Comparative genomics of beetle-vectored fungal pathogens reveals a reduction in genome size and independent evolution of pathogenicity of two tree pathogens.

Taruna A Schuelke[1], Anthony Westbrook[2], Keith Woeste[3], David C. Plachetzki[1], Kirk Broders[4], Matthew D. MacManes[1]

[1]Department of Molecular, Cellular, & Biomedical Sciences, University of New Hampshire, 105 Main Street, Durham, NH 03824; [2]Department of Computer Science, University of New Hampshire, 105 Main Street, Durham, NH 03824; [3]USDA Forest Service Hardwood Tree Improvement and Regeneration Center, Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907; [4]Department of Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, CO 80523

Authors for correspondence

Kirk Broders

Tel: +1 970 491 0850

Email: kirk.broders@colostate.edu

Total word count (excluding summary, references and legends)	5043	No. of figures	3 (all in color)
Summary	200	No. of tables	6
Introduction	341	No of supporting information files	4 (Methods S1, Table S1-S4)
Materials and Methods	1511		
Results	2402		
Discussion	677		
Acknowledgements	112		

Summary

- *Geosmithia morbida* is an emerging fungal pathogen which serves as a paradigm for examining the evolutionary processes behind pathogenicity because it is one of two known pathogens within a genus of mostly saprophytic, beetle-associated, fungi. This pathogen causes thousand cankers disease in black walnut trees and is vectored into the host via the walnut twig beetle. *G. morbida* was first detected in western US and currently threatens the timber industry concentrated in eastern US.
- We sequenced the genomes of *G. morbida* and two non-pathogenic *Geosmithia* species and compared these species to other fungal pathogens and nonpathogens to identify genes under positive selection in *G. morbida* that may be associated with pathogenicity.
- *G. morbida* possesses one of the smallest genomes among the fungal species observed in this study, and one of the smallest fungal pathogen genomes to date. The enzymatic profile is this pathogen is very similar to its relatives.
- Our findings indicate that genome reduction is an important adaptation during the evolution of a specialized lifestyle in fungal species that occupy a specific niche, such as beetle vectored tree pathogens. We also present potential genes under selection in *G. morbida* that could be important for adaptation to a pathogenic lifestyle.

Key words

Geosmithia morbida, pathogenicity, genome reduction, tree pathogen, thousand cankers disease.

Introduction

Uncovering the specific genetic and molecular events behind the evolution of novel traits such as pathogenicity in fungal species has been a long-standing objective of pathologists. *Geosmithia* (Ascomycota: Hypocreales), a genus first proposed in 1979 for fungi that were formerly placed in genus *Penicillium* (Pitt, 1979), serves a paradigm for examining the processes contributing to the evolution of pathogenicity. *Geosmithia* species are filamentous fungi that most commonly associate with phloeophagous bark beetles (Kolarik *et al.*, 2005; Kolarik *et al.*, 2011), although some *Geosmithia* fungi, such as *G. eupagioceri* and *G. microcorthyli*, are known to affiliate with ambrosia beetles (Kolarik & Jankowiak 2013). *Geosmithia* species and their beetle associates occupy a variety of hosts, including pines, oaks, junipers, *and* walnut trees (Kolarik *et al.*, 2007; Kolarik & Kirkendall 2010; Kolarik & Jankowiak 2013). The ecology and diversity of symbiosis between these fungi and their beetle associates is poorly understood, but investigators are beginning to explore such relationships (Kolarik *et al.*, 2007; Kolarik & Jankowiak 2013). While most species in *Geosmithia* are saprotrophic, two species were recently determined to be pathogenic—*G. pallida* (Lynch *et al.*, 2014) and *G. morbida* (Tisserat *et al.*, 2009), on coast live oak (*Quercus agrifolia*) and black walnut (*Juglans nigra*), respectively. However, both of these species live saprophytically in association with bark beetles and other tree hosts. It is still unclear what mechanisms allow these species of *Geosmithia* to be pathogenic to a new host while other members of the genus remain saprobes.

Geosmithia morbida causes thousand cankers disease (TCD) in *Juglans nigra* (eastern black walnut). Although no evidence of TCD has been detected in other *Juglans* to date, several species, such as *J. californica*, *J. cinerea*, *J. hindsii*, *J. regia*, are also susceptible to the pathogen (Utley *et al.*, 2013). The fungus is most often vectored into its hosts by *Pityophthorus juglandis*, commonly known as the walnut twig beetle (WTB) (Kolarik *et al.*, 2011). Unusual mortality of *J. nigra* was first noted in Colorado, US in 2001. Since then, nine western states (CO, WA, OR, ID, NV, UT, CA, NM, AZ) and seven eastern states (PA, OH, IN, MD, VA, TN,

NC) have reported TCD in one or more locations (Zerillo *et al.*, 2014). This increase in TCD is likely a consequence of the expansion of WTB's geographic range. WTB was present in only four counties of California, Arizona and New Mexico in the 1960s, however, as of 2014, the beetle has been detected in over 115 counties in the western and eastern US (Rugman-Jones *et al.*, 2014).

The origin of this pathogen is not clear. However, it has been hypothesized that *G. morbida* may have undergone a host shift from *J. major* (Arizona black walnut) to a more naïve host, *J. nigra*, because the fungus does not cause disease in the Arizona black walnut, and neither WTB nor *G. morbida* were observed in the native range of *J. nigra* until 2010 (Zerillo *et al.*, 2014). *J. nigra* is not indigenous to western US but was planted throughout the region as an ornamental species. An alternative prediction based on *G. morbida* population genetic data is that the origin of *G. morbida* and WTB are the walnut populations of southern California, where the pathogen has been isolated from both healthy and diseased *J. californica* trees (Zerillo *et al.*, 2014).

Early symptoms of infection by *G. morbida* include yellowing, wilting and thinning of the foliage followed by branch dieback and tree death within 2-3 years after the initial infestation (Tisserat *et al.*, 2009; Kolarik *et al.*, 2011). Little is known about the specific means *G. morbida* employs for initiating and maintaining the infection, or what benefits, if any, the fungus imparts to the WTB vector. However, previous studies have demonstrated that fungal pathogens that occupy ecological niches similar to *G. morbida* must be capable of enduring and combating toxic host environments used by plants to resist infection. For instance, *Grosmannia clavigera*, a fungal symbiont of the mountain pine beetle (*Dendroctonus ponderosae)*, can detoxify metabolites such as terpenoids and phenolics produced by the host as defense mechanisms (DiGuistini *et al.*, 2011).

We recently developed a reference genome of *Geosmithia morbida* that consisted of 73 scaffolds totaling 26.5 Mbp in length (Schuelke *et al.*, 2016). This genome represents one of the

smaller fungal tree pathogen genomes reported to date. Rapid changes in genome size have accompanied dramatic biological changes in newly emerged fungal and oomycete species (Raffaele & Kamoun 2012; Adhikari *et al.*, 2013). In fungi, a link has been observed between genome expansion and evolution of pathogenicity (Raffaele & Kamoun 2012). Genome expansions were associated with parasitism in general and increased pathogenicity and virulence in several fungal lineages (Spanu *et al.*, 2010). Previous genome sequencing of *Geosmithia morbida* (Schuelke *et al.*, 2016) showed that this newly emerged fungal pathogen has a smaller genome than several of its closely related nonpathogenic relatives in the Hypocreales. Hence, *Geosmithia* has taken an evolutionary path to pathogenicity that has not been characterized previously in plant-associated fungi.

The evolution of pathogenicity via genome size reduction is not understood, and although it is contrary to our current expectation of how pathogenicity develops in non-pathogens, it might be common, particularly in beetle-associated symbionts. In fact, in the last decade several beetle-associated fungi have emerged as plant pathogens, including *Grosmannia clavigera*, which is also a tree infecting, beetle-vectored fungus, and its genome size is only 29.8 Mb (Diguistini *et al.*, 2011). The arrival of new pathogens, frequently referred to as Black Swan events due to their perceived unpredictability, represent a significant threat to native and agriculturally important tree species (Pleotz *et al.*, 2013). Thus, beetle-associated symbionts that have switched to pathogens represent excellent models for investigating the evolution of pathogenicity and its relationship to genome size. Although the genus *Geosmithia* is distributed worldwide, *G. morbida*, and more recently, *G. pallida,* are the first members of the genus to be described as plant pathogens among the 60 known nonpathogenic species (Kolarik & Kirkendall 2010; Kolarik *et al.*, 2011; Lynch *et al.*, 2014).

