 Verification of variant calls

To estimate the robustness of SAMtools results, genomic variants were also assessed using the SNV calling algorithm implemented in GATKv1.40 (McKenna et al. 2010). This method also uses a Bayesian model to estimate the likelihood of a site harboring an alternative allele for each sample. GATK was run on the same BAM files as SAMtools, using default parameters. GATK raw-variant calls were filtered in the

same manner as the SAMtools calls (see above). To estimate SNV false positives in our dataset, we generated Illumina paired-end reads for the slkw1407 reference strain (Appendix B.4) and assessed variant calls for these reads mapped against their own published genome, and identified 1,796 high-quality SNVs. Because the alternate base was present in all eleven genomes and also in Illumina read alignments from the slkw1407, a large percentage (93.2%) of these changes likely represent errors in the reference genome assembly. We removed these ambiguous calls from the final SNV dataset. For sixteen additional Gc and Gs isolates (Table 4.1), we also used PCR and Sanger sequencing to validate SNPs in the nine candidate genes listed in Appendix B.5. For the eleven *Grosmannia* strains, we aligned homologous contigs of the candidate genes to those of slkw1407 genome and gene models using progressive Mauve 2.3.1 (Darling et al. 2010). Primers were designed based on the alignment using Geneious 5.1 (Biomatters Ltd, New Zealand). Amplicons were purified and sequenced at the Nucleic Acid Protein Service Unit at the University of British Columbia (Vancouver, Canada). All sequences are available at GenBank (accession nos.).

4.2.6 Functional annotations for SNVs

SNVs were annotated as coding (synonymous and non-synonymous), intronic, flanking and intergenic, with SNPeffect v.2.0.5 (Reumers et al. 2006), using the slkw1407 genome's sequence and predicted gene models. We assigned flanking regions (i.e. UTRs and putative regulatory regions) of 1,000 nt upstream and

downstream of the initiation/termination codons of the annotated slkw1407 gene models, unless neighboring gene sequences were within this range; for such cases we truncated the assigned regions. We also characterized, as a set of variants, SNVs that result in the loss or gain of a stop codon, which likely affect the integrity of the protein products. Orthologous genes containing a premature stop codon were labeled as pseudogenes. We assessed the accuracy of stop-codon variant calls using expressed sequence tag libraries (EST) and RNA-seq data from slkw1407 (DiGuistini et al. 2009) as well as Illumina reads from more than one strain within each Gs and Gc group. Finally, we applied Gene ontology (GO) functional enrichment analysis (Molecular function or Biological process) on pseudogene candidates. The GO term associations were determined for each slkw1407 reference gene models using Blast2Go v2.5.0 with the default parameters (Conesa et al. 2005). Blast2GO was also used for a GO functional enrichment analysis; for that we performed the Fisher's exact test with a FDR correction to obtain an adjusted p-value between the candidate genes and the whole genome annotation.

4.2.7 SNVs clustering and phylogenomics

We used the genome-wide SNV data to determine phylogenomic relationships and the nucleotide divergence among *Grosmannia* genomes. To assess genome-wide-SNV clusters among a relatively small number of *Grosmannia* strains, we used the non-parametric AWclust (Gao and Starmer 2008) R package, because it requires no model

assumptions (e.g. Hardy–Weinberg equilibrium) and is based on hierarchical clustering of a distance matrix rather than on allele frequency variation. We compared the clustering results with inferences from maximum parsimony (MP) and Bayesian phylogenetic analyses, for which we concatenated the genomic SNV dataset into one continuous sequence for each strain (total character=103,430). MP trees were identified using PAUP* 4.0b10 (Swofford 2003) by heuristic searches with TBR branch-swapping and the MULPARS option, and 100 random sequence additions. Bayesian analyses used MrBayes 3.2 (Ronquist and Huelsenbeck 2003), under the best-fit substitution model selected by the Akaike information criterion (AIC) implemented in JModelTest 0.1.1 (Posada 2008). Each run consisted of four incrementally heated Markov chains, using default-heating values. The chains were initiated from a random tree and were run for 2 million generations with sampling every 1000 generations. To assess the confidence of phylogenomic analyses, MP bootstrap (BS) values were calculated with 1000 replicates and the heuristic option (Felsenstein 1985) using PAUP*, and Bayesian posterior probabilities (PP) were inferred with a 50% majority-rule consensus tree that was sampled after the likelihood scores had converged, using MrBayes. The stationarity of likelihood scores for sampled trees was assessed in Tracer v1.5 (Rambaut and Drummond 2009), and the convergence was assessed using cumulative posterior probability plots in AWTY (Nylander et al. 2008) to assess split frequency within and between MCMC runs. The roots of the resulting trees were inferred by midpoint rooting. Mean nucleotide divergence (Dxy) was calculated using the Maximum Composite Likelihood method implemented in Mega 5.0 (Tamura et al. 2011), and was averaged

across 1,000 bootstrap replicates. The SNP character matrix used in the cluster and phylogenetic analyses is deposited in TreeBASE (TB..: http://www.treebase.org).

4.2.8 Gene genealogies and concatenated data phylogeny

To assess the monophyly of phylogenetic clades (or to assess biogeographic traits) resolved using the genome-wide SNP dataset, we randomly selected nine gene loci (Appendix B.5) that showed putative fixed differences between distinct Gs and Gc populations and sequenced them in 16 additional strains (Table 4.1). For each of the nine gene datasets, we generated MP and statistical-parsimony genealogies using PAUP and TCS v. 1.13 (Clement et al. 2000). Gaps were treated as missing data and no weighting was introduced in the single-gene analyses. The nine gene loci were concatenated to conduct maximum likelihood (ML) analysis (with 1000 nonparametric replicates bootstrap) using RAxML-VI-HPC 7.0.4 (Stamatakis 2006), as well as the weighted parsimony, with the weighting inversely proportional to the number of parsimony informative characters at each locus. We also performed Bayesian analyses for each gene and for the combined dataset. For Bayesian and MP analyses and for assessing their confidence and best-fit model of sequence evolution, we used the same criteria as those applied to construct SNP phylogenies. Monophylies supported by both BS ≥70% and PP≥95% were considered significant. The multigene dataset is deposited in TreeBase (TB.. http://www.treebase.org).

4.2.9 Detecting signature of selection and rate of protein evolution

For selection analyses with Gs-Gc multiple alignments, we first searched for genes that were orthologous to the 8,312 gene models of the reference strain slkw1407. We found an average of 8,064 orthologs for the eleven assemblies, ranging from 7,876 to 8,222 in Gs and 7,973 to 8,198 in Gc (Appendix B.6). We retained orthologs to 7,340 slkw1407 genes that matched at least four of the eight Gs and/or at least two of the three Gc genomes (n=972) and removed 3,864 of these because they had either fewer than two coding differences (n=3,377) or zero divergence (D_N+D_S=0; n=487). The selection analyses included the remaining 3,476 orthologs, which contained 19,616 nucleotide differences in coding sequences with a median size of 1,749 aligned bases, after excluding the gaps (Appendix B.7). The average number of Gs–Gc genomes in the aligned datasets was n=8.6.

To detect signatures of positive selection in *Grosmannia* we applied different methods to Gs and Gc gene predictions that were orthologous to the slkw1407 gene models. First we compared polymorphisms within Gs (n=4–8 strains) with fixed substitutions (i.e. divergence) between Gs and Gc sequences. We considered synonymous and nonsynonymous differences and used two Gc strains from *P. jeffreyi* and *P. ponderosa* as the outgroup taxa. We used Gs-Gc multiple alignments of all genes with at least two coding differences that were aligned by MAFFTv7.023 (Katoh et al. 2002) for their entire coding regions, and applied the McDonald–Kreitman tests (MK;

McDonald and Kreitman 1991) implemented in the MK.pl script (Holloway et al. 2007). We assessed whether the ratio of nonsynonymous and synonymous was statistically independent of differences being polymorphic (P_N:P_S) or divergent (D_N:D_S), using Fisher's exact test. For each gene, MK results for the direction and degree of departure from neutrality were summarized using the neutrality index (NI; Rand and Kann 1996), after adding one pseudocount to each mutation class to eliminate zero counts. We also reported an unbiased NI estimate for differences across all the genes (NITG; Stoletzki and Eyre-Walker 2011).

Next, we applied maximum likelihood methods implemented in the Codeml from PAMLV4.0 (Yang 2007). We estimated Gs–Gc pairwise distances at nonsynonymous (dN) and synonymous (dS) sites for each gene, by setting parameters as follow: seqtype = 1, CodonFreq = 2, Runmode = -2 and the transition-transversion ratio κ estimated from the data (Goldman and Yang 1994). For this test, we used pairwise alignments of single coding sequences from each species, generated for all Gs and Gc strains; the number of pairwise comparisons ranged from 8 to 27 per gene. Threshold dS values were determined by plotting dN as a function of dS, excluding outliers from the main distribution. To test for further evidence of positive selection, we applied the "site-specific" models M1a/ M2a and M7/M8 (Nielsen and Yang 1998). Only gene alignments displaying more than three fixed ($D_N+D_S \geq 3$) and/or synonymous ($D_S+P_S \geq 3$) differences were considered for this additional test (Stoletzki and Eyre-Walker 2010). M1a assumes that codons contain only $0<dN/dS<1$ or $dN/dS=1$. We compared this with the alternative model M2a, which allows dN/dS for a site to be less than, equal to or

greater than 1. If dN/dS is significantly greater than 1, then adaptive substitutions are assumed to have occurred to fix nonsynonymous differences between species. If dN/dS<1, adaptive evolution may still have occurred on some fraction of all differences, but cannot be inferred with certainty. We also compared the null model M7, which assumes a beta distribution of 0≤ dN/dS≤ 1 across sites with the alternative model M8, which allows for positive selection . The log-likelihoods for the null and alternative models were used to calculate a likelihood ratio test statistic, which was then compared against the χ^2 df = 2 distribution (Yang 2007). The positive selection hypothesis was accepted if both alternative models M2a and M8 provided a statistically significant better fit to the data. For all the analyses, we removed low frequency polymorphisms (singletons) to avoid biases caused by slightly deleterious mutations regarding the prevalence of adaptive divergence (Fay et al. 2001; Li et al. 2008).

4.2.10 Physiological assessments

We characterized the monoterpene utilization of (+)-limonene as a carbon source by *Grosmannia* strains from three different pine trees *P. contorta*, *P. jeffreyi* and *P. ponderosa.* For this experiment, we selected five Gs and Gc strains from independent samples of each tree species (total n=15, Table 4.1). The three-day fungal cultures actively growing on MEA were transferred to glass plates containing yeast nitrogen base minimal medium (0.17% YNB, 1.5% granulated agar), where 200 μl of (+)-limonene (Sigma, Oakville, ON) were added onto two (2 × 4 cm) strip filter papers that

were placed inside the lid of each glass plate. The plates were sealed with DuraSeal film (Laboratory Sealing Film; VWR, Mississauga, Ontario, Canada) and incubated at 22°C in a sealed glass container. Limonene was re-supplied biweekly; after six weeks, the mycelial plugs treated with limonene were transferred to normal MEA plates to check whether the fungus was killed or survived the chemical treatment. The control was YNB minimal medium without monoterpene.

4.3 Results

4.3.1 Genome assembly, orthologs determination and SNV variants

For the eleven *Grosmannia* strains, we obtained assemblies ranging from 27.7 to 32.4 Mb (Appendix B.2). We found no significant evidence of genome rearrangements for any of the sequenced strains compared to the slkw1407 reference genome (Appendices B.2 and B.13). The eleven strains shared more than 8000 genes with an average sequence identify of 98±0.4% between Gs and Gc genomes. On average only 3% of genes were missing or highly divergent (<70% sequence identity) relative to the reference gene models (Appendix B.6 and figure 4.1b).

Assessing coverage for variant calling, we noted that between 86 and 99% of the filtered reads mapped to the slkw1407 genome sequence, providing an average read depth between 22x to 58x per strain (Table 4.2). On average, 94.1% of the slkw1407

genome (i.e. ~27.4 Mbp) was covered by ≥ 5 uniquely mapped reads, with a range of 90.0% to 97.8% coverage across the eleven genomes. Coverage for slkw1407 gene-prediction models on average showed slightly higher depth than for the whole genomes (Table 4.2).

We compared the variants called by SAMtools and GATK, which showed a high percentage of overlapping SNVs (n=91,763) between the two methods, and used the SAMtools results because it generated fewer unique calls (12.7% of total 105,104) than GATK (21.9% of total 117,449). Of 198,362 putative variants, 105,104 SNVs and 9,907 indels passed quality control and filtering, yielding 115,011 high-confidence differences across the twelve *Grosmannia* genomes. After removing ambiguous calls that are likely to represent errors in the reference genome assembly, we obtain 103,430 SNV sites with a mean transition-to-transversion ratio of 3.4 (Appendix B.9). We estimated a false positive rate of $4.4 \cdot 10^{-6}$ or one per 24,590 nts and a false negative rate of 0.046% for the sequenced regions.

4.3.2 Functional classification of genomic variants

We classified nucleotide variants for their potential functional and/or adaptive significance by characterizing the level of intra- and interspecific differences in different genomic regions. From 103,430 SNVs across twelve *Grosmannia* genomes, we identified 36,017 variants within the slkw1407 gene models. Of these genic variants,

5,826 were intronic and 30,191 were in coding exons, 14,889 of which were synonymous and 15,302 nonsynonymous (Appendix B.9). Of the nongenic variants, 24,589 were located in our predicted ~6000kb gene-flanking regions and 42,880 were intergenic. Because gene models in slkw1407 can overlap (DiGuistini et al. 2009), 56 of the genic SNVs were identified in more than one gene region (e.g. a variant in a coding region as well as in an intronic region). Among the coding variants, 262 variants in 218 genes in the total *Grosmannia* genomes were predicted either to cause a premature stop codon (n=226) or to eliminate a stop codon (n=36). Of these 262 variants, 92 that had truncated proteins and 3 that had lost a stop codon occurred in only one genome, 155 were found in at least two genomes, and 12 were observed in all eleven genomes. The latter 12 variants may indicate that the slkw1407 genome sequence has an error or a low frequency allele in these positions. For this analysis we removed the 12 variants that occurred in all eleven genomes, as well as the 95 SNVs that were found in only one genome, which removed 85 genes. Of the remaining 133 genes, 85 were slkw1407 gene-models with known functions (n=86 for 71 genes with premature stop SNV; n=14 for 13 genes with stop-loss SNVs; and one gene showed both a stop-gain and a stop-loss SNV; Appendix B.10). Blast2Go enrichment analysis of genes with known functions identified enrichment of stop-codon variants for members with oxidoreductase activity (31%, $p < 0.001$) within both biological process (BP) and molecular function (MF) classifications, followed by genes involved in transmembrane transporters (16.9%) and nucleotide binding activities (18%) in BP and MF, respectively (Appendix B.11). Some of the enriched oxidoreductases belonged to gene families with known roles in detoxification (Appendices B.10–11). For example a flavoprotein monooxygenase

(CMQ-6740) in the slkw1407-gene cluster (figure 4.2a–b), which was reported to have a role in detoxification and/or utilization of host-tree defense chemicals (DiGuistini et al. 2011), showed a stop-codon in both Gc strains from *P. jeffreyi.* For this gene, we confirmed the variant by slkw1407 EST and RNA-seq data (Appendix B.10), as well as by an independent PCR validation of additional strains (total Gc n=16, Gs=12 and two other species n=4), showing that this mutation is unique to the Gc strains from *P. jeffreyi* (figure 4.2c).

4.3.3 Divergence classification of genomic variants

Across the twelve *Grosmannia* genomes, approximately 67% (n=70,018) of the total number of SNVs were parsimony informative in that multiple strains contained alternate nucleotide bases. The remaining SNVs (n=33,412) were unique differences (i.e. singletons) in that only one strain showed the alternate nucleotide base. To characterize intra and interspecific variants, we assigned the informative polymorphisms to three classes: fixed, exclusive and shared (Table 4.3). Most SNPs were either fixed (n=37,712), or were exclusive to the nine Gs (18,871) or the three Gc (n=9,685) strains; the rest (n=3,750) were present in both species. Within Gs, the eight re-sequenced genomes differed from the slkw1407 reference genome by an average of 12,859 SNPs and 3,315 short indels, corresponding to one SNP per 2,133 nucleotides in the ~27.4 Mbp covered regions (Tables 4.2, Appendix B.9). In contrast, the three Gc strains showed an average of 61,512 SNPs and 6,878 short indels, corresponding to one SNP

per 446 nucleotides. The mean single nucleotide divergence between the Gs and Gc genomes was estimated at 1.66 (± 0.006), which was respectively ~7 and 11 times higher than mean intraspecific divergence for Gc (0.24 ± 0.002) and Gs (0.15 ± 0.0006).

4.3.4 Clustering and phylogenomic analysis of SNVs

We assessed genetic distance and phylogenetic relationships among *Grosmannia* genomes by the AWclust non-parametric clustering (Gao and Starmer 2008) and phylogenetic analyses of SNV data. The AWClust resolved *Grosmannia* genomes into four clusters corresponding to Gs and Gc lineages that each formed additional sub-clusters according to the geographic regions and host tree associates of the fungal taxa (Appendix B.14). The maximum parsimony (MP) and Bayesian phylogenetic trees supported the results from cluster analysis and showed identical tree topologies that only differed in the placement of the slkw1407 reference strain either within the Gs isolates from Alberta or those from Rocky Mountains. Here we only showed the MP (figure 4.1a), which was best described by a single unrooted tree with consistency index of 0.79, and 0.97 when including only Gs (GsRef, GsB3, GsC1) and Gc (GcC2 and GcB1) from distinct populations. The MP tree provided high statistical support (BS = 100% and PP = 1.0) for the positioning of Gs and Gc strains into two divergent clades, and for additional subclades within each clade. As expected, slkw1407 grouped within the *P. contorta*–infesting Gs strains, which formed a distinct clade from the Gc strains. In the Gc clade, the *G. clavigera* holotype that had been isolated from MPB-infested *P.*

ponderosae was in a different subclade than the two *P. jeffreyi* associates. Within the Gs, strains from MPB-epidemic regions in BC, Alberta and Rocky Mountains were significantly separated from the two strains from the localized California population. This pattern was also consistent with the SNP density for the latter two genomes, which showed almost twice as many differences as the reference strain and the epidemic strains (Appendix B.9).

4.3.5 Ecological and physiological assessments

To support SNP phylogenetic relationships among the twelve *Grosmannia* genomes and to assess the host and distribution ranges of distinct lineages, we sequenced nine gene loci (Appendix B.5) in 16 additional strains from localized populations of MPBs and JPBs in their respective host trees *P. contorta, P. ponderosa* and *P. jeffreyi* (Table 4.1). Genealogies from each of these genes (Appendix B.15), as well as, the concatenated phylogeny (figure 4.3a) confirmed the genome-wide SNV results noted above. The single-locus genealogies supported the monophyly of the entire Gc and Gs clades in eight of the nine gene trees (BS ≥70% and PP ≥0.95, Appendix B.15). Within Gc, seven gene trees separated the taxa associated with JPBs (n=10) in California from the MPB associates infesting *P. ponderosa* trees in BC, California and South Dakota (n=6). The Gc–*P. ponderosa* subclade was statistically supported in the concatenated phylogeny (figure 4.3a), as well as in one single-gene tree (CMQ6965–ABC.C, Appendix B.15). The phylogeny from concatenated loci was

also consistent with geographic isolation within Gs, with five strains from the localized population in California forming a monophyletic clade separated from the epidemic strains; but no single-gene tree or the concatenated dataset showed statistical support for the California localized group. The concatenated matrix of nine gene sequences resulted in 14,308 aligned nucleotide positions, 296 variable sites and 143 informative characters (Appendix B.5). The nine-gene species tree showed identical topology based on ML, MP and Bayesian analyses with minor differences in the placement of terminal taxa (ML, figure 4.3a).

Consistent with results from the genome-wide SNP analyses and nine-gene phylogenies, we showed that while *P. ponderosa* and *P. jeffreyi* associates are genetically very close they can be characterized with distinct pattern of (+)-limonene utilization. Consistent with *P. jeffreyi* producing a lower level of limonene than *P. contorta* and *P. ponderosa,* we found that no Gc isolates from *P. jeffreyi* grew on (+)-limonene minimum media, in contrast to all Gc isolates from *P. ponderosa*, as well as to all Gs and the closely related species from *P. contorta*, which did grow (figure 4.3b).

4.3.6 Signature of positive and purifying selections in *Grosmannia*

To test for positive adaptive selection in Gs–Gc orthologs, we compared the ratio of SNPs within Gc with the sequence divergence between Gs and Gc at nonsynonymous and synonymous positions (figure 4.4a–c). Under a neutral model of molecular

evolution, the ratio of polymorphism to divergence should be comparable between these two site classes. We used the neutrality index (NI) to quantify deviations from the neutral expectation in 3,476 *Grosmannia* orthologs. Assuming that synonymous positions are selectively neutral, positive selection should result in an excess of amino acid divergence ($-\log_{10}NI>0$) while purifying selection will show excess of nonsynonymous polymorphisms ($-\log_{10}NI<0$), indicating weakly deleterious variations segregating within species. For Gs–Gc orthologs, we obtained a median $-\log_{10}NI$ value of less than zero (–0.05; Appendices B.7, B.16), which suggested that the majority of genes (n=1,755) in our dataset are subject to weak purifying selection. We also detected a statistically significant ($p < 10^{-05}$) signal of purifying selection in the pooled analysis of all 3,476 genes ($-\log_{10}NI_{TG} = -0.11$, pooled $P_N = 3,834$, $D_N = 5,903$, $P_S = 3,267$, $D_S = 6,612$). But only five genes showed significant evidence of purifying selection on a per-gene basis. While 1,215 genes showed $-\log_{10}NI > 0$, indicating fewer amino acid polymorphisms within Gs relative to those between Gs and Gc, we only found 11 genes with statistically significant ($p \leq 0.05$) excess of protein divergence between the two species. Six of the 11 genes were among the 42 *Grosmannia* orthologs showing an excess of nonsynonymous fixed differences between Gs and Gc (i.e. the 1.2% of the 3,476 Gs–Gc polymorphic genes having $D_N \geq 9$; Appendix B.7 and figures 4.4a–c). Among genes exhibiting the strongest evidence for positive selection (i.e. $-\log_{10}NI > 0$), we noted polyketide synthases (PKS; CMQ_4392, _5323, _5095 and _2677), a non-ribosomal peptide synthase (NRPS; CMQ_3566), ABC transporters (CMQ_6634,_6965, _6960), oxidoreductases (CMQ_1999, _ 5949) and an heterokaryon incompatibility gene (CMQ_742) (Table 4.4 and Appendix B.7). However,

no genes were significant for either positive or purifying selection after correction for multiple testing (Benjamini and Hochberg 1995).

For detecting positive selection in our dataset, we also applied codon-based models and likelihood estimates of dN, dS and dN/dS (ω) ratios. Divergence estimates were made from Gs-Gc pairwise alignments of the 3,476 orthologs using a codon substitution model that takes into account possible biases such as codon preference and nucleotide composition (Yang and Nielsen 2000). We estimated mean dS $0.0032 \pm 8.9 \cdot 10^{-19}$, corresponding to an average of one mutation per 312 synonymous sites between Gs and Gc since the common ancestor. This number was lower than the genome-wide average (one mutation per 446 nts, relative to the slkw1407 reference strain), presumably due to selective constraints in the coding regions. The mean pairwise dN was lower than dS ($0.0011 \pm 4.8 \cdot 10^{-5}$), reflecting the expected stronger constraints on substitutions that changed amino acids. The overall mean for dN/dS in the 3,476 orthologous genes (i.e. excluding 289 genes with dS equal to zero) was 0.3 ± 0.005. This value was similar to the $-\log_{10} NI_{TG} = -0.11$ value obtained for the MK test, suggesting that a large majority of genes are conserved and evolve with dN/dS less than one (Appendices B.7, B.16).

The pairwise dN/dS ratio is a measure of the overall evolutionary constraint averaged across the sequences of the gene and may be too conservative for detecting positively selected sites along a gene. Thus we applied “site-specific” models to test for

further evidence of positive selection within a more divergent subset of the 3,476 orthologous genes, removing 2,263 genes that had fewer than three fixed (n=1,567) and/or synonymous (n=696) differences. For the remaining 1,213 genes, the site-based approach identified 82 genes statistically significant for the positively selected sites (ω>1; $p \leq 0.05$). For the majority of these significant genes (n=46), the MK test also estimated a summary statistic of positive selection $-\log_{10}NI > 0$, indicating an excess of protein divergence by both methods (Appendices B.7–8). The genes exhibiting the strongest evidence for positively selected sites include polyketide synthases (PKS; CMQ_5095, _2687, _2677), an ABC transporter (CMQ_6965), CYP450s (CMQ_3491, _ 6107, _4067), oxidoreductases (CMQ_277, _5685), ankyrin-repeat containing proteins (CMQ_1651, _ 569), a heat repeat protein (CMQ_7934) and an autophagy protein (CMQ_7167) (Table 4.4). The summary statistics on selection from MK and from PAML generally agreed (Appendix B.8). However, the two methods both identified significant signal for positive selection in only one gene (PKS_5095, Table 4.4). Another two PKS genes (CMQ_4392 and _2677) showed a significant or marginally significant signal for positive selection with both methods before correction for multiple tests. After correction for multiple tests, the signal was no longer significant (PAML P=0.07 and MK P=0.09). The number of 46 significant genes (4%) in our dataset is lower than the conventionally accepted significance level of 5% because majority of genes are conserved and evolve with ω less than one. Nonetheless, after correction for multiple testing, we identified at least seven genes that evolved with ω greater than one (P<0.0001). This indicated that even though the level of divergence between the Gs and Gc was low, there is statistically significant evidence for site-specific positive selection between *Grosmannia*

species. Results for all the genes are available in Appendices B.7–8; Table 4.4 shows only the genes with the strongest evidence of positive selection using both PAML “site-model” and the MK test.

4.4 Discussion

In this study, to identify features common across distinct *Grosmannia* populations and species, we compare the genomes of twelve *Grosmannia clavigera sensu lato* strains, representing two known sibling species that have different ecological characteristics (Massoumi Alamouti et al. 2011). We first used genome assemblies to assess changes in genomic structure such as rearrangements and gene gains/losses, and then focused on variation at the gene and nucleotide levels. We identified a number of functional variants in genes potentially involved in secondary metabolism and chemical detoxification, reflecting fungal adaptation to the specific chemistries of host trees. The data and results generated are a resource for assessing and characterizing fungal populations in the present MPB epidemic as it continues to spread into new habitats, including the *P. banksiana* boreal forest as well as in future MPB outbreaks. The approach described here can also be applied to other insect-vectored/tree-colonizing fungi.

4.4.1 *Grosmannia* draft genomes

Using Illumina sequencing, we assembled draft genomes from eleven *Grosmannia* strains that represent distinct populations of the two known cryptic species: Gs and Gc. We showed that the *de novo* assemblies in these fungi could be mapped over a large fraction of the *Grosmannia* reference genome, suggesting that the majority of the assembled contigs, and the genes they contain, lie in regions that are collinear within and between the cryptic species. The extensive similarities in gene content (large-scale synteny) and order (colinearity) (Hane et al. 2011) within a large fraction of aligned contigs suggested that the morphologically cryptic *Grosmannia* species have diverged recently (Appendix B.13). This is consistent with the previous gene genealogies of *Grosmannia clavigera sense lato* and few other close relatives, which suggested that these pine-infesting, beetle-associated taxa have yet to reach a reciprocal monophyly for all the loci (Massoumi Alamouti et al. 2011). While teleomorphs (i.e. sexual structures) have been found rarely in Gs and not yet in Gc from *P. jeffreyi*, evidence of repeat-induced point mutation (RIP) in the reference genome and recombination in the population, as well as the extensive synteny reported here suggest that both fungi have a sexual phase (DiGuistini et al. 2011; Massoumi Alamouti et al. 2011; Tsui et al. 2012).

Large scale structural changes can exceed nucleotide evolution in plant pathogens like *Mycosphaerella* and *Fusarium* spp., which retain lineage-specific chromosomal islands or even entire lineage-specific chromosomes (Cuomo et al. 2007; Stukenbrock

et al. 2010; Klosterman et al. 2011). In filamentous ascomycetes such structural changes may be attributed to relatively long divergence times or horizontal gene transfer (Hane et al. 2007; Desjardins et al. 2011; Hane et al. 2011; Klosterman et al. 2011). Here major structural changes that would uniquely distinguish the cryptic *Grosmannia* species were not evident in our draft assemblies. Instead, the distinct ecological differences and host preferences in these fungi appear to be driven mainly by local nucleotide changes.

4.4.2 *Grosmannia* genome-wide SNVs

Single nucleotide variations (SNVs) are the most abundant type of genetic variation reported for eukaryotic genomes (Brumfield et al. 2003; Morin et al. 2004). Detecting genome-wide nucleotide variants within and between species using high-throughput sequencing technologies depends on two factors: a) whether the non-reference allele are present in the strains sequenced, and b) the number of high-quality and accurately-mapped reads that overlap the variant sites. The greater than 100,000 novel SNVs that we identified occurred in similar densities in intergenic, regulatory and coding regions across the eleven strains, and provide the first comprehensive assessment of genome-wide intra- and interspecific nucleotide variants for this group of beetle-vectored fungal symbionts. These SNV calls likely somewhat underestimate the total nucleotide differences within and between *Grosmannia* genomes, given that at least 10% of the

reference genome had less than 5x read coverage – a limitation expected for Illumina sequencing of repetitive and GC-rich genomic regions (Li et al. 2008; Wang et al. 2011).

The genome-wide frequencies of nucleotide variants within *Grosmannia* species were lower than in other filamentous ascomycetes, including the plant pathogens *Magnaporthe oryzae*, *Mycosphaerella graminicola*, *Sclerotinia sclerotiorum* and different *Verticillium* and *Cochliobolus* species, as well as human pathogens in the genera *Coccidioides* and *Paracoccidioides*, and the generalist saprophyte *Neurospora crassa* and species in the genus *Aspergillus* (Lambreghts et al. 2009; Ma et al. 2010; Neafsey et al. 2010; Amselem et al. 2011; Andersen et al. 2011; Desjardins et al. 2011; Klosterman et al. 2011; McCluskey et al. 2011; Stukenbrock et al. 2011; Xue et al. 2012; Condon et al. 2013). In these fungal species whole-genome intraspecific SNV densities range from one per 865 nucleotides in the corn pathogen *C. heterostrophus* to one per 132 bases in the human pathogen *P. brasiliensis*. These numbers are higher than the *Grosmannia* intraspecific variant frequency of one per 2,133 nucleotides and often higher than nucleotide divergence between *Grosmannia* sister species (i.e. one per 446 bases). Our intraspecific SNV frequencies were comparable to those of *Fusarium graminearum*, a global pathogen of cereal crops (Cuomo et al. 2007). This pathogen is a sordariomycete like the ophiostomatoid fungi; it differs from other filamentous ascomycetes, including *G. clavigera*, because it is homothallic (i.e. self-fertile) and rarely out-crosses (Cuomo et al. 2007; Tsui et al. 2013). *F. graminearum*'s inbreeding may be associated with lower nucleotide diversity, as is the case in other fungal and Oömycetes genomes (Tyler et al. 2006; Cuomo et al. 2007). The frequency

of genome-wide SNVs in the opportunistic human pathogen *A. fumigatus* is similar to that for *Grosmannia*, and is surprisingly low compared to its close relatives (Rydholm et al. 2006; Rokas et al. 2007). *A. fumigatus*' low nucleotide variance and its lack of population structure globally have been explained by the worldwide spread of this fungus having occurred too recently for mutations to have accumulated within and between populations (Rydholm et al. 2006).

Differences in genome-wide frequency of SNVs among filamentous fungi may be in part due to differences in their life histories and dispersal processes. Ascomycetes comparative genomics have largely focused on saprotrophs that have broad host ranges and on pathogens that have the ability to survive for extended periods as free-living saprophytes without a specific host. Such fungi tend to have more stable population sizes and higher genetic variation in natural populations (Thompson 1994; Barrett et al. 2008). In contrast, fungal symbionts like *Grosmannia* have limited and specific ecological niches (beetle vectors and host trees) and are more likely to experience local population outbreaks, crashes and re-colonization than generalist and saprophytic fungi (Thompson 1994; Six and Paine 1999; Carroll et al. 2006; Smith et al. 2010; Roe et al. 2011; Tsui et al. 2012). After such crashes, long periods of low endemic population sizes are expected for both the beetle and its associated fungi. Such cycles promote loss of genetic variance within populations and generate between-population genetic differences, through genetic drift and adaptive selection. Consistent with the above, our results show that while *Grosmannia* fungi have lower overall genome-wide frequencies of nucleotide variants than other filamentous fungi, their

SNVs support distinguishing two cryptic species and also suggest phylogenetically and biogeographically structured lineages that may include at least one additional species.