In this work, we compare the reference genome of the pathogenic and host specific species *G. morbida* with two closely related non-pathogenic generalist species, *G. flava* and *G. putterillii*. Based on this comparison, we identify genes under positive selection that may be

involved in the specialization of a pathogenic life strategy that depends on a single beetle vector and a narrow, but potentially expanding, host range. We also present a species phylogeny estimated using single-copy orthologs that confirms the placement of *Geosmithia* species in the order Hypocreales, and that their closest fungal relative is *Acremonium chrysogenum*. The primary goal of this study was to gain insight into the evolution of pathogenicity within *G. morbida*. We also investigated the presence of convergent evolution in *G. morbida* and *Grosmannia clavigera*, two tree pathogens vectored into their hosts via beetle symbionts.

Materials and Methods

DNA extraction and sequencing

The CTAB method delineated by the Joint Genome Institute was used to extract DNA for genome sequencing from lyophilized mycelium of *Geosmithia flava* and *Geosmithia putterillii* (Kohler *et al.*, 2011). Table 1 lists genetic, geographic, and host information for each *Geosmithia* species used in this study. Total DNA concentration was measured with Nanodrop, and DNA sequencing was conducted at Purdue University Genomics Core Facility in West Lafayette, Indiana. DNA libraries were prepared using the paired-end Illumina Truseq protocol and sequenced on an Illumina HiSeq 2500 using a single lane. Mean insert sizes for *G. flava* and *G. putterillii* were 477bp and 513bp, correspondingly. The remaining sequencing statistics can be found in Table 2.

Preprocessing sequence data

The raw paired-end reads for *G. flava* and *G. putterillii* were corrected using BFC (version r181) (Li 2015). BFC utilizes a combination of hash table and bloom-filter to count *k*-mers for a given read and correct errors in that read based on the *k*-mer support. Because BFC requires interleaved reads as input, khmer 1.1 was leveraged to interleave as well as split the paired-end reads before and after the error correction stage, respectively (Crusoe *et al.*, 2015). Next, low quality bases and adapters in error corrected reads were trimmed with Trimmomatic, version 0.32, using a Phred threshold of 4 (Bolger *et al.*, 2014).

Assembly construction

Genome assemblies were constructed with ABySS 1.9.0 using four k-mer sizes (61, 71, 81, and 91) (Simpson *et al.*, 2009). The resulting assemblies were evaluated using BUSCO (v1.1b1) (Simão *et al.*, 2015), which assess completeness based on the presence of universal single-copy orthologs within fungi. Length-based statistics were generated with QUAST v2.3 (Gurevich *et al.*, 2013). Final assemblies were manually chosen based on length-based and genome completeness statistics. Furthermore, the raw reads of *G. flava* and *G. putterillii* were mapped back to their corresponding genomes using BWA version 0.7.9a-r786 (Li & Durbin, 2009) to assess the quality of the chosen assemblies.

Structural and Functional Annotation

We utilized the automated annotation software Maker version 2.31.8 (Cantarel *et al.*, 2008) to functionally annotate the genomes of *G. flava* and *G. putterillii*. We used two of the three gene prediction tools available within the pipeline SNAP (released 2013, Korf 2004) and Augustus 2.5.5 (Stanke *et al.*, 2006). SNAP was trained using gff files generated by CEGMA v2.5 (a program similar to BUSCO) (Parra *et al.*, 2007). Augustus was trained with *Fusarium solani* protein models (v2.0.26) downloaded from Ensembl Fungi (Kersey *et al.*, 2016). The protein sequences generated by the structural annotation were blasted against the Swiss-Prot database (Boutet *et al.*, 2016) to functionally annotate the genomes of *G. flava* and *G. putterillii*.

Assessing repetitive elements profile

To evaluate the repetitive elements profile of *G. flava* and *G. putterillii*, we masked the interspersed repeats within the assembled genomes with RepeatMasker 4.0.5 (Smit *et al.*, 1996) using the sensitive mode and default values as arguments.

Identifying putative genes involved in host-pathogen interactions

To search for putative genes contributing to pathogenicity, we conducted a BLASTp (v2.2.28+) (Altschul *et al.*, 1990) search with an e-value threshold of 1e-6 against the PHI-base 4.0 database (Winnenburg *et al.*, 2006) that includes known genes implicated in pathogenicity.

Additionally, we identified proteins that contain signal peptides and lack transmembrane domains in each *Geosmithia* species as well as their close relative *Acremonium chrysogenum* with SignalP 4.1 and TMHMM 2.0 using default parameters (Krogh *et al.*, 2001; Peterson *et al.*, 2011).

Identifying species specific genes

To identify unique genes present in *Geosmithia* morbida, we performed an all-versus-all BLASTp search among the three *Geosmithia* species and *A. chrysogenum* with Orthofinder version 0.3.0 (Emms & Kelly 2015). Using a custom Python script, we analyzed homology among the four fungal species.

Identifying carbohydrate-active proteins and peptidases

To identify enzymes capable of degrading carbohydrate molecules in species belonging to Hypocreales and *G. clavigera*, we performed a HMMER 3.1b1 (Eddy 1998) search against the CAZy database (Lombard *et al.*, 2014) released July 2015 and filtered the results following the developer's recommendations. Lastly, we profiled the proteolytic enzymes present in species using the *MEROPS* database 10.0 (Rawlings *et al.*, 2016).

Phylogenetic analysis

Taxon Sampling

In order to determine phylogenetic position of *Geosmithia*, we combined the predicted peptide sequences from three *Geosmithia* species described here with the predicted peptide sequences of an additional 17 fungal genomes that represent the breadth of pathogens and non-pathogens within Ascomycota. Our dataset contained eleven pathogens and nine non-pathogens (Table 3).

Inferring Orthology

Orthologous peptide sequences among the 20 fungal genomes were determined using OrthoFinder version 0.3.0 (Emms & Kelly 2015). OrthoFinder performs an all-versus-all BLASTp (2.2.28+, Altschul *et al.*, 1990) search among a set of protein coding genes to infer orthogroups and aligns them using MAFFT (v7.123b, Katoh & Standley 2013). These orthogroups may

contain paralogs as well as orthologs, and because datasets rich in paralogs can confound phylogenomic analysis, the alignment files produced by OrthoFinder were parsed to recover only those orthogroups that contained single-copy orthologs from each of the 20 species. This resulted in 1,916 total orthogroups with 100% taxon occupancy.

Trimming Alignments

For each alignment, regions that contained gap rich sites were removed using *–gappout* option in trimAl v1.4.rev15 (Capella-Gutiérrez *et al.*, 2009). Next, all files containing orthogroups were renamed so the respective headers among these files were identical and individual alignments were concatenated. Concatenation resulted in a single fasta file containing all 1,916 partitions with 1,054,662 sites at 100% taxon occupancy. This initial alignment was further filtered using MARE (v.0.1.2) (Misof *et al.*, 2013), which reduced the character matrix to 247,627 sites. This reduced fasta alignment was converted into a partitioned phylip formatted file. Next, the best-fit substitution models for each partition and a global partitioning scheme were determined with PartitionFinder (v1.1.1) using hcluster clustering algorithm and default parameters (Lanfear *et al.*, 2014).

Constructing Phylogeny

Maximum likelihood (ML) analysis was conducted in RaxML v 8.1.20 (Stamatakis 2014) leveraging the partitioning scheme determined by PartitionFinder. The ML tree and 200 bootstrap replicates were performed in a single analysis using the *–f a* option. In addition, we conducted Bayesian Markov Chain Monte Carlo (BMCMC) analysis in MrBayes 3.2.6 (Ronquist *et al.*, 2012). For MrBayes analysis, we specified the mixed amino acid model prior and ran the fully partitioned tree search for 215,000 generations. A consensus tree was then generated after discarding 50% of the run as burnin. The nexus file, including MrBayes block, provides other details of the MrBayes analysis (Methods S1).

Detecting genes under positive selection

To identify genes under positive selection in *G. morbida*, we compared *G. morbida* with all non-pathogens from the aforementioned 20 fungi used to estimate the species tree. Among this batch of 10 fungal species, we detected 22,908 protein orthogroups using OrthoFinder that contained paralogs as well as orthologs. Of these, only 9,560 orthogroups were alignable with MAFFT because many groups consisted of only one sequence from a single species (Katoh & Standley, 2013). A total of 3,327 orthogroups, composed of single-copy orthologs, were filtered and corresponding coding DNA sequences for each peptide in these partitions were extracted using custom scripts that can be found online.

The coding DNA sequences were then aligned with MACSE v1.01.b (Ranwez *et al.*, 2011). This Java-based utility accounts for frameshifts and premature stop codons in coding sequences during the alignment process and outputs aligned protein and nucleotide sequences. In order to filter out alignments with frameshifts and internal stop codons, we utilized a program called PAL2NAL v14 (Suyama *et al.*, 2006). This software searches for complementary regions between multiple protein alignments and the corresponding coding DNA sequences, and omits any problematic codons from the output file. This cleaning step reduced the number of 3,327 orthogroups to 2,798 that were used for detecting genes under selective pressures.

We used the branch-site model (BSM) in the CodeML program of package PAML v4.8 for selection analysis (Yang 2007). BSM permits ω (dN/dS) to vary among sites and branches permitting the identification of specific branches and sites subjected to selection. We computed two models in order to calculate and compare the likelihood values: a null model with a fixed ω value of 1 and an alternative model that estimates ω in the foreground branch, which is *G. morbida* in our case. In the effort to reduce false positives, we implemented the Benjamini-Hochberg correction method when comparing likelihood ratios for null and alternative models using a *P*-value threshold of 0.05. We performed similar BLAST searches as mentioned previously to characterize the functions of these proteins and identify proteins with signal peptides and transmembrane domains.

We repeated the above procedures for detecting genes under selection in *Grosmannia clavigera* because this fungal pathogen plays an ecological role similar to *G. morbida*. By performing these analyses, we sought to uncover genes under adaptive evolution in both beetle-vectored tree pathogens.

Sequence data and code availability

The raw reads and assembled genomes reported in this paper are available at European Nucleotide Archive under Project Number PRJEB13066. The in silico genereated transcript and protein files are being deposited at Dryad. The code is available at Github (https://github.com/tarunaaggarwal/G.morbida.Comp.Gen)

Results

Assembly features

We recently assembled a reference genome for a *G. morbida* strain isolated from *Juglans californica* in Southern California (Schuelke *et al*., 2016). The reference contained 73 scaffolds with an estimated size of 26.5 Mbp. By using the MAKER annotation pipeline, we predicted 6,273 protein models in this reference in-silico (Cantarel *et al.*, 2008). In this work, we sequenced strains of *G. flava* and *G. putterillii* at approximately 102x and 131x coverage, respectively. The *G. flava* assembly was composed of 1,819 scaffolds totaling 29.47 Mbp in length, and the *G. putterillii* genome contained 320 scaffolds extending 29.99 Mbp. *G. flava* and *G. putterillii* totaled 6,976 and 7,086 protein models, respectively. Both genomes contained 98% of the single-copy orthologs present in more than 90% of the fungal species. Nearly all of the raw reads (97% and 98%) mapped back to *G. flava* and *G. putterillii* genome assemblies, respectively (Table 4). These statistics indicated that our genome assemblies are high quality and complete.

An estimated 0.80% of *G. morbida* reference genome sequence represented repeats, whereas 0.63% and 0.64% of the sequences in *G. flava* and *G. putterillii* consisted of repetitive

276 elements. There were 60, 42, and 15 DNA transposons in *G. morbida*, *G. flava*, and *G. putterillii*,
277 respectively. Furthermore, *G. morbida* possessed only 152 retroelements, whereas *G. flava* and
278 *G. putterillii* had 401 and 214 of such elements, correspondingly (Table 5).
279 Although, the extent to which mobile genetic elements affect genome evolution in
280 *Geosmithia* is unknown, mobile genetic elements may be influential drivers of adaptive evolution
281 in *G. morbida*. They are known to be responsible for genomic rearrangements and expansion,
282 horizontal gene transfer and generation of new genes (Casacuberta & Gonzalez 2013;
283 Stukenbrock & Croll 2014). For example, *Fusarium oxysporum* has a genome nearly 60 Mbp in
284 length and contains 16.83 Mbp of repetitive sequences. *F. oxysporum* also contained four more
285 chromosomes than closely related species, and the added chromosomes were rich in
286 transposons and genes such as putative effectors, necrosis and ethylene-inducing proteins and
287 carbohydrate binding enzymes (Ma *et al.*, 2010). Although *Geosmithia morbida* harbors fewer
288 mobile genetic elements than fungal species such as *F. oxysporum*, it is possible that such
289 elements have contributed to the evolution of pathogenicity in *Geosmithia* via gene expansion
290 and/or horizontal gene transfer. Understanding the role of mobile genetic elements within genus
291 *Geosmithia* may be key in discovering the genetic basis behind the evolution of pathogenicity.
292 **Identifying putative genes involved in pathogenicity**
293 Approximately 32%, 34%, and 35% of the total proteins in *G. morbida*, *G. flava* and *G. putterillii*
294 respectively shared significant homology with protein sequences in the database. The number
295 of unknown proteins with hits in the PHI-base database was similar for *G. morbida* (26), *G. flava*
296 (28), and *G. putterillii* (36). The full BLASTp search results against the PHI-base database for *G.*
297 *morbida*, *G. flava*, and *G. putterillii* are available in the supporting material (Table S1).
298 **Identifying species-specific genes**
299 The three *Geosmithia* species and *A. chrysogenum* contained a total of 9065 orthologs and
300 paralogs. Among the set of homologous genes there were 4,655 single copy orthologs. *A.*
301 *chrysogenum* contained 2338 species-specific genes, of which seven genes were paralogous.

G. morbida possessed 76 unique genes whereas, *G. putterilli* and *G. flava* had 161 and 146 species-specific genes. The two nonpathogenic *Geosmithia* species did not contain any paralogs, however *G. morbida* had three unique genes present in multiple copies. Based on a functional search against NCBI's non-redundant database, the three genes encode hydantoinase B/oxoprolinase, aldehyde dehydrogenase, and ABC-2 type transporter.

These findings are significant because all three of these proteins are involved in stress responses that can be induced by the host immune system during the infection process. For example, aldehyde dehydrogenases are part of a large protein family that detoxify aldehydes and alcohols in all organisms including fungal species (Asiimwe *et al.*, 2012). Hydantoinase B/oxoprolinase is involved in the synthesis of glutathione, a compound essential for basic cellular functions but also important in cellular defense against oxidative stress (Pocsi *et al.*, 2004). Glutathione has been shown to chelate damaging metal ions by inhibiting their spread in the cell (Pocsi *et al.*, 2004), and to prevent the accumulation of H_2O_2 in *Paxillus involutus* (Ott *et al.*, 2002). Lastly, ATP-binding cassette (ABC) proteins belong to an especially large family of proteins that regulates transport of substances across the cellular membrane. In pathogenic fungi, they are involved in drug resistance and in the production of defense molecules (Krattinger *et al.*, 2009; Wang *et al.*, 2013; Karlsson *et al.*, 2015).

Identifying putative secreted proteins

A total of 349, 403, and 395 proteins in *G. morbida*, *G. flava*, and *G. putterillii* contained signal peptides respectively. Of these putative signal peptide-containing proteins in *G. morbida*, 27 (7.7%) encoded proteins with unknown function, whereas *G. flava* and *G. putterillii* contained 29 (7.2%) and 30 (7.6%) unknown proteins, respectively. The difference in percent of unknown proteins with signal peptides was minimal among the three genomes. For each species, proteins containing signal peptides were subjected to a membrane protein topology search using TMHMM v2.0. There were 237, 281, and 283 proteins in *G. morbida*, *G. flava*, and *G.*

putterillii that lacked any transmembrane protein domains. Again, these numbers were not significantly different.

Profiling carbohydrate active enzymes and peptidases

CAZymes are carbohydrate active enzymes that break down plant structural components, enabling initiation and establishment of infection. We assessed the CAZymatic profile of all species in the order *Hypocreales*, *Geosmithia* species, and *Grosmannia clavigera* (Figure 1). The glycoside hydrolase (GH) family members dominated all protein models, followed by glycosyltransferase (GT) family. The two most prominent families among all fungal species were GH3 and GH16 (Table S2). GH3 hydrolases are involved in cell wall degradation and overcoming the host immune system, and GH16 enzymes fulfill a wide range of cellular functions including transporting amino acids. The third most representative family was GH18; however *G. morbida* only contained 4 of these enzymes. In contrast, this number for other species ranges from 9 to 31 enzymes. Along with acetylglucosaminidases, family GH18 harbors chitinases that assist in the production of carbon and nitrogen. In terms of other CAZyme families, all fungi except *F. solani* express a similar overall distribution. *Fusarium solani* contains more CAZymes than any other pathogen or non-pathogen. This *Fusarium* species is a generalist necrotrophic pathogen that is believed to possess more CAZymes than biotrophic and hemibiotrophic fungi. This discrepancy may be due to the fact that necrotrophic pathogens require an extensive toolkit to promote host cell death as quickly as possible; whereas biotrophs need to keep the host alive, and dispensing large number of degradation enzymes can be detrimental to that aim (Zhao *et al.*, 2013).