4.4.3 *Grosmannia* SNV-phylogenomics

Using genome-wide SNVs, we generated a high-resolution phylogeny that separated the twelve *Grosmannia* strains into two divergent monophyletic clades, confirming our previous gene genealogy discrimination of the Gs and Gc sister cryptic species (Massoumi Alamoutiet al. 2011). If the two species share extensive polymorphism through introgression, or incomplete lineage sorting due to a recent split from a common ancestor, we would expect that inter- and intraspecific nucleotide differences would be correlated (Avise 2004; Kulathinal et al. 2009). Here, no such correlation was evident; interspecific nucleotide divergence was significantly ($p<0.01$) higher than the mean intraspecific variation within Gc and Gs, suggesting that gene flow between *Grosmannia* cryptic species was weak or absent. This was consistent with the low level of homoplasy in our SNV-phylogeny (CI=0.97), and with our previous gene genealogies using population-level samples (Massoumi Alamoutiet al. 2011). The statistical support for each *Grosmannia* SNV-phylogenetic group indicates that we can detect lineage-specific variants, and so may be able to identify functional variants that are likely important to Gs and Gc adaptation to distinct ecological niches or to divergence of other phylogenetic groups resolved here.

Our SNV-phylogeny divided the epidemic Gs strains into well-supported phylogenetic groups that were also identified previously using AFLP and microsatellite markers (Lee et al. 2007; Tsui et al. 2012). In addition, within Gs, we found a more divergent subclade, separating the strains from localized populations in California from the epidemic BC subpopulations. The average pairwise nucleotide divergence between California and epidemic phylogenetic groups were more than twice as large than divergences within and among epidemic groups, likely due to California location being distant, in the southernmost part of the species' range, along the Great Basin Desert (Wood 1982). While genetic structures within localized Gs populations have not been documented before, they have been reported for the MPB populations using AFLP markers (Mock et al. 2007). MPB populations in California were more divergent compared to those from other epidemic and most of the localized populations, consistent with our results on the fungal associate. This consistency reflects the co-evolutionary association between the beetle and the fungus, as suggested for other similar insect-fungal associations (Marin et al. 2009). MPB divergence based on AFLP makers was not significantly higher than expected for the isolation by distance, and it was suggested to correlate with the phylogenetic pattern of *P. contorta* trees experiencing a northward expansion into British Columbia and the Northwest Territories since the last glaciation period (Marshall et al. 2002; Mock et al. 2007). For the fungal associate, whether or not the Gs-California lineage warrants recognition as a species would require sampling additional isolates from the localized populations infesting *P. contorta* trees in the southern and eastern portion of the species' range, preferably using SNV makers optimized for this application (Morin et al. 2004). Our previous

network analysis on a 15-gene concatenated dataset of the population-level samples from California and epidemic regions had shown incongrurence among gene genealogies, inferring the evidence of either incomplete lineage sorting or recombination. Either of these processes could be occurring in Gs populations. They may well have resulted from a recent species divergence maintaining high population size during the ongoing epidemics, a typical scenario in incomplete lineage sorting. Recombination is also likely and indicative of the potential lack of species structures within Gs when phylogenomic analyses are applied to population-level samples.

Within the Gc clade, our whole-genome SNV-phylogeny indicates host-specific differentiation in *Grosmannia* by separating the JPB associate from the holotype isolated from MPB-infested *P. ponderosa* (PP) in BC (Robinson-Jeffrey and Davidson 1968). Consistent with these results, the protein-coding combined phylogeny of additional Gc strains suggested that one lineage (Gc-JP) is exclusively associated with the JPB infesting the host tree *P. jeffreyi* in California whereas the other (Gc-PP) was only found on MPBs infesting *P. ponderosa* trees. The Gc from *P. ponderosa* host species in different geographic areas (i.e. BC, South Dakota and California) was genetically closer than those collected from different host species (*P. jeffreyi*) in the same geographic region in California. While our data from *P. ponderosa* trees were limited, preventing us from assessing the extent of host-specificity across the MPB-localized USA populations, or the role of geographical isolation in speciation, overall, our results suggest that speciation process in these fungi can be attributed to the host-

tree species and the geographic isolation of the host species from the current epidemics.

The genome-wide SNV divergence between the Gc-JP and Gc-PP was only twice as large as the intraspecific differences, reflecting the recent divergence of these lineages. A recently diverged population may represent an early stage in speciation, which begins when populations become genetically separated through geographical isolation or through ecological selection, and when adaptation acts as barrier to gene flow, and leads to genetically cohesive populations that are called species because they are ‘segments of separately evolving lineages’ (de Queiroz 2007). The genealogical nondiscordance criterion (Dettman et al. 2003) and the phylogeny of nine informative (i.e. genes randomly selected because of their potential fixed differences between the *P. ponderosa* and *P. jeffreyi* associates) protein-coding loci suggest that Gc-JP and Gc-PP are independent evolutionary lineages. The SNV-phylogeny and gene genealogies were further supported by our current ecological data showing that each lineage was associated with distinct beetle and tree host species. Further characterization of lineage-specific SNVs at a population level would strengthen evidence for the work reported here, which suggests that lineages within Gc likely warrant recognition as genealogical and ecological species.

4.4.4 *Grosmannia* genes involved in host adaptation and ecological divergence

A combination of life-history traits and selection imposed by host trees may have promoted speciation and ecological divergence in *Grosmannia* lineages, as shown for many plant pathogenic fungi (Giraud et al. 2006; Stukenbrock and McDonald 2008; Giraud et al. 2010). In pine trees, phenolics and terpenoids from oleoresin are key constitutive and inducible chemical defenses (Boone et al. 2011). While monoterpenes (e.g. β-phellandrene and limonene) and heptane (a straight-chain alkane) are toxic to many pathogens and insects, beetle-fungal complexes have evolved efficient mechanisms to survive and become established in such environments (DiGuistini et al. 2011; Wang et al. 2013). For Gs, functional genomics and transcriptomic data suggest that ABC transporters, genes associated with oxidative stress responses and fatty acid β-oxidation pathways, and gene clusters that contain cytochrome P450s, dehydrogenases, and monooxygenases are involved in overcoming tree defenses (Hesse-Orce et al. 2010; DiGuistini et al. 2011; Wang et al. 2013). However, chemical defense systems differ quantitatively and qualitatively between species of pine and between different populations within a pine species (Keeling and Bohlmann 2006; Gerson et al. 2009; Boone et al. 2011). For example, *P. jeffreyi* has lower level of limonene and higher level of heptane than *P.* contorta (Mirov and Hasbrouck 1976; Paine and Hanlon 1994; Smith 2000). Limonene is one of the most toxic defense chemicals for bark beetle-fungal complexes (Raffa 2001; Raffa et al. 2005); it influences MPB-attack density in epidemic regions and it is found at high concentrations in *P. ponderosa* populations that have been subject to beetle-fungal outbreaks (Sturgeon

1979; Clark et al. 2010). Given this, host preferences among *Grosmannia* lineages may reflect different abilities to survive and adapt to host chemicals or to other biotic and abiotic stresses inside the host.

Changes in gene contents and in gene products are central mechanisms in fungal genome evolution. Genes lost or in the process of being lost through pseudogenization have been shown in plant pathogens (Stukenbrock et al. 2010; Marcet-Houben et al. 2012; Raffaele and Kamoun 2012; de Wit et al. 2012) and in the closely-related human-pathogenic yeasts *Candida albicans* and *C. dubliniensis* (Moran et al. 2011). Similarly, 1.3% of the protein-coding genes in the Gs and Gc genomes contain premature stop codons, indicating that the genes may have been pseudogenized. Twenty-two of these genes have oxidoreductase activity, including those with known roles in stress response and detoxification like short-chain dehydrogenases, cytochrome P450s and monooxygenases. Among those, twenty appear to have been lost in both Gc–PP and Gc–JP, or in only one of these lineages. For example, a flavoprotein monooxygenase identified in the Gs gene cluster potentially involved in terpenoid detoxification and/or utilization has been pseudogenized in all the Gc-JP strains tested (n=10). Sequencing of additional Gc-PP and Gs strains and two related species confirmed that the stop codon is unique to the *P. jeffreyi* associates. Consistent with our genomic and sequencing results, all *Grosmannia* fungi including three species from *P. contorta* (Gs, *L. longiclavatum* and *L. terebrantis*) and Gc strains from *P. ponderosa* were able to grow on limonene as a sole carbon source, but none of the Gc strains from *P. jeffreyi* grew or survived in this condition. Combined, these results suggest that a number of

genes with potential roles in Gs host-adaptation are inactivated or are evolving to become pseudogenes in Gc lineages. For example, because *P. jeffreyi* produces lower levels of monoterpenes (including limonene) than pine species in epidemic and localized populations, the Gc-JP lineage may no longer require certain genes. The Gc lineages likely have more pseudogenes than we report here, because we have characterized only those caused by stop-codons, and not those due to indels and/or frameshift mutations.

We assessed both purifying and positive selection in *Grosmannia* protein-coding genes. Under the assumption that synonymous changes are neutral, purifying selection is inferred when the ratio of nonsynonymous to synonymous substitutions is less than 1, while positive selection pressure is usually inferred when the ratio is greater than 1 (Wright and Andolfatto 2008). Our genome-wide characterization of Gs–Gc protein-coding evolution showed that most genes evolve under purifying selection ($dN/dS=0.3\pm0.005$). Similar results have been reported for other filamentous ascomycetes, reflecting overall evolutionary constraints on protein-coding genes (Gu et al. 2005; Nielsen et al. 2005; Rokas 2009; Sharpton et al. 2009; Stukenbrock et al. 2010; 2011). In contrast, among all the variable *Grosmannia* genes, only 46 showed significant evidence for positive selection (i.e. before correction for multiple testing, $p<0.05$), which is not surprising given the close similarity between Gs and Gc orthologs ($dS=0.0032$). We note, however, that current divergence-based selection methods have limited statistical power for closely related species (Li et al. 2008; Oleksyk et al. 2010), and consequently we may have missed some genes with weaker signs of selection. For

instance, sequence diversity and divergence in our data suggested that 1,215 candidate genes were showing some signs of adaptive selection, but the test was only significant for eleven genes ($p<0.05$).

The most significant examples of evidence for positive selection were the four PKSs, one NRPS and three ABC transporters. The PKSs and NRPS families are key enzymes for producing secondary metabolites, which are involved in fungal host-colonization and pathogenicity (Kroken et al. 2003; Collemare et al. 2008; de Wit et al. 2012). Based on our ABC domain and phylogenetic analyses (data not shown), the three membrane transporters are classified in the ABC-C subfamily, and so have potential roles in either host-chemical defenses or secondary metabolite export (Kovalchuk and Driessen 2010). Other genes with putative functions in chemical detoxification or utilization included an oxidoreductase, an isoflavone reductase and three cytochrome P450s (DiGuistini et al. 2011; Lah et al. 2013). We also found that some of these genes had putative role in nutrient uptake (a ferric reductase and a mono-carboxylate permease). Other genes were potentially involved in cell signaling (e.g. histidine kinase, phospholipase), fungal development and growth (e.g. membrane copper amine oxidase, cell morphogenesis, autophagy protein, heat-repeat protein, hit finger domain protein); and a few with putative roles in protein-protein interactions or self/nonself recognition (e.g. two ankyrin repeat proteins and a heterokaryon incompatibility protein) (Luhtala 2004; Fedorova et al. 2005; Bahn et al. 2006; Kohler et al. 2006; Liu and Gelli 2008; Soanes et al. 2008; Pollack et al. 2009). In summary, for *Grosmannia* lineages that are adapted to different pine trees, our results suggest that

many of the genes that are evolving under positive selection are involved in secondary metabolite synthesis and secretion, host-chemical detoxification and stress responses, nutrient uptake from the host plants, and hyphal growth and differentiation.

In conclusion, we have used the *Grosmannia* genomes to show relationships between ecology and biological functions that are maintained or that diverge during colonization of a range of pine host trees, which are themselves adapting to changing environmental conditions. Although the fungal population has expanded and contracted repeatedly over at least several hundred years, large-scale synteny, with conserved gene content and order, suggests that these closely related strains adapt to different pine hosts largely through local nucleotide changes. This initial genome-wide SNV dataset is a phylogenetic resource that can be extended into a more comprehensive characterization of *Grosmannia* ecology and population structure.

4.5 Tables and figures

Table 4.1 Fungal strains used in this study

Fungal species	Beetle associate	Host tree	Collection site (Map no. [a])	Source [b]	Code [c]
***Grosmannia* sp. (Gs)**	*Dendroctonus ponderosae*	*Pinus contorta* (*Pc*)	Canada, BC, Kamloops (1)	**(UAMH 11150)***	GsB1
			BC, Houston (2)	**UAMH 11153**	GsB2
			.	**(UAMH 11348)**	GsB3
		Pc x P. banksiana	Canada, Alberta, Fox Creek (3)	**(UAMH 11353)**	GsA1
			.	**(UAMH 11354)**	GsA2
		Pc	Alberta, Cypress Hills (4)	**(UAMH 11347)**	GsA3
			USA, Montana, Hidden Valley (5)	**(UAMH 11156)**	GsM1
			USA, California, Sierra Nevada (6)	**(UAMH 11349)**	GsC1
			.	**UAMH 11350**	GsC2
			.	CB 67F21	GsC3
			.	UAMH 11361	GsC4
			.	UAMH 11362	GsC5
***Grosmannia clavigera* (Gc)**	*D. ponderosae*	*Pinus ponderosae* (*Pp*)	BC, Cache Creek (7)	**(ATCC 18086)**	GcB1
			USA, South Dakota, Black Hills (8)	UAMH 11369	GcS1
			.	(UAMH 11370)	GcS2
			.	(UAMH 11371)	GcS3
			.	(CB 15B29C2)	GcS4
			.	CB 23B110C5	GcS5
			.	CB 32B85C10	GcS6
			USA, California, Sierra Nevada (6)	(UAMH 11372)	GcC11
			California, Lassen (9)	UAMH 11373	GcC12
	D. jeffreyi	*Pinus jeffreyi* (Pj)	California, San Bernardino (10)	**(UAMH 11351)**	GcC1
			.	**(UAMH 11352)**	GcC2
			.	DLS 1560	GcC3
			.	DLS 1595	GcC4
			California, Lassen (9)	UAMH 11377	GcC6

Fungal species	Beetle associate	Host tree	Collection site (Map no. [a])	Source [b]	Code [c]
			.	(DLS 210)	GcC7
			California, Sierra Nevada (6)	(C 843)	GcC5
			.	(UAMH 11375)	GcC8
			.	DLS 681	GcC9
		Pj x Pp	California, Lassen (9)	UAMH 11374	GcC10
Leptographium terebrantis	*D. ponderosae*	*Pc*	BC	UAMH9722	
		Pc x P. banksiana	Canada, Saskatchewan	AU 123-113	
L. longiclavatum	*D. ponderosae*	*Pc*	BC	CB LKG+T2B	
	D. ponderosae	*Pc*	BC	UAMH 4876	

[a] General map location of collection sites corresponding to Appendix B.12

[b] *, The fungal strain used for the reference genome published by DiGuistini et al. 2011; Strains selected for Illumina sequencing are bolded and fungal strains used for the physiological assessment are shown in parentheses; Source of isolates: UAMH, University of Alberta Microfungus Collection and Herbarium, Canada; ATCC, American Type Culture Collection, USA; Isolates beginning with CB, DLS, AU, and C are from culture collections of C. Breuil, University of British Columbia, Canada; D.L. Six, University of Montana, USA; A. Uzunovic, FPInnovations, Canada; and T.C. Harrington, Iowa State University, USA, respectively;

[c] Letters indicate the location and numbers indicate the number of isolates from each location.

Table 4.2 Summary of the genomic and gene coverage data in the eleven sequenced genomes

IDs*	Covered genomic bases	Genomic coverage	Covered gene bases	Gene coverage	Unmapped reads
GsB1 (control)	28,791,583 (98.8%)	70×	15,507,591 (99.9%)	47×	0.7%
GsB2	27,370,438 (94.0%)	27×	15,333,848 (98.8%)	30×	1.2%
GsB3	28,464,740 (97.7%)	48×	15,316,500 (98.7%)	48×	2.6%
GsA1	28,488,061 (97.8%)	58×	15,334,490 (98.8%)	61×	3.2%
GsA2	26,491,256 (91.0%)	35×	13,533,085 (87.2%)	78×	5.0%
GsA3	27,146,569 (93.2%)	24×	15,203,277 (98.0%)	29×	0.8%
GsM1	27,197,413 (93.4%)	22×	15,170,992 (97.8%)	33×	1.3%
GsC1	28,153,881 (96.6%)	50×	15,092,814 (97.3%)	57×	4.1%
GsC2	28,084,478 (96.4%)	49×	15,031,784 (96.9%)	63×	4.0%
GcB1.a	26,267,089 (90.1%)	14×	13,636,725 (87.9%)	20×	6.8%
GcB1.b	28,360,091 (97.4%)	55×	15,399,793 (99.3%)	50×	10.1 %
GcB1.ab	28,455,969 (97.7%)	79×	15,435,487 (99.5%)	77×	9.2%
GcC1	27,001,022 (92.7%)	25×	15,204,933 (98.0%)	29×	2.4%
GcC2.a	26,377,060 (90.5%)	36×	13,414,391 (86.5%)	54×	13.4 %
GcC2.b	27,130,509 (93.2%)	42×	14,137,799 (91.1%)	56×	12.1%
GcC2.ab	27,940,266 (95.9%)	72×	14,868,092 (95.8%)	87×	13.6 %
Average	27,425,586 (94.1%)	37×	14,701,029 (94.7%)	47×	5.1%

*IDs, "a" and "b" are results from two independent sequence lanes for the same strain, "ab" results from two sequence runs combined for the same strain. The estimated coverage are based on filtered reads mapped to the slkw1407 reference genome sequence, which is ~ 29.1 Mbp after excluding gaps.

Table 4.3 Genome-wide characterization of fixed and shared polymorphisms between Gs and Gc lineages

Genomic Regions	Fixed [a]	Shared [b]	Exclusive [c] to Gs (parsimony-informative)	Exclusive to Gc (parsimony-informative)	Total (parsimony-informative)	Dxy (± SD)
Total	37,712	3,750	35,765 (18,871)	26,203 (9,685)	103,430 (70,018)	1.66 (±0.006)
Intergenic	10,818	2,458	14,811	14,793	42,880	nc
Flanking regions	11,148	559	8,308	4,574	24,589	
Intronic	2,755	112	2,001	958	5,826	
Synonymous	6,808	353	4,984	2,744	14,889	
Nonsynonymous	6,116	264	5,601	3,059	15,040	
stop gain – stop lost	73 – 18	4 – 1	80 – 14	69 – 3	226 – 36	

[a] Fixed polymorphisms are nucleotide sites, at which all Gs strains differ from all strains of Gc.

[b] Shared polymorphisms are sites for which multiple nucleotides are found in both Gs and Gc strains.

[c] Exclusive polymorphisms are those that are polymorphic in one species and invariant in the other.

nc, not calculated.

Table 4.4 The top 42 genes showing evidence of positive selection [a]

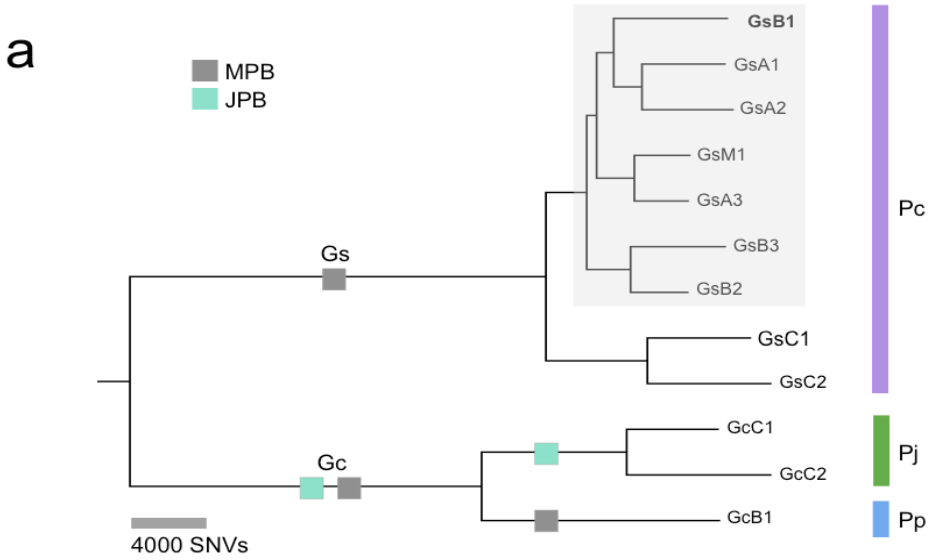
Genes	Gene description	MK [b]					PAML [c]			
		P_N	D_N	P_S	D_S	p-value	M1a–M2a	M7–M8	p-value	dN/dS (%)
CMQ_6634	ABC transporter	1	10	3	1	*	ns	ns	ns	ns
CMQ_1999	Ferric reductase	0	5	2	0	*	.	.	.	.
CMQ_5949	Putative oxidoreductase	0	5	3	1	*	.	.	.	.
CMQ_742	Heterokaryon incompatibility	1	9	3	1	*	.	.	.	.
CMQ_3665	Thermotolerance protein	1	4	4	0	*	.	.	.	.
CMQ_3566	NRPS	0	21	4	13	*	.	.	.	.
CMQ_5323	Polyketide synthase	1	17	4	5	*	.	.	.	.
CMQ_5095	**.**	**1**	**20**	**4**	**9**	*	**21**	**26**	**	55 (1.3)
CMQ_4392	**.**	**0**	**29**	**6**	**8**	* *	**5**	**5**	**p=0.07**	5.3 (na)
CMQ_2677	**.**	**3**	**8**	**2**	**5**	**p=0.09**	**26**	**26**	**	157 (0.18)
CMQ_2687	**.**	**3**	**7**	**2**	**2**	**ns**	**30**	**31**	**	577 (0.24)
CMQ_1651	**Ankyrin repeat protein**	**1**	**27**	**0**	**9**	**.**	**27**	**23**	**	35 (4.70)
CMQ_569	.	1	49	3	31	.	12	13	**	6 (12.5)
CMQ_5699	**Peroxin**	**1**	**5**	**1**	**2**	**.**	**35**	**37**	**	**999 (0.08)**
CMQ_7934	**Heat repeat protein**	**2**	**4**	**3**	**2**	**.**	**18**	**19**	**	**306 (0.20)**
CMQ_8021	Heat shock transcription factor	1	3	4	0	.	8	8	*	158 (0.60)
CMQ_6965	ABC transporter	1	10	3	3	.	8	8	**	164 (0.60)
CMQ_1224	Membrane copper amine oxidase	0	4	2	2	.	7	7	*	75 (0.70)
CMQ_277	**Isoflavone reductase**	**1**	**3**	**4**	**2**	**.**	**32**	**32**	**	**58 (0.70)**
CMQ_5685	2-dehydropantoate 2-reductase	1	1	3	2	.	10	10	**	453 (0.06)
CMQ_6107	Cytochrome p450 monooxygenase	0	5	1	2	.	10	16	**	67 (1.15)
CMQ_3491	.	1	4	1	2	.	7	7	*	61 (1.13)
CMQ_4067	.	0	5	1	4	.	7	7	*	230 (0.06)
CMQ_1868	C2h2 finger domain containing protein	1	3	2	2	.	7	7	*	81 (0.48)
CMQ_938	GTP cyclohydrolase	1	2	2	3	.	7	7	*	80 (0.40)
CMQ_5377	Glycoside hydrolase	3	4	2	1	.	7	7	*	113 (0.60)
CMQ_1846	C-terminal hydrolase	1	2	2	3	.	7	7	*	252 (0.18)
CMQ_4864	Cell morphogenesis protein	1	7	3	4	.	7	7	*	73 (0.70)
CMQ_7167	Autophagy protein	1	6	1	3	.	8	8	*	152 (0.60)
CMQ_555	Death-box RNA helicase	2	4	2	1	.	8	8	*	147 (0.97)
CMQ_3835	Dash complex subunit	0	5	1	2	.	7	7	*	74 (0.80)
CMQ_6415	tRNA-guanine transglycosylase	1	2	2	2	.	12	12	**	935 (0.03)
CMQ_906	Monocarboxylate permease	1	5	1	2	.	7	6	*	72 (2.20)

Genes	Gene description	MK [b]					PAML [c]			
		P_N	D_N	P_S	D_S	p-value	M1a–M2a	M7–M8	p-value	dN/dS (%)
CMQ_4470	Sensor histidine kinase response	0	8	1	9	.	7	7	.	821 (0.01)
CMQ_2502	Phospholipase	0	2	1	2	.	11	11	**	884 (0.03)
CMQ_3494	Ribosome biogenesis	0	4	1	1	.	9	9	.	393 (0.4)
CMQ_3485	mtDNA inheritance protein	1	4	1	3	.	8	8	*	156 (0.33)
CMQ_1204	Inositol polyphosphate phosphatase	2	2	3	2	.	8	8	.	148 (0.56)
CMQ_3564	Dihydrodipicolinate synthetase	0	3	1	1	.	17	18	**	889 (0.12)
CMQ_4773	ATP-dependent DNA helicase	2	2	3	2	.	7	7	*	170 (0.20)
CMQ_599	ATP-binding endoribonuclease	3	2	4	2	.	14	14	**	103 (0.50)
CMQ_3121	60S ribosomal protein	1	6	1	3	.	7	7	*	82 (0.80)

[a] List of genes that showed the strongest evidence of positive selection using MK test and PAML “site-model”; Genes that were found significant for positive selection by both methods and those that after correction for multiple testing remained significant are highlighted; Results are only shown for genes with a putative function; NS, not significant; NA, not applicable;

[b] P_N, number of nonsynonymous polymorphisms within Gs; D_N, number of nonsynonymous differences between Gs and Gc; P_S, number of synonymous polymorphisms within Gs; D_S, number of synonymous differences between Gs and Gc; P-value, significant excess of nonsynonymous divergence in MK test using the Fisher exact test (*P*<0.05* or *P*<0.01**).

[c] M1a-M2a and M7-M8, twice the difference in the natural logs of the likelihoods of the two models being compared. This value is used in a likelihood ratio test along with 2 degrees of freedom; p-value, is an uncorrected value from $\chi 2$ distribution to indicates the confidence with, which the null model can be rejected; dN/dS, nonsynonymous /synonymous substitution ratio (ω = dN/dS) under M8 model of variable ω ratios among sites, and the percent of codons placed in that class. Amino acid positions identified in the class of codons evolving under positive selection in M8 (posterior probability>0.90) are listed in Appendix B.8.


a
MPB
JPB
GsB1
GsA1
GsA2
GsM1
GsA3
GsB3
GsB2
Gs
GsC1
GsC2
Pc
GcC1
Gc
GcC2
Pj
GcB1
Pp
4000 SNVs

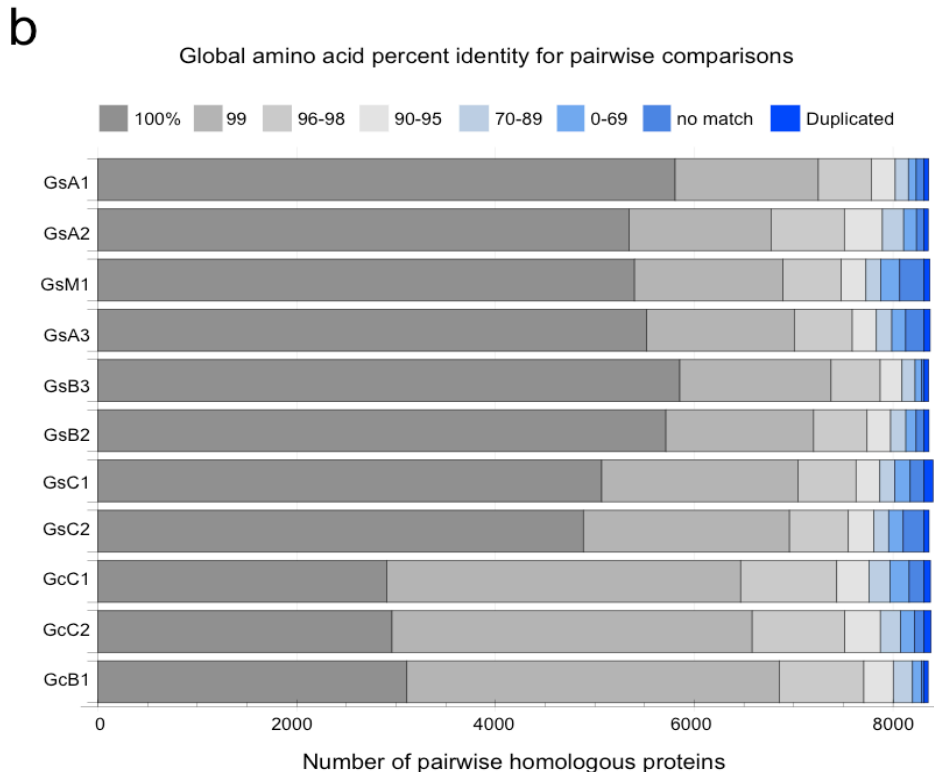
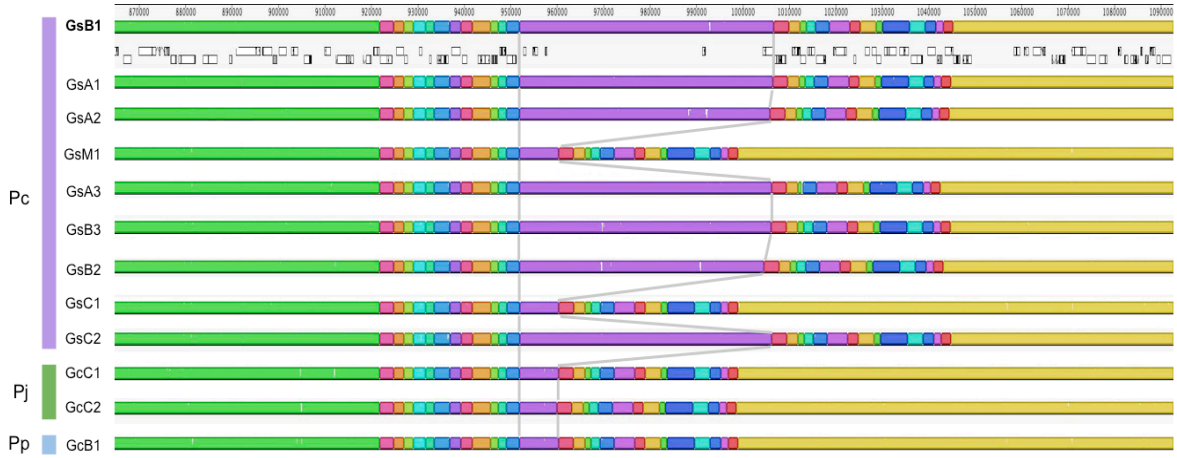

b
Global amino acid percent identity for pairwise comparisons
100%
99
96-98
90-95
70-89
0-69
no match
Duplicated
GsA1
GsA2
GsM1
GsA3
GsB3
GsB2
GsC1
GsC2
GcC1
GcC2
GcB1
0
2000
4000
6000
8000
Number of pairwise homologous proteins

Figure 4.1 *Grosmannia* SNP-phylogenomics, gene content and amino acid similarity

a) Maximum parsimony analysis of 103,430 SNPs among 12 *Grosmannia* genomes. All branches have 100% bootstrap support and posterior probabilities of 1.0. The scale bar indicates the number of SNPs along each branch. A, B, C and M are the collection sites. **GsB1** is the reference genome. Pc, Pj and Pp are the host tree species (Table 4.1). The grey box highlights Gs strains from epidemic regions. **b)** Genome-wide pairwise amino acid identity between 8,312 *Grosmannia* reference gene models and homologous proteins in the 11 other strains. Homologous proteins with high amino acid identity are likely orthologs.

a



b

	P_N	D_N	P_S	D_S
aconitase a/isopropylmalate dehydratase	0	3	1	1
hexachloro-cyclohexane acid phosphatase	.	.	.	.
hypothetical protein w. transcriptional support	.	.	.	.
collagen superfamily protein	1	1	1	3
f-box domain containing protein	.	.	.	.
multidrug resistance protein	.	.	.	.
arylsulfatase	.	.	.	.
hypothetical protein w. transcriptional support	.	.	.	.
zinc ion binding transcription factor	.	.	.	.
short chain dehydrogenase reductase	.	.	.	.
lipase esterase family protein	0	3	0	1
duf1446 domain containing protein	0	5	0	2
hypothetical protein	.	.	.	.
methyltransferase type 12	.	.	.	.
hypothetical protein w. transcriptional support	.	.	.	.
hypothetical protein	.	.	.	.
zinc ion binding transcription factor	3	1	2	2
Acyl-CoA ligase/synthetase	0	1	3	2
major facilitator superfamily transporter sugar	1	4	0	2
zinc-type alcohol dehydrogenase	2	1	0	4
aldehyde dehydrogenase mitochondrial precursor	.	.	.	.
hypothetical protein w. transcriptional support	1	2	1	2
putative arsenite efflux transporter	.	.	.	.
transcription factor	.	.	.	.
peptidoglycan-binding lysin protein	.	.	.	.
NADP oxidoreductase, coenzyme f420-dependent	.	.	.	.
epoxide hydrolase	.	.	.	.
flavoprotein monooxygenase	1	3	2	2
tripeptidyl peptidase a	.	.	.	.
zinc ion binding transcription factor	3	4	1	4
FMO-like monooxygenase w. lipocalin signature	.	.	.	.
short-chain dehydrogenase/reductase	.	.	.	.
FMO-like monooxygenase	0	5	1	1

c

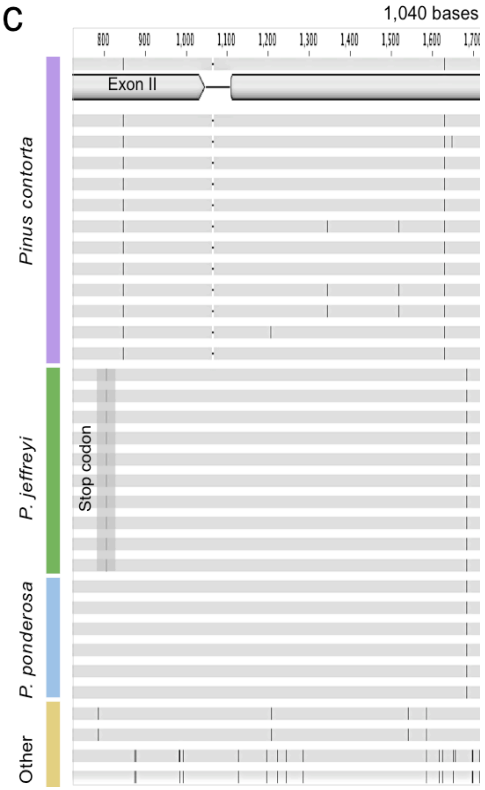


Figure 4.2 Intra-/interspecific variants in the terpenoid-processing gene cluster

a) Alignment of 12 homologous supercontigs shows complete synteny and colinearity for all *Grosmannia* strains. **GsB1**: *Grosmannia* reference genome. Locally collinear blocks (LCB, shown in the same color) have similar sizes among strains, and the 33 gene models that are potentially involved in limonene utilization or detoxification are in complete synteny. The apparent indel (purple LCB) is located in an intergenic region that may have been subjected to assembly error for a few Gs and Gc strains**.** To rule out the possibility of assembly error, a contig longer than the indel (puple LCB) is required, but these data were missing for all the strains showing the deletion. **b)** The 33 orthologous genes in the terpenoid-processing cluster showed relatively low numbers of polymorphisms (non-synonymous, P_N; synonymous: P_S) and divergence (non-synonymous, D_N; synonymous, D_S). Dots represent genes with less than two coding nucleotide differences. **c)** The flavoprotein monooxygenase (CMQ-6740) has a stop codon in its second exon that is unique to the Gc isolates from *P. jefferyi.*

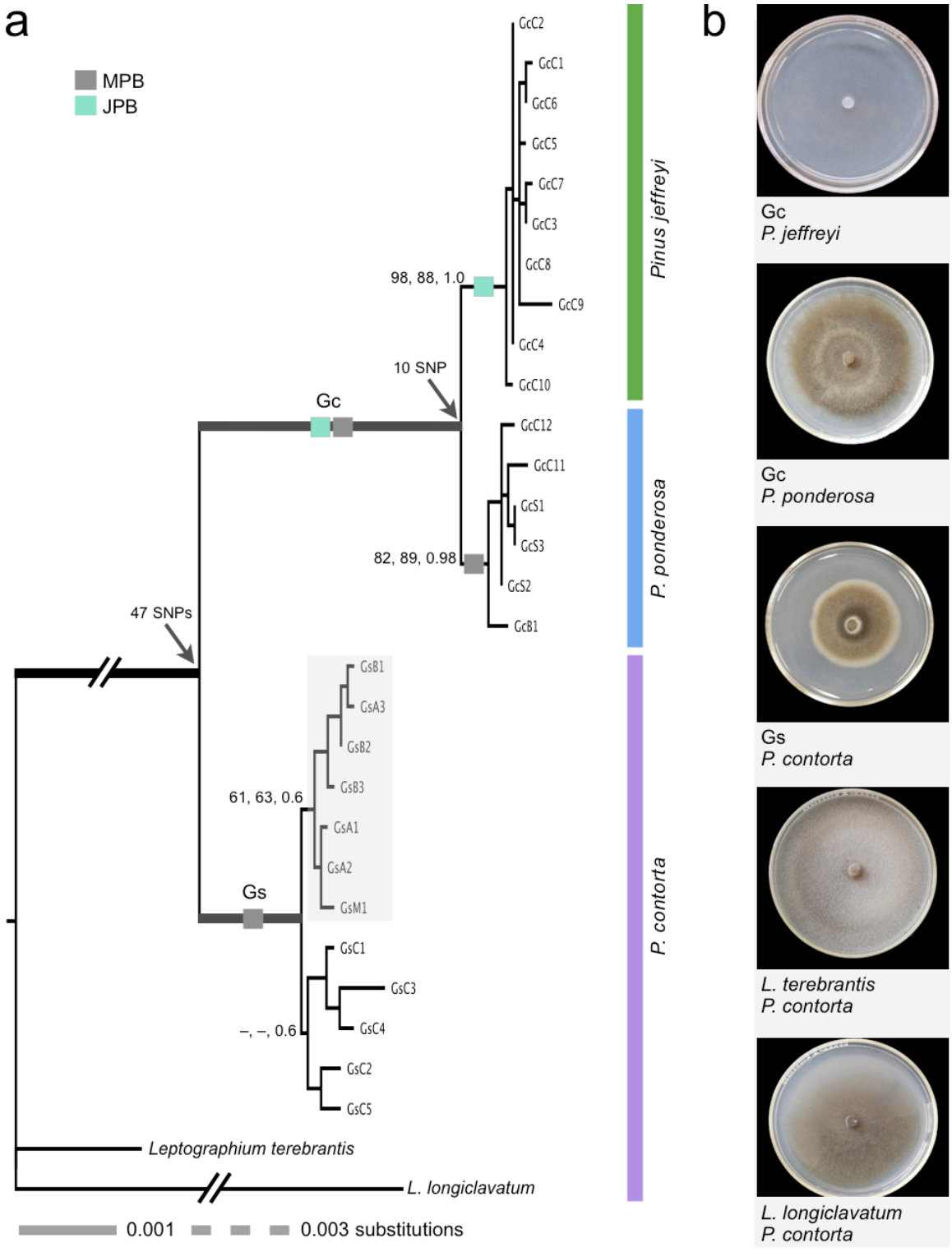

a
MPB
JPB
GcC2
GcC1
GcC6
GcC5
GcC7
GcC3
GcC8
GcC9
GcC4
GcC10
98, 88, 1.0
10 SNP
Gc
GcC12
GcC11
GcS1
GcS3
82, 89, 0.98
GcS2
GcB1
47 SNPs
GsB1
GsA3
GsB2
GsB3
61, 63, 0.6
GsA1
GsA2
Gs
GsM1
GsC1
GsC3
GsC4
–, –, 0.6
GsC2
GsC5
Leptographium terebrantis
L. longiclavatum
0.001
0.003 substitutions
Pinus jeffreyi
P. ponderosa
P. contorta
b
Gc
P. jeffreyi
Gc
P. ponderosa
Gs
P. contorta
L. terebrantis
P. contorta
L. longiclavatum
P. contorta

Figure 4.3 *Grosmannia* species phylogeny correlated with the host-tree species of different phylogenetic lineages and fungal lineage tolerance to a related host-defence chemical

a) Maximum likelihood (ML) analysis of a concatenated nine-gene dataset subdivides Gs and Gc strains into three well-supported clades according to host tree species. Thick branches indicate nodes with 100% support from ML, maximum parsimony and Bayesian analyses. The grey box highlights Gs strains from epidemic regions. Arrows indicate total numbers of fixed differences between Gs-Gc, Gc–*P. ponderosa* and Gc–*P. jeffreyi* lineages. The tree is rooted with the outgroup taxa *L. longiclavatum* and *L. terebrantis*. Dashed line indicates an adjustment of scale. **b)** Physiological assessment comparing (+)-limonene utilization as a carbon source in Gs and Gc lineages and close relatives. The growth of the fungal lineages from *P. contorta* and *P. ponderosa* indicates their ability to tolerate and utilize this toxic chemical.

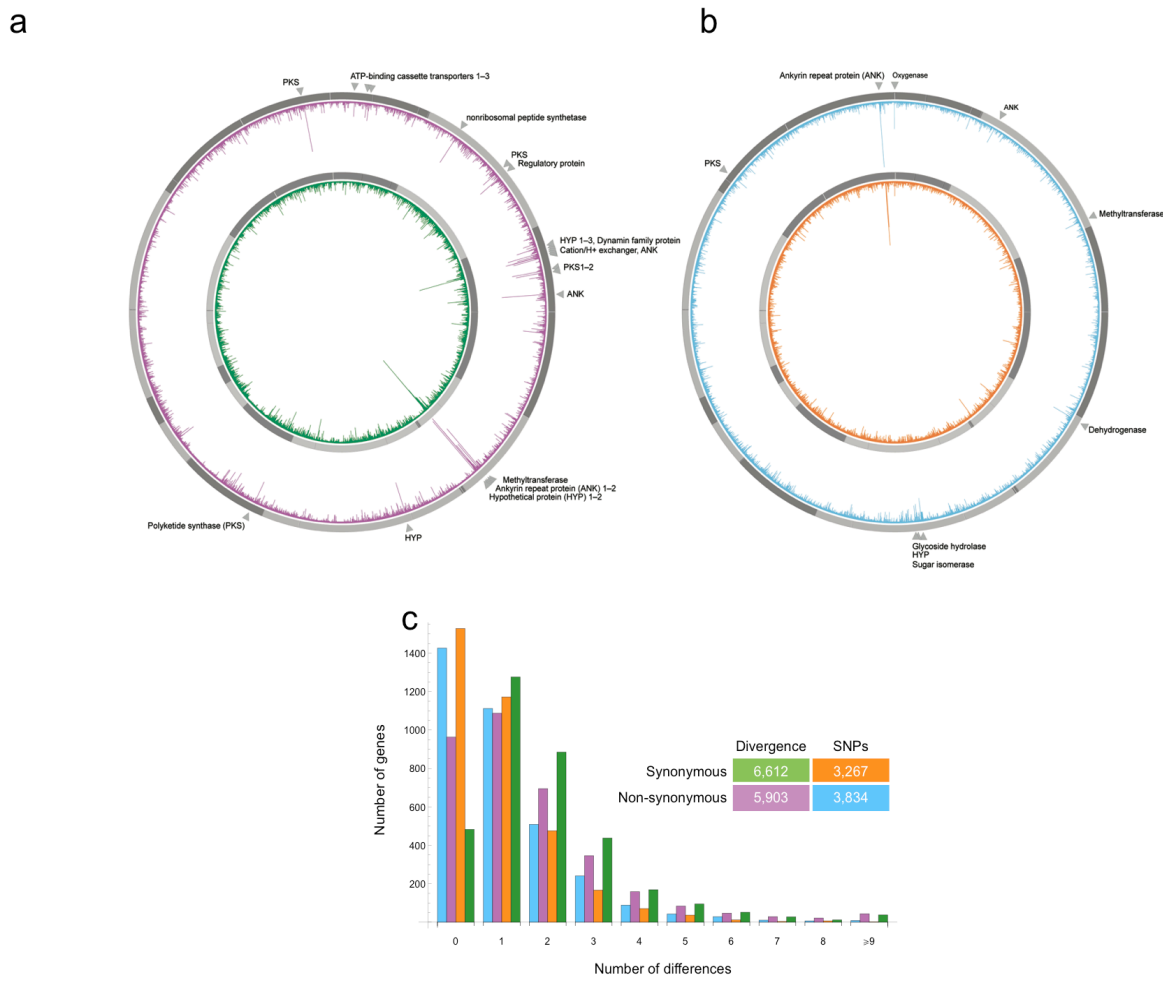


Figure 4.4 Comparison of divergence and polymorphism

a) Number of synonymous and non-synonymous Gs–Gc fixed differences (divergences) in 7,340 orthologous gene models. **b)** Number of synonymous and non-synonymous SNPs within Gs strains in 7,340 orthologous gene models. Circular bands with alternating shades of grey represent 36 scaffolds in which the gene models are located. Grey triangles mark genes with the largest numbers of non-synonymous divergences (e.g. PKS, ANK, ABC in **a**) and polymorphisms (e.g. ANK in **b**). **c)** Summary distributions of McDonald-Kreitman (MK) cell entries for the fixed differences and SNPs in 3,476 variable genes (Appendix B.7).

Chapter 5 Conclusions

This thesis highlights the importance of studying genomes in ophiostomatoid fungi, particularly for species of the genus *Grosmannia*, which include the most common associates of bark beetles and important tree pathogens like *G. clavigera*. Until now phylogenomic data in the form of sequences of many individual loci or fully sequenced genomes had not been used for establishing relationship among species, defining species boundaries, or assessing host adaptation and ecological divergence in the beetle-tree associated fungi. Here, we show that genome data are particularly relevant for assessing the systematics of this fungal group, which remains problematic at both interspecific and generic levels. Further, these large datasets can generate insights into the ecological and biological attributes of these organisms and evolutionary patterns that affect species diversity, and relationships of species with their vectors and host trees.

Sequenced genomes support comparing patterns of nucleotide substitutions for orthologous regions within and between closely related species, and so can provide evidence for evolutionary forces and the processes that influence the evolution of fungal pathogens. When nearly complete genome sequences are available for many species, these comparisons are powerful. Until recently, genomic resources for the beetle-tree associated fungi were available only for *G. clavigera*. When we made the reference genome of this species available, one of the many research opportunities created was

to characterize evolutionary divergence between *Grosmannia* strains. For this, we generated draft genomes for eleven additional *G. clavigera* strains from distinct populations including different beetle vectors and host trees. These genomes provided a unique opportunity for comparative analysis to establish species and population boundaries, and to reveal the potential role of host-tree adaptation and population dynamics in fungal evolution and divergence. Our analyses focus on generating a comprehensive assessment of genome-wide single nucleotide variations within and between the newly recognized species, and using this information to identify specific genes that are potentially involved in species divergence and host adaptation. Note that the work reported here, which demonstrates an approach for understanding the effects of genome divergence, should be extended to populations over wider geographic regions.

While sequence data for such work is now practical to generate, there is a need to adapt or develop analysis tools to such genomic data in order that they can be used efficiently and reliably to identify species diversity, population structures and processes affecting the evolution of the tree pathogens in their natural ecosystem. The work describe here is an important step toward this objective. We developed new genomic resources that provide a foundation for future development of phylogenomic and lineage-specific markers to assess species diversity and relationships within the ophiostomatoid fungi. As well, we identified functional variants that are potentially involved in host colonization and progressive divergence within and between the *Grosmannia* species. The resources that we created provide basic information for future

research on comparative and functional genomics of other species of this important group of fungi, and for related ophiostomatoid genera. A longer-term program that extends the demonstration reported here could be part of a multidisciplinary approach towards developing and implementing sustainable forestry management practices that reduce the intensity of beetle-fungal outbreaks.

5.1 Ophiostomatoids systematics limitation and status of the genus *Grosmannia*

Taxonomy has been a dynamic and progressive discipline for many fungal groups (Hawksworth et al. 2013); however, for beetle-tree associated fungi (ophiostomatoids), it has needed further development. In chapter 2, I used morphological characterization and a multigene-phylogenetic approach to evaluate the ophiostomatoid systematics by defining genera and species boundaries, and I revised the status of some *Ophiostomatales* genera, including *Grosmannia.* The results demonstrate the limitations of using classical morphological traits and molecular analyses based on single genes to address the taxonomy of this fungal group. The sexual form is less frequently found for ophiostomatoids and thus many taxa are primarily classified based on the asexual or anamorph structures (Upadhyay 1981; Wingfield and Seifert 1993). I show that no anamorph can unambiguously support the phylogenetic groups resolved by our multigene datasets. Molecular systematics based on multiple informative loci integrated with a more natural taxonomic scheme separate species of the genera *Grosmannia*,

Ambrosiella and *Raffaelea* into different clades corresponding to their distinct ecological niches and vector associates, i.e. bark- versus wood-boring beetles.

For this work I generated a multigene dataset for sixty-seven taxa that represent a diverse set of ophiostomatoid teleomorphs and anamorph genera. The results present the phylogenetic relationships of ophiostomatoids at a higher taxonomic level, including different genera (i.e. *Ceratocystiopsis*, *Grosmannia* and *Ophiostoma*) that are associated with the *Dendroctonus ponderosae* (MPB)-fungal outbreak in western North America (Lee et al. 2006a). The high resolution of our multigene phylogeny also confirms the monophyletic status of the genus *Grosmannia*, which was my primary question, since species of this genus were previously mixed with the ecological group of *Ophiostomatales*, which was collectively called ambrosia fungi. Finally, I show that ambrosia fungi, including the genera *Ambrosiella* and *Raffaelea*, are polyphyletic, forming at least six distinct phylogenetic clades, each of which I propose should be reassigned to a new genus. These results will influence future research on the *Ophiostomatales*' taxonomy, especially for the novel clades that need additional support with an expanded collection of fungal taxa and gene loci.

5.2 Defining species boundaries in *Grosmannia clavigera*

Cryptic species are increasingly reported for plant pathogens. The results in chapter 3 provide additional examples of cryptic species within *G. clavigera*, the well-known

fungal pathogen of *P. contorta* forests in western North America and the symbiont of two sister bark beetles: MPB and JPB. We found that genetic relatedness in *Grosmannia* lineages mainly corresponds with the host tree species, and that host specificity and/or preferences are more evident in the localized populations than in epidemic regions.

5.2.1 Polymorphism discovery and species recognition

Cataloging species diversity is a first step toward developing an understanding of how various organisms interact with their environment, which is a key factor in establishing organisms' roles in the ecosystem. Despite previous systematics efforts, defining species in the *G. clavigera* complex was a challenge and required developing new molecular resources. I began this work by characterizing nucleotide variations in 67 gene loci and then sequencing 15 loci across 53 population samples. Using “phylogenetic species recognition by genealogical concordance” and different population genetic analyses of this dataset, I identified two cryptic sibling species within the pathogen: Gc and Gs. The data suggested a potential history of recombination within both species, which in addition with our ecological data (i.e. occasional observance of sexual stage inside the older galleries of the beetle associates) allowed for names according to the teleomorphs genus. I retained the nomenclatural name *G. clavigera* for the lineage (Gc) that is genetically and ecologically represented by the holotype (Robinson-Jeffrey and Davidson 1968), while *Grosmannia* species (Gs) is a

newly recognized species that remains to be described. This work also generated protein-coding sequences that were used for developing target-specific PCR-primers for the *Grosmannia* species and their close relatives (Khadempour et al. 2010).

5.2.2 Ecologically distinguishable lineags

Sequencing one of the informative loci across 166 isolates indicated that Gc is more specific to the closely related tree species *P. jeffreyi* and *P. ponderosa* in the USA, where these pines are infested by localized populations of respective beetle associates. In contrast, Gs is an exclusive associate of MPB and its primary host-tree *P. contorta*; however, in the current epidemic areas, it is also found in other pine species. This is an important finding because it suggests that both host tree species and beetle population dynamics are important factors in the evolution and divergence of these fungi. In regions with localized population, although the beetle-fungal population level is low and difficult to sample (Smith et al. 2010), we show that further investigation from different host trees (*P. contorta*, *P. ponderosa* and *P. flexilis*) in eastern and southern portion of the MPB range i.e., in areas that have not been reached by the current epidemics, is necessary for understanding the biological and ecological interaction between beetle-fungal complexes and the host trees. These interactions are not evident in the epidemic regions since increase in the beetle-fungal population levels has enabled these organisms to spread from their native range to other suitable habitat and to establish in new host trees (e.g. *P. banksiana*).

5.3 Population genomics in *Grosmannia*

While sequencing one genome per species allows insights into an organism's biology, inferring genetic variations that lead to adaptation in different environments, and to biological/physiological differences between populations and species requires more than one genome sequence for a species. The most comprehensive view on such variations is not gained by sequencing a few candidate genes, but by genome-wide studies of variants within and between closely related species, i.e. population genomics. This has become routine in non-model organisms with the advent of NGS systems like those from Illumina. In chapter 4, we use these technical advances to generate additional genomic resources from *Grosmannia* strains representing the two newly recognized species: Gs and Gc. Despite their close phylogenetic relationship these fungi inhabit distinct ecological niches, and I believe that adaptation to the specific chemistries of host tree is an important feature in their evolutionary divergence. My analysis of the new genomic resources indicates that this hypothesis likely holds true, or at least warrants further investigation for a set of genes potentially involved in host-colonization and/or pathogenicity.

5.3.1 *Grosmannia* draft genomes and genome-wide characterization of SNVs

Because Illumina systems generate deep sequencing data, the genomes for each of the eleven *Grosmannia* strains provided high coverage for much of the reference

genome, allowing for a comprehensive characterization of genome-wide sequence variation in these strains, which were isolated from a large geographic area and distinct ecological niches. In contrast to other eukaryotes, most filamentous fungi have compact genomes with relatively few repetitive sequences; this makes generating genome assemblies from short sequence reads affordable and technically less challenging than for higher eukaryotes (Nowrousian 2010). While finished assemblies tend to be essential for thoroughly characterizing lineage-specific differences, and for detecting recently evolved or rapidly evolving parts of genomes, draft genomes support identifying major structural changes and local changes like SNVs and short indels.

Large scale structural changes have been shown to exceed nucleotide evolution in plant pathogens like *Mycosphaerella* and *Fusarium* spp., which have lineage-specific chromosomal islands or even entire lineage-specific chromosomes (Cuomo et al. 2007; Stukenbrock et al. 2010; Klosterman et al. 2011). Such structural changes may be attributed to relatively long divergence times or horizontal gene transfer (Hane et al. 2011). In this work; however, draft assemblies showed no evidence of major structural changes in the genomes of the eleven *Grosmannia* isolates, suggesting that these fungi have diverged relatively recently, consistent with the gene genealogies that we describe in the third chapter of this thesis. We found more than 100,000 SNVs, suggesting that distinct ecological differences and host preferences in these fungi may be driven mainly by local nucleotide changes, rather than by large-scale rearrangements.

5.3.2 Genome-wide SNV-phylogeny

The genome-wide phylogeny that I derived from the SNV dataset confirmed chapter three's genealogical study of *G. clavigera* populations, which separate the twelve strains into divergent Gs and Gc monophyletic clades; and it further divides each clade into potentially distinct phylogenetic groups that had not been previously reported. Within Gs, I suggested that strains from epidemic regions and localized populations might represent two evolutionary independent lineages that before applying the species criteria require further sampling in southern and eastern portions of the species' range and preferably using new molecular markers from our SNV dataset. Further, within Gc, we found additional evidence of host preference and/or specificity in the localized populations showing that JPB associates (JP) form a separate clade from the holotype isolated from MPB-infested *P. ponderosa* (PP) in BC (Robinson-Jeffrey and Davidson 1968). I examined the monophyly of Gc phylogenetic groups by genealogies of additional *P. ponderosa* associates collected from California (n=2), South Dakota (n=3) and the holotype as the only remainder from previous epidemics in BC. The "genealogical nondiscordance criterion" (Dettman et al. 2003) suggested that Gc-JP and Gc-PP may warrant recognition as genealogical and ecological species. Further characterization of *P. ponderosa* associates at the population level using informative SNV markers should clarify the extent of host-specificity across the MPB-localized USA populations and/or the role of geographical isolation in their divergence.

Our network analysis in figure 3.3a of chapter 3 may seem to disagree with the SNV subdivisions within Gs and Gc. While the network analysis allows visualizing conflicts among the 15 gene datasets, it cannot distinguish between the causes of conflicts, which may be recombination or incomplete lineage sorting (i.e. retention of ancestral polymorphisms through multiple speciation events). Further, none of the 15 gene trees were informative in resolving either Gs-California or Gc-*P. ponderosa* groups, except for a peptidase locus that separated Gc strains from *P. ponderosa*, but with no statistical support. However, for recently diverged species the lack of diagnosable characters at some loci and a reticulate pattern of incomplete lineage sorting are not unexpected. Neutral coalescent theory indicates that the expected time to monophyly is often long even for a single nuclear gene, suggesting that many species exist, but have not yet achieve monophyly even for a few genes (Hudson and Coyne 2002; Knowles and Carstens 2007; Shaffer and Thomson 2007). Nevertheless, potential signs of recombination, particularly within Gs, can provide important clues for avoiding over-estimating species diversity in these fungi. Additional analyses at the population level should be carried out using the informative SNP markers reported here.

5.3.3 Functional characterization of SNVs and their adaptive contribution

Nucleotide substitutions that generate adaptive variations can occur in both regulatory and coding sequences. In this thesis, while our analysis focuses on protein-coding genes, we also identified intra- and interspecific SNVs in potential intergenic and

regulatory regions. Both the coding and noncoding variants that I report should be good starting points for future studies on the functional genomics of the *Grosmannia* species and on the role of positive selection in the evolution of gene regulatory elements. For instance, SNVs characterized in regions upstream and downstream of *Grosmannia* gene models may cause adaptive variations, especially if they cause substitutions in regulatory sequences like promoters and terminators, and so they may activate or suppress gene expression. However, because the annotation of regulatory regions in *Grosmannia* genomes is currently putative, we only present the results for adaptive variants in coding regions, which provide a rich resource for future investigations of adaptive evolution and functional variation in the pathogens.

Most studies that seek to detect adaptive evolution in fungal pathogens have focused on specific likely candidate genes. Recently, however, genomic data has allowed 'reverse ecology' studies that search for genes under positive selection without *a priori* expectations (Li et al. 2008; Ellison et al. 2011; Gladieux et al. 2013). Here, we use this approach to carry out a genome-wide assessment of polymorphism and divergence in 12 Gs–Gc gene orthologs. Only 1.2% of 3,476 informative genes showed rapid amino acid evolution (i.e. non-synonymous fixed differences between Gs and Gc). Considering these genes, and other variable coding regions, we found evidence of positive selection in at least 46 genes (p-value<0.05). Some of these genes were involved in secondary metabolite synthesis and secretion (PKS). Others had putative roles in exporting host-chemical defense chemicals or secondary metabolites (ABC–C

transporters), and in protein-protein interactions or self/nonself recognition (ankyrin repeat proteins and a heterokaryon incompatibility protein).

The between-species selection methods applied here (i.e. MK and dN/dS tests) showed limited statistical power for inferring positive selection between *Grosmannia* siblings. Generally, these methods can be used to identify non-recent selective forces, but they require a large number of nucleotide variants to exceed the background of mutational drift over long periods of species differentiation (Li et al. 2008; Oleksyk et al. 2010). Despite low divergence level between Gs and Gc lineages, we find statistically significant evidence of site-specific positive selection on a small set of genes, including those with a putative function in host-colonization and pathogenicity (e.g. PKSs).

Characterizing the coding variants, we also identified a number of truncated or potential pseudogenes within *Grosmannia* gene models. The majority of these genes (n=22) have potential oxidoreductase activity, including those with known roles in stress response and detoxification like cytochrome P450s, short-chain dehydrogenases and monooxygenases. These results are consistent with our physiological assessment, and suggest that the stop-codon variants may also reflect adaptation of Gs and Gc lineages to the specific chemistries of *Pinus contorta*, *ponderosa*, and *jeffreyi* trees. Both the potential *Grosmannia* pseudogenes and the genes showing evidence of positive selection (e.g. PKS and ankyrin repeat proteins) are important candidates for being

functionally important in divergence and/or ecological adaptation of *Grosmannia* fungi. These warrant further validation and investigation.

5.4 Perspective on future work

This thesis provides a foundation for future work on the taxonomy and systematics of the *Ophiostomatales*. Different phylogenetic groups were redefined and newly suggested genera and novel species should be incorporated in future studies, to expand the collection of fungi isolated from different ecological and geographical sources, and to characterize their systematics using the new genome datasets. Genome-wide scans for positively selected genes that provide insight into adaptive evolution in *Grosmannia* lineages, the genetic basis of differences between and within species, and putative functions of some genes should be expanded. For example, the potential pseudogenes in *P. jeffreyi* associates or PKS genes with positive selection signatures might be important in host chemical detoxification and/or host colonization, and can be the focus of future comparative studies or experimental functional characterization. Adding other closely related species such as *L. terebrantis* and *L. longiclavatum* would increase the phylogenetic depth of the genome datasets, which can improve statistical power of the selection analyses and may permit new lineage- and clade-specific questions to be tested.

Very recently, in Alberta, the MPB-fungal complexes have succeeded in colonizing a new host-tree species, *P. banksiana*, adapting to the new chemical and physical environment present in this host. While *P. banksiana* is more closely related to *P. contorta* than to *P. ponderosa* or *P. jeffreyi*, environmental conditions prevailing in the northern boreal forest could allow the symbiotic partners to continue to adapt and to spread eastward, with potentially important ecological and economic consequences. To infer recent selection pressures on the beetle-pathogen complexes as they exploit new hosts requires polymorphism data and population genomic-based analysis– the excess of non-synonymous to synonymous differences would not be great enough to be statistically significant using divergence-based selection analysis. The genomic datasets and informative markers for landscape population genomics provide a comprehensive resource for work on ongoing adaptation in this vector-pathogen complex, and the approach reported here could be extended to other beetle-tree associates.

References

Agapow P-M, Burt A. 2001. Indices of multilocus linkage disequilibrium. Mol. Ecol. Notes 1:101–102.

Aguileta G, Lengelle J, Marthey S, et al. 2009. Finding candidate genes under positive selection in Non-model species: examples of genes involved in host specialization in pathogens. Mol. Ecol. 19:292–306.

Alfaro RI, Humble LM, Gonzalez P, Villaverde R, Allegro G. 2007. The threat of the ambrosia beetle Megaplatypus mutatus (Chapuis) (=Platypus mutatus Chapuis) to world poplar resources. Forestry 80:471–479.

Amselem J, Cuomo CA, van Kan JAL, et al. 2011. Genomic Analysis of the Necrotrophic Fungal Pathogens Sclerotinia sclerotiorum and Botrytis cinerea. PLoS Genet 7:e1002230.

Andersen MR, Salazar MP, Schaap PJ, et al. 2011. Comparative genomics of citric-acid-producing Aspergillus niger ATCC 1015 versus enzyme-producing CBS 513.88. Genome Res. 21:885–897.

Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P. 2004. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. Trends Ecol. Evol. 19:535–544.

Von Arx JA, Hennebert GL. 1965. Deux champignons ambrosia. Mycopathol. Mycol. Appl. 25:309–315.

Avise JC, Ball RM. 1990. Principles of genealogical concordance in species concepts and biological taxonomy. In: Futuyma D, Antonovics J, editors. Oxford Surveys in Evolutionary Biology. Vol. 7. Oxford: Oxford University Press. p. 45–67.