In addition to profiling CAZymes, we also performed a BLAST search against the peptidase database—Merops v10.0 (Rawlings *et al.*, 2016)-- for each *Hypocreales*, *Ceratocystis platani*, and *G. clavigera*. Among the pathogens, *G. morbida* has the third highest percent of predicted proteases after *Cordyceps militaris* (insect pathogen) and *G. clavigera* (Figure 2, Table S3). Moreover, *Geosmithia flava* and *G. putterillii* have the largest percent of peptidases

among the nonpathogenic fungi. All three *Geosmithia* species illustrate similar proteolytic profiles and contain no glutamic and mixed peptidases. These results were expected because all three *Geosmithia* species are closely related. Furthermore, given that these species are plant affiliates (except *Cordyceps militaris*), the ability to degrade lignin and cellulose is an important life history trait that is conserved throughout fungal pathogens, but perhaps did not give rise to pathogenicity in *G. morbida*.

Inferring phylogeny

Even though *Geosmithia* was first established as a genus in 1979, it has only recently been described in depth. One of the main objectives in this study was to uncover the phylogenetic relationship between *Geosmithia* species and other fungal pathogens using coding DNA sequence data. In order to determine the broader evolutionary history of *Geosmithia* species, we constructed maximum likelihood (ML) and Bayesian Markov Chain Monte Carlo (BMCMC) phylogenies using 1,916 single-copy orthologs from *G. morbida*, *G. putterillii*, *G. flava*, and 17 additional fungal taxa (Table 3). Our final dataset consisted of 11 pathogens and 9 non-pathogens. After trimming and filtering, our 1,916 orthogroups contained approximately $1e10^6$ amino acid sites in total. The topologies of trees generated under ML and BMCMC were identical, and all nodes in all analyses received bootstrap support of 100% (ML) and posterior probabilities of 1.0 (BMCMC). The analyses resulted in a single, identical tree topology (Figure 3) that was supported by previous research (Fitzpatrick *et al.*, 2006, Wang *et al.*, 2009). *Geosmithia* species form a monophyletic clade with two nonpathogenic fungi—*Acremonium chrysogenum* and *Stanjemonium griseum*—indicating that the common ancestor shared among these species was not a pathogen.

Genes under positive selection

In order to understand the molecular basis of pathogenicity in *G. morbida*, we sought to detect genes under positive selection. For this, we first searched for all single-copy orthologs shared among the 9 non-pathogens and *G. morbida* using OrthoFinder (v0.3.0). We extracted the

379 corresponding coding sequences for each protein in the 3,327 orthogroups containing 1:1
380 orthologs using a custom python script. These orthogroups were aligned using MACSE v1.01b
381 and cleaned with PAL2NAL v14. After alignment and cleaning of orthogroups there were 2,798
382 multiple sequence alignments that were used for selection analysis.
383 To identify coding sequences and sites under selection, we leveraged the branch-site
384 model in PAML's codeml program (4.8). *Geosmithia morbida* was selected as the foreground
385 branch. Our results showed 38 genes to be under positive selection using an adjusted *P*-value <
386 0.05. Next, we performed a functional search for each protein by blasting the peptide sequences
387 against the NCBI non-redundant and pfam databases. We determined that several were
388 involved in catabolic activity, gene regulation, and cellular transport (Table 6, Table S4).
389 For instance, a cullin3-like protein was predicted to be under positive selection. Cullin3-
390 like proteins belong to a group of structurally similar molecules involved in protein degradation,
391 such as the Skp-Cullin-F-box (SCF) ubiquitin ligase complex, was predicted to be under positive
392 selection (Cardozo & Pagano 2004; Pintard *et al.*, 2004). Furthermore, a ubiquitin-conjugating
393 enzyme (E2) that interacts with cullin3 to prepare substrate for degradation, also had a dn/ds >
394 1, indicating that both genes are under positive selection within *G. morbida*. Although little is
395 known regarding the precise functional abilities of these complexes, it is possible these proteins
396 are involved in pathogenicity of *G. morbida*. Previous studies have also implicated ubiquitin
397 ligase complexes in infection and disease development (Duyvesteijn *et al.*, 2005; Han *et al.*,
398 2007).
399 Our analysis also revealed a regulatory protein homologous to basic leucine zipper
400 (bZIP) transcription factors was under selection. The bZIP proteins are similar to AP-1
401 transcription factors and monitor several developmental and physiological processes including
402 oxidative stress responses in eukaryotes (Corrêa *et al.*, 2008). Fungal pathogens such as the
403 rice blast fungus *Magnaporthe oryzae* express AP1-like transcription factor called MoAP1 that
404 contains bZIP domain. MoAP1 is highly active during infection and is translocated from the

cytoplasm to the nucleus in response to oxidative stress induced by H_2O_2 (Guo *et al.*, 2011). MoAP1 regulates enzymes such as laccase and glutamate decarboxylase that are involved in lignin breakdown and metabolism of γ-aminobutyric acid, respectively (Solomon & Oliver 2002; Baldrian 2005; Janusz *et al.*, 2013). Some of the other positively selected genes include ABC transporter, proteases, proteins involved in apoptosis, and proteins related to DNA replication and repair. As previously mentioned, ABC transporters are important mediators that aid in protection against plant defenses as well as natural toxic compounds (Krattinger *et al.*, 2009; Wang *et al.*, 2013; Karlsson *et al.*, 2015; Lo Presti *et al.*, 2015). Apoptosis or programmed cell death helps establish resistance during host-microbe interactions, helps organisms cope with oxidative environments, and may even be essential for infection (Veneault-Fourrey *et al.*, 2006; Kabbage *et al.*, 2013). In fungal species, proteins involved in DNA replication and repair are essential for the formation and penetration of appressorial structures into the host cell (Son *et al.*, 2016). Only five of the 38 genes with evidence of selection encoded proteins with unknown functions. These positively selected genes might be involved in the evolution and adaptation of *G. morbida*.

Transmembrane protein and effector genes

Transmembrane proteins are important mediators between a host and its pathogens during microbial invasion. Fungal pathogens either penetrate a surface or enter the host through a wound or opening such as stomata in order to gain access to the nutrients in the plant (Chisholm *et al.*, 2006). Once the infiltration process is completed, pathogens are exposed to host plasma membrane receptors that detect pathogen-associated molecular patterns (PAMP) and induce PAMP-triggered immunity (PTI) to prevent further proliferation of the microbe. Transmembrane proteins expressed by a fungal pathogen are crucial during PTI because they are responsible for suppressing PTI directly or by secreting effector molecules, which contain signal peptides necessary for proper targeting and transport (Chisholm *et al.*, 2006; Boller & He 2009). Our analysis of the 38 proteins under positive selection showed that 11 of these possess

at least one or more transmembrane domains. Although nearly 30% of the positively selected genes identified were membrane bound, a similar proportion of non-selected genes in *G. morbida* were membrane associated, so this result is not strong evidence that interactions with the host surface are drivers of evolution within *G. morbida*. Among proteins under selection we found no protein that contained a signal peptide, indicating none of these proteins are secretory.

Genes under adaptive evolution in beetle-vectored fungal pathogens

In addition to detecting genes under selective pressures in *G. morbida*, we performed the same selection analysis for *Grosmannia clavigera* to identify overlapping proteins that may help explain adaptations leading to the ecological role these two beetle-vectored fungi play. We found that *G. clavigera* possessed 42 positively selected genes that shared protein domains with only two of the 38 genes predicted to be under selection in *G. morbida*. The two overlapping motifs are methyltransferase and protein kinase domains. Our KEGG analysis exhibited no common pathways between *G. morbida* and *G. clavigera*. These findings emphasize that evolutionary forces act differently on divergent populations. All fungal pathogens face dissimilar environmental challenges and associate with different hosts both spatially and temporally. Even closely related organisms can be highly distinct molecularly. For instance, the fungi responsible for the Dutch elm disease—*Ophiostoma ulmi* and *O. novo-ulmi*—differ in their genetic composition and virulence despite their strong evolutionary relationship (Brasier 2001; Khoshraftar *et al.*, 2013; Comeau *et al.*, 2015).

Discussion

This study aims to provide insight into the evolution of pathogenicity within *Geosmithia morbida*, a beetle vectored pathogen that is the causal agent of Thousand Cankers Disease in *Julgans* species. Here, we present *de novo* genome assemblies of two nonpathogenic *Geosmithia* species, *G. flava* and *G.* putterillii, and employ comparative genomics approach to uncover the molecular factors contributing to pathogenicity in *G. morbida*.

G. flava and *G. putterillii* have estimated genome sizes of 29.6 Mbp and 30.0 Mbp, correspondingly. These assemblies are larger than the genome of *G. morbida*, which measures 26.5 Mbp in length. In contrast to other species in the phylogeny (Figure 3), fungi associated with trees either as pathogens or saprophytes (*Geosmithia* species, *G. clavigera*, and *C. platani*) had reduced genomes and gene content. We predict this genome and gene content reduction is a result of evolving specialized lifestyles to occupy a specialized niche. For instance, all three *Geosmithia* species and *G. clavigera* are vectored into their respective hosts via bark beetles, which may result in strong selection on the genetic variability of the fungi because they must adapt to their vectors and hosts simultaneously. Moreover, possessing genes that are not essential for this specialized lifestyle may impose a fitness disadvantage on the pathogen, as they may represent potential targets for host resistance genes. A recent study characterizing the genome of mycoparasite *Escovopsis weberi* showed that specialized pathogens tend to have smaller genomes and predicted protein sets because they lack genes that are not required beyond their restricted niche when compared to closely related generalists (de Man *et al.*, 2016). Our results agree with this finding because *G. morbida* has a more specialized beetle vector (*P. juglandis*) and plant host range (*Juglans* species) in comparison to *G putterilli* and *G. flava* which can be found on a variety of trees species including both gymnosperms and angiosperms (Kolařík *et al.*, 2004; Kolařík & Jankowiak 2013), and can be vectored by multiple beetle species (Kolařík *et al.*, 2008.) This represents a significant contrast to previous reports that have documented the importance of genome expansion with the evolution of pathogenicity (Adhikari *et al.*, 2014; Raffaele & Kamoun 2012). Furthermore, our results are supported by prior findings which showed that gene loss and gain can lead to a more specialized lifestyle in bacterial and eukaryotic lineages (Ochman & Moran 2001, Lawrence 2005). Another example is a study by Nagendran *et al.*, 2009 that showed *Amanita bisporigera*, which is an obligate ectomycorrhizal symbiont, lacks many plant cell-wall-degrading enzymes suggesting that these genes may no longer be required for *A. bisporigera's* specialized lifestyle.

Genome reduction is an important evolutionary mechanism that propels divergence of species and more often than not enables adaptation to specific environments. Although genome reduction is more frequent in prokaryotes, it is not uncommon among eukaryotes including fungal species (Ochman and Moran 2001, Nagendran *et al.*, 2009, Spanu *et al.*, 2010).

Although one might expect the pathogen *G. morbida* to possess more carbohydrate binding enzymes and peptidases than its non-pathogenic relatives, our results indicated that all three species had similar enzymatic profiles (Figures 1 and 2). Despite these similarities, our PAML analysis identified 38 genes under positive selection in *G. morbida* when compared to other nonpathogens within the order Hypocreales including non-pathogenic *Geosmithia* species. These genes encode for proteins that have been implicated in pathogenicity in other fungal pathogens such as *Magnaporthe oryzae*. Additionally, we found peptides with protein kinase and methyltransferase domains that are under positive selection in both *G. morbida* and *G. clavigera.* Proteins kinases were previously shown to be under strong positive selection in *G. clavigera* (Alamouti *et al.*, 2014). This result was especially important given the key contributions that protein kinases make in initiating signal transduction pathways during pathogen host interactions. Our study identified a small set of genes potentially involved in the evolution of pathogenicity in the genus *Geosmithia*. Functional experiments and analyses of the expression levels of these genes during infection as compared to gene expression of a non-pathogen would shed light on the mechanisms influencing pathogen evolution and genome reduction.

Acknowledgments

The use of trade names is for the information and convenience of the reader and does not imply official endorsement or approval by the United States Department of Agriculture or the Forest Service of any product to the exclusion of others that may be suitable. Partial funding was provided by the New Hampshire Agricultural Experiment Station. Special thanks to Dr. Miroslav Kolarik for providing isolates of *Geosmithia flava* and *Geosmithia putterilli*. We are also grateful to Dr. Joseph Spatafora and his team for giving us permission to utilize sequence data for *Stanjemonium griseum* and *Myrothecium inundatum*. Lastly, we thank the 1000 Fungal Genomes Project for being a valuable source of genetic data.

Author contribution

TAS conceived, designed and performed the experiments and wrote the manuscript. AW assembled and annotated the genomes. KB conceived and designed the study, wrote and reviewed the manuscript. KB also conceived funding. DCP designed and implemented the phylogenetic methods in this study and reviewed the manuscript. KW conveived funding and designed the experiments. KW also wrote and reviewed the manuscript. MDM conceived and designed the study, developed analyses pipelines and edited the manuscript.

References

Adhikari BN, Hamilton JP, Zerillo MM, Tisserat N, Lévesque CA, Buell CR. 2013. Comparative genomics reveals insight into virulence strategies of plant pathogenic oomycetes. *PLoS One* **8:** e75072.

Alamouti SM, Haridas S, Feau N, Robertson G, Bohlmann J, Breuil C. 2014. Comparative genomics of the pine pathogens and beetle symbionts in the genus *Grosmannia*. *Molecular biology and evolution* **31:** 1454-1474.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215:** 403–410.

Amselem J, Cuomo CA, Van Kan JA, Viaud M, Benito EP, Couloux A, Coutinho PM, De Vries RP, Dyer PS, Fillinger S *et al.* 2011. Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genetics* **7:** e1002230.

536 **Asiimwe T, Krause K, Schlunk I, Kothe E. 2012.** Modulation of ethanol stress tolerance by
537 aldehyde dehydrogenase in the mycorrhizal fungus *Tricholoma vaccinum*. *Mycorrhiza*
538 **22:** 471-484.
539 **Baldrian P. 2005.** Fungal laccases-occurrence and properties. *FEMS Microbiology Reviews*
540 **30:** 215-242.
541 **Belbahri L. 2015.** Genome sequence of *Ceratocystis platani*, a major pathogen of plane
542 trees. [WWW document] URL http://www.ncbi.nlm.nih.gov/nuccore/814603118.
543 **Berka RM, Grigoriev IV, Otillar R, Salamov A, Grimwood J, Reid I, Ishmael N, John T,**
544 **Darmond C, Moisan MC *et al.* 2011.** Comparative genomic analysis of the thermophilic
545 biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. *Nature*
546 *Biotechnology* **29:** 922-927.
547 **Blanco-Ulate B, Rolshausen PE, Cantu D. 2013.** Draft Genome Sequence of the
548 Grapevine Dieback Fungus *Eutypa lata* UCR-EL1. *Genome Announcements* **1:** e00228-
549 13.
550 **Bolger AM, Lohse M, Usadel B. 2014.** Trimmomatic: a flexible trimmer for Illumina
551 sequence data. *Bioinformatics* **30:** 2114-2120.
552 **Boller T, He SY. 2009.** Innate immunity in plants: An arms race between pattern recognition
553 receptors in plants and effectors in microbial pathogens. *Science* **324:** 742-744.
554 **Boutet, E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, Poux S,**
555 **Bougueleret L, Xenarios I. 2016.** UniProtKB/Swiss-Prot, the manually annotated
556 section of the UniProt KnowledgeBase: how to use the entry view. *Plant Bioinformatics:*
557 *Methods and Protocols* 23-54. [WWW document] URL http://www.uniprot.org/.
558 [accessed 6 May 2015].
559 **Brasier CM. 2001.** Rapid Evolution of Introduced Plant Pathogens via Interspecific
560 Hybridization Hybridization is leading to rapid evolution of Dutch elm disease and other
561 fungal plant pathogens. *Bioscience* **51:** 123-133.
562 **Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B. 2008.** MAKER: An easy-to-
563 use annotation pipeline designed for emerging model organism genomes. *Genome*
564 *Research* **18:** 188-196
565 **Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009.** trimAl: a tool for automated
566 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25:** 1972-1973.
567 **Cardozo T, Pagano M. 2004.** The SCF ubiquitin ligase: insights into a molecular machine.
568 *Nature Reviews Cell Biology* **5:** 739-751.