Avise JC. 2004. Molecular markers, natural history, and evolution. 2nd Revised edition. Sunderland, MA: Sinauer Associates

Ayres MP, Lombardero MJ. 2000. Assessing the consequences of global change for forest disturbance from herbivores and pathogens. Sci. Total Environ. 262:263–286.

Bahn Y-S, Kojima K, Cox GM, Heitman J. 2006. A Unique Fungal Two-Component System Regulates Stress Responses, Drug Sensitivity, Sexual Development, and Virulence of Cryptococcus neoformans. Mol. Biol. Cell 17:3122–3135.

Barrett LG, Thrall PH, Burdon JJ, Linde CC. 2008. Life history determines genetic structure and evolutionary potential of host-parasite interactions. Trends Ecol. Evol. 23:678–685.

Batra L. 1963. Ecology of Ambrosia Fungi and Their Dissemination by Beetles. Trans. Kans. Acad. Sci. 66:213–236.

Batra L. 1967. Ambrosia Fungi: A Taxonomic Revision, and Nutritional Studies of Some Species. Mycologia 59:976–1017.

Batra LR. 1966. Ambrosia Fungi: Extent of Specificity to Ambrosia Beetles. Science 153:193–195.

Batra LR. 1972. Ectosymbiosis between ambrosia fungi and beetles. i. Indian J. Mycol. Plant Pathol. 2:165–169.

Baum D, Shaw KL. 1995. Experimental approaches to plant systematics - Monographs in Botany from the Missouri Botanical Garden 53. (Hoch PC, Stephenson AG, editors.). St, Louis: Missouri Botanical Garden

De Beer ZW, Seifert KA, Wingfield MJ. 2013. A nomenclator for ophiostomatoid genera and species in the Ophiostomatales and Microascales. In: Seifert KA, de Beer ZW, Wingfield MJ, editors. The Ophiostomatoid Fungi: Expanding Frontiers. CBS Biodiversity Series. Utrecht, Netherlands: CBS-KNAW Fungal Biodiversity Centre. p. 245–322.

De Beer ZW, Wingfield MJ. 2013. Emerging lineages in the Ophiostomatales. In: Seifert KA, De Beer ZW, Wingfield MJ, editors. Emerging lineages in the Ophiostomatales. 12. Utrecht, Netherlands: CBS-KNAW Fungal Biodiversity Centre. p. 21–46.

Benade E, Wingfield MJ, Wyk PSV. 1995. Conidium Development in the Hyalorhinocladiella Anamorph of Ophiostoma ips. Mycologia 87:298.

Benade E, Wingfield MJ, Wyk PSV. 1996. Conidium development in the Hyalorhinocladiella anamorphs of Ceratocystiopsis minuta-bicolor and Ophiostoma minus. Can. J. Bot. 74:891–897.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. 57:289–300.

Bentz BJ, Régnière J, Fettig CJ, Hansen EM, Hayes JL, Hicke JA, Kelsey RG, Negrón JF, Seybold SJ. 2010. Climate Change and Bark Beetles of the Western United States and Canada: Direct and Indirect Effects. BioScience 60:602–613.

Berbee ML, Taylor JW. 2001. Fungal Molecular Evolution: Gene Trees and Geologic Time. In: McLaughlin DJ, McLaughlin EG, Lemke PA, editors. The Mycota: a comprehensive

treatise on fungi as experimental systems for basic and applied research. Volume VII: Systematics and Evolution, Part B. Berlin: Springer. p. 229–245.

Bickford CP, Kolb TE, Geils BW. 2005. Host physiological condition regulates parasitic plant performance: Arceuthobium vaginatum subsp. cryptopodum on Pinus ponderosa. Oecologia 146:179–189.

Bleiker K, Six D. 2009. Competition and Coexistence In a Multi-partner Mutualism: Interactions Between two Fungal Symbionts of the Mountain Pine Beetle In Beetle-attacked Trees. Microb. Ecol. 57:191–202.

Bohlmann J. 2012. Pine terpenoid defences in the mountain pine beetle epidemic and in other conifer pest interactions: specialized enemies are eating holes into a diverse, dynamic and durable defence system. Tree Physiol. 32:943–945.

Boone CK, Aukema BH, Bohlmann J, Carroll AL, Raffa KF. 2011. Efficacy of tree defense physiology varies with bark beetle population density: a basis for positive feedback in eruptive species. Can. J. For. Res. 41:1174–1188.

Borden JH. 1984. Semiochemical-mediated aggregation and dispersal in the Coleoptera. In: Lewis T, editor. Insect Communication. London: Academic Press. p. 123–149.

Brasier CM. 1987. The dynamics of fungal speciation. In: Rayner AD., Brasier CM, Moore D, editors. Evolutionary Biology of the Fungi. Cambridge, UK: Cambridge University Press. p. 231–260.

Bruen TC, Philippe H, Bryant D. 2006. A Simple and Robust Statistical Test for Detecting the Presence of Recombination. Genetics 172:2665–2681.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV. 2003. The utility of single nucleotide polymorphisms in inferences of population history. Trends Ecol. Evol. 18:249–256.

Bryant D, Moulton V. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol. Biol. Evol. 21:255–265.

Burnett J. 2003. Fungal Population and Species. Oxford, UK: Oxford University Press

Burt A, Carter DA, Koenig GL, White TJ, Taylor JW. 1996. Molecular markers reveal cryptic sex in the human pathogen Coccidioides immitis. Proc. Natl. Acad. Sci. U. S. A. 93:770–773.

Butcher D. 1995. Muller's ratchet, epistasis and mutation effects. Genetics 141:431–437.

Carroll A, Aukema B, Raffa K, Linton DA, Smith G, Lindgren B. 2006. Mountain pine beetle outbreak development: The endemic--incipient epidemic transition. Victoria, BC: Canadian Forest Service

Cassar S, Blackwell M. 1996. Convergent Origins of Ambrosia Fungi. Mycologia 88 (4):596–601.

Clark EL, Carroll AL, Huber DPW. 2010. Differences in the Constitutive Terpene Profile of Lodgepole Pine Across a Geographical Range in British Columbia, and Correlation with Historical Attack by Mountain Pine Beetle. Can. Entomol. 142:557–573.

Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9:1657–1659.

Coleman JJ, Rounsley SD, Rodriguez-Carres M, et al. 2009. The Genome of Nectria haematococca: Contribution of Supernumerary Chromosomes to Gene Expansion. PLoS Genet 5:e1000618.

Collemare J, Billard A, Böhnert HU, Lebrun M-H. 2008. Biosynthesis of secondary metabolites in the rice blast fungus Magnaporthe grisea: the role of hybrid PKS-NRPS in pathogenicity. Mycol. Res. 112:207–215.

Collins FS, Brooks LD, Chakravarti A. 1998. A DNA polymorphism discovery resource for research on human genetic variation. Genome Res. 8:1229–1231.

Condon BJ, Leng Y, Wu D, et al. 2013. Comparative Genome Structure, Secondary Metabolite, and Effector Coding Capacity across Cochliobolus Pathogens. PLoS Genet 9:e1003233.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676.

Croll D, McDonald BA. 2012. The Accessory Genome as a Cradle for Adaptive Evolution in Pathogens. PLoS Pathog. 8.

Cullingham CI, Cooke JEK, Dang S, Davis CS, Cooke BJ, Coltman DW. 2011. Mountain pine beetle host-range expansion threatens the boreal forest. Mol. Ecol. 20:2157–2171.

Cuomo CA, Güldener U, Xu J-R, et al. 2007. The Fusarium graminearum Genome Reveals a Link Between Localized Polymorphism and Pathogen Specialization. Science 317:1400–1402.

Darling AE, Mau B, Perna NT. 2010. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. PLoS ONE 5:e11147.

Dean RA, Talbot NJ, Ebbole DJ, et al. 2005. The genome sequence of the rice blast fungus Magnaporthe grisea. Nature 434:980–986.

Delorme L, Lieutier F. 1990. Monoterpene composition of the preformed and induced resins of Scots pine, and their effect on bark beetles and associated fungi. Eur. J. For. Pathol. 20:304–316.

Desjardins CA, Champion MD, Holder JW, et al. 2011. Comparative Genomic Analysis of Human Fungal Pathogens Causing Paracoccidioidomycosis. PLoS Genet 7:e1002345.

Dettman JR, Jacobson DJ, Taylor JW. 2003. A Multilocus Genealogical Approach to Phylogenetic Species Recognition in the Model Eukaryote Neurospora. Evolution 57:2703–2720.

DiGuistini S, Liao NY, Platt D, et al. 2009. De novo genome sequence assembly of a filamentous fungus using Sanger, 454 and Illumina sequence data. Genome Biol. 10:R94.

DiGuistini S, Ralph SG, Lim YW, Holt R, Jones S, Bohlmann J, Breuil C. 2007. Generation and annotation of lodgepole pine and oleoresin-induced expressed sequences from the blue-stain fungus Ophiostoma clavigerum, a Mountain Pine Beetle-associated pathogen. FEMS Microbiol. Lett. 267:151–158.

DiGuistini S, Wang Y, Liao NY, et al. 2011. Genome and transcriptome analyses of the mountain pine beetle-fungal symbiont Grosmannia clavigera, a lodgepole pine pathogen. Proc. Natl. Acad. Sci. 108:2504–2509.

Duong TA, Beer ZW de, Wingfield BD, Wingfield MJ. 2012. Phylogeny and taxonomy of species in the Grosmannia serpens complex. Mycologia 104:715–732.

Eck SH, Benet-Pagès A, Flisikowski K, Meitinger T, Fries R, Strom TM. 2009. Whole genome sequencing of a single Bos taurus animal for single nucleotide polymorphism discovery. Genome Biol. 10:R82.

Ellison CE, Hall C, Kowbel D, Welch J, Brem RB, Glass NL, Taylor JW. 2011. Population genomics and local adaptation in wild isolates of a model microbial eukaryote. Proc. Natl. Acad. Sci. U. S. A. 108:2831–2836.

Engel MS, Grimaldi DA. 2004. The first Mesozoic stephanid wasp (Hymenoptera: Stephanidae). J. Paleontol. 78:1192–1197.

Farrell B, Sequeira A, O'Meara B, Normark B, Chung J, Jordal B. 2001. The evolution of agriculture in beetles (Curculionidae: Scolytinae and Platypodinae). Evolution 55:2011–2027.

Fay JC, Wyckoff GJ, Wu CI. 2001. Positive and negative selection on the human genome. Genetics 158:1227–1234.

Fedorova ND, Badger JH, Robson GD, Wortman JR, Nierman WC. 2005. Comparative analysis of programmed cell death pathways in filamentous fungi. BMC Genomics 6:177.

Fedorova ND, Khaldi N, Joardar VS, et al. 2008. Genomic islands in the pathogenic filamentous fungus Aspergillus fumigatus. PLoS Genet. 4:e1000046.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.

Forgetta V, Leveque G, Dias J, et al. 2013. Sequencing of the Dutch Elm Disease Fungus Genome Using the Roche/454 GS-FLX Titanium System in a Comparison of Multiple Genomics Core Facilities. J. Biomol. Tech. JBT 24:39–49.

Forrest GI. 1980. Genotypic variation among native scots pine populations in scotland based on monoterpene analysis. Forestry 53:101 –128.

Fraedrich SW, Harrington TC, Rabaglia RJ, Ulyshen MD, Mayfield AE, Hanula JL, Eickwort JM, Miller DR. 2008. A Fungal Symbiont of the Redbay Ambrosia Beetle Causes a Lethal Wilt in Redbay and Other Lauraceae in the Southeastern United States. Plant Dis. 92:215–224.

Francke-Grosmann H. 1952. Uber die Ambrosiazucht der Kiefernborkenkafer Myelophilus minor Htg. und Ips acuminatus Gyll. Medd. Statens Skogsforskning Inst. Stockh. 41:1–52.

Francke-Grosmann H. 1967. Symbiosis. In: Henry SM, editor. Ectosymbiosis in Wood-Inhabiting Insects. Vol. 2. Academic, New York, London: Academic Press. p. 141–205.

Funk A. 1970. Fungal symbionts of the ambrosia beetle Gnathotrichus sulcatus. Can. J. Bot. 48:1445–1448.

Galagan JE, Henn MR, Ma L-J, Cuomo CA, Birren B. 2005. Genomics of the fungal kingdom: Insights into eukaryotic biology. Genome Res. 15:1620–1631.

Gao X, Starmer JD. 2008. AWclust: point-and-click software for non-parametric population structure analysis. BMC Bioinformatics 9:77.

Gayathri Samarasekera GDN, Bartell NV, Lindgren BS, Cooke JEK, Davis CS, James PMA, Coltman DW, Mock KE, Murray BW. 2012. Spatial genetic structure of the mountain pine beetle (Dendroctonus ponderosae) outbreak in western Canada: historical patterns and contemporary dispersal. Mol. Ecol. 21:2931–2948.

Gebhardt, Begerow, Oberwinkler. 2004. Identification of the ambrosia fungus of Xyleborus monographus and X. dryographus (Coleoptera: Curculionidae, Scolytinae). Mycol. Prog. 3:95–102.

Gebhardt H, Oberwinkler F. 2005. Conidial development in selected ambrosial species of the genus Raffaelea. Antonie Van Leeuwenhoek 88:61–65.

Gebhardt H, Weiss M, Oberwinkler F. 2005. Dryadomyces amasae: a nutritional fungus associated with ambrosia beetles of the genus Amasa (Coleoptera: Curculionidae, Scolytinae). Mycol. Res. 109:687–696.

Geiser DM, Pitt JI, Taylor JW. 1998. Cryptic speciation and recombination in the aflatoxin-producing fungus Aspergillus flavus. Proc. Natl. Acad. Sci. U. S. A. 95:388–393.

Gernandt DS, Hernández-León S, Salgado-Hernández E, Pérez de La Rosa JA. 2009. Phylogenetic Relationships of Pinus Subsection Ponderosae Inferred from Rapidly Evolving cpDNA Regions. Syst. Bot. 34:481–491.

Gerson EA, Kelsey RG, St Clair JB. 2009. Genetic variation of piperidine alkaloids in Pinus ponderosa: a common garden study. Ann. Bot. 103:447 –457.

Giraud T, Gladieux P, Gavrilets S. 2010. Linking the emergence of fungal plant diseases with ecological speciation. Trends Ecol. Evol. 25:387–395.

Giraud T, Refrégier G, Le Gac M, de Vienne DM, Hood ME. 2008. Speciation in fungi. Fungal Genet. Biol. 45:791–802.

Giraud T, Villaréal LMMA, Austerlitz F, Le Gac M, Lavigne C. 2006. Importance of the life cycle in sympatric host race formation and speciation of pathogens. Phytopathology 96:280–287.

Gladieux P, Devier B, Aguileta G, Cruaud C, Giraud T. 2013. Purifying selection after episodes of recurrent adaptive diversification in fungal pathogens. Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis. 17:123–131.

Goffeau A, Barrell BG, Bussey H, et al. 1996. Life with 6000 Genes. Science 274:546–567.

Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. 11:725–736.

Goodwin SB, Ben M'Barek S, Dhillon B, et al. 2011. Finished Genome of the Fungal Wheat Pathogen Mycosphaerella graminicola Reveals Dispensome Structure, Chromosome Plasticity, and Stealth Pathogenesis. PLoS Genet 7:e1002070.

Gu Z, David L, Petrov D, Jones T, Davis RW, Steinmetz LM. 2005. Elevated evolutionary rates in the laboratory strain of Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA. 102:1092–1097.

Guerrero R. 1966. Una nueva especie de hongo inperfecto asociado con el coleoptero Platypus sulcatus Chapius. Revta Investnes Agropecu. Ser. 5 3:97–103.

Hane JK, Lowe RGT, Solomon PS, et al. 2007. Dothideomycete–Plant Interactions Illuminated by Genome Sequencing and EST Analysis of the Wheat Pathogen Stagonospora nodorum. Plant Cell Online 19:3347–3368.

Hane JK, Rouxel T, Howlett BJ, Kema GH, Goodwin SB, Oliver RP. 2011. A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. Genome Biol. 12:R45.

Haridas S, Breuill C, Bohlmann J, Hsiang T. 2011. A biologist's guide to de novo genome assembly using next-generation sequence data: A test with fungal genomes. J. Microbiol. Methods 86:368–375.

Haridas S, Wang Y, Lim L, et al. 2013. The genome and transcriptome of the pine saprophyte Ophiostoma piceae, and a comparison with the bark beetle-associated pine pathogen Grosmannia clavigera. BMC Genomics 14:373.

Harrington T, McNew D, Steimel J, Hofstra D, Farrell R. 2001. Phylogeny and Taxonomy of the Ophiostoma piceae Complex and the Dutch Elm Disease Fungi. Mycologia 93:111–136.

Harrington T. 2005. Ecology and evolution of mycophagous bark beetles and their fungal partners. In: Vega F, Blackwell M, editors. Ecological and evolutionary advances in insect–fungal associations. New York, NY: Oxford University Press. p. 257–291.

Harrington TC, Fraedrich SW, Aghayeva D. 2008. Raffaelea lauricola, a new ambrosia beetle symbiont and pathogen on the Lauraceae. Mycotaxon 104:399–404.

Hartig T. 1844. Ambrosia des Bostrychibs dispar. Allegemeine Jagd-Ztg. NS:73–74.

Hartl DL, Clark AG. 2007. Principles of population genetics. Sunderland, Mass.: Sinauer Associates

Hausner G, Reid J, Klassen G. 2000. On the phylogeny of members of Ceratocystis ss and Ophiostoma that possess different anamorphic states, with emphasis on the anamorph genus Leptographium, based on partial ribosomal DNA sequences. Can. J. Bot. 78:903–916.

Hausner G, Reid J, Klassen GR. 1993. On the phylogeny of Ophiostoma, Ceratocystis s.s., and Microascus, and relationships within Ophiostoma based on partial ribosomal DNA sequences. Can. J. Bot. 71:1249–1265.

Hawksworth DL, Kirk PM, Sutton BC, Pegler DN. 1995. Ainsworth & Bisby's dictionary of the fungi. Wallingford, Oxon, UK: CAB International

Hawksworth DL, McNeill J, de Beer ZW, Wingfield MJ. 2013. Names of fungal species with the same epithet applied to different morphs: how to treat them. IMA Fungus 4:53–56.

Hawksworth DL. 2001. The magnitude of fungal diversity: the 1.5 million species estimate revisited. Mycol. Res. 105:1422–1432.

Hawksworth DL. 2011. A new dawn for the naming of fungi: impacts of decisions made in Melbourne in July 2011 on the future publication and regulation of fungal names. IMA Fungus 2:155–162.

Hay CJ. 1956. Experimental crossing of mountain pine beetle with Black Hills beetle. Ann. Entomol. Soc. Am. 49:567–571.

Heckman DS, Geiser DM, Eidell BR, Stauffer RL, Kardos NL, Hedges SB. 2001. Molecular Evidence for the Early Colonization of Land by Fungi and Plants. Science 293:1129–1133.

Hesse-Orce U, DiGuistini S, Keeling CI, et al. 2010. Gene discovery for the bark beetle-vectored fungal tree pathogen Grosmannia clavigera. BMC Genomics 11:536.

Hey J, Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D. persimilis. Genetics 167:747–760.

Higby PK, Stock MW. 1982. Genetic Relationships Between Two Sibling Species of Bark Beetle (Coleoptera: Scolytidae), Jeffrey Pine Beetle and Mountain Pine Beetle, in Northern California. Ann. Entomol. Soc. Am. 75:668–674.

Hintz W, Pinchback M, de la Bastide P, Burgess S, Jacobi V, Hamelin R, Breuil C, Bernier L. 2011. Functional categorization of unique expressed sequence tags obtained from the yeast-like growth phase of the elm pathogen Ophiostoma novo-ulmi. BMC Genomics 12:431.

Hofstetter RW, Mahfouz JB, Klepzig KD, Ayres MP. 2005. Effects of tree phytochemistry on the interactions among endophloedic fungi associated with the southern pine beetle. J. Chem. Ecol. 31:539–560.

Hogenhout SA, Van der Hoorn RAL, Terauchi R, Kamoun S. 2009. Emerging concepts in effector biology of plant-associated organisms. Mol. Plant-Microbe Interactions MPMI 22:115–122.

Holloway AK, Lawniczak MKN, Mezey JG, Begun DJ, Jones CD. 2007. Adaptive Gene Expression Divergence Inferred from Population Genomics. PLoS Genet 3:e187.

Hopkin AD. 1902. On the study of forest entomology in America. U.S. Department of Agriculture, Division of Entomology

Hopkins A. 1909. Practical Information on the Scolytid Beetles on North American Forests : Part I - Barkbeetles of the Genus Dendroctonus. Washington D.C.

Hsiau PTW, Harrington TC. 2003. Phylogenetics and adaptations of basidiomycetous fungi fed upon by bark beetles (Coleoptera: Scolytidae). Symbiosis 34:111–131.

Huber DPW, Aukema BH, Hodgkinson RS, Lindgren BS. 2009. Successful colonization, reproduction, and new generation emergence in live interior hybrid spruce Picea engelmannii glauca by mountain pine beetle Dendroctonus ponderosae. Agric. For. Entomol. 11:83–89.

Hudson RR, Coyne JA. 2002. Mathematical consequences of the genealogical species concept. Evol. Int. J. Org. Evol. 56:1557–1565.

Hulcr J, Kolarik M, Kirkendall LR. 2007. A new record of fungus-beetle symbiosis in Scolytodes bark beetles (Scolytinae, Curculionidae, Coleoptera). Symbiosis 43:151–159.

Hull CM, Raisner RM, Johnson AD. 2000. Evidence for mating of the “asexual” yeast Candida albicans in a mammalian host. Science 289:307–310.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254–267.

Huson DH, Lockhart PJ, Steel MA. 2005. Reconstruction of reticulate networks from gene trees. In: In Proceedings of the Ninth International Conference on Research in Computational Molecular Biology (RECOMB). Springer. p. 233–249.

Hutchison LJ, Reid J. 1988. Taxonomy of some potential wood-staining fungi from New Zealand 1. Ophiostomataceae. New Zealand J. Bot. 26:63–81.

Hynum BG, Berryman AA. 1980. Dendroctonus ponderosae (Coleoptera: Scolytidae) pre-aggregation landing and gallery initiation on lodgepole pine. Can. Entomol. 112:185–191.

Jacobs K, Kirisits T, Wingfield MJ. 2003. Taxonomic re-evaluation of three related species of Graphium, based on morphology, ecology and phylogeny. Mycologia 95:714–727.

Jacobs K, Wingfield MJ. 2001. Leptographium species: tree pathogens, insect associates, and agents of blue-stain. St. Paul, MN: American Phytopathological Society Press

James PMA, Coltman DW, Murray BW, Hamelin RC, Sperling FAH. 2011. Spatial Genetic Structure of a Symbiotic Beetle-Fungal System: Toward Multi-Taxa Integrated Landscape Genetics. PLoS ONE 6:e25359.

Jones K, Blackwell M. 1998. Phylogenetic analysis of ambrosial species in the genus Raffaelea based on 18S rDNA sequences. Mycol. Res. 102 (6):661–665.

Jordal BH, Cognato AI. 2012. Molecular phylogeny of bark and ambrosia beetles reveals multiple origins of fungus farming during periods of global warming. BMC Evol. Biol. 12:133.

Kasuga T, White TJ, Koenig G, et al. 2003. Phylogeography of the fungal pathogen Histoplasma capsulatum. Mol. Ecol. 12:3383–3401.

Kasuga T, White TJ, Taylor JW. 2002. Estimation of Nucleotide Substitution Rates in Eurotiomycete Fungi. Mol. Biol. Evol. 19:2318 –2324.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.

Katsir L, Chung HS, Koo AJK, Howe GA. 2008. Jasmonate signaling: a conserved mechanism of hormone sensing. Curr. Opin. Plant Biol. 11:428–435.

Keeling CI, Bohlmann J. 2006. Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens. New Phytol. 170:657–675.

Keeling CI, Yuen MM, Liao NY, et al. 2013. Draft genome of the mountain pine beetle, Dendroctonus ponderosae Hopkins, a major forest pest. Genome Biol. 14:R27.

Kelley S, Farrell B. 1998. Is specialization a dead end? The phylogeny of host use in Dendroctonus bark beetles (Scolytidae). Evolution 52:1731–1743.

Kelley ST, Farrell BD, Mitton JB. 2000. Effects of specialization on genetic differentiation in sister species of bark beetles. Heredity 84:218–227.

Kemen E, Gardiner A, Schultz-Larsen T, et al. 2011. Gene Gain and Loss during Evolution of Obligate Parasitism in the White Rust Pathogen of Arabidopsis thaliana. PLoS Biol 9:e1001094.

Khadempour L, Massoumi Alamouti S, Hamelin R, Bohlmann J, Breuil C. 2010. Target-specific PCR primers can detect and differentiate ophiostomatoid fungi from microbial communities associated with the mountain pine beetle Dendroctonus ponderosae. Fungal Biol. 114:825–833.

Khoshraftar S, Hung S, Khan S, Gong Y, Tyagi V, Parkinson J, Sain M, Moses AM, Christendat D. 2013. Sequencing and annotation of the Ophiostoma ulmi genome. BMC Genomics 14:162.

Kim J-J, Lim YW, Wingfield MJ, Breuil C, Kim G-H. 2004. Leptographium bistatum sp. nov., a new species with a Sporothrix synanamorph from Pinus radiata in Korea. Mycol. Res. 108:699–706.

Kirisits T. 2004. Fungal Associates of European Bark Beetles With Special Emphasis on the Ophiostomatoid Fungi. In: Lieutier F, Day KR, Battisti A, Grégoire J-C, Evans HF, editors. Bark and Wood Boring Insects in Living Trees in Europe, a Synthesis. Springer Netherlands. p. 181–236.

Kirk PM, Cannon PF, Minter DW, Stalpers JA. 2008. Ainsworth and Bisby's Dictionary of the Fungi. Wallingford, UK: CABI

Klepzig K, Wilkens R. 1997. Competitive Interactions among Symbiotic Fungi of the Southern Pine Beetle. Appl. Environ. Microbiol. 63:621–627.

Klepzig KD, Six DL. 2004. Bark beetle fungal symbioses: Context dependency in complex associations. Symbiosis 37:189–206.

Klosterman SJ, Subbarao KV, Kang S, et al. 2011. Comparative Genomics Yields Insights into Niche Adaptation of Plant Vascular Wilt Pathogens. PLoS Pathog 7:e1002137.

Knowles L, Carstens B. 2007. Delimiting species without monophyletic gene trees. Syst. Biol. 56:887–895.

Kohler GA, Brenot A, Haas-Stapleton E, Agabian N, Deva R, Nigam S. 2006. Phospholipase A2 and Phospholipase B Activities in Fungi. Biochim. Biophys. Acta 1761:1391–1399.

Korbel JO, Urban AE, Affourtit JP, et al. 2007. Paired-end mapping reveals extensive structural variation in the human genome. Science 318:420–426.

Kovalchuk A, Driessen A. 2010. Phylogenetic analysis of fungal ABC transporters. BMC Genomics 11:177.

Kroken S, Glass NL, Taylor JW, Yoder OC, Turgeon BG. 2003. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. Proc. Natl. Acad. Sci. 100:15670–15675.

Krokene P, Solheim H. 1996. Fungal associates of five bark beetle species colonizing Norway spruce. Can. J. For. Res. 26:2115–2122.

Krupkin AB, Liston A, Strauss SH. 1996. Phylogenetic Analysis of the Hard Pines (Pinus subgenus Pinus, Pinaceae) from Chloroplast DNA Restriction Site Analysis. Am. J. Bot. 83:489–498.

Kubono T, Ito S. 2002. Raffaelea quercivora sp. nov. associated with mass mortality of Japanese oak, and the ambrosia beetle (Platypus quercivorus). Mycoscience 43:0255–0260.

Kuhnholz S. 2004. Chemical Ecology and Mechanisms of Reproductive Isolation in Ambrosia Beetles.

Kulathinal RJ, Stevison LS, Noor MAF. 2009. The Genomics of Speciation in Drosophila: Diversity, Divergence, and Introgression Estimated Using Low-Coverage Genome Sequencing. PLoS Genet 5:e1000550.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol. 5:R12–R12.

Kurz WA, Dymond CC, Stinson G, Rampley GJ, Neilson ET, Carroll AL, Ebata T, Safranyik L. 2008. Mountain pine beetle and forest carbon feedback to climate change. Nature 452:987–990.

Lagerberg T, Lundberg G, Melin E. 1927. Biological and practical researches into blueing in pine and spruce. Sven. Skogsvardsforen 25:145–272.

Lah L, Haridas S, Bohlmann J, Breuil C. 2013. The cytochromes P450 of Grosmannia clavigera: Genome organization, phylogeny, and expression in response to pine host chemicals. Fungal Genet. Biol. 50:72–81.

Lambreghts R, Shi M, Belden WJ, et al. 2009. A high-density single nucleotide polymorphism map for Neurospora crassa. Genetics 181:767–781.

Langor DW, Spence JR, Pohl GR. 1990. Host effects on fertility and reproductive success of Dendroctonus ponderosae Hopkins (Coleoptera: Scolytidae). Evolution 44:609.

Langor DW. 1989. Host effects on the phenology, development, and mortality of field populations of the mountain pine beetle, Dendroctonus Ponderosae Hopkins (Coleoptera:Scolytidae). Can. Entomol. 121:149–157.

Lanier GN, Wood, DL. 1968. Controlled mating, caryology, morphology and sex ratio in the Dendroctonus ponderosae complex. Ann. Entomol. Soc. Am. 61:517–526.

Latta RG, Linhart YB, Snyder MA, Lundquist L. 2003. Patterns of variation and correlation in the monoterpene composition of xylem oleoresin within populations of ponderosa pine. Biochem. Syst. Ecol. 31:451–465.

Leach JG, Orr G, Christense C. 1934. The interrelationships of bark beetles and blue-staining fungi in felled Norway pine timber. J. Agric. Res. 49:315–341.

Lee S, Hamelin RC, Six DL, Breuil C. 2007. Genetic diversity and the presence of two distinct groups in Ophiostoma clavigerum associated with Dendroctonus ponderosae in British Columbia and the northern Rocky Mountains. Phytopathology 97:1177–1185.

Lee S, Kim J, Breuil C. 2005. Leptographium longiclavatum sp. nov., a new species associated with the mountain pine beetle, Dendroctonus ponderosae. Mycol. Res. 109:1162–1170.

Lee S, Kim J, Fung S, Breuil C. 2003. A PCR-RFLP marker distinguishing Ophiostoma clavigerum from morphologically similar Leptographium species associated with bark beetles. Can. J. Bot. 81:1104–1112.

Lee S, Kim J-J, Breuil C. 2006a. Diversity of fungi associated with mountain pine beetle, Dendroctonus ponderosae, and infested lodgepole pines in British Columbia. Fungal Divers. 22:91–105.

Lee S, Kim J-J, Breuil C. 2006b. Pathogenicity of Leptographium longiclavatum associated with Dendroctonus ponderosae to Pinus contorta. Can. J. For. Res. 36:2864–2872.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760.

Li H, Handsaker B, Wysoker A, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078 –2079.

Li H, Ruan J, Durbin R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. 18:1851–1858.

Li R, Fan W, Tian G, et al. 2010. The sequence and de novo assembly of the giant panda genome. Nature 463:311–317.

Li YF, Costello JC, Holloway AK, Hahn MW. 2008. “Reverse ecology” and the power of population genomics. Evol. Int. J. Org. Evol. 62:2984–2994.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451 –1452.

Lieutier F, Yart A, Salle A. 2009. Stimulation of tree defenses by Ophiostomatoid fungi can explain attack success of bark beetles on conifers. Ann. For. Sci. 66:801–801.

Lim Y, Alamouti S, Kim J, Lee S, Breuil C. 2004. Multigene phylogenies of Ophiostoma clavigerum and closely related species from bark beetle-attacked Pinus in North America. FEMS Microbiol. Lett. 237:89–96.

Liu M, Gelli A. 2008. Elongation Factor 3, EF3, Associates with the Calcium Channel Cch1 and Targets Cch1 to the Plasma Membrane in Cryptococcus neoformans. Eukaryot. Cell 7:1118–1126.