569 **Casacuberta E, González J. 2013.** The impact of transposable elements in environmental
570 adaptation. *Molecular ecology* **22:** 1503-1517.
571 **Chisholm ST, Coaker G, Day B, Staskawicz BJ. 2006.** Host-microbe interactions: shaping
572 the evolution of the plant immune response. *Cell* **124:** 803-814.
573 **Coleman JJ, Rounsley SD, Rodriguez-Carres M, Kuo A, Wasmann CC, Grimwood J,**
574 **Schmutz J, Taga M, White GJ, Zhou S *et al.* 2009.** The genome of *Nectria*
575 *haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS*
576 *Genetics* **5:** e1000618.
577 **Comeau AM, Dufour J, Bouvet GF, Jacobi V, Nigg M, Henrissat B, Laroche J,**
578 **Levesque RC, Bernier L. 2015.** Functional annotation of the *Ophiostoma novo-ulmi*
579 genome: insights into the phytopathogenicity of the fungal agent of Dutch elm disease.
580 *Genome biology and evolution* **7:** 410-430.
581 **Corrêa LG, Riaño-Pachón DM, Schrago CG, dos Santos RV, Mueller-Roeber B,**
582 **Vincentz M. 2008.** The Role of bZIP Transcription Factors in Green Plant Evolution:
583 Adaptive Features Emerging from Four Founder Genes. *PLoS One* **3:** e2944.
584 **Crusoe MR, Alameldin HF, Awad S, Boucher E, Caldwell A, Cartwright R,**
585 **Charbonneau A, Constantinides B, Edvenson G, Fay S *et al*. 2015.** The khmer
586 software package: enabling efficient nucleotide sequence analysis. *F1000Research*
587 **4:** 900.
588 **Cuomo CA, Güldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma LJ,**
589 **Baker SE, Rep M *et al.* 2007.** The Fusarium graminearum genome reveals a link
590 between localized polymorphism and pathogen specialization. *Science* **317:** 1400-1402.
591 **de Man TJ, Stajich JE, Kubicek CP, Teiling C, Chenthamara K, Atanasova L,**
592 **Druzhinina IS, Levenkova N, Birnbaum SS, Barribeau SM *et al.* 2016**. Small genome
593 of the fungus *Escovopsis weberi*, a specialized disease agent of ant agriculture.
594 *Proceedings of the National Academy of Sciences, USA* **113:** 3567-3572.
595 **DiGuistini S, Wang Y, Liao NY, Taylor G, Tanguay P, Feau N, Henrissat B, Chan SK,**
596 **Hesse-Orce U, Alamouti SM *et al*. 2011.** Genome and transcriptome analyses of the
597 mountain pine beetle-fungal symbiont *Grosmannia clavigera*, a lodgepole pine pathogen.
598 *Proceedings of the National Academy of Sciences, USA* **108:** 2504-2509.
599 **Duyvesteijn RG, Van Wijk R, Boer Y, Rep M, Cornelissen BJ, Haring MA. 2005.** Frp1 is
600 a *Fusarium oxysporum* F□box protein required for pathogenicity on tomato. *Molecular*
601 *microbiology* **57:** 1051-1063.
602 **Eddy SR. 1998.** Profile hidden Markov models. *Bioinformatics* **14:** 755-63.

603 **Emms DM, Kelly S. 2015.** OrthoFinder: solving fundamental biases in whole genome
604 comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16:**
605 2-14.
606 **Fitzpatrick DA, Logue ME, Stajich JE, Butler G. 2006**. A fungal phylogeny based on 42
607 complete genomes derived from supertree and combined gene analysis. *BMC*
608 *Evolutionary Biology* **6:** 99.
609 **Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma**
610 **LJ, Smirnov S, Purcell S *et al.* 2003.** The genome sequence of the filamentous fungus
611 *Neurospora crassa*. *Nature* **422:** 859-868.
612 **Guo M, Chen Y, Du Y, Dong Y, Guo W, Zhai S, Zhang H, Dong S, Zhang Z, Wang Y *et***
613 ***al.* 2011.** The bZIP Transcription Factor MoAP1 Mediates the Oxidative Stress
614 Response and Is Critical for Pathogenicity of the Rice Blast Fungus *Magnaporthe*
615 *oryzae*. *PLoS Pathogens* **7:** e1001302.
616 **Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013.** QUAST: quality assessment tool for
617 genome assemblies. *Bioinformatics* **29:** 1072–1075.
618 **Han YK, Kim MD, Lee SH, Yun SH, Lee YW. 2007.** A novel F-box protein involved in
619 sexual development and pathogenesis in *Gibberella zeae*. *Molecular Microbiology* **63:**
620 768-779.
621 **Janusz G, Kucharzyk KH, Pawlik A, Staszczak M, Paszczynski AJ. 2013.** Fungal
622 laccase, manganese peroxidase and lignin peroxidase: Gene expression and regulation.
623 *Enzyme and Microbial Technology* **52:** 1-12.
624 **Kabbage M, Williams B, Dickman MB. 2013.** Cell death control: the interplay of apoptosis
625 and autophagy in the pathogenicity of *Sclerotinia sclerotiorum*. *PLoS Pathogens* **9:**
626 e1003287.
627 **Karlsson M, Durling MB, Choi J, Kosawang C, Lackner G, Tzelepis GD, Nygren K,**
628 **Dubey MK, Kamou N, Levasseur A, Zapparata A. 2015.** Insights on the evolution of
629 mycoparasitism from the genome of *Clonostachys rosea*. *Genome biology and*
630 *evolution*. **7:** 465-480.
631 **Katoh K, Standley DM. 2013.** MAFFT Multiple Sequence Alignment Software Version 7:
632 Improvements in Performance and Usability. *Molecular Biology Evolution* **30:** 772-780.
633 **Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M,**
634 **Davis P, Falin LJ, Grabmueller C *et al.* 2016.** Ensembl Genomes 2016: more
635 genomes, more complexity. *Nucleic acids research* **44:** D574-80. [WWW document]
636 URL http://fungi.ensembl.org/index.html. [accessed 14 November 2015].

637 **Khoshraftar S, Hung S, Khan S, Gong Y, Tyagi V, Parkinson J, Sain M, Moses AM,**
638 **Christendat D. 2013.** Sequencing and annotation of the *Ophiostoma ulmi genome*.
639 *BMC genomics* **14:** 162.
640 **Kohler A, Francis M, Costa M. 2011.** High quality genomic DNA extraction using CTAB
641 and Qiagen genomic-tip (version 2). [WWW document] URL
642 http://1000.fungalgenomes.org/home/wp-
643 content/uploads/2013/02/genomicDNAProtocol-AK0511.pdf [accessed 12 December
644 2015].
645 **Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canbäck B, Choi C, Cichocki**
646 **N, Clum A *et al.* 2015.** Convergent losses of decay mechanisms and rapid turnover of
647 symbiosis genes in mycorrhizal mutualists. *Nature Genetics* **47:** 410-415.
648 **Kolarik M, Freeland E, Utley C, Tisserat N. 2011.** *Geosmithia morbida* sp. nov., a new
649 phytopathogenic species living in symbiosis with the walnut twig beetle (*Pityophthorus*
650 *juglandis*) on Juglans in USA. *Mycologia* **103:** 325–332.
651 **Kolarik M, Jankowiak R. 2013.** Vector Affinity and Diversity of Geosmithia Fungi Living on
652 Subcortical Insects Inhabiting *Pinaceae* Species in Central and Northeastern Europe.
653 *Microbial Ecology* **66:** 682–700.
654 **Kolarik M, Kirkendall LR. 2010.** Evidence for a new lineage of primary ambrosia fungi in
655 *Geosmithia Pitt* (Ascomycota: Hypocreales). *Fungal Biology* **114:** 676–689.
656 **Kolarik M, Kostovcik M, Pazoutova S. 2007.** Host range and diversity of the genus
657 *Geosmithia* (Ascomycota: Hypocreales) living in association with bark beetles in the
658 Mediterranean area. *Mycological Research* **111:** 1298–1310.
659 **Kolarik M, Kubatova A, van Cepicka I, Pazoutova S, Srutka P. 2005.** A complex of three
660 new white-spored, sympatric, and host range limited *Geosmithia* species. *Mycological*
661 *Research* **109:**1323–1336.
662 **Korf I. 2004.** Gene finding in novel genomes. *BMC bioinformatics* **5:** 1.
663 **Krattinger SG, Lagudah ES, Spielmeyer W, Singh RP, Huerta-Espino J, McFadden H,**
664 **Bossolini E, Selter LL, Keller B. 2009.** A putative ABC transporter confers durable
665 resistance to multiple fungal pathogens in wheat. *Science* **323:** 1360-1363.
666 **Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001.** Predicting transmembrane
667 protein topology with a hidden Markov model: Application to complete genomes. *Journal*
668 *of Molecular Biology* **305:** 567-580.
669 **Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez DA, Druzhinina IS, Thon M,**
670 **Zeilinger S, Casas-Flores S, Horwitz BA, Mukherjee PK, Mukherjee M. 2011.**