Logan JA, Powell JA. 2001. Ghost forests, global warming, and the mountain pine beetle (Coleoptera: Scolytidae). Am. Entomol. 47:160–173.

Luhtala N. 2004. Bro1 coordinates deubiquitination in the multivesicular body pathway by recruiting Doa4 to endosomes. J. Cell Biol. 166:717–729.

Lynch M, Bürger R, Butcher D, Gabriel W. 1993. The Mutational Meltdown in Asexual Populations. J. Hered. 84:339–344.

Ma L-J, van der Does H Charlotte, Borkovich KA, et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. Nature 464:367–373.

Ma L-J, van der Does H. Charlotte, Borkovich KA, et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. Nature 464:367–373.

Maddison WP, Knowles LL. 2006. Inferring Phylogeny Despite Incomplete Lineage Sorting. Syst. Biol. 55:21–30.

Magee BB, Magee PT. 2000. Induction of mating in Candida albicans by construction of MTLa and MTLalpha strains. Science 289:310–313.

Malloch D, Blackwell M. 1993. Dispersal biology of the Ophiostomatoid Fungi. In: Wingfield MJ, Seifert KA, Webber JF, editors. Ceratocystis and ophiostoma: taxonomy, ecology, and pathogenicity. St. Paul, Minnesota: American Phytopathological Society Press. p. 195–206.

Manning VA, Pandelova I, Dhillon B, et al. 2013. Comparative Genomics of a Plant-Pathogenic Fungus, Pyrenophora tritici-repentis, Reveals Transduplication and the Impact of Repeat

Elements on Pathogenicity and Population Divergence. G3 GenesGenomesGenetics 3:41–63.

Marcet-Houben M, Ballester A-R, Fuente B de la, Harries E, Marcos JF, González-Candelas L, Gabaldón T. 2012. Genome sequence of the necrotrophic fungus Penicillium digitatum, the main postharvest pathogen of citrus. BMC Genomics 13:646.

Mardis ER. 2008. Next-generation DNA sequencing methods. Annu. Rev. Genomics Hum. Genet. 9:387–402.

Marin M, Preisig O, Wingfield BD, Kirisits T, Wingfield MJ. 2009. Single sequence repeat markers reflect diversity and geographic barriers in Eurasian populations of the conifer pathogen Ceratocystis polonica. For. Pathol. 39:249–265.

Marshall DH, Newton C, Ritland K. 2002. Chloroplast phylogeography and evolution of highly polymorphic microsatellites in lodgepole pine (Pinus contorta). TAG Theor. Appl. Genet. Theor. Angew. Genet. 104:367–378.

Marvaldi AE, Sequeira AS, O'Brien CW, Farrell BD. 2002. Molecular and morphological phylogenetics of weevils (coleoptera, curculionoidea): do niche shifts accompany diversification? Syst. Biol. 51:761–785.

Massoumi Alamouti S, Kim J-J, Humble LM, Uzunovic A, Breuil C. 2007. Ophiostomatoid fungi associated with the northern spruce engraver, Ips perturbatus, in western Canada. Antonie Van Leeuwenhoek 91:19–34.

Massoumi Alamouti S, Tsui CKM, Breuil C. 2009. Multigene phylogeny of filamentous ambrosia fungi associated with ambrosia and bark beetles. Mycol. Res. 113:822–835.

Massoumi Alamouti S, Wang V, DiGuistini S, Six DL, Bohlmann J, Hamelin RC, Feau N, Breuil C. 2011. Gene genealogies reveal cryptic species and host preferences for the pine fungal pathogen Grosmannia clavigera. Mol. Ecol. 20:2581–2602.

Matute DR, McEwen JG, Puccia R, et al. 2006. Cryptic Speciation and Recombination in the Fungus Paracoccidioides brasiliensis as Revealed by Gene Genealogies. Mol Biol Evol 23:65–73.

Mayden RL. 1997. The denouement in the saga of the species problem. In: Claridge MF, Dawah HA, Wilson MR, editors. Species: The Units of Biodiversity. London: Chapman and Hall. p. 381–424.

McCluskey K, Wiest AE, Grigoriev IV, Lipzen A, Martin J, Schackwitz W, Baker SE. 2011. Rediscovery by Whole Genome Sequencing: Classical Mutations and Genome Polymorphisms in Neurospora crassa. G3 Genes Genomes Genet. 1:303–316.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652–654.

McKenna A, Hanna M, Banks E, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.

Metzker ML. 2010. Sequencing technologies - the next generation. Nat. Rev. Genet. 11:31–46.

Miller JR, Koren S, Sutton G. 2010. Assembly algorithms for next-generation sequencing data. Genomics 95:315–327.

Minter DW, Kirk PM, Sutton BC. 1982. Holoblastic phialides. Trans. Br. Mycol. Soc. 79:75–93.

Minter DW, Kirk PM, Sutton BC. 1983. Thallic phialides. Trans. Br. Mycol. Soc. 80:39–66.

Mirov NT, Hasbrouck J. 1976. The Story of Pines. First Edition. Bloomington, IN: Indiana University Press

Mirov NT. 1948. The terpenes (in relation to the biology of genus Pinus). Annu. Rev. Biochem. 17:521–540.

Mock KE, Bentz BJ, O'neill EM, Chong JP, Orwin J, Pfrender ME. 2007. Landscape-scale genetic variation in a forest outbreak species, the mountain pine beetle (Dendroctonus ponderosae). Mol. Ecol. 16:553–568.

Möller EM, Bahnweg G, Sandermann H, Geiger HH. 1992. A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues. Nucleic Acids Res. 20:6115–6116.

Moran GP, Coleman DC, Sullivan DJ. 2011. Comparative genomics and the evolution of pathogenicity in human pathogenic fungi. Eukaryot. Cell 10:34–42.

Morin PA, Luikart G, Wayne RK, the SNP workshop group. 2004. SNPs in ecology, evolution and conservation. Trends Ecol. Evol. 19:208–216.

Münch E. 1907. Die Blaufaule des Nadelholzes. Naturw Ztschr F Forst U Landw 5:531–573.

Neafsey DE, Barker BM, Sharpton TJ, et al. 2010. Population genomic sequencing of Coccidioides fungi reveals recent hybridization and transposon control. Genome Res. 20:938–946.

Nielsen R, Bustamante C, Clark AG, et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. 3:e170.

Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929–936.

Nowrousian M, Stajich JE, Chu M, et al. 2010. De novo Assembly of a 40 Mb Eukaryotic Genome from Short Sequence Reads: Sordaria macrospora, a Model Organism for Fungal Morphogenesis. PLoS Genet 6:e1000891.

Nowrousian M. 2010. Next-Generation Sequencing Techniques for Eukaryotic Microorganisms: Sequencing-Based Solutions to Biological Problems. Eukaryot. Cell 9:1300–1310.

Nylander JAA, Wilgenbusch JC, Warren DL, Swofford DL. 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24:581–583.

O'Donnell K. 1992. Ribosomal DNA internal transcribed spacers are highly divergent in the phytopathogenic ascomycete Fusarium sambucinum (Gibberella pulicaris). Curr. Genet. 22:213–220.

Okada G, Seifert K, Takematsu A, Yamaoka Y. 1998. A molecular phylogenetic reappraisal of the Graphium complex based on 18S rDNA sequences. Can J Bot.

Oleksyk TK, Smith MW, O'Brien SJ. 2010. Genome-wide scans for footprints of natural selection. Philos. Trans. R. Soc. B Biol. Sci. 365:185 –205.

Paine TD, Hanlon CC. 1994. Influence of oleoresin constituents from Pinus ponderosa and Pinus jeffreyi on growth of mycangial fungi from Dendroctonus ponderosae and Dendroctonus jeffreyi. J. Chem. Ecol. 20:2551–2563.

Paine TD, Raffa KF, Harrington TC. 1997. Interactions Among Scolytid Bark Beetles, Their Associated Fungi, and Live Host Conifers. Annu. Rev. Entomol. 42:179–206.

Paulin-Mahady AE, Harrington TC, McNew D. 2002. Phylogenetic and taxonomic evaluation of Chalara, Chalaropsis, and Thielaviopsis anamorphs associated with Ceratocystis. Mycologia 94:62–72.

Pawlowska TE, Taylor JW. 2004. Organization of genetic variation in individuals of arbuscular mycorrhizal fungi. Nature 427:733–737.

Pollack JK, Harris SD, Marten MR. 2009. Autophagy in filamentous fungi. Fungal Genet. Biol. 46:1–8.

Posada D, Crandall K. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818.

Posada D. 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25:1253 –1256.

Pringle A, Baker DM, Platt JL, Wares JP, Latgé J-P, Taylor JW. 2005. cryptic speciation in the cosmopolitan and human pathogenic fungus Aspergillus fumigatus. Evolution 59:1886–1899.

De Queiroz K. 1998. The general lineage concept of species, species criteria, and the process of speciation. In: Howard DJ, Berlocher SH, editors. Endless forms: species and speciation. New York: Oxford University Press. p. 57–75.

De Queiroz K. 1999. The general lineage concept of species and the defining properties of the species category.

De Queiroz K. 2007. Species Concepts and Species Delimitation. Syst. Biol. 56:879–886.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinforma. Oxf. Engl. 26:841–842.

Raffa, Aukema B, Erbilgin N, Klepzig KD, Wallin KF. 2005. Recent advances in phytochemistry. Toronto, Canada: Elsevier

Raffa K. 1988. The mountain pine beetle in western North America. In: Berryman A, editor. Dynamics of forest insect populations patterns, causes, implications. New York, NY: Plenum Press. p. 506–550.

Raffa K. 2001. Mixed Messages Across Multiple Trophic Levels: The Ecology of Bark Beetle Chemical Communication Systems. Chemoecology 11:49–65.

Raffa KF, Berryman AA. 1982. Gustatory cues in the orientation of Dendroctonus ponderosae (Coleoptera: Scolytidae) to host trees. Can. Entomol. 114:97–104.

Raffa KF, Berryman AA. 1983. The role of host plant resistance in the colonization behavior and ecology of Bark Beetles (Coleoptera: Scolytidae). Ecol. Monogr. 53:27–49.

Raffaele S, Farrer RA, Cano LM, et al. 2010. Genome Evolution Following Host Jumps in the Irish Potato Famine Pathogen Lineage. Science 330:1540–1543.

Raffaele S, Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. Nat. Rev. Microbiol. 10:417–430.

Rand DM, Kann LM. 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Mol. Biol. Evol. 13:735–748.

Reid RW. 1963. Biology of the Mountain Pine Beetle, Dendroctonus monticolae Hopkins, in the East Kootenay Region of British Columbia.: III. Interaction between the Beetle and its Host, with Emphasis on Brood Mortality and Survival. Can. Entomol. 95:225–238.

Reinhardt JA, Baltrus DA, Nishimura MT, Jeck WR, Jones CD, Dangl JL. 2009. De novo assembly using low-coverage short read sequence data from the rice pathogen Pseudomonas syringae pv. oryzae. Genome Res. 19:294–305.

Renwick JAA, Pitman GB. 1979. An attractant isolated from female Jeffrey pine beetles, Dendroctonus jeffreyi. Environ. Entomol. 8:40–41.

Reumers J, Maurer-Stroh S, Schymkowitz J, Rousseau F. 2006. SNPeffect v2.0: a new step in investigating the molecular phenotypic effects of human non-synonymous SNPs. Bioinformatics 22:2183–2185.

Rice AV, Thormann Markus N., Langor DW. 2007. Mountain pine beetle associated blue-stain fungi cause lesions on jack pine, lodgepole pine, and lodgepole × jack pine hybrids in Alberta. Can. J. Bot. 85:307–315.

Rice AV, Thormann M.N., Langor DW. 2007. Virulence of, and interactions among, mountain pine beetle associated blue-stain fungi on two pine species and their hybrids in Alberta. Can. J. Bot. 85:316–323.

Richardson DM. 2000. Ecology and biogeography of Pinus. Cambridge, England: Cambridge University Press

Robinson RC. 1962. Blue stain fungi in lodgepole pine (Pinus contorta Doug. var. latifolia Engelm.) infested by the mountain pine beetle (Dendroctonus monticolae Hopk.).

Robinson-Jeffrey RC, Davidson RW. 1968. Three new Europhium species with Verticicladiella imperfect states on blue-stained pine. Can. J. Bot. 46:1523–1527.

Roe AD, Rice AV, Bromilow SE, Cooke JEK, Sperling FAH. 2010. Multilocus species identification and fungal DNA barcoding: insights from blue stain fungal symbionts of the mountain pine beetle. Mol. Ecol. Resour. 10:946–959.

Roe AD, Rice AV, Coltman DW, Cooke JEK, Sperling FAH. 2011. Comparative phylogeography, genetic differentiation and contrasting reproductive modes in three fungal symbionts of a multipartite bark beetle symbiosis. Mol. Ecol. 20:584–600.

Roeper RA. 1995. Patterns of mycetophagy in Michigan ambrosia beetles. Mich. Acad. 26:153–161.

Rokas A, Payne G, Fedorova ND, et al. 2007. What can comparative genomics tell us about species concepts in the genus Aspergillus? Stud. Mycol. 59:11–17.

Rokas A. 2009. The effect of domestication on the fungal proteome. Trends Genet. 25:60–63.

Rollins F, Jones K, Krokene P, Solheim H. 2001. Phylogeny of Asexual Fungi Associated with Bark and Ambrosia Beetles. Mycologia 93 (5):991–996.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinforma. Oxf. Engl. 19:1572–1574.

Rumbold CT. 1941. A blue stain fungus, Ceratostomella montium n. sp., and some yeasts associated with two species of Dendroctonus. J. Agric. Res. 62:589–60l.

Rydholm C, Szakacs G, Lutzoni F. 2006. Low Genetic Variation and No Detectable Population Structure in Aspergillus fumigatus Compared to Closely Related Neosartorya Species. Eukaryot. Cell 5:650–657.

Safranyik L, Carroll A l., Régnière J, et al. 2010. Potential for range expansion of mountain pine beetle into the boreal forest of North America. Can. Entomol. 142:415–442.

Safranyik L, Carroll A. 2006. The biology and epidemiology of the mountain pine beetle in lodgepole pine forests. In: Safranyik L, Wilson B, editors. The mountain pine beetle: a synthesis of biology, management, and impacts on lodgepole pine. Victoria, BC: Canadian Forest Service. p. 3–66.

Safranyik L, Linton DA. 1983. Brood production by three spp. of Dendroctonus (Coleoptera: Scolytidae) in bolts from host and nonhost trees. Entomol. Soc. Br. Columbia 80:10–13.

Safranyik L, Wilson WR. 2006. The mountain pine beetle: a synthesis of biology, management, and impacts on lodgepole pine. Victoria, BC: Pacific Forestry Centre

Savluchinske Feio S, Gigante B, Roseiro JC, Marcelo-Curto MJ. 1999. Antimicrobial activity of diterpene resin acid derivatives. J. Microbiol. Methods 35:201–206.

Schirawski J, Mannhaupt G, Münch K, et al. 2010. Pathogenicity Determinants in Smut Fungi Revealed by Genome Comparison. Science 330:1546–1548.

Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinforma. Oxf. Engl. 27:863–864.

Scott DB, Du Toit JW. 1970. Three new Raffaelea species. Trans. Br. Mycol. Soc. 55:181–186.

Seifert KA, Okada G. 1993. Graphium anamorphs of Ophiostoma species and similar anamorphs of other ascomycetes. In: Wingfield MJ, Seifert Keith A., Webber JF, editors.

Ceratocystis and Ophiostoma: Taxonomy, Ecology and Pathology. St. Paul, MN: American Phytopathological Society Press. p. 27–41.

Seifert KA, Samuels GJ. 2000. How should we look at anamorphs. Stud. Mycol. 45:5–18.

Selosse M-A, Le Tacon F. 1998. The land flora: a phototroph-fungus partnership? Trends Ecol. Evol. 13:15–20.

Seybold SJ, Bohlmann J, Raffa KF. 2000. Biosynthesis of coniferophagous bark beetle pheromones and conifer isoprenoids: evolutionary perspective and synthesis. Can. Entomol. 132:697–753.

Shaffer HB, Thomson RC. 2007. Delimiting Species in Recent Radiations. Syst. Biol. 56:896–906.

Sharpton TJ, Stajich JE, Rounsley SD, et al. 2009. Comparative genomic analyses of the human fungal pathogens Coccidioides and their relatives. Genome Res. 19:1722–1731.

She R, Chu JS-C, Uyar B, Wang J, Wang K, Chen N. 2011. genBlastG: using BLAST searches to build homologous gene models. Bioinformatics 27:2141–2143.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. Genome Res. 19:1117–1123.

Six D, Harrington TC, Steimel J, McNew D, Paine TD. 2003. Genetic relationships among Leptographium terebrantis and the mycangial fungi of three western Dendroctonus bark beetles. Mycologia 95:781–792.

Six DL, Paine TD, Hare JD. 1999. Allozyme diversity and gene flow in the bark beetle, Dendroctonus jeffieyi (Coleoptera : Scolytidae). Can. J. For. Res. 29:315–323.

Six DL, Paine TD. 1997. Ophiostoma clavigerum Is the Mycangial Fungus of the Jeffrey Pine Beetle, Dendroctonus jeffreyi. Mycologia 89:858–866.

Six DL, Paine TD. 1998. Effects of mycangial fungi and host tree species on progeny survival and emergence of Dendroctonus ponderosae (Coleoptera: Scolytidae). Environ. Entomol. 27:1393–1401.

Six DL, Paine TD. 1999. Allozyme diversity and gene flow in Ophiostoma clavigerum (Ophiostomatales: Ophiostomataceae), the mycangial fungus of the Jeffrey pine beetle, Dendroctonus jeffreyi. Can. J. For. Res. 29:324.

Six DL, Wingfield MJ. 2011. The role of phytopathogenicity in bark beetle-fungus symbioses: a challenge to the classic paradigm. Annu. Rev. Entomol. 56:255–272.

Six DL. 2003. Bark Beetle-Fungus Symbioses. In: Bourtzis K, Miller TA, editors. Insect Symbioses. CRC Press. p. 97–114.

Slatkin M, Maddison WP. 1990. Detecting isolation by distance using phylogenies of genes. Genetics 126:249–260.

Smith GD, Carroll AL, Lindgren BS. 2010. Facilitation in bark beetles: endemic mountain pine beetle gets a helping hand. Agric. For. Entomol. 13:37–43.

Smith JM, Smith NH, O'Rourke M, Spratt BG. 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. U. S. A. 90:4384–4388.

Smith RH. 2000. Xylem monoterpenes of pines: distribution, variation, genetics, function. Berkeley, Calif: USDA

Soanes DM, Alam I, Cornell M, et al. 2008. Comparative Genome Analysis of Filamentous Fungi Reveals Gene Family Expansions Associated with Fungal Pathogenesis. PLoS ONE 3:e2300.

Solheim H, Krokene P. 1998. Growth and virulence of mountain pine beetle associated blue-stain fungi, Ophiostoma clavigerum and Ophiostoma montium. Can. J. Bot. 76:561–566.

Spanu PD, Abbott JC, Amselem J, et al. 2010. Genome Expansion and Gene Loss in Powdery Mildew Fungi Reveal Tradeoffs in Extreme Parasitism. Science 330:1543–1546.

Spatafora J, Blackwell M. 1994. The polyphyletic origins of ophiostomatoid fungi. Mycol. Res. 98 (1):1–9.

Spooner B, Roberts P. 2010. Fungi (Collins New Naturalist Library, Book 96). HarperCollins UK

Stajich JE, Berbee ML, Blackwell M, Hibbett DS, James TY, Spatafora JW, Taylor JW. 2009. Primer -- The Fungi. Curr. Biol. CB 19:R840–R845.

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

Stock MW, Amman GD, Higby PK. 1984. Genetic Variation Among Mountain Pine Beetle (Dendroctonus ponderosae) (Coleoptera: Scolytidae) Populations from Seven Western States. Ann. Entomol. Soc. Am. 77:760–764.

Stock MW, Amman GD. 1980. Genetic differentiation among mountain pine beetle populations from lodgepole pine and ponderosa pine in northeast Utah. Ann. Entomol. Soc. Am. 73:472–478.

Stock MW, Guenter JD. 1979. Isozyme Variation Among Mountain Pine Beetle (Dendroctonus ponderosae) Populations in the Pacific Northwest. Environ. Entomol. 8:889–893.

Stoletzki N, Eyre-Walker A. 2010. The positive correlation between dN/dS and dS in mammals is due to runs of adjacent substitutions. Mol. Biol. Evol. [Internet]. Available from: http://mbe.oxfordjournals.org/content/early/2010/11/29/molbev.msq320

Stoletzki N, Eyre-Walker A. 2011. Estimation of the Neutrality Index. Mol. Biol. Evol. 28:63–70.

Stukenbrock Eva H., Bataillon T, Dutheil JY, Hansen TT, Li R, Zala M, McDonald BA, Wang J, Schierup MH. 2011. The making of a new pathogen: Insights from comparative population genomics of the domesticated wheat pathogen Mycosphaerella graminicola and its wild sister species. Genome Res.

Stukenbrock E. H., Bataillon T, Dutheil JY, Hansen TT, Li R, Zala M, McDonald BA, Wang J, Schierup MH. 2011. The making of a new pathogen: Insights from comparative population genomics of the domesticated wheat pathogen Mycosphaerella graminicola and its wild sister species. Genome Res. 21:2157–2166.

Stukenbrock EH, Bataillon T. 2012. A Population Genomics Perspective on the Emergence and Adaptation of New Plant Pathogens in Agro-Ecosystems.Heitman J, editor. PLoS Pathog. 8:e1002893.

Stukenbrock EH, Christiansen FB, Hansen TT, Dutheil JY, Schierup MH. 2012. Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. Proc. Natl. Acad. Sci. U. S. A. 109:10954–10959.

Stukenbrock EH, Jørgensen FG, Zala M, Hansen TT, McDonald BA, Schierup MH. 2010. Whole-Genome and Chromosome Evolution Associated with Host Adaptation and Speciation of the Wheat Pathogen Mycosphaerella graminicola. PLoS Genet. 6.

Stukenbrock EH, McDonald BA. 2008. The Origins of Plant Pathogens in Agro-Ecosystems. Annu. Rev. Phytopathol. 46:75–100.

Stukenbrock EH. 2013. Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. New Phytol. 199:895–907.

Sturgeon KB, Mitton JB. 1982. Evolution of bark beetle communities. In: Mitton JB, Sturgeon K. B., editors. Bark Beetles in North American Conifers. Austin, TX: Texas University Press. p. 350–384.

Sturgeon KB, Mitton JB. 1986a. Allozyme and Morphological Differentiation of Mountain Pine Beetles Dendroctonus ponderosae Hopkins (Coleoptera: scolytidae) Associated with Host Tree. Evolution 40:290–302.

Sturgeon KB, Mitton JB. 1986b. Allozyme and Morphological Differentiation of Mountain Pine Beetles Dendroctonus ponderosae Hopkins (Coleoptera: scolytidae) Associated with Host Tree. Evolution 40:290–302.

Sturgeon KB. 1979. Monoterpene Variation in Ponderosa Pine Xylem Resin Related to Western Pine Beetle Predation. Evolution 33:803.

Swofford D, Olsen G, Waddell P, Hillis D. 1996. Phylogenetic inference. In: Hillis D, Moritz C, Mable B, editors. Molecular Systematics. Sunderland, MA: Sinauer Associates. p. 407–514.

Swofford D. 2003. PAUP*: phylogenetic analysis using parsimony, version 4.0b10.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Mol. Biol. Evol. 24:1596 –1599.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28:2731–2739.

Taylor J, Jacobson D, Fisher M. 1999. The evolution of asexual fungi: Reproduction, Speciation and Classification. Annu. Rev. Phytopathol. 37:197–246.

Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, Fisher MC. 2000. Phylogenetic species recognition and species concepts in fungi. Fungal Genet. Biol. 31:21–32.

Taylor TN, Hass H, Remy W. 1992. Devonian Fungi: Interactions with the Green Alga Palaeonitella. Mycologia 84:901.

Taylor TN, Osborn JM. 1996. The importance of fungi in shaping the paleoecosystem. Rev. Palaeobot. Palynol. 90:249–262.

Templeton AR. 1983. Phylogenetic Inference From Restriction Endonuclease Cleavage Site Maps with Particular Reference to the Evolution of Humans and the Apes. Evolution 37:221.

Thompson JN. 1994. The coevolutionary process. Chicago, IL: University of Chicago Press

Trapnell C, Salzberg SL. 2009. How to map billions of short reads onto genomes. Nat. Biotechnol. 27:455–457.

Tsui CK-M, DiGuistini S, Wang Y, Feau N, Dhillon B, Bohlmann J, Hamelin RC. 2013. Unequal Recombination and Evolution of the Mating-Type (MAT) Loci in the Pathogenic Fungus Grosmannia clavigera and Relatives. G3 Genes Genomes Genet. 3:465–480.

Tsui CKM, Feau N, Ritland CE, Massoumi Alamouti S, DiGuistini S, Khadempour L, Bohlmann J, Breuil C, Hamelin RC. 2009. Characterization of microsatellite loci in the fungus, Grosmannia clavigera, a pine pathogen associated with the mountain pine beetle. Mol. Ecol. Resour. 9:1500–1503.

Tsui CKM, Roe AD, El-Kassaby YA, Rice AV, Alamouti SM, Sperling F a. H, Cooke JEK, Bohlmann J, Hamelin RC. 2012. Population structure and migration pattern of a conifer pathogen, Grosmannia clavigera, as influenced by its symbiont, the mountain pine beetle. Mol. Ecol. 21:71–86.

Tsuneda A, Currah R. 2006. Toward a deeper understanding of the nature of pleomorphism in conidial fungi. Reports Tottori Mycol. Inst. Jpn. 44:1–52.

Tsuneda A, Hiratsuka Y. 1984. Sympodial and annellidic conidiation in Ceratocystis clavigera. Can. J. Bot. 62:2618–2624.

Tyler BM, Tripathy S, Zhang X, et al. 2006. Phytophthora Genome Sequences Uncover Evolutionary Origins and Mechanisms of Pathogenesis. Science 313:1261–1266.

Upadhyay H, Kendrick W. 1974. A New Graphium-like Genus (Conidial State of Ceratocystis). Mycologia 66:181–13.

Upadhyay HP. 1981. Monograph of Ceratocystis and Ceratocystiopsis. Athens: University of Georgia Press

Upadhyay HP. 1993. Classification of the ophiostomatoid fungi. In: Wingfield MJ, Seifert KA, Webber .JF., editors. Ceratocystis and ophiostoma: taxonomy, ecology, and pathogenicity. St. Paul, Mirmesota.: American Phytopathological Society Press. p. 7–13.

Verrall AF. 1943. Fungi associated with certain ambrosia beetles. Agric. Res. 66:135–144.

Vilgalys R, Hester M. 1990. Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several Cryptococcus species. J. Bacteriol. 172:4238–4246.

Wang J, Wang W, Li R, et al. 2008. The diploid genome sequence of an Asian individual. Nature 456:60–65.

Wang W, Wei Z, Lam T-W, Wang J. 2011. Next generation sequencing has lower sequence coverage and poorer SNP-detection capability in the regulatory regions. Sci. Reports [Internet] 1. Available from: http://www.nature.com/srep/2011/110805/srep00055/full/srep00055.html

Wang Y, Lim L, DiGuistini S, Robertson G, Bohlmann J, Breuil C. 2013. A specialized ABC efflux transporter GcABC-G1 confers monoterpene resistance to Grosmannia clavigera, a bark beetle-associated fungal pathogen of pine trees. New Phytol. 197:886–898.

Wheeler DA, Srinivasan M, Egholm M, et al. 2008. The complete genome of an individual by massively parallel DNA sequencing. Nature 452:872–876.

Whitney H. 1971. Association of Dendroctonus ponderosae (Coleoptera: Scolytidae) with blue stain fungi and yeasts during brood development in lodgepole pine. Can. Entomol. 103:1495–1503.

Whitney HS, Farris SH. 1970. Maxillary Mycangium in the Mountain Pine Beetle. Science 167:54–55.

William J, Ballard O. 1996. Combining data in phylogenetic analysis. Trends Ecol. Evol. 11:334.

Wingfield MJ, De Beer ZW, Slippers B, Wingfield BD, Groenewald JZ, Lombard L, Crous PW. 2012. One fungus, one name promotes progressive plant pathology. Mol. Plant Pathol. 13:604–613.

Wingfield MJ, Seifert KA, Webber .J.F. eds. 1993. Ceratocystis and ophiostoma: taxonomy, ecology, and pathogenicity. St. Paul, Mirmesota: American Phytopathological Society Press

Wingfield MJ, Seifert KA. 1993. Ceratocystis and Ophiostoma: Taxonomy, Ecology, and Pathogenicity. (Webber JF, editor.). Amer Phytopathological Society

De Wit PJGM, van der Burgt A, Ökmen B, et al. 2012. The Genomes of the Fungal Plant Pathogens Cladosporium fulvum and Dothistroma septosporum Reveal Adaptation to Different Hosts and Lifestyles But Also Signatures of Common Ancestry. PLoS Genet 8:e1003088.

Wood S. 1963. A revision of the bark beetle genus. Dendroctonus Erichson (Coleoptera: Scolytidae). Gt. Basin Nat 23:1–117.

Wood S. 1982. The Bark and ambrosia beetles of North and Central America (Coleoptera, Scolytidae): a taxonomic monograph. Great Basin naturalist memoirs, no. 6

Wright SI, Andolfatto P. 2008. The Impact of Natural Selection on the Genome: Emerging Patterns in Drosophila and Arabidopsis. Annu. Rev. Ecol. Evol. Syst. 39:193–213.

Xue M, Yang J, Li Z, et al. 2012. Comparative Analysis of the Genomes of Two Field Isolates of the Rice Blast Fungus Magnaporthe oryzae. PLoS Genet 8:e1002869.

Yamaoka Y, Hiratsuka Y, Maruyama PJ. 1995. The ability of Ophiostoma clavigerum to kill mature lodgepolepine trees. Eur. J. For. Pathol. 25:401–404.

Yamaoka Y, Swanson RH, Hiratsuka Y. 1990. Inoculation of lodgepole pine with four blue-stain fungi associated with mountain pine beetle, monitored by a heat pulse velocity (HPV) instrument. Can. J. For. Res. 20:31–36.

Yang Z, Nielsen R. 2000. Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. Mol. Biol. Evol. 17:32–43.

Yang Z, Swanson WJ. 2002. Codon-Substitution Models to Detect Adaptive Evolution that Account for Heterogeneous Selective Pressures Among Site Classes. Mol. Biol. Evol. 19:49–57.

Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. 24:1586–1591.

Zambino PJ, Harrington TC. 1992. Correspondence of Isozyme Characterization with Morphology in the Asexual Genus Leptographium and Taxonomic Implications. Mycologia 84:12.

Zipfel RD, de Beer ZW, Jacobs K, Wingfield BD, Wingfield MJ. 2006. Multi-gene phylogenies define Ceratocystiopsis and Grosmannia distinct from Ophiostoma. Stud. Mycol. 55:75–97.

Zúñiga G, Cisneros R, Hayes JL, Macias-samano J. 2002. Karyology, Geographic Distribution, and Origin of the Genus Dendroctonus Erichson (Coleoptera: Scolytidae). Ann. Entomol. Soc. Am. 95:267–275.