671 Comparative genome sequence analysis underscores mycoparasitism as the ancestral
672 life style of *Trichoderma*. *Genome Biology* **12:** R40.
673 **Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A. 2014.** Selecting optimal
674 partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology* **14:** 82.
675 **Lawrence JG. 2005.** Common themes in the genome strategies of pathogens. *Current*
676 *opinion in genetics & development* **15:** 584-588.
677 **Li H, Durbin R. 2009.** Fast and accurate short read alignment with Burrows-Wheeler
678 transform. *Bioinformatics* **25:** 1754–1760.
679 **Li H. 2015.** BFC: correcting Illumina sequencing errors. *Bioinformatics* **31:** 2885-2887.
680 **Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A,**
681 **Reissmann S, Kahmann R. 2015.** Fungal effectors and plant susceptibility. *Annual*
682 *review of plant biology* **66:** 513-545.
683 **Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The**
684 **Carbohydrate-active enzymes database (CAZy) in 2013.** *Nucleic Acids Research* **42:**
685 D490-495.
686 **Lynch SC, Wang DH, Mayorquin JS, Rugman-Jones PF, Stouthamer R, Eskalen E.**
687 **2014.** First Report of *Geosmithia pallida* Causing Foamy Bark Canker, a new disease on
688 coast live oak (*Quercus agrifolia*), in association with *Pseudopityophthorus pubipennis* in
689 California. *Plant Disease* **98:** 1276.
690 **Ma LJ, Van Der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A,**
691 **Dufresne M, Freitag M, Grabherr M, Henrissat B *et al.* 2010.** Comparative genomics
692 reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464:** 367-373.
693 **Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J,**
694 **Chertkov O, Coutinho PM, Cullen D *et al.* 2008.** Genome sequencing and analysis of
695 the biomass-degrading fungus *Trichoderma reesei* (syn. Hypocrea jecorina). *Nature*
696 *Biotechnology* **26:** 553-560.
697 **Misof B, Meyer B, von Reumont BM, Kuck P, Misof K, Meusemann K. 2013.** Selecting
698 informative subsets of sparse supermatrices increases the chance to find correct trees.
699 *BMC Bioinformatics* **14:** 348.
700 **Nagendran S, Hallen-Adams HE, Paper JM, Aslam N, Walton JD. 2009.** Reduced
701 genomic potential for secreted plant cell-wall-degrading enzymes in the ectomycorrhizal
702 fungus *Amanita bisporigera*, based on the secretome of *Trichoderma reesei*. *Fungal*
703 *Genetics and Biology* **46:** 427-435.
704 **Ochman H, Moran NA. 2001.** Genes lost and genes found: evolution of bacterial

pathogenesis and symbiosis. *Science* **292:** 1096-1099.

Ott T, Fritz E, Polle A, Schützendübel A. 2002. Characterisation of antioxidative systems in the ectomycorrhiza-building basidiomycete *Paxillus involutus* (Bartsch) Fr. and its reaction to cadmium. *FEMS microbiology ecology* **42:** 359-66.

Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23:** 1061-1067.

Peterson TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8:** 785-786.

Pintard L, Willems A, Peter M. 2004. Cullin□based ubiquitin ligases: Cul3–BTB complexes join the family. *EMBO journal* **23:** 1681-1687.

Pitt JI. 1979. *Geosmithia*, gen. nov. for *Penicillium lavendulum* and related species. *Canadian Journal of Botany* **57:** 2021-2030.

Ploetz RC, Hulcr J, Wingfield MJ, De Beer ZW. 2013. Destructive tree diseases associated with ambrosia and bark beetles: black swan events in tree pathology? *Plant Disease* **97:** 856-872.

Pócsi I, Prade RA, Penninckx MJ. 2004. Glutathione, altruistic metabolite in fungi. *Advances in microbial physiology* **49:** 1-76.

Raffaele S, Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology* **10:** 417-430.

Ranwez V, Harispe S, Delsuc F, Douzery EJP. 2011. MACSE: Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons. *PLoS One* **6:** e22594.

Rawlings ND, Barrett AJ, Finn RD. 2016. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research* **44:** D343-D350.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61:** 539–542.

Rugman-Jones PF, Seybold SJ, Graves AD, Stouthamer R. 2015. Phylogeography of the walnut twig beetle, *Pityophthorus juglandis*, the vector of thousand cankers disease in North American walnut trees. *PLoS ONE* **10:** e118264.

Schuelke TA, Westbrook A, Broders K, Woeste K, MacManes MD. 2016. De novo genome assembly of *Geosmithia morbida*, the causal agent of thousand cankers disease. *PeerJ* **4:** e1952.

739 **Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015.** BUSCO:
740 assessing genome assembly and annotation completeness with single-copy orthologs.
741 *Bioinformatics* 1–3.
742 **Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. 2009.** ABySS: A
743 parallel assembler for short read sequence data. *Genome Research* **19:** 1117-1123.
744 **Smit AFA, Hubley R, Green P. 1996.** RepeatMasker. [WWW document] URL
745 http://www.repeatmasker.org.
746 **Solomon PS, Oliver RP. 2002.** Evidence that γ-aminobutyric acid is a major nitrogen
747 source during *Cladosporium fulvum* infection of tomato. *Planta* **214:** 414-420.
748 **Son H, Fu M, Lee Y, Lim JY, Min K, Kim JC, Choi GJ, Lee YW. 2016.** A novel
749 transcription factor gene FHS1 is involved in the DNA damage response in *Fusarium*
750 *graminearum*. *Scientific reports* **6:** 21572.
751 **Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, van Themaat EV,**
752 **Brown JK, Butcher SA, Gurr SJ *et al*. 2010.** Genome expansion and gene loss in
753 powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **330:** 1543-1546.
754 **Staats M, van Kan JA. 2012.** Genome update of *Botrytis cinerea* strains B05.10 and T4.
755 *Eukaryotic Cell* **11:** 1413-1414.
756 **Stamatakis A. 2014.** RAxML Version 8: A tool for Phylogenetic Analysis and Post-analysis
757 of Large Phylogenies. *Bioinformatics* **30:** 1312–1313.
758 **Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006.** AUGUSTUS:
759 ab initio prediction of alternative transcripts. *Nucleic acids research* **34:** W435-W439.
760 **Stukenbrock EH, Croll D. 2014.** The evolving fungal genome. *Fungal Biology Reviews* **28:**
761 1-2.
762 **Suyama M, Torrents D, Bork P. 2006.** PAL2NAL: robust conversion of protein sequence
763 alignments into the corresponding codon alignments. *Nucleic Acids Research* **34:** W609-
764 W612.
765 **Terfehr D, Dahlmann TA, Specht T, Zadra I, Kurnsteiner H, Kuck U. 2014.** Genome
766 Sequence and Annotation of *Acremonium chrysogenum*, Producer of the β-Lactam
767 Antibiotic Cephalosporin C. *Genome Announcements* **2:** e00948-14.
768 **Tisserat N, Cranshaw W, Leatherman D, Utley C, Alexander K. 2009.** Black walnut
769 mortality in colorado caused by the walnut twig beetle and thousand cankers disease.
770 *Plant Health Progress* 1–10.

771 **Trail F, Xu JR, San Miguel P, Halgren RG, Kistler HC. 2003.** Analysis of expressed
772 sequence tags from *Gibberella zeae* (anamorph *Fusarium graminearum*). *Fungal*
773 *Genetics and Biology* **38:** 187-197.
774 **Utley C, Nguyen T, Roubtsova T, Coggeshall M, Ford TM, Grauke LJ, Graves AD,**
775 **Leslie CA, McKenna J, Woeste K *et al*. 2013.** Susceptibility of walnut and hickory
776 species to *Geosmithia morbida*. *Plant Disease* **97:** 601–607.
777 **Veneault-Fourrey C, Barooah M, Egan M, Wakley G, Talbot NJ. 2006.** Autophagic fungal
778 cell death is necessary for infection by the rice blast fungus. *Science* **312:** 580-583.
779 **Wang H, Xu Z, Gao L, Hao B. 2009.** A fungal phylogeny based on 82 complete genomes
780 using the composition vector method. *BMC Evolutionary Biology* **9:** 195.
781 **Wang Y, Lim L, DiGuistini S, Robertson G, Bohlmann J, Breuil C. 2013.** A specialized
782 ABC efflux transporter GcABC□G1 confers monoterpene resistance to *Grosmannia*
783 *clavigera*, a bark beetle□associated fungal pathogen of pine trees. *New Phytologist*.
784 **197:** 886-898.
785 **Winnenburg R, Baldwin TK, Urban M, Rawlings C, Köhler J, Hammond-Kosack KE.**
786 **2006.** PHI-base: a new database for pathogen host interactions. *Nucleic acids research*.
787 **34:** D459-464. [WWW document] URL http://www.phi-base.org/. [accessed 22
788 November 2015].
789 **Yang Z. 2007.** PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology*
790 *Evolution* **24:** 1586-1591.
791 **Zerillo MM, Caballero JI, Woeste K, Graves AD, Hartel C, Pscheidt JW, Tonos J,**
792 **Broders K, Cranshaw W, Seybold SJ *et al.* 2014.** Population structure of *Geosmithia*
793 *morbida*, the causal agent of thousand cankers disease of walnut trees in the United
794 States. *PLoS ONE* **9:** e112847.
795 **Zhang Y, Zhang K, Fang A, Han Y, Yang J, Xue M, Bao J, Hu D, Zhou B, Sun X *et al*.**
796 **2014.** Specific adaptation of *Ustilaginoidea virens* in occupying host florets revealed by
797 comparative and functional genomics. *Nature communications* **5:** 3849
798 **Zhao Z, Liu H, Wang C, Xu J. 2013.** Erratum to: Comparative analysis of fungal genomes
799 reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* 15: 6.
800 **Zheng P, Xia Y, Xiao G, Xiong C, Hu X, Zhang S, Zheng H, Huang Y, Zhou Y, Wang S *et***
801 ***al.* 2011.** Genome sequence of the insect pathogenic fungus *Cordyceps militaris*, a
802 valued traditional Chinese medicine. *Genome Biology* **12:** R116.
803
804