Appendices

Appendix A: Supplementary information supporting chapter 3

A.1 List of isolates, sampling locations, collection resources and deposited culture collections

Fungal pecies	Isolate	Other culture collectios	Host tree/substrate	Beetle associate	Location	Collection site
Grosmannia sp. (Gs clade)	**UAMH 4585**		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Riske Creek
	NOF 1280		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Terry Fox
	CB H55	UAMH 11153	*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Houston
	CB H18		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Houston
	CB H19	UAMH 11348	*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Houston
	CB H21		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Houston
	CB H22		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Houston
	CB H41		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Houston
	CB H42		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Houston
	CB H43		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Houston
	CB H48		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Houston
	CB H50		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Houston
	CB SLA11		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Tweedsmuir Park
	CB W14		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Williams lake
	CB W6-1		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Williams lake
	CB SLKW1407	UAMH 11150	*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Kamloops
	CB 200-1-14	UAMH 11151	*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Kamloops
	CB KDW4		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Kelowna
	CB DPKGT1B	UAMH 11152	*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Kelowna
	CB M6		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Manning Park
	CB M3		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Manning Park
	CB M11		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Manning Park
	CB M44		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Manning Park
	CB M46		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Manning Park
	UAMH 4818		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Westcastle
	NOF 842		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Carbondale
	NOF 2893		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Blairmore
	CB B5	UAMH 11154	*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Banff
	CB B6		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Banff
	CB B10	UAMH 11155	*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Banff
	CB B14		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Banff
	CB B20		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Banff
	CB BW26		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Banff
	CB B1		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Banff
	CB B16		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Banff
	CB B17		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Banff
	CB B19		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Banff
	CB B21		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Banff
	CB BW22		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Banff
	CB BW27		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Banff
	CB BW28		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Banff
	CB B101		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Banff
	CB CHMC3	UAMH 11347	*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Cypress Hills
	CB CHDSC7	UAMH 11355	*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Cypress Hills
	CB CHEBC10	UAMH 11356	*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Cypress Hills
	CB CHIHC2		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Cypress Hills
	CB CHMPB8		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, Alberta	Cypress Hills

Fungal pecies	Isolate	Other culture collectios	Host tree/substrate	Beetle associate	Location	Collection site
	CB HV14		*Pinus contorta*	*Dendroctonus ponderosae*	USA, Montana	Hidden Valley
	CB HV30	UAMH 11357	*Pinus contorta*	*Dendroctonus ponderosae*	USA, Montana	Hidden Valley
	CB HV4		*Pinus contorta*	*Dendroctonus ponderosae*	USA, Montana	Hidden Valley
	CB HV6		*Pinus contorta*	*Dendroctonus ponderosae*	USA, Montana	Hidden Valley
	CB HV8		*Pinus contorta*	*Dendroctonus ponderosae*	USA, Montana	Hidden Valley
	CB HV9	UAMH 11358	*Pinus contorta*	*Dendroctonus ponderosae*	USA, Montana	Hidden Valley
	CB HV15	UAMH 11156	*Pinus contorta*	*Dendroctonus ponderosae*	USA, Montana	Hidden Valley
	CB HV20		*Pinus contorta*	*Dendroctonus ponderosae*	USA, Montana	Hidden Valley
	CB HV21		*Pinus contorta*	*Dendroctonus ponderosae*	USA, Montana	Hidden Valley
	CB HV24		*Pinus contorta*	*Dendroctonus ponderosae*	USA, Montana	Hidden Valley
	CB D1128	UAMH 11359	*Pinus contorta*	*Dendroctonus ponderosae*	USA, Idaho	Hell roaring
	CB D1151	UAMH 11360	*Pinus contorta*	*Dendroctonus ponderosae*	USA, Idaho	Hell roaring
	CB D1129		*Pinus contorta*	*Dendroctonus ponderosae*	USA, Idaho	Hell roaring
	CB D1131		*Pinus contorta*	*Dendroctonus ponderosae*	USA, Idaho	Hell roaring
	CB D1135		*Pinus contorta*	*Dendroctonus ponderosae*	USA, Idaho	Hell roaring
	CB D1137		*Pinus contorta*	*Dendroctonus ponderosae*	USA, Idaho	Hell roaring
	CB D1146		*Pinus contorta*	*Dendroctonus ponderosae*	USA, Idaho	Hell roaring
	CB D1153		*Pinus contorta*	*Dendroctonus ponderosae*	USA, Idaho	Hell roaring
	CB HR7-20(2)		*Pinus contorta*	*Dendroctonus ponderosae*	USA, Idaho	Hell roaring
	CB HR8-5		*Pinus contorta*	*Dendroctonus ponderosae*	USA, Idaho	Hell roaring
	DLS 1061		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	DLS 1037		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 12G13	UAMH 11349	*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 23G23		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 55B11	UAMH 11350	*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 68B21	UAMH 11361	*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 710G16	UAMH 11362	*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 11S26		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 11L24		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 23L11		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 23S11		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 24S12		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 24G22		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 24G24		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 55L22		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 55S11		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 55G23		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 56G14		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 67F21		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 68S13		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 68G12		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 79B34		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 79G31		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 79L21		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB 710G23		*Pinus contorta*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	CB Pa–9	UAMH 11363	*Pinus albicaulis*	*Dendroctonus ponderosae*	Canada, British Columbia	Nelson
	CB Pa–6	UAMH 11364	*Pinus albicaulis*	*Dendroctonus ponderosae*	Canada, British Columbia	Nelson
	CB Pa–1		*Pinus albicaulis*	*Dendroctonus ponderosae*	Canada, British Columbia	Nelson
	CB Pa–5		*Pinus albicaulis*	*Dendroctonus ponderosae*	Canada, British Columbia	Nelson
	CB Pa-10		*Pinus albicaulis*	*Dendroctonus ponderosae*	Canada, British Columbia	Nelson
	CB GCA02	UAMH 11365	*Pinus strobiformi*	*Dendroctonus ponderosae*	USA, Arizona	Pinaleño Mountains
	CB GCA04	UAMH 11366	*Pinus strobiformi*	*Dendroctonus ponderosae*	USA, Arizona	Pinaleño Mountains

Fungal pecies	Isolate	Other culture collectios	Host tree/substrate	Beetle associate	Location	Collection site
	CB GCA05		*Pinus strobiformi*	*Dendroctonus ponderosae*	USA, Arizona	Pinaleño Mountains
	CB GCA07		*Pinus strobiformi*	*Dendroctonus ponderosae*	USA, Arizona	Pinaleño Mountains
	CB GCA08		*Pinus strobiformi*	*Dendroctonus ponderosae*	USA, Arizona	Pinaleño Mountains
	CB GCA13		*Pinus strobiformi*	*Dendroctonus ponderosae*	USA, Arizona	Pinaleño Mountains
	CB GCA14		*Pinus strobiformi*	*Dendroctonus ponderosae*	USA, Arizona	Pinaleño Mountains
	CB PY2–3b	UAMH 11367	*Pinus ponderosae*	*Dendroctonus ponderosae*	Canada, British Columbia	Kamloops
	CB PY8–8	UAMH 11368	*Pinus ponderosae*	*Dendroctonus ponderosae*	Canada, British Columbia	Kamloops
	CB PY1–4		*Pinus ponderosae*	*Dendroctonus ponderosae*	Canada, British Columbia	Kamloops
	CB PY1(G1-5)A		*Pinus ponderosae*	*Dendroctonus ponderosae*	Canada, British Columbia	Kamloops
	CB PY2(G1-4)A		*Pinus ponderosae*	*Dendroctonus ponderosae*	Canada, British Columbia	Kamloops
	CB PY4–4		*Pinus ponderosae*	*Dendroctonus ponderosae*	Canada, British Columbia	Kamloops
	CB PY4(G5-8)D		*Pinus ponderosae*	*Dendroctonus ponderosae*	Canada, British Columbia	Kamloops
	CB PY8–7		*Pinus ponderosae*	*Dendroctonus ponderosae*	Canada, British Columbia	Kamloops
	CB KDPT5		*Pinus ponderosae*	*Dendroctonus ponderosae*	Canada, British Columbia	Kelowna
	CB KDLPT3		*Pinus ponderosae*	*Dendroctonus ponderosae*	Canada, British Columbia	Kelowna
	CB KGBPT2		*Pinus ponderosae*	*Dendroctonus ponderosae*	Canada, British Columbia	Kelowna
	DPLKGAPT6		*Pinus ponderosae*	*Dendroctonus ponderosae*	Canada, British Columbia	Kelowna
	CB KGW5		*Pinus ponderosae*	*Dendroctonus ponderosae*	Canada, British Columbia	Kelowna
Grosmannia clavigera (Gc clade)	**ATCC18086**		*Pinus ponderosae*	*Dendroctonus ponderosae*	Canada, British Columbia	Cache Creek
	CB 15B29C1	UAMH 11369	*Pinus ponderosae*	*Dendroctonus ponderosae*	USA, South Dakota	Black Hills
	CB 15B29C2		*Pinus ponderosae*	*Dendroctonus ponderosae*	USA, South Dakota	Black Hills
	CB 16B17C3		*Pinus ponderosae*	*Dendroctonus ponderosae*	USA, South Dakota	Black Hills
	CB 16B24C4		*Pinus ponderosae*	*Dendroctonus ponderosae*	USA, South Dakota	Black Hills
	CB 23B110C5		*Pinus ponderosae*	*Dendroctonus ponderosae*	USA, South Dakota	Black Hills
	CB 34B94C6	UAMH 11370	*Pinus ponderosae*	*Dendroctonus ponderosae*	USA, South Dakota	Black Hills
	CB 24B166C7		*Pinus ponderosae*	*Dendroctonus ponderosae*	USA, South Dakota	Black Hills
	CB 24B166C8	UAMH 11371	*Pinus ponderosae*	*Dendroctonus ponderosae*	USA, South Dakota	Black Hills
	CB 32B85C9		*Pinus ponderosae*	*Dendroctonus ponderosae*	USA, South Dakota	Black Hills
	CB 32B85C10		*Pinus ponderosae*	*Dendroctonus ponderosae*	USA, South Dakota	Black Hills
	CB 34B94C11		*Pinus ponderosae*	*Dendroctonus ponderosae*	USA, South Dakota	Black Hills
	CB 34B94C12		*Pinus ponderosae*	*Dendroctonus ponderosae*	USA, South Dakota	Black Hills
	CB 11B8C13		*Pinus ponderosae*	*Dendroctonus ponderosae*	USA, South Dakota	Black Hills
	CB 11B23C14		*Pinus ponderosae*	*Dendroctonus ponderosae*	USA, South Dakota	Black Hills
	CB 21B69C15		*Pinus ponderosae*	*Dendroctonus ponderosae*	USA, South Dakota	Black Hills
	DLS 15	UAMH 11372	*Pinus ponderosae*	*Dendroctonus ponderosae*	USA, California	Sierra Nevada
	DLS 24	UAMH 11373	*Pinus ponderosae*	*Dendroctonus ponderosae*	USA, California	Lassen
	DLS 56	UAMH 11374	*Pinus ponderosae x P. Jeffreyi*	*Dendroctonus ponderosae*	USA, California	Lassen
	C843		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Sierra Nevada
	DLS 554	CMW 15398 / UAMH 11375	*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Sierra Nevada
	DLS 776	CMW 15785	*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Sierra Nevada
	DLS 833	UAMH 11376	*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Sierra Nevada
	DLS 681		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Sierra Nevada
	DLS 690		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Sierra Nevada
	DLS 771		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Sierra Nevada
	DLS 417		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Sierra Nevada
	DLS 651		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Sierra Nevada
	DLS 792		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Sierra Nevada
	DLS 108		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Lassen
	DLS 120		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Lassen
	DLS 122	CMW 15394	*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Lassen
	DLS 126		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Lassen
	DLS 140		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Lassen

Fungal pecies	Isolate	Other culture collectios	Host tree/substrate	Beetle associate	Location	Collection site
	DLS 173	CMW 15395 / UAMH 11377	*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Lassen
	DLS 190		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Lassen
	DLS 210		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Lassen
	DLS 235		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Lassen
	DLS 237	UAMH 11378	*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Lassen
	DLS 52	CMW 15783	*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	San Bernardino Mountains
	DLS 1560		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	San Bernardino Mountains
	DLS 1565		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	San Bernardino Mountains
	DLS 1575	UAMH 11351	*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	San Bernardino Mountains
	DLS 1588	UAMH 11352	*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	San Bernardino Mountains
	DLS 1595		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	San Bernardino Mountains
	DLS 1561		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	San Bernardino Mountains
	DLS 1571		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	San Bernardino Mountains
	DLS 1581		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	San Bernardino Mountains
	DLS 1591		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	San Bernardino Mountains
Grosmannia aurea	**CBS 438.69**		*Pinus contorta*	*Dendroctonus* sp.	Canada, British Columbia	Invermere
Leptographium longiclavatum	**CB SLKW1436**		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Kamloops
	C 845		*Pinus jeffreyi*	*Dendroctonus jeffreyi*	USA, California	Sierra Nevada
Leptographium terebrantis	**CB 878 AW 1-2**		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Kamloops
	CB LPKRLT-3		*Pinus contorta*	*Dendroctonus ponderosae*	Canada, British Columbia	Kamloops
	C 418		*Pinus ponderosae*	*D. brevicomis*	USA, California	Sierra Nevada
Leptographium wingfieldii	**CBS 645.89**		*P. sylvestris*	*Tomicus piniperda*	France,	Orléans
	CBS 648.89		*P. brutia*	NA	Greece,	Thessaloniki

***G. clavigera* host tree**	**Total number**
Pinus contorta	92
Pinus albicaulis	5
Pinus strobiformi	7
Pinus ponderosae	32
Dendroctonus ponderosae	136
Dendroctonus jeffreyi/Pinus jeffreyi	30
Total number of isolates	**166**

Isolates selected for the phylogenetic analysis are bolded.

A.2 Primer sequences and gene descriptions for 67 *G. clavigera* loci screened for polymorphisms

Locus no.	Gene no.	Gene description (abbreviation)	Primers sequence	Target scaffold (Length of target	Target gene (Length of target	Length of sequence
1	**1**	**ABC transporter 1 (ABC)**	**ATGTGCAGGGTGGCGAGCGAA** **GAATACCGCTCCGCTCGCACA**	GCLVSC_144 (2,496,812)	GLEAN_6117 (6,801)	549
2	2	ABC transporter 2	GTTGACGTGCTTGATGAGG AGACGGACGAATAGGGAG	GCLVSC_82 (1,001,357)	GLEAN_1593 (2,205)	857
3			AGGAAGAGCGAAGAGGCA AGGATGGGCTGGTACGGA	GCLVSC_113 (3,089,927)	GLEAN_3921 (3,354)	810
4	**3**	ABC transporter 3	AATCCCGATGCTTGCCCT CTGCCTCTTCGCTCTTCCT	GCLVSC_113 (3,089,927)	GLEAN_3921 (3,354)	804
5			AAACTCCGCCTTCAGCCA ATCACCGCTTCTCCACCA	GCLVSC_113 (3,089,927)	GLEAN_3921 (3,354)	830
6			CTCTGTTTGCATCACTTCTC ATCATGGCTCTACCTCCT	GCLVSC_144 (2,496,812)	GLEAN_1678 (4,563)	865
7	4	Lipases 1	GTGTACCCCAAAGCCACC CAAAATCGCCACAGTCCTC	GCLVSC_144 (2,496,812)	GLEAN_1678 (4,563)	770
8			TCTCGTTGTCTATGTCGCT AAAACTCCTGCCCAAACC	GCLVSC_144 (2,496,812)	GLEAN_1678 (4,563)	1,297
9	5	Lipases 2	CGATGTTTTGGCCGTTGT CTCTGCGTGTCCTGGTTT	GCLVSC_113 (3,089,927)	GLEAN_4044 (2,286)	872
10			ATGCGCCTCCACTTGTCC ACTTCTACACCGACTTCCTCCT	GCLVSC_173 (2,341,815)	GLEAN_5805 (2,748)	720
11	**6**	**Lipid acyl hydrolase (LAH)**	**CACACGGACCAACGACGA** **CTCTCCTGCCCCTCTTCTC**	GCLVSC_173 (2,341,815)	GLEAN_5805 (2,748)	1,125
12	7	Multidrug facilitator superfamily transporter 1	GGAGACAGCGGGTATAGAG CAAGCAGATATGGAAACGGA	GCLVSC_89 (1,124,797)	GLEAN_4905 (516)	857
13			TCCCAAACAAACAGGCCA ATCCACACACAGCATCAG	GCLVSC_132 (1,063,191)	GLEAN_851 (1,695)	769
14	8	Multidrug facilitator superfamily transporter 2	GTGGTTCTCAGTCTTCTCGT TGATGCTGTGTGTGGATGG	GCLVSC_132 (1,063,191)	GLEAN_851 (1,695)	641
15			**GACATTGTAGAGGGCAGC** **AGATGGGAGGTTGGAGAG**	GCLVSC_113 (3,089,927)	GLEAN_8113 (1,641)	847
16	**9**	**Cytochrome P450 1 (P450 I)**	**AGTAGAACACCGCCGACAG** **CCGACCAAACACACCGCA**	GCLVSC_113 (3,089,927)	GLEAN_8113 (1,641)	792
17	**10**	**Cytochrome P450 2 (P450 II)**	**TGCAGCAATGGGACCGGATGA** **TCGTCACGTTCTCCCAGCGCT**	GCLVSC_173 (2,341,815)	GLEAN_1091 (1,578)	710
18	**11**	Cytochrome P450 3	TGAACGAGACGCTGCGTCTGATG TGCAGCTTCAGCGTCAACAGGGT	GCLVSC_167 (1,996,734)	GLEAN_4745 (2,010)	560
19			CTTTGTCATCCACCCGCT CATCCATCTCCTCGGCCT	GCLVSC_132 (1,063,191)	GLEAN_5485 (1,569)	575
20	12	Cytochrome P450 4	GATCGCAGACATTGGCCT CCTCCTCCACACCTTCAC	GCLVSC_132 (1,063,191)	GLEAN_5485 (1,569)	818
21	13	**Metallo-peptidase (MPEP)**	**TAAGGAAAGGGAGGGCGGT** **TGGGTGCGTGATGAGCGA**	GCLVSC_113 (3,089,927)	GLEAN_8241 (1,485)	913

Locus no.	Gene no.	Gene description (abbreviation)	Primers sequence	Target scaffold (Length of target	Target gene (Length of target	Length of sequence
22			**ATTCCCCTCCCCTACTCC** **CTTCCATGTCCTCCTTCC**	GCLVSC_113 (3,089,927)	GLEAN_8241 (1,485)	780
23			**TGATTCGACTTTCCCCCT** **CGTCGAACACAAACTCCT**	GCLVSC_144 (2,496,812)	GLEAN_1886 (2,325)	886
24	14	**Anthranilate synthase (TRPG)**	**GGAGTTTGTGTTCGACGAG** **GAATGACAAGGCTATGAAGGGA**	GCLVSC_144 (2,496,812)	GLEAN_1886 (2,325)	975
25	15	beta-tubulin	GCTGTCATGTCTGGCGTGTCCA TCTTCGCCATACGCCTCCTCG	GCLVSC_141 (93,906)	GLEAN_4834 (1,341)	595
26	16	**alpha-tubulin**	**TCCAGACGAACCTGGTGCCGT** **CAGGCGTCATCGAGCAAGCGA**	GCLVSC_108 (1,834,804)	GLEAN_6984 (1,353)	641
27	17	**40S ribosomal protein S3 (40SRP)**	**TCACGCCCACCGTTACCGACA** **TGGAAATGGTCGGTGCCGAGGT**	GCLVSC_140 (4,437,291)	GLEAN_2722 (795)	752
28	18	**Phosphatidylinositol transferase (PLT)**	**CGGTCGCCCGCTCTACATTGA** **CTCAGCCTCTAAGCCGTTGCCT**	GCLVSC_168 (797,636)	GLEAN_6502 (1,017)	570
29	19	Acetyltransferase	TGGCGGCAAGATCACGGAGGA CCATTCGCTCGTGCTCTCGCTT	GCLVSC_140 (4,437,291)	GLEAN_2685 (1,404)	600
30	20	Glyceraldehyde 3-phosphate dehydrogenase	GAAGGGTGGTGCCAAGAAGGTGA TCTTCCGCCCCTCCCTCGTAGT	GCLVSC_108 (1,834,804)	GLEAN_7049 (954)	549
31	21	Alcohol dehydrogenase	CGTGACGAAGAGCGGCATCGA TGCAGCACGAGCTTGCCGAACT	GCLVSC_140 (4,437,291)	GLEAN_2576 (1,170)	489
32	22	**Lysophospholipase**	**TGCTGTCGAGAACTGGAGGCGT** **CGGCAGGACCTGGAACAGGAA**	GCLVSC_156 (1,463,716)	GLEAN_6592 (726)	568
33	23	putative septin	TCGAGGAGAACGCCGAGCTGA TCGAGCGAACGCTGGTACCACT	GCLVSC_160 (595,886)	GLEAN_3045 (1,137)	666
34	24	Thiol-specific antioxidant	TCACTGCCACCACCGTCTTTCCT GCACGAAGAAAGCACGCAGGA	GCLVSC_132 (1,063,191)	GLEAN_5524 (642)	648
35	25	Superoxide dismutase	CATGCGTGTTCGGCCTTCTGCT TGCCGACCCAACTGGCTTAGCCT	GCLVSC_97 (2,272,097)	GLEAN_5060 (687)	726
36	26	Aspartic proteinase	AGATCGAGTCCACGTCACCGCA GGTTGCGCCATTGAAGACGACA	GCLVSC_160 (595,886)	GLEAN_3069 (1,263)	758
37	27	Polyketide synthase 1	CGTCGTTCCTGCCGCTGATTGA ACCGTCCTCCACTCCGATCAGCT	GCLVSC_144 (2,496,812)	GLEAN_1845 (6,522)	515
38	28	Polyketide synthase 2	TGGAAACCGAGGACAGGCGAACT ACGCCGGTCTCACGAGAAATCCA	GCLVSC_160 (595,886)	GLEAN_3011 (6,633)	564
39	29	**Peroxisomal-coenzyme A synthetase (PCAS)**	**TGCCGACAAGGTGGCCAAGTTC** **GCGCAGCGCAACATTGACGACT**	GCLVSC_167 (1,996,734)	GLEAN_0350 (1575)	687
40	30	**Common fungal extracellular membrane protein 1 (CFEM I)**	**CTCTTCTTTGCCGGCCTTGCTGT** **CGCAACGCAAACGCCAGAAGA**	GCLVSC_173 (2,341,815)	GLEAN_5868 (582)	667
41	31	**CFEM 2 (CFEM II)**	**GCGTCCATTGATCGGCGTGATGT** **AACCGCCAACATGGCAACGG**	GCLVSC_156 (1,463,716)	GLEAN_3507 (564)	491
42	32	CFEM 3	TGTTCCGTGCCTTCGATGCCCT AGGTATCGCTGACCACATCCCGA	GCLVSC_168 (797,636)	GLEAN_1411 (921)	824
43	33	CFEM 4	CTCTCTTTCGTCCTGGCTCTCGG CAAGACTCCGAGTCCGCCATTG	GCLVSC_161 (1,111,432)	GLEAN_4535 (366)	260
44	34	CFEM 5	TGCTTCTGCTGCTCTGCGTGGT TGGACCGACGTAGTTGTGGCCGT	GCLVSC_132 (1,063,191)	GLEAN_827 (1,551)	1,229

Locus no.	Gene no.	Gene description (abbreviation)	Primers sequence	Target scaffold (Length of target	Target gene (Length of target	Length of sequence
45	35	CFEM 6	ATCGCACCCTTGACTACCGCCA GGTGACAGAAAAGAGGACGCGCT	GCLVSC_142 (425,798)	GLEAN_5021 (627)	608
46	36	Anonymous 1	TCACGCCATACCAGATCC AAACTCCCTCCTTCATCCA	GCLVSC_179 (1,213,309)	GLEAN_4283 (3,915)	833
47			AGGAGTACGAGGCCACAG GCCACAGAACAGATAACGA	GCLVSC_179 (1,213,309)	GLEAN_4283 (3,915)	607
48	37	Anonymous 2	CCATCCGACAACAACACC TCCTCCACGTCATCCTCCA	GCLVSC_140 (4,437,291)	GLEAN_2807 (603)	871
49	38	Anonymous 3	CTCTATTCCCCTTCAACTCTCTC CTTCTCCCACCCCTTGAC	GCLVSC_144 (2,496,812)	GLEAN_1731 (765)	825
50	39	Anonymous 4	TTTCGTGTTGATCGCCTG CCACCACACAATTGCTCC	GCLVSC_161 (1,111,432)	GLEAN_3159 (717)	594
51	40	Anonymous 5	CAAAGCAACCCGCAAAACAC ACGTAGCGACCGACAACC	GCLVSC_108 (1,834,804)	GLEAN_6881 (1,062)	878
52	41	Anonymous 6	TCACATACCGGAGCCACCA AAAACAAAGAACGGCACCCA	GCLVSC_144 (2,496,812)	GLEAN_1625 (765)	820
53	42	Anonymous 7	GAGGGAAGGAAAAGGGGA CTGGCTAAATCTCTCTTCGT	GCLVSC_140 (4,437,291)	GLEAN_2380 (342)	629
54	43	Anonymous 8	CATCCTCACCCTTCAACC ACTTAAACCACGCCAGAC	GCLVSC_140 (4,437,291)	GLEAN_7708 (546)	810
55			CTTCGCCCAACGGTACCA CAGCCAAAAAAGGGGGCCA	GCLVSC_140 (4,437,291)	GLEAN_2314 (1,362)	912
56	44	Anonymous 9	ATCTTCCTCTCTTTCCACAC CACTTTGTTTCCGCTCCT	GCLVSC_173 (2,341,815)	GLEAN_1181 (618)	740
57	46	Anonymous 11	TATTGTTGAGATTGGCCCTG TTGGATATCTTGTGCTGCTG	GCLVSC_113 (3,089,927)	GLEAN_4069 (5,592)	830
58			CCATCGCTGCCAAATGTC CTCCTGGCACGCTATCTC	GCLVSC_113 (3,089,927)	GLEAN_4069 (5,592)	810
59			AGGTGTCGTTGTTGATGTGG GTTTGTGCTGGTGTCGGT	GCLVSC_113 (3,089,927)	GLEAN_4069 (5,592)	730
60			ACCGCATTTTCCTCATCAC CCCTCTCACAATCTCCCA	GCLVSC_113 (3,089,927)	GLEAN_7855 (6,297)	900
61			CTTCTTCGCGTTCTTGCT GGCATTTCTGGCTTGGGT	GCLVSC_113 (3,089,927)	GLEAN_7855 (6,297)	790
62			AGAGGGTGAAGTGAAAAGG TGAAGAAGGGTGACGAGG	GCLVSC_113 (3,089,927)	GLEAN_7855 (6,297)	907
63			TATCCCAGTTCGCCAGCA AAGAACAACCCGAAGGAG	GCLVSC_113 (3,089,927)	GLEAN_7855 (6,297)	783
64	**47**	**Anonymous 12 (Anonymous I)**	**CACGACGACGAACTCCTCTCCCA** **CAGGATGCCCTCGGCCTCTAAC**	GCLVSC_167 (1,996,734)	GLEAN_4597 (381)	458
65	**48**	**Anonymous 13 (Anonymous II)**	**TGCCAGACTGGTCCACATCTGCA** **ACGCCGGCAAGACCTACACCA**	GCLVSC_140 (4,437,291)	GLEAN_7505 (432)	805
66	49	Anonymous 14	TGGTCCACAGTACCATCCCGTCA TGCATGTGCCCACGTCCACGA	GCLVSC_113 (3,089,927)	GLEAN_3737 (603)	661
67	50	Anonymous 15	ACCGCGAAGATGCCAGGCAAACA GAGAACGCAACAGAACGCAGCCA	GCLVSC_173 (2,341,815)	GLEAN_5862 (870)	631

Genes and primer sequences selected for further phylogenetic and population genetic analyses are bolded

TreeBASE URL for concatenated alignment of the 67 loci (49,853 base pairs): http://purl.org/phylo/treebase/phylows/study/TB2:S11355

A.3 Polymorphism summaries and diversity indices within the two monophyletic clades in *G. clavigera*

Gene predictions	Species clade	Number of single nucleotide polymorphisms					Indels (length in base pairs)	Diversity indices [a]			
		Total	Singleton	Noncoding	Synonymous	Replacement		Number of haplotypes	Gene diversity (H)	Diversity π (10^{-3})	Diversity θ (10^{-3})
40SRP	Gs	2	1	0	1	1	0	3	0.27	0.37	0.63
	Gc	1	1	0	1	0	0	2	0.09	0.12	0.37
alpha tubulin	Gs	0	0	0	0	0	0	1	0.00	0.00	0.00
	Gc	0	0	0	0	0	0	1	0.00	0.00	0.00
ABC	Gs	1	1*	0	0	1	0	2	0.05	0.09	0.43
	Gc	0	0	0	0	0	0	1	0.00	0.00	0.00
TRPG	Gs	6	1*	0	1	5	0	4	0.65	1.30	0.73
	Gc	1	1	0	1	0	0	2	0.09	0.05	0.14
MPEP	Gs	2	1	0	1	1	0	3	0.14	0.09	0.28
	Gc	2	1*	0	1	1	0	3	0.51	0.33	0.33
P450 I	Gs	3	1*	0	1	2	0	4	0.38	0.26	0.44
	Gc	3	2*	0	0	3	0	4	0.39	0.27	0.33
P450 II	Gs	2	1	0	2	0	0	3	0.23	0.39	0.66
	Gc	3	1	0	2	1	0	4	0.50	1.01	1.16
LAH	Gs	0	0	0	0	0	0	1	0.00	0.00	0.00
	Gc	1	0	0	0	1	0	2	0.24	0.22	0.24
CFEM I	Gs	5	1	2	3	0	0	4	0.58	2.53	1.77
	Gc	1	0	0	0	1	0	2	0.42	0.62	0.41
CFEM II	Gs	4	0	2	1	1	0	4	0.73	3.78	1.92
	Gc	5	0	1	2	2	0	4	0.73	3.60	2.81
LPL	Gs	2	1	0	1	1	0	3	0.50	0.91	0.83
	Gc	0	0	0	0	0	0	1	0.00	0.00	0.00
PLT	Gs	2	0	0	2	0	0	3	0.43	0.80	0.83
	Gc	1	1	0	1	0	0	2	0.09	0.16	0.48
PCAS	Gs	1	1	1	0	0	0	2	0.05	0.07	0.34
	Gc	2	2	2	0	0	0	3	0.17	0.27	0.80
Anonymous I	Gs	0	0	0	0	0	0	1	0.00	0.00	0.00
	Gc	1	1	1	0	0	0	2	0.09	0.20	0.60
Anonymous II	Gs	3	1	2	1	0	2 (3)	3	0.39	0.96	0.88
	Gc	3	0	1	2	0	0	4	0.65	0.99	1.02
Total Within *G. clavigera*	62	86	18	18	33	35	2 (6)	58	0.99	2.00	1.37
Total within Gs	40	33	10	7	14	12	2 (6)	36	0.99	0.68	0.59
Total within Gc	22	24	10	5	10	9	0	22	1.00	0.41	0.50

*Coding regions

[a] H, haplotype/gene diversity which corresponds to the probability that two randomly chosen individuals are different at a chosen locus; π, nucleotide diversity based on average number of nucleotide differences per site between two sequences (Nei 1987); θ, nucleotide diversity based on the number of segregating sites (Watterson 1975)

A.4 Gs morphology compared with those of the *G. clavigera* holotype

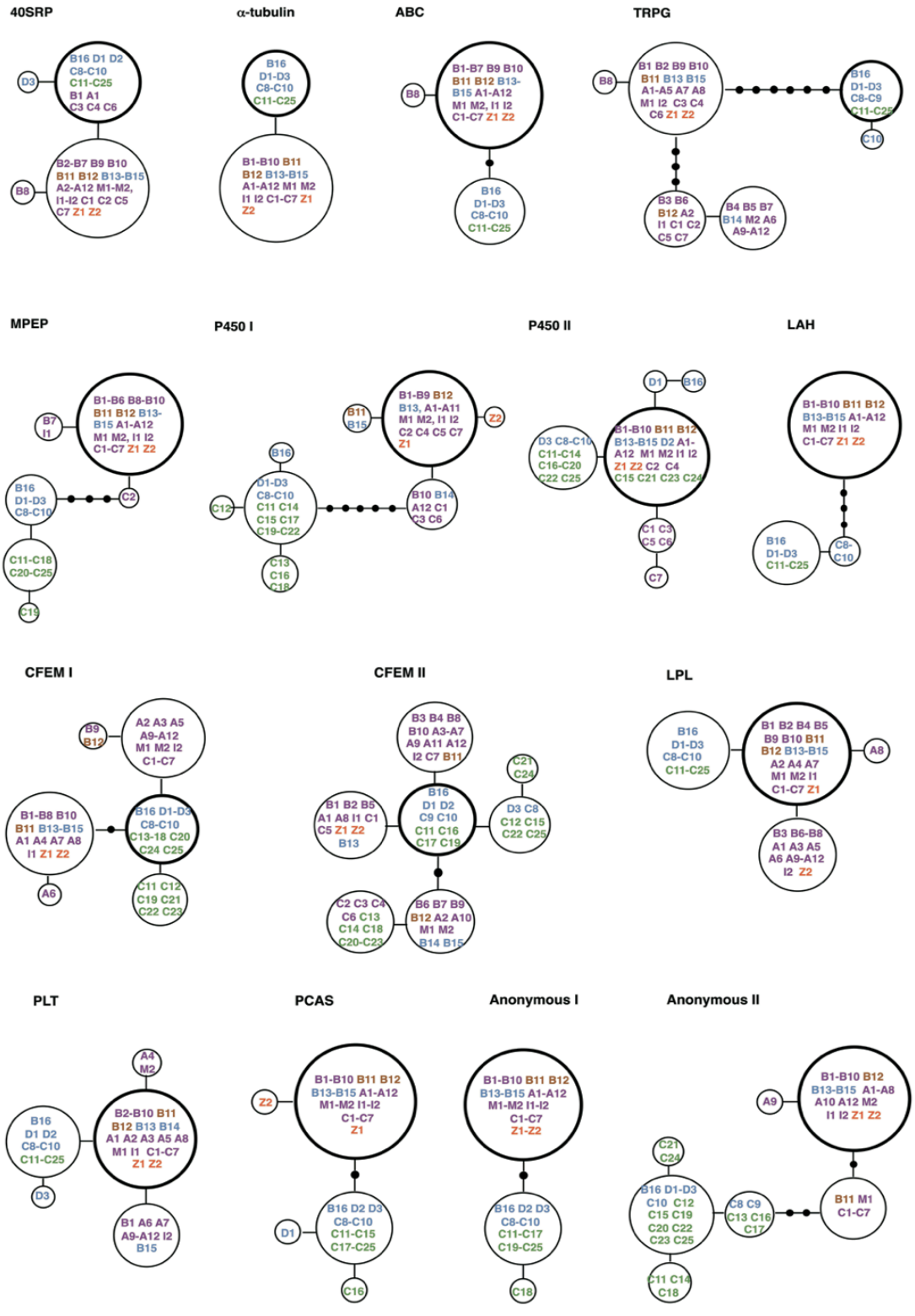
Fungal isolates	*Grosmannia* sp. (Gs clade)	*G. clavigera* (Gc clade)a
Host tree	*Pinus contorta*	*Pinus ponderosa*
Insect associate	*Dendroctonus ponderosae*	*Dendroctonus ponderosae*
Anamorph (asexual reproduction)	Mononematous *Leptographium* and synnematous conidiophores	Mononematous *Leptographium* and synnematous conidiophores
Conidiophore length including conidiogenous apparatus (µm)	80–1150	100–1040
Conidium shape	Clavate and cylindrical to oblong	Clavate and cylindrical to obclavate
Conidium size (clavate µm) / (others µm)	16.5–62 × 4–6 / 14.8 × 2.3	35–68 × 4.2–5.6 / 12.6 × 3.8
Teleomorph (sexual reproduction)	*Grosmannia*	*Grosmannia*
Ascocarp shape (color)	spherical cleistothecia b (black)	spherical cleistothecia (black)
Ascocarp size	245–570	250–640
Ascospore shape	reniform (cucullate in side view)	reniform (cucullate in side view)
Ascospore size	3.5–5.0 × 2.5–4.0	3.5–5.6 × 2.8–4.2

[a] Observations by Robinson and Davidson (1968) and Six and Paine (1997)

Lee et al. (2003) have observed ascocarps with short necks ranging from 20–65 µm

A.5 Haplotype network

Traditional phylogenetic methods that produce bifurcating gene trees cannot accurately portray genealogical relationships at the intraspecific level, due to different population processes such as recombination and/or presence of ancestral alleles. Therefore, for each of the 15 gene datasets, we generated parsimony networks of *G. clavigera* haplotypes using TCS 1.21 (Clement et al. 2000). This method estimates the unrooted tree and provides a 95% plausible set for parsimonious relationships between all haplotypes. Network approaches can account for homoplasy by introducing loops where recombination or frequent mutation could have occurred in the history of the genealogy.


40SRP
B16 D1 D2 C8-C10 C11-C25 B1 A1 C3 C4 C6
D3
B2-B7 B9 B10 B11 B12 B13-B15 A2-A12 M1-M2, I1-I2 C1 C2 C5 C7 Z1 Z2
B8
α-tubulin
B16 D1-D3 C8-C10 C11-C25
B1-B10 B11 B12 B13-B15 A1-A12 M1 M2 I1 I2 C1-C7 Z1 Z2
ABC
B1-B7 B9 B10 B11 B12 B13-B15 A1-A12 M1 M2, I1 I2 C1-C7 Z1 Z2
B8
B16 D1-D3 C8-C10 C11-C25
TRPG
B1 B2 B9 B10 B11 B13 B15 A1-A5 A7 A8 M1 I2 C3 C4 C6 Z1 Z2
B8
B16 D1-D3 C8-C9 C11-C25
C10
B3 B6 B12 A2 I1 C1 C2 C5 C7
B4 B5 B7 B14 M2 A6 A9-A12
MPEP
B1-B6 B8-B10 B11 B12 B13-B15 A1-A12 M1 M2, I1 I2 C1-C7 Z1 Z2
B7 I1
B16 D1-D3 C8-C10
C2
C11-C18 C20-C25
C19
P450 I
B1-B9 B12 B13, A1-A11 M1 M2, I1 I2 C2 C4 C5 C7 Z1
B11 B15
Z2
B16
D1-D3 C8-C10 C11 C14 C15 C17 C19-C22
C12
B10 B14 A12 C1 C3 C6
C13 C16 C18
P450 II
D1
B16
D3 C8-C10 C11-C14 C16-C20 C22 C25
B1-B10 B11 B12 B13-B15 D2 A1-A12 M1 M2 I1 I2 Z1 Z2 C2 C4 C15 C21 C23 C24
C1 C3 C5 C6
C7
LAH
B1-B10 B11 B12 B13-B15 A1-A12 M1 M2 I1 I2 C1-C7 Z1 Z2
B16 D1-D3 C11-C25
C8-C10
CFEM I
B9 B12
A2 A3 A5 A9-A12 M1 M2 I2 C1-C7
B1-B8 B10 B11 B13-B15 A1 A4 A7 A8 I1 Z1 Z2
B16 D1-D3 C8-C10 C13-18 C20 C24 C25
A6
C11 C12 C19 C21 C22 C23
CFEM II
B3 B4 B8 B10 A3-A7 A9 A11 A12 I2 C7 B11
C21 C24
B1 B2 B8 A1 A8 I1 C1 C5 Z1 Z2 B13
B16 D1 D2 C9 C10 C11 C16 C17 C19
D3 C8 C12 C15 C22 C25
C2 C3 C4 C6 C13 C14 C18 C20-C23
B6 B7 B9 B12 A2 A10 M1 M2 B14 B15
LPL
B16 D1-D3 C8-C10 C11-C25
B1 B2 B4 B5 B9 B10 B11 B12 B13-B15 A2 A4 A7 M1 M2 I1 C1-C7 Z1
A8
B3 B6-B8 A1 A3 A5 A6 A9-A12 I2 Z2
PLT
A4 M2
B16 D1 D2 C8-C10 C11-C25
D3
B2-B10 B11 B12 B13 B14 A1 A2 A3 A5 A8 M1 I1 C1-C7 Z1 Z2
B1 A6 A7 A9-A12 I2 B15
PCAS
B1-B10 B11 B12 B13-B15 A1-A12 M1-M2 I1-I2 C1-C7 Z1
Z2
B16 D2 D3 C8-C10 C11-C15 C17-C25
D1
C16
Anonymous I
B1-B10 B11 B12 B13-B15 A1-A12 M1-M2 I1-I2 C1-C7 Z1-Z2
B16 D2 D3 C8-C10 C11-C17 C19-C25
C18
Anonymous II
B1-B10 B12 B13-B15 A1-A8 A10 A12 M2 I1 I2 Z1 Z2
A9
C21 C24
B16 D1-D3 C10 C12 C15 C19 C20 C22 C23 C25
C8 C9 C13 C16 C17
B11 M1 C1-C7
C11 C14 C18

Appendix B: Supplementary information supporting chapter 4

B.1 Main features of primary genome sequence data

StrainID	Sequencing Platform	Read Length	Low Kmer	High Kmer
GsB2	Illumina Genome Analyzer II	50	25	35
GsB3	Illumina Genome Analyzer IIx	75	31	61
GsA1	Illumina Genome Analyzer IIx	75	31	61
GsA2	Illumina Genome Analyzer IIx	75	31	61
GsA3	Illumina Genome Analyzer II	50	25	35
GsM1	Illumina Genome Analyzer II	50	25	35
GsC1	Illumina Genome Analyzer IIx	75	31	61
GsC2	Illumina Genome Analyzer IIx	75	31	61
GcB1.a	Illumina Genome Analyzer II	50	25	35
GcB1.b	Illumina Genome Analyzer IIx	75	31	61
GcC1	Illumina Genome Analyzer II	50	25	35
GcC2	Illumina Genome Analyzer IIx	75	31	61

B.2 Assembly statistics of *Grosmannia* genomes

Strain [a]	*de novo* contigs						after ordering and orientating *de novo* contigs against the refernece sequence [b]						Aligned contigs [c]	
	Assembly Size	Number of Scaffolds	N50	N90	Gaps	Percent Gaps	Assembly Size	Number of Scaffolds	N50	N90	Gaps	Percent Gaps	bp length	%
GsB2	27,738,425	1,348	95,703	22,052	40,805	0.15	28,453,282	91	2,219,886	787,389	2,460,073	8.6	27,931,310	101
GsB3	30,914,973	2,924	163,110	10,928	15,386	0.05	29,213,916	140	1,984,429	594,737	940,778	3.2	30,794,128	100
GsA1	29,882,447	2,597	159,048	10,755	12,448	0.04	29,066,433	144	2,242,497	588,345	1,068,256	3.7	29,957,163	100
GsA2	29,997,851	2,327	111,753	8,908	20,875	0.07	28,802,836	143	2,220,345	594,681	919,606	3.2	29,952,967	100
GsA3	27,996,749	1,376	92,960	19,371	47,237	0.17	28,153,243	85	2,260,542	655,482	2,617,547	9.3	28,041,899	100
GsM1	28,888,169	1,414	102,139	21,466	49,831	0.17	27,600,675	96	2,166,132	756,581	2,504,286	9.1	28,985,091	100
GsC1	31,128,695	2,942	177,573	9,769	31,673	0.10	28,880,088	146	1,981,110	594,384	1,338,407	4.6	30,521,551	98
GsC2	32,243,025	2,397	221,770	13,462	20,957	0.06	28,560,494	146	1,834,826	588,324	1,217,577	4.3	31,761,788	99
GcB1.a	28,694,637	4,000	32,813	6,628	28,776	0.10	28,615,521	127	2,180,445	581,727	2,158,396	7.5	27,719,482	97
GcB1.b	30,843,415	1,409	133,367	17,320	459	0.00	29,424,681	123	1,983,195	591,692	1,320,203	4.5	29,143,054	94
GcC1	28,399,467	1,353	98,789	20,245	42,002	0.15	27,978,235	70	2,216,111	744,378	2,583,449	9.2	27,851,123	98
GcC2	32,445,864	2,787	155,238	6,569	20,119	0.06	28,845,018	136	1,993,212	591,042	1,263,041	4.4	30,323,732	93

[a] IDs, "a" and "b" are results from two independent sequence lanes for the same strain.

[b] The reference genome published by DiGuistini et al. 2011

[c] The total length and percentage of de-novo contigs aligned to the reference sequence over at least 100 bp with 95% identity. For strains with multiple contigs mapped to the same location in the reference genome, the total length and percentage of aligned contigs are higher than original de novo assembly values, representing a measure of redundancy and divergence from the reference genome

B.3 Summary of genBlastG output used for gene annotations of each Grosmannia draft genome, with pairwise homology (PID) with reference gene models

(Data are available upon request)

B.4 Total number of sequence reads and filtering steps used for SNV[a] calling

Fungal strain	IDs[b]	Raw paired-end reads ([c])	Total number of Paired-end and single-end reads after removing		Total bp	Estimated coverage
			Low–quality reads	Duplicate reads (%)		
UAMH 11150	*GsB1 (control)*	41,327,658 (50)	31,374,695	19,483,575 – 7,523,362 (26)	2,092,073,040	70×
UAMH 11153	GsB2	18,812,843 (50)	11,248,893	7,089,534 – 2,317,892 (27)	742,363,200	25×
UAMH 11348	GsB3	25,340,644 (76)	14,919,663	8,110,381 – 3,200,348 (35)	1,378,898,810	47×
UAMH 11353	GsA1	35,991,059 (76)	18,087,186	10,091,552 – 3,643,716 (34)	1,691,704,220	58×
UAMH 11354	GsA2	37,723,693 (76)	13,259,161	5,973,601 – 2,517,982 (45)	1,027,028,064	35×
UAMH 11347	GsA3	16,682,452 (50)	10,109,303	6,492,213 – 2,053,087 (26)	676,688,085	23×
UAMH 11156	GsM1	14,617,945 (50)	9,092,931	5,931,623 – 1,668,522 (26)	608,929,560	21×
UAMH 11349	GsC1	33,385,453 (76)	18,585,344	8,446,641 – 3,848,853 (44)	1,472,691,585	51×
UAMH 11350	GsC2	33,834,154 (76)	19,419,436	8,090,204 – 3,814,108 (49)	1,419,610,636	49×
ATCC 18086	GcB1.a	19,701,172 (50)	5,851,822	4,307,999 – 833,338 (19)	425,220,120	15×
–	GcB1.b	32,988,736 (76)	19,737,773	10,016,836 – 4,232,013 (38)	1,722,863,635	59×
–	GcB1.ab	52,689,908	25,589,595	14,324,835 – 5,065,351 (34)	2,148,083,755	74×
UAMH 11351	GcC1	17,815,964 (50)	10,334,042	6,606,133 – 2,075,845 (26)	687,964,995	24×
UAMH 11352	GcC2.a	33,250,787 (76)	14,324,332	6,807,253 – 2,752,816 (43)	1,162,079,862	40×
–	GcC2.b	34,219,918 (76)	17,494,252	7,647,398 – 3,452,387 (46)	1,331,049,993	46×
–	GcC2.ab	67470705 (76)	31,818,584	14,454,651 – 6,205,203 (48)	2,326,984,246	80×

[a] We use the term SNV instead of the common term SNP (Single Nucleotide Polymorphism) because SNPs are normally defined relative to a population and imply a minimum minor allele frequency whereas we are interested in finding all sequence variants that do not match the reference genome sequence, regardless of their frequency in the population.

[b] IDs, "a" and "b" are results from two independent sequence lanes for the same strain, "ab" results from two sequence runs combined for the same strain. The estimated coverage are based on filtered reads mapped to the slkw1407 reference genome sequence, which is ~ 29.1 Mbp after excluding gaps.

[c] nucleotide bases

B.5 Primer sequences used in the SNV validation and in the phylogenetic and population genetic analyses

Gene	Primer sequence 5' —> 3'		sequence length in bp*			GenBank accession no.	Gene description (abbreviation)	Phylogenetic information [a]			
	Forward	Reverse	Total	Exon	Intron			Total characters	Variable sites	Parsimony informative	MP trees
CMQ_5562	CGAACGCCTCTGGCTCTCCATTG	GCCTGCCGGAAATGTGACGTTG	1,258	3	2	xxxxxx	ABC transporter (subfamily C)	2,843	32	18	1
	ATCCGTGGTCCAGCAGTC	AGCACGACCTTCTCCAGC	1,722	1	0	xxxxxx					
CMQ_6993.6634	AGACCTCTCATGGCCTACTGCG	AACATGCGGCAGCAGTCCCAAC	1,064	4	4	xxxxxx	—	2,140	52	23	1
	ACACGGGCGAGCTTCTCAACAG	TTCCACCACCTCCCAGAGTCCCA	1,182	3	2	xxxxxx					
CMQ_6965	CGCGAGCTTTGGTGTTTCTGCC	ACCAGCAGGCTGACTAGCGACA	1,074	1	0	xxxxxx	—	1,826	36	18	1
	TTCATTGGAGAGCGAGGCGCTG	TGCGCAGCCAGTCGACCAATAC	943	2	1	xxxxxx					
CMQ_861	GGCCAGCAACGGCATCTTTGAC	TGCCACAGACCAAAGCCTGGAC	563	2	1	xxxxxx	ABC transporter (subfamily G)	494	19	11	3
CMQ_4184	AAGCACGCCCGCTATGCACTC	TGCTGACAGTTGTTGGCTGCCG	1,087	1	0	xxxxxx	—	1,055	18	8	1
CMQ_3826	TAAGGAAAGGGAGGGCGGT	GGAGTAGGGGAGGGGAAT	913	1	1	xxxxxx	Metallo-peptidase	1,585	18	5	1
	TCGCTCATCACGCACCCA	CTTCCATGTCCTCCTTCC	780	1	0	xxxxxx					
CMQ_5095	TGTGTCCAAGGCATTCCCCGAC	CACAAACAGCGGCGTCGAGTTG	1,414	1	0	xxxxxx	Polyketide synthase	1,786	44	27	1
	AAGGTGCTCCTGATGATGCGGC	AATGGCAGCCGATGTGGCAGAG	1,008	2	1	xxxxxx					
CMQ_5323	ACAGTCCGGAGAGCGTGACCATC	AGGTTGGCGAACAAGTCCTGGG	1,213	1	0	xxxxxx	Polyketide synthase	1,539	47	25	1
	AGTCCTTCTCTCGCGGCATCTCC	AACCAGCATGTTCCGCACCTCG	1,331	2	1	xxxxxx					
CMQ_6740	TTTGGCATCTCCAAGCCCTGCG	GCGTCATCCAGACGGTCATCAGC	1,191	2	1	xxxxxx	Flavoprotein monooxygenase	1,040	30	8	1
Nine-gene combined dataset						To be submitted		14,308	296	143	40

[a] Phylogenetic information for each and combined nine gene aligned dataset sequenced from additional Gs and Gc strains including the outgroup taxa

B.6 *Grosmannia* gene-model summaries

Fungal IDs	Homologous genes	100%	99%	98–96%	95–90%	89-70%	PID<70%	Genes with no hit	duplicates/paralogs
GsB2	8,129	5,715	1,486	537	238	153	103	80	49
GsB3	8,222	5,855	1,521	495	219	132	63	27	46
GsA1	8,155	5,808	1,440	534	240	133	78	79	46
GsA2	8,107	5,345	1,428	738	382	214	129	76	42
GsA3	7,988	5,522	1,485	583	242	156	139	185	61
GsM1	7,876	5,399	1,493	587	248	149	191	245	58
GsC1	8,018	5,069	1,974	585	239	151	154	140	91
GsC2	7,956	4,885	2,073	592	257	149	146	210	50
GcB1	8,198	3,108	3,747	851	297	193	89	27	41
GcC1	7,973	2,907	3,560	966	326	210	193	150	64
GcC2	8,079	2,959	3,624	931	359	202	140	97	70
Average	**8,064**						**130**	**120**	

B.7 McDonald-Kreitman (MK) test results and the mean Gs-Gc pairwise rate of protein-coding divergence (dN/dS)

(Data are available upon request)

B.8 PAML “site-model” test of positive selection for 1,213 Gs-Gc orthologous genes

(Data are available upon request)

B.9 Variants identified using the published *Grosmannia* genome as the reference sequence

Fungal IDs	Single nucleotide variations (SNV)								Indels
	All	Ts/Tv	Intergenic	Intronic	Flanking	Synonymous coding	Nonsynonymous coding	Nonsynoymous stop gain – stop lost	
GsB1 (control)	1,796		297	202	526	223	540	03 – 05	
GsB2	10,242	2.7	3,866	609	2,458	1,576	1,706	23 – 09	3,117
GsB3	12,196	3.1	5,573	630	2,521	1,607	1,837	23 – 10	3,240
GsA1	10,405	3.1	4,405	567	2,333	1,479	1,602	16 – 09	3,170
GsA2	10,332	3.0	4,482	589	2,333	1,338	1,568	15 – 10	3,154
GsA3	9,941	2.8	3,919	616	2,384	1,442	1,565	14 – 09	3,095
GsM1	10,139	2.8	3,804	677	2,474	1,566	1,596	17 – 10	3,082
GsC1	19,098	3.4	8,867	980	4,100	2,453	2,668	22 – 13	3,781
GsC2	20,521	3.4	9,362	1,110	4,503	2,651	2,858	35 – 09	3,882
Average	12,859	3.0	5,535	722	2,888	1,764	1,925		3,315
GcB1.a	58,663	3.9	19,918	3,921	15,819	9,754	9,150	103 – 33	6,751
GcB1.b	59,926	3.8	21,169	3,918	15,806	9,763	9,170	103 – 33	6,853
GcB1.ab	60,435	3.9	21,617	3,925	15,832	9,782	9,179	103 – 33	6,862
GcC1	60,808	3.8	20,968	3,989	16,318	9,895	9,522	125 – 29	6,814
GcC2.a	63,204	3.8	23,883	3,915	16,230	9,694	9,375	117 – 29	6,934
GcC2.b	63,219	3.7	23,886	3,913	16,230	9,700	9,382	118 – 29	6,951
GcC2.ab	63,294	3.7	23,945	3,916	16,234	9,705	9,386	118 – 29	6,957
Average	61,512	3.8	22,177	3,943	16,128	9,794	9,362		6,878
Total	**103,430**		**42,880**	**5,826**	**24,589**	**14,889**	**15,040**	**226 – 36**	**9,906**

The reference genome published by DiGuistini et al. 2011

B.10 Gene models containing intra- and interspecific stop-codon variants

Gene	Gene description	Scaffold			Stop codon			RNA evidence [a]		SNP class
		ID	from	to	gain	lost	position	RNAseq and/or EST	annotation	
CMQ_7030	salicylate hydroxylase	SC_105	16,604	20,349	X	-	19,630			exclusive to Gs
CMQ_6551	siderophore biosynthesis protein	SC_108	1,570,163	1,573,981	X	-	1,573,138	Yes	confirmed	fixed between Gs and Gc
CMQ_6582	autophagy protein	SC_108	1,362,950	1,364,821	X	-	1,363,994	Yes	confirmed	fixed between Gs and Gc
CMQ_6740	**flavoprotein monooxygenase**	**SC_108**	**1,030,358**	**1,032,964**	**X**	**-**	**1,031,164**	**Yes**	**confirmed**	**exclusive to Gc from JP**
CMQ_6988	tol-like protein	SC_108	1,250,926	1,254,937	X	-	1,254,412	No	nc	exclusive to Gs
CMQ_3349	arginyl-tRNA synthetase	SC_113	2,283,134	2,284,072	-	X	2,284,072	Yes	confirmed	fixed between Gs and Gc
CMQ_3907	cytochrome p450 monooxygenase	SC_113	2,304,173	2,308,001	X	-	2,037,150	Yes	confirmed	fixed between Gs and Gc
CMQ_3907	cytochrome p450 monooxygenase	SC_113	2,304,173	2,308,001	X	-	2,037,706	Yes	confirmed	exclusive to Gc from JP
CMQ_3795	siderochrome-iron transporter	SC_113	1,184,097	1,186,007	X	-	1,184,320	Yes	confirmed	fixed between Gs and Gc
CMQ_3862	tyrosinase	SC_113	1,090,713	1,093,110	X	-	1,091,358	Yes	confirmed	fixed between Gs and Gc
CMQ_3890	pre-mRNA splicing factor	SC_113	388,304	389,549	-	X	388,307	Yes	confirmed	fixed between Gs and Gc
CMQ_3894	major facilitator superfamily transporter	SC_113	721,228	722,896	X	-	722,513	Yes	confirmed	exclusive to Gs
CMQ_3894	major facilitator superfamily transporter	SC_113	721,228	722,896	X	-	722,575	Yes	confirmed	fixed between Gs and Gc
CMQ_4004	esterase/lipase	SC_113	1,182,730	1,183,390	X	-	1,182,746	Yes	confirmed	fixed between Gs and Gc
CMQ_4014	stomatin family protein	SC_113	946,286	949,981	X	-	948,073	Yes	confirmed	fixed between Gs and Gc
CMQ_4014	stomatin family protein	SC_113	946,286	949,981	X	-	948,145	Yes	confirmed	fixed between Gs and Gc
CMQ_4014	stomatin family protein	SC_113	946,286	949,981	X	-	948,163	Yes	confirmed	fixed between Gs and Gc
CMQ_4097	extracellular scp domain containing protein	SC_113	851,732	852,407	X	-	852,126	Yes	confirmed	fixed between Gs and Gc
CMQ_4102	duf1680 domain containing protein	SC_113	1,009,005	1,011,413	X	-	1,010,651	Yes	confirmed	exclusive to Gc from JP
CMQ_4148	conidiophore development protein hyma	SC_113	1,266,837	1,270,034	X	-	1,266,939	Yes	confirmed	fixed between Gs and Gc
CMQ_1788	amino acid permease	SC_124	96,510	98,566	X	-	96,879	Yes	nc	exclusive to Gs
CMQ_8025	class 5 chitinase 1	SC_125	37,549	43,430	X	-	39,034	Yes	nc	exclusive to Gc from JP
CMQ_8025	class 5 chitinase 1	SC_125	37,549	43,430	X	-	39,035	Yes	nc	exclusive to Gc from JP
CMQ_8025	class 5 chitinase 1	SC_125	37,549	43,430	X	-	39,920	Yes	nc	exclusive to Gc from JP
CMQ_8025	class 5 chitinase 1	SC_125	37,549	43,430	X	-	41,011	Yes	nc	exclusive to Gc from JP
CMQ_8027	FAD-binding domain containing protein	SC_125	7,180	8,998	X	-	8,569	Yes	confirmed	exclusive to Gc from JP
CMQ_5120	ankyrin repeat-containing protein	SC_132	856,946	865,104	X	-	864,916	Yes	confirmed	fixed between Gs and Gc
CMQ_5120	ankyrin repeat-containing protein	SC_132	856,946	865,104	X	-	865,045	Yes	confirmed	fixed between Gs and Gc
CMQ_5339	sodium/phosphate symporter	SC_132	246,066	250,738	X	-	249,907	Yes	confirmed	fixed between Gs and Gc
CMQ_5339	sodium/phosphate symporter	SC_133	246,066	250,738	X	-	249,904	Yes	confirmed	fixed between Gs and Gc
CMQ_5356	steroid monooxygenase	SC_132	332,989	334,789	X	-	334,270	Yes	confirmed	exclusive to Gc from JP
CMQ_5374	dynamin family protein	SC_132	485,507	487,916	X	-	487,815	Yes	confirmed	fixed between Gs and Gc
CMQ_1184	intracellular serine protease	SC_140	4,027,986	4,031,141	X	-	4,028,662	Yes	confirmed	fixed between Gs and Gc
CMQ_1274	ankyrin repeat protein	SC_140	3,535,897	3,538,054	-	X	3,535,900	Yes	confirmed	fixed between Gs and Gc
CMQ_1290	nacht and tpr domain containing protein	SC_140	261,976	266,833	X	-	262,855	Yes	nc	exclusive to Gc
CMQ_1290	nacht and tpr domain containing protein	SC_140	261,976	266,833	X	-	263,124	Yes	nc	exclusive to Gc
CMQ_1340	cript family protein	SC_140	2,471,746	2,472,341	X	-	2,472,335	Yes	nc	exclusive to Gs
CMQ_1422	cellulase family protein	SC_140	1,492,574	1,493,183	X	-	1,492,647	Yes	confirmed	fixed between Gs and Gc
CMQ_1422	cellulase family protein	SC_140	1,492,574	1,493,183	X	-	1,492,889	Yes	confirmed	fixed between Gs and Gc

Gene	Gene description	Scaffold			Stop codon			RNA evidence [a]		SNP class
		ID	from	to	gain	lost	position	RNAseq and/or EST	annotation	
CMQ_1455	isp4 protein	SC_140	3,859,995	3,862,532	X	-	3,860,274	Yes	nc	exclusive to Gc from JP
CMQ_1644	4-coumarate-CoA ligase	SC_140	2,219,024	2,220,788	X	-	2,219,571	Yes	nc	exclusive to Gc from JP
CMQ_7268	nb-arc and tpr domain containing protein	SC_173	757,407	760,211	X	-	256,915	Yes	nc	fixed between Gs and Gc
CMQ_369	major facilitator superfamily transporter	SC_140	632,447	636,531	X	-	633,792	Yes	confirmed	fixed between Gs and Gc
CMQ_407	cytochrome p450 monooxygenase	SC_140	317,966	319,318	X	-	319,213	Yes	confirmed	fixed between Gs and Gc
CMQ_547	fungal specific transcription factor	SC_140	4,238,452	4,240,119	X	-	4,239,599	Yes	confirmed	exclusive to Gc from JP
CMQ_547	fungal specific transcription factor	SC_140	4,238,452	4,240,119	X	-	4,239,602	Yes	confirmed	exclusive to Gc from JP
CMQ_761	periodic tryptophan protein 2	SC_140	3,047,416	3,050,257	-	X	3,050,255	Yes	nc	fixed between Gs and Gc
CMQ_781	acetate kinase	SC_140	44,718	47,502	X	-	47,091	Yes	nc	exclusive to Gc
CMQ_984	duf1479 domain containing protein	SC_140	565,877	566,990	X	-	566,939	Yes	confirmed	fixed between Gs and Gc
CMQ_836, CMQ_1107	vacuolar ATPase, hypothetical protein	SC_140 SC_140	3494217, 3494740	3495330, 3494989	X	-	3,494,963	Yes	confirmed	fixed between Gs and Gc
CMQ_5397	zinc finger, mynd-type domain containing protein	SC_142	89,680	96,004	X	-	95,970	No	nc	exclusive to Gc from JP
CMQ_1833	polyketide synthase	SC_180	46,088	51,288	X	-	1,290,261	No	nc	exclusive to Gs
CMQ_2716	dead deah box DNA helicase	SC_144	360,004	368,123	X	-	360,280	Yes	confirmed	fixed between Gs and Gc
CMQ_2761	glycosyl transferase	SC_144	271,187	274,967	X	-	271,221	Yes	confirmed	fixed between Gs and Gc
CMQ_2880	major facilitator superfamily transporter toxin efflux pump	SC_144	763,037	767,468	X	-	763,265	Yes	confirmed	fixed between Gs and Gc
CMQ_2915	FAD-binding domain containing protein	SC_144	1,301,553	1,302,916	X	-	1,301,799	Yes	confirmed	exclusive to Gs
CMQ_3057	metallo-beta-lactamase superfamily protein	SC_144	1,296,201	1,298,402	X	-	1,297,842	Yes	confirmed	exclusive to Gs
CMQ_5826	kinase	SC_146	235,947	236,928	X	-	236,799	No	nc	exclusive to Gc from JP
CMQ_8029	integral membrane protein	SC_150	4,571	5,455	X	-	5,257	Yes	nc	exclusive to Gc from JP
CMQ_2296	succinate dehydrogenase cytochrome b560 subunit	SC_156	524,590	527,376	X	-	525,351	Yes	nc	exclusive to Gs
CMQ_2316	bacilysin biosynthesis oxidoreductase	SC_156	1,202,328	1,203,381	-	X	1,202,329	Yes	confirmed	fixed between Gs and Gc
CMQ_2320	ubiquitin thiolesterase	SC_156	1,373,084	1,376,324	X	-	1,374,505	Yes	confirmed	fixed between Gs and Gc
CMQ_2390	alpha beta hydrolase fold protein	SC_156	865,934	868,702	X	-	866,838	Yes	confirmed	exclusive to Gs
CMQ_8064	major facilitator superfamily transporter	SC_161	426,923	428,356	-	X	428,354	Yes	confirmed	fixed between Gs and Gc
CMQ_8079	hlh transcription factor	SC_161	650,035	654,037	X	-	650,646	Yes	confirmed	shared
CMQ_8216	ankyrin unc44	SC_161	662,004	663,579	-	X	663,579	Yes	nc	shared
CMQ_8227	FAD dependent oxidoreductase superfamily	SC_161	332,380	336,708	X	-	335,539	Yes	confirmed	fixed between Gs and Gc
CMQ_8251	short chain dehydrogenase reductase	SC_161	157,508	158,403	X	-	158,250	No	nc	fixed between Gs and Gc
CMQ_5958	glycerol-3-phosphate acyltransferase	SC_167	344,180	349,668	X	-	346,655	Yes	confirmed	exclusive to Gs
CMQ_6209	kinesin light chain	SC_167	432,239	434,567	-	X	434,565	Yes	confirmed	exclusive to Gs
CMQ_6468	salicylate hydroxylase	SC_167	212,906	214,272	X	-	213,363	Yes	confirmed	exclusive to Gs
CMQ_1913	major facilitator superfamily transporter multidrug resistance	SC_168	293,517	295,417	-	X	293,520	Yes	confirmed	fixed between Gs and Gc
CMQ_1913	major facilitator superfamily transporter multidrug resistance	SC_168	293,517	295,417	X	-	295,077	Yes	confirmed	fixed between Gs and Gc
CMQ_7076	2og-Fe oxygenase family protein	SC_173	179,330	180,443	X	-	180,436	Yes	confirmed	fixed between Gs and Gc
CMQ_7358	beta-glucosidase	SC_173	1,697,690	1,706,226	X	-	1,703,615	Yes	confirmed	fixed between Gs and Gc
CMQ_7517	short chain dehydrogenase reductase	SC_173	1,949,855	1,950,845	X	-	1,950,824	No	nc	fixed between Gs and Gc
CMQ_7566	short chain dehydrogenase reductase	SC_173	2,141,595	2,145,557	X	-	2,142,360	Yes	confirmed	fixed between Gs and Gc
CMQ_7662	d-isomer specific 2-hydroxyacid dehydrogenase	SC_173	144,461	148,338	X	-	145,983	Yes	confirmed	fixed between Gs and Gc
CMQ_235	ankyrin repeat-containing protein	SC_179	557,417	562,149	X	-	560,393	Yes	confirmed	fixed between Gs and Gc
CMQ_319	duf221 domain protein	SC_179	940,757	945,050	X	-	943,747	Yes	nc	shared
CMQ_5561	methyltransferase type 11	SC_89	557,677	558,334	X	-	558,206	Yes	confirmed	fixed between Gs and Gc
CMQ_5634	ribonuclease p complex subunit	SC_89	810,058	817,060	X	-	811,177	Yes	confirmed	exclusive to Gs

Gene	Gene description	Scaffold			Stop codon			RNA evidence [a]		SNP class
		ID	from	to	gain	lost	position	RNAseq and/or EST	annotation	
CMQ_5658	ABC transporter	SC_89	370,126	372,064	-	X	372,064	Yes	confirmed	exclusive to Gs
CMQ_5691	c6 zinc finger domain containing protein	SC_89	200,560	206,396	X	-	201,113	No	nc	fixed between Gs and Gc
CMQ_4272	aristolochene synthase	SC_97	2,165,605	2,166,601	-	X	2,165,607	Yes	confirmed	exclusive to Gs
CMQ_4308	tpr domain containing protein	SC_97	2,188,283	2,189,729	X	-	2,188,551	Yes	confirmed	fixed between Gs and Gc
CMQ_4308	tpr domain containing protein	SC_97	2,188,283	2,189,729	X	-	2,189,422	Yes	confirmed	fixed between Gs and Gc
CMQ_4310	glycoside hydrolase family 3 domain containing protein	SC_97	677,205	680,705	X	-	677,855	Yes	confirmed	fixed between Gs and Gc
CMQ_4346	duf636 domain containing protein	SC_97	2,014,860	2,016,111	X	-	2,015,359	Yes	confirmed	exclusive to Gs
CMQ_4381	major facilitator superfamily transporter multidrug resistance	SC_97	1,383,778	1,385,732	X	-	1,384,324	Yes	nc	exclusive to Gc from JP
CMQ_4594	kinesin light chain	SC_97	1,054,003	1,057,778	X	-	1,055,683	Yes	confirmed	fixed between Gs and Gc
CMQ_4594	kinesin light chain	SC_97	1,054,003	1,057,778	X	-	1,056,579	Yes	confirmed	fixed between Gs and Gc
CMQ_4620	cytochrome p450 monooxygenase	SC_97	2,147,761	2,149,239	-	X	2,147,764	Yes	confirmed	exclusive to Gs
CMQ_4642	cytochrome c oxidase assembly protein	SC_97	522,663	528,069	X	-	527,727	Yes	confirmed	fixed between Gs and Gc
CMQ_4686	nonribosomal peptide synthetase 11	SC_97	2,154,440	2,154,770	-	X	2,154,442	Yes	confirmed	exclusive to Gs
CMQ_4704	DNA mismatch repair protein	SC_97	1,765,673	1,767,902	X	-	1,767,188	Yes	confirmed	fixed between Gs and Gc
CMQ_4750	methyltransferase type 11 domain containing protein	SC_97	2,156,931	2,157,678	-	X	2,157,677	Yes	confirmed	fixed between Gs and Gc
CMQ_4813	alcohol dehydrogenase	SC_97	3,984	4,977	X	-	4,371	Yes	confirmed	fixed between Gs and Gc
CMQ_4844	ABC multidrug transporter mdr1	SC_97	2,158,740	2,162,491	X	-	2,160,614	Yes	confirmed	exclusive to Gs
CMQ_4844	ABC multidrug transporter mdr1	SC_97	2,158,740	2,162,491	X	-	2,160,719	Yes	confirmed	exclusive to Gs
CMQ_7031	hypothetical protein	SC_105	22,586	22,826	X	-	22,783	Yes	confirmed	fixed between Gs and Gc
CMQ_6504	hypothetical protein	SC_108	1,425,750	1,427,958	X	-	1,425,851	Yes	confirmed	fixed between Gs and Gc
CMQ_6899	hypothetical protein	SC_108	310,115	312,256	X	-	311,047	Yes	confirmed	exclusive to Gs
CMQ_6899	hypothetical protein	SC_108	310,115	312,256	X	-	311,676	Yes	confirmed	exclusive to Gc
CMQ_6945	hypothetical protein	SC_108	1,263,658	1,265,139	X	-	1,263,896	Yes	confirmed	exclusive to Gs
CMQ_7010	hypothetical protein	SC_108	647,790	649,796	X	-	649,512	Yes	confirmed	exclusive to Gs
CMQ_3377	hypothetical protein	SC_113	2,351,277	2,352,385	-	X	2,351,280	Yes	confirmed	fixed between Gs and Gc
CMQ_3672	hypothetical protein	SC_113	2,178,566	2,181,847	X	-	2,179,332	Yes	confirmed	exclusive to Gc
CMQ_3744	hypothetical protein	SC_113	894,884	896,097	X	-	894,970	Yes	confirmed	fixed between Gs and Gc
CMQ_4129	hypothetical protein	SC_113	313,835	314,557	X	-	313,942	Yes	confirmed	exclusive to Gs
CMQ_4151	hypothetical protein	SC_113	2,455,279	2,456,106	X	-	2,456,102	Yes	confirmed	fixed between Gs and Gc
CMQ_4166	hypothetical protein	SC_113	2,045,229	2,046,555	X	-	2,045,411	Yes	confirmed	fixed between Gs and Gc
CMQ_5187	hypothetical protein	SC_132	468,567	468,927	-	X	468,570	Yes	confirmed	fixed between Gs and Gc
CMQ_5224	hypothetical protein	SC_132	313,184	313,927	X	-	313,734	Yes	confirmed	fixed between Gs and Gc
CMQ_5224	hypothetical protein	SC_132	313,184	313,927	X	-	313,735	Yes	confirmed	fixed between Gs and Gc
CMQ_5230	hypothetical protein	SC_132	179,393	180,065	X	-	179,732	Yes	confirmed	fixed between Gs and Gc
CMQ_5252	hypothetical protein	SC_132	427,645	429,724	X	-	429,116	Yes	confirmed	fixed between Gs and Gc
CMQ_5384	hypothetical protein	SC_132	450,192	451,290	X	-	450,859	Yes	confirmed	fixed between Gs and Gc
CMQ_1572	hypothetical protein	SC_140	352,991	357,307	X	-	354,072	Yes	confirmed	exclusive to Gc from JP
CMQ_1658	hypothetical protein	SC_140	390,422	390,623	-	X	390,425	Yes	confirmed	fixed between Gs and Gc
CMQ_1731	hypothetical protein	SC_140	4,121,328	4,121,775	X	-	4,121,733	Yes	confirmed	fixed between Gs and Gc
CMQ_711	hypothetical protein	SC_140	4,139,826	4,149,737	X	-	4,149,687	Yes	confirmed	fixed between Gs and Gc
CMQ_866	hypothetical protein	SC_140	2,426,461	2,427,679	X	-	2,427,191	Yes	confirmed	fixed between Gs and Gc
CMQ_2694	hypothetical protein	SC_144	2,093,750	2,095,583	X	-	2,094,251	Yes	confirmed	exclusive to Gs
CMQ_2694	hypothetical protein	SC_144	2,093,750	2,095,583	X	-	2,095,268	Yes	confirmed	exclusive to Gs

Gene	Gene description	Scaffold			Stop codon			RNA evidence [a]		SNP class
		ID	from	to	gain	lost	position	RNAseq and/or EST	annotation	
CMQ_2733	hypothetical protein	SC_144	1,342,629	1,343,439	X	-	1,343,213	Yes	confirmed	fixed between Gs and Gc
CMQ_2784	hypothetical protein	SC_144	2,286,211	2,290,184	-	X	2,286,212	Yes	confirmed	fixed between Gs and Gc
CMQ_2875	hypothetical protein	SC_144	329,491	330,599	-	X	329,494	Yes	confirmed	fixed between Gs and Gc
CMQ_2883	hypothetical protein	SC_144	2,061,319	2,062,634	X	-	2,061,788	Yes	confirmed	fixed between Gs and Gc
CMQ_2982	hypothetical protein	SC_144	2,056,629	2,058,582	X	-	2,057,866	Yes	confirmed	fixed between Gs and Gc
CMQ_2982	hypothetical protein	SC_144	2,056,629	2,058,582	X	-	2,057,867	Yes	confirmed	fixed between Gs and Gc
CMQ_2982	hypothetical protein	SC_144	2,056,629	2,058,582	X	-	2,058,446	Yes	confirmed	fixed between Gs and Gc
CMQ_3076	hypothetical protein	SC_144	780,569	781,924	X	-	781,919	Yes	confirmed	fixed between Gs and Gc
CMQ_3086	hypothetical protein	SC_144	382,354	382,837	-	X	382,356	Yes	confirmed	fixed between Gs and Gc
CMQ_5913	hypothetical protein	SC_146	331,122	332,570	X	-	331,261	Yes	confirmed	exclusive to Gc from JP
CMQ_2379	hypothetical protein	SC_156	723,985	724,628	X	-	724,035	Yes	confirmed	fixed between Gs and Gc
CMQ_2493	hypothetical protein	SC_156	1,033,692	1,034,459	X	-	1,034,016	Yes	confirmed	exclusive to Gc from JP
CMQ_4924	hypothetical protein	SC_160	134,153	134,900	X	-	134,885	Yes	confirmed	fixed between Gs and Gc
CMQ_8178	hypothetical protein	SC_161	453,820	457,067	-	X	457,065	Yes	confirmed	fixed between Gs and Gc
CMQ_8230	hypothetical protein	SC_161	661,156	661,746	X	-	661,650	Yes	confirmed	shared
CMQ_6397	hypothetical protein	SC_167	1,836,776	1,838,110	X	-	1,838,012	Yes	confirmed	fixed between Gs and Gc
CMQ_6467	hypothetical protein	SC_167	1,776,729	1,777,100	-	X	1,776,732	Yes	confirmed	fixed between Gs and Gc
CMQ_1922	hypothetical protein	SC_168	446,150	447,728	X	-	446,547	Yes	confirmed	exclusive to Gs
CMQ_1945	hypothetical protein	SC_168	51,526	52,066	X	-	51,893	Yes	confirmed	fixed between Gs and Gc
CMQ_7312	hypothetical protein	SC_173	165,197	166,754	X	-	165,928	Yes	confirmed	exclusive to Gc from JP
CMQ_7312	hypothetical protein	SC_173	165,197	166,754	X	-	165,934	Yes	confirmed	exclusive to Gc from JP
CMQ_7448	hypothetical protein	SC_173	1,451,403	1,452,607	X	-	1,451,846	Yes	confirmed	exclusive to Gs
CMQ_7466	hypothetical protein	SC_173	1,851,470	1,853,295	X	-	1,852,771	Yes	confirmed	exclusive to Gs
CMQ_7570	hypothetical protein	SC_173	1,691,446	1,694,174	X	-	1,693,931	Yes	confirmed	fixed between Gs and Gc
CMQ_7578	hypothetical protein	SC_173	1,644,717	1,645,174	X	-	1,645,140	Yes	confirmed	fixed between Gs and Gc
CMQ_7685	hypothetical protein	SC_173	1,556,619	1,556,994	-	X	1,556,992	Yes	confirmed	fixed between Gs and Gc
CMQ_215	hypothetical protein	SC_179	179,159	180,043	X	-	179,801	Yes	confirmed	exclusive to Gc from JP
CMQ_215	hypothetical protein	SC_179	179,159	180,043	X	-	180,025	Yes	confirmed	exclusive to Gs
CMQ_1832	hypothetical protein	SC_180	62,260	65,724	-	X	65,723	Yes	confirmed	exclusive to Gs
CMQ_4677	hypothetical protein	SC_97	1,408,128	1,408,410	-	X	1,408,131	Yes	confirmed	fixed between Gs and Gc

[a] To avoid stop-codon variants due to erroneous annotation, exon-intron boundaries were check using expressed sequence tag libraries (EST) and transcriptome data from slkw1407 reference strain; nc, indicate genes for which EST and/or RNA-seq data were missing or didnot provide enough coverage to confirm the gene annotations.

The stop-codon variant confirmed by direct sequencing is bolded

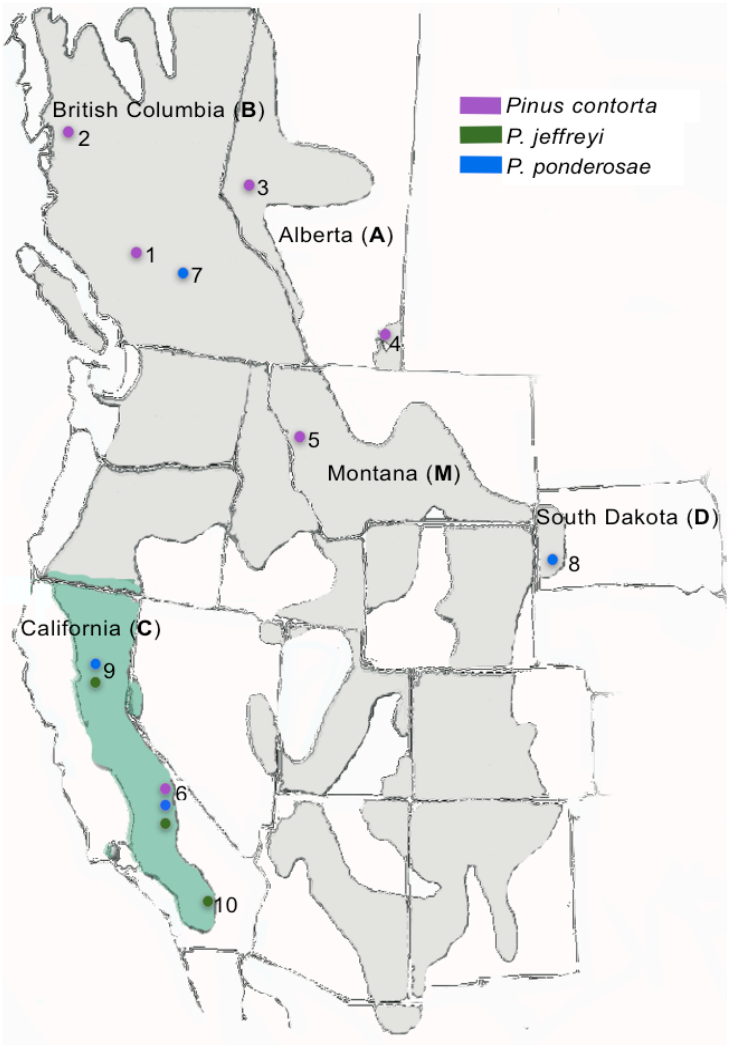
B. 11 GO functional enrichment analysis of potential pseudogenes in Gs and Gc strains

GO-ID	Term	Number of annotated test set	Number of annotated Reference set	Test set%	Referece set%	Category[a]	FDR	P-Value	#Test	#Ref
GO:0055114	oxidation-reduction process	71	5400	31.0	13.9	P	0.27	1.91E-04	22	750
GO:0016491	oxidoreductase activity	71	5400	31.0	14.4	F	0.27	3.11E-04	22	776
GO:0055085	transmembrane transport	71	5400	16.9	8.2	P	1	0.013374468	12	444
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	71	5400	7.0	2.0	F	1	0.014654786	5	107
GO:0004497	monooxygenase activity	71	5400	7.0	2.0	F	1	0.015181917	5	108
GO:0000272	polysaccharide catabolic process	71	5400	4.2	0.7	P	1	0.017840778	3	40
GO:0016798	hydrolase activity, acting on glycosyl bonds	71	5400	7.0	2.1	F	1	0.019863043	5	116
GO:0042537	benzene-containing compound metabolic process	71	5400	8.5	0.3	P	1	0.024626681	2	17
GO:0000166	nucleotide binding	71	5400	18.3	20.3	F	1	0.041854134	21	1096

[a] F, Molecular Functions; P, Biological Process

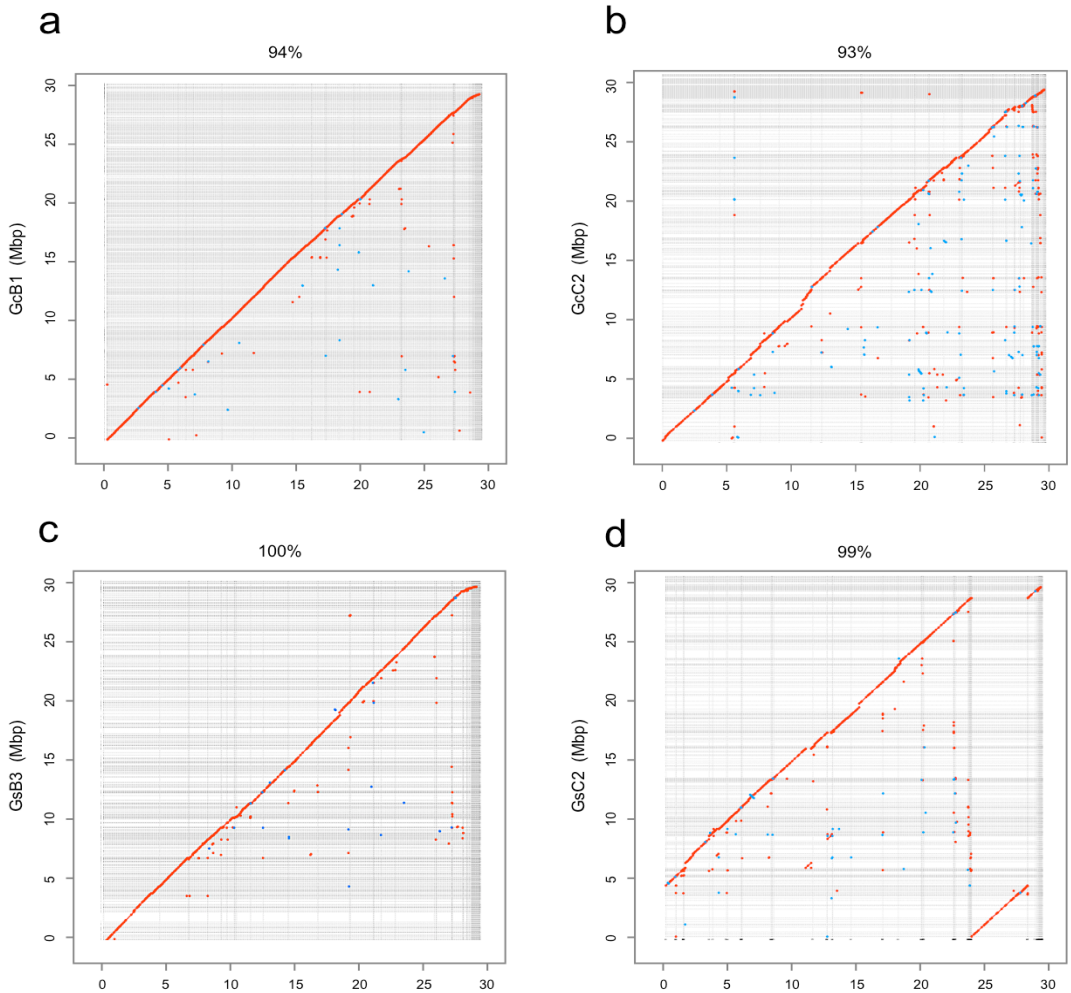
B.12 Map of western North America representing the mountain pine beetle (MPB) distribution.

In the green region the MPB occurs with its sister species, the jeffrey pine beetle. Numbers indicate the fungal collection sites (Table 4.1). Colored dots represent the locations and host tree species from which the fungi were isolated.



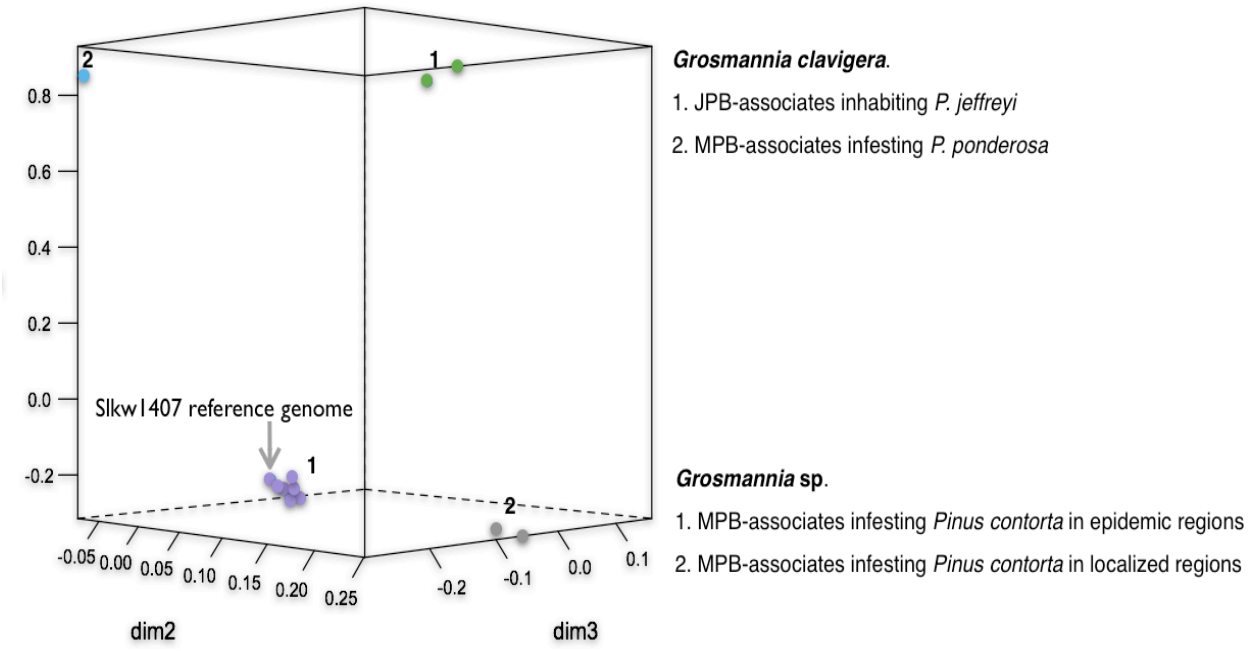
B.13 Large-scale synteny between *Grosmannia* genomes

The MUMmer dotplots indicate matching sequences and show co-directional regions of synteny in red. The genome of the reference strain with a total length of 29.8 Mb is represented on the X-axis and those from other strains including Gc from *P. ponderosa* (**a**) and *P. jeffreyi* (**b**), and Gs from northern BC (**c**) and California (**d**) are shown on the Y-axis. The values above the graph show the percentage of the Gc and Gs assembled contig length that was syntenic with the reference genome (See appendix B.2).



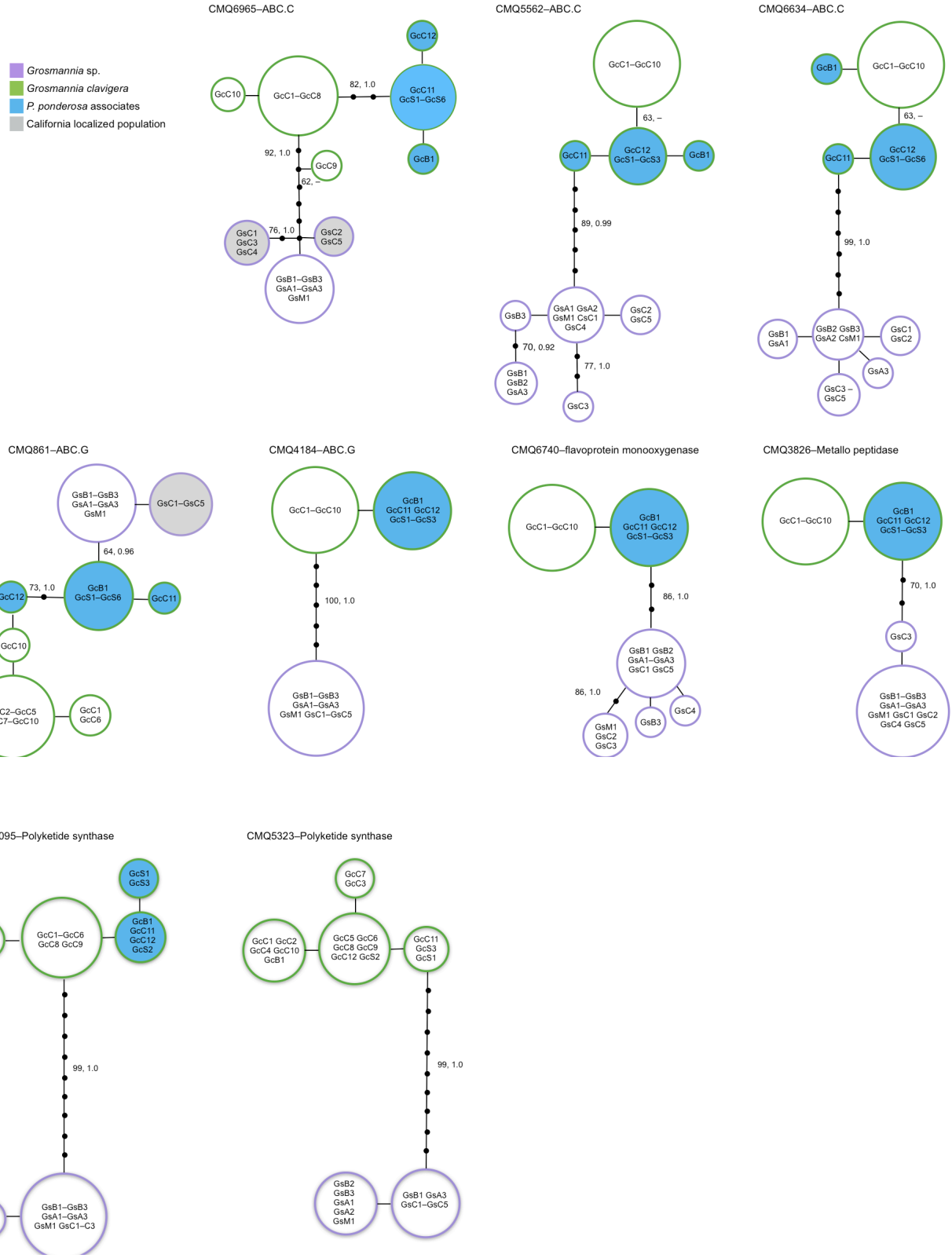
B.14 Genetic distance between the twelve *Grosmannia* genomes.

The multidimensional scaling-plot represents the genetic distance between genomes calculated from 103,430 single nucleotide variants.



B.15 Haplotype networks and genealogies of nine gene regions sequenced in 28 *Grosmannia* strains.

The genealogy of each of the nine-gene regions was best described by a single, unrooted, most-parsimonious tree, except for the positioning of Gc–10 in the CMQ861–ABC.G transporter tree, which shows paraphyly between Gc strains from *P. ponderosa* and those from *P. jeffreyi*. Relationships among the clades were topologically identical in the network method as well as the tree-based methods using maximum parsimony (MP) and Bayesian analyses. Here, the parsimony networks are displayed with the MP bootstrap support and Bayesian posterior probability given for the main branches. The purple circles represent the Gs haplotypes and the green represents the Gc haplotypes. A connecting line between haplotypes represents one mutation and small black dots represent missing (inferred) haplotypes. Blue-filled circles are haplotypes found only in Gc strains from *P. ponderosa,* and the gray-filled circles are haplotypes in Gs strains from the California localized populations.


Grosmannia sp.
Grosmannia clavigera
P. ponderosa associates
California localized population
CMQ6965–ABC.C
GcC12
GcC10
GcC1–GcC8
82, 1.0
GcC11
GcS1–GcS6
GcB1
92, 1.0
GcC9
62, –
GsC1
GsC3
GsC4
76, 1.0
GsC2
GsC5
GsB1–GsB3
GsA1–GsA3
GsM1
CMQ5562–ABC.C
GcC1–GcC10
63, –
GcC11
GcC12
GcS1–GcS3
GcB1
89, 0.99
GsB3
GsA1 GsA2
GsM1 CsC1
GcC4
GsC2
GsC5
70, 0.92
77, 1.0
GsB1
GsB2
GsA3
GsC3
CMQ6634–ABC.C
GcB1
GcC1–GcC10
63, –
GcC11
GcC12
GcS1–GcS6
99, 1.0
GsB1
GsA1
GsB2 GsB3
GsA2 GsM1
GsC1
GsC2
GsA3
GsC3 –
GsC5
CMQ861–ABC.G
GsB1–GsB3
GsA1–GsA3
GsM1
GsC1–GsC5
64, 0.96
GcC12
73, 1.0
GcB1
GcS1–GcS6
GcC11
GcC10
GcC2–GcC5
GcC7–GcC10
GcC1
GcC6
CMQ4184–ABC.G
GcC1–GcC10
GcB1
GcC11 GcC12
GcS1–GcS3
100, 1.0
GsB1–GsB3
GsA1–GsA3
GsM1 GsC1–GsC5
CMQ6740–flavoprotein monooxygenase
GcC1–GcC10
GcB1
GcC11 GcC12
GcS1–GcS3
86, 1.0
GsB1 GsB2
GsA1–GsA3
GsC1 GsC5
86, 1.0
GsM1
GsC2
GsC3
GsB3
GsC4
CMQ3826–Metallo peptidase
GcC1–GcC10
GcB1
GcC11 GcC12
GcS1–GcS3
70, 1.0
GsC3
GsB1–GsB3
GsA1–GsA3
GsM1 GsC1 GsC2
GsC4 GsC5
CMQ5095–Polyketide synthase
GcS1
GcS3
GcC7
GcC9
GcC1–GcC6
GcC8 GcC9
GcB1
GcC11
GcC12
GcS2
99, 1.0
GsC4
GsC5
GsB1–GsB3
GsA1–GsA3
GsM1 GsC1–C3
CMQ5323–Polyketide synthase
GcC7
GcC3
GcC1 GcC2
GcC4 GcC10
GcB1
GcC5 GcC6
GcC8 GcC9
GcC12 GcS2
GcC11
GcS3
GcS1
99, 1.0
GsB2
GsB3
GsA1
GsA2
GsM1
GsB1 GsA3
GsC1–GsC5

B.16 Distributions of polymorphism-to-divergence ratios (NI) and the rate of protein-coding evolution (dN/dS) across 3,476 Gs-Gc orthologous gene models.

a) From the distribution of the neutrality index (NI), the 1,215 protein-coding genes with $-\log_{10}NI > 0$ show evidence of adaptive evolution, but the majority of the genes appear to be under weak purifying selection ($-\log_{10}NI < 0$). **b)** The distribution of the estimated average rate of protein-coding evolution in pairwise Gs–Gc comparisons show that the majority of genes have $dN/dS \leq 1$, indicating that most of these genes are either under purifying selection or positive selection at specific sites. For 228 protein-coding genes, $dN/dS > 1$ suggests adaptive evolution or at least relaxed selection constraints.