Figure legends

Figure 1 Carbohydrate active enzymes (CAZymes) distribution for *Geosmithia* species, other Hypocreales, and *Ceratocystis platani*. The species in red are pathogens, while the names in black are nonpathogens. CAZymes were identified with HMMer searches of dbCAN peptide models. GH: glycoside hydrolases, GT: glycosyltransferases, PL: polysaccharide lyases, CE: carbohydrate esterases, AA: auxiliary activities enzymes, and CBM: carbohydrate-binding molecules.

Figure 2 Proteolytic enzymes distribution for *Geosmithia* species, other Hypocreales, and *Ceratocystis platani*. The species in red are pathogens, while the names in black are nonpathogens. Proteases were identified using BLASTp searches against the MEROPs database v10. S: serine, M: metallo, C: cysteine, A: aspartic, T:threonine, I: inhibitors, P: mixed, G: glutamic.

Figure 3 The Bayesian Markov Chain Monte Carlo (BMCMC) phylogeny was estimated using the mixed amino acid model in MrBayes (Ronquist *et al.*, 2012) on a dataset containing 89,999 positions. This topology is identical to partitioned analyses conducted in RAxML (Stamatakis 2014). All nodes in BMCMC and ML analyses receive maximum support. The black circles symbolize classes. The color-shaded boxes at the right of the figure denote the orders within each class. The first and second numbers in parentheses represent the genome sizes in Mbp and the number of predict protein models, respectively. Black and red branches correspond to non-pathogens and pathogens, respectively, which span multiple orders.

Support Information

The following Supporting Information is available for this article:

Table S1 Complete BLASTp results against the Phibase4 database for *Geosmithia morbida, Geosmithia flava* and *Geosmithia putterillii.*

Table S2 CAZymes hits data for 20 fungal species used in this study.

Table S3 Merops hits data for 20 fungal species used in this study

Table S4 Complete BLASTp results against NCBI's nr database for 38 genes found to be under selection in *Geosmithia morbida*.

Methods S1 A modified nexus file illustrating Mr. Bayes block used for constructing phylogeny.

Table 1 Species, geographic origins, and host information for *Geosmithia morbida*, *Geosmithia flava*, and *Geosmithia putterillii*.

Species	Pathogen	Isolate	Geographic origins	Host
*G. morbida**	Yes	1262	California	*Juglans californica*
G. flava	No	CCF3333	Czech Republic	*Castenea sativa*
G. putterillii	No	CCF4204	California	*Juglans californica*

*This isolate is the reference genome. The details of assembly for this genome are discussed in Schuelke *et al.*, 2016.

Table 2 Statistics for sequence data from isolates of *Geosmithia morbida*, *Geosmithia flava* and *Geosmithia putterillii*.

Species	Total read pairs		Est. coverage	
G. morbida	14,013,863*	20,674,289**	109x*	160x**
G. flava	16,183,281		102x	
G. putterillii	19,711,745		131x	

*These values are for paired-end read data for *G. morbida* from Schuelke *et al.*, 2016. **These values are for mate-pair read data for *G. morbida* from Schuelke *et al.*, 2016.

Table 3 Fungal species used for phylogenetic analysis in this study. The species in bold were utilized for positive selection analysis.

Species	Class	Order	Ecological role	Download source	References
Geosmithia morbida	Sordariomycetes	Hypocreales	Pathogen	-	Schuelke *et al.*, 2016
Geosmithia flava	Sordariomycetes	Hypocreales	Non-pathogen	-	-
Geosmithia putterillii	Sordariomycetes	Hypocreales	Non-pathogen	-	-
Acremonium chrysogenum	Sordariomycetes	Hypocreales	Beneficial	FungalEnsembl	Terfehr *et al.*, 2014
Stanjemonium griseum	Sordariomycetes	Hypocreales	Saprotrophic	JGI	Used with permission
Trichoderma virens	Sordariomycetes	Hypocreales	Mycoparasite	JGI	Kubicek *et al.*, 2011
Trichoderma reesei	Sordariomycetes	Hypocreales	Saprotrophic	FungalEnsembl	Martinez *et al.*, 2008
Escovopsis weberi	Sordariomycetes	Hypocreales	Mycoparasite	EnsemblGenomes	de Man *et al.*, 2016
Ustilaginoidea virens	Sordariomycetes	Hypocreales	Biotrophic pathogen	FungalEnsembl	Zhang *et al.*, 2014
Cordyceps militaris	Sordariomycetes	Hypocreales	Insect pathogen	FungalEnsembl	Zheng *et al.*, 2011
Myrothecium inundatum	Sordariomycetes	Hypocreales	Saprotrophic	JGI	Used with permission
Fusarium solani	Sordariomycetes	Hypocreales	Necrotrophic pathogen	FungalEnsembl	Coleman *et al.*, 2009
Fusarium graminearum	Sordariomycetes	Hypocreales	Necrotrophic pathogen	FungalEnsembl	Trail *et al.*, 2003; Cuomo *et al.*, 2007; Ma *et al.*, 2010
Ceratocystis platani	Sordariomycetes	Microascales	Pathogen	FungalEnsembl	Belbahri 2015
Neurospora crassa	Sordariomycetes	Sordariales	Saprotrophic	FungalEnsembl	Galagan *et al.*, 2003
Chaetomium globosum	Sordariomycetes	Sordariales	Saprotrophic	JGI	Berka *et al.*, 2011
Grosmannia clavigera	Sordariomycetes	Ophiostomatales	Pathogen	FungalEnsembl	DiGuistini *et al.*, 2011
Eutypa lata	Sordariomycetes	Xylariales	Pathogen	JGI	Blanco-Ulate *et al.*, 2013

Botrytis cinerea	Leotiomycetes	Helotiales	Necrotrophic pathogen	FungalEmsebl	Amselem *et al.*, 2011, Staats & van Kan 2012
Oidiodendron maius	Leotiomycetes	Incertae sedis	Mycorrhizal	JGI	Kohler *et al.*, 2015

Table 4 Length-based statistics for *Geosmithia morbida*, *Geosmithia flava*, and *Geosmithia putterillii* generated with QUAST v2.3. The average GC content for *G. morbida*, *G. flava*, and *G. putterillii* equals 54%, 52%, and 55.5% respectively. All genome completeness values were produced with BUSCO v1.1b1. These percentages represent genes that are complete and not duplicated or fragmented.

Species	Est. genome size (Mbp)	*k*-mer for ABySS assembly	Scaffold count	Largest scaffold	NG50*	LG50*	Genome completeness	Predicted Proteins
G. morbida	26.5	NA[1]	73	2,597,956	1,305,468	7	98	6,273
G. flava	29.6	91	1,819	1,534,325	460,430	22	98	6,976
G. putterillii	30.0	91	320	2,758,267	1,379,352	9	98	7,086

[1]Genome assembly for *G. morbida* was constructed using AllPaths-LG (v49414). See Schuelke *et al.*, 2016 for further details. *NG50 is the scaffold length such that considering scaffolds of equal or longer length produce 50% of the bases of the reference genome. LG50 is the number of scaffolds with length NG50.

	Genome size (Mbp)	GC (%)	Bases masked (%)	Retroelements (N)	DNA transposons (N)
G. morbida	26.5	54.30	0.81	152	60
G. flava	29.6	51.87	0.63	401	42
G. putterillii	30.0	55.47	0.64	214	15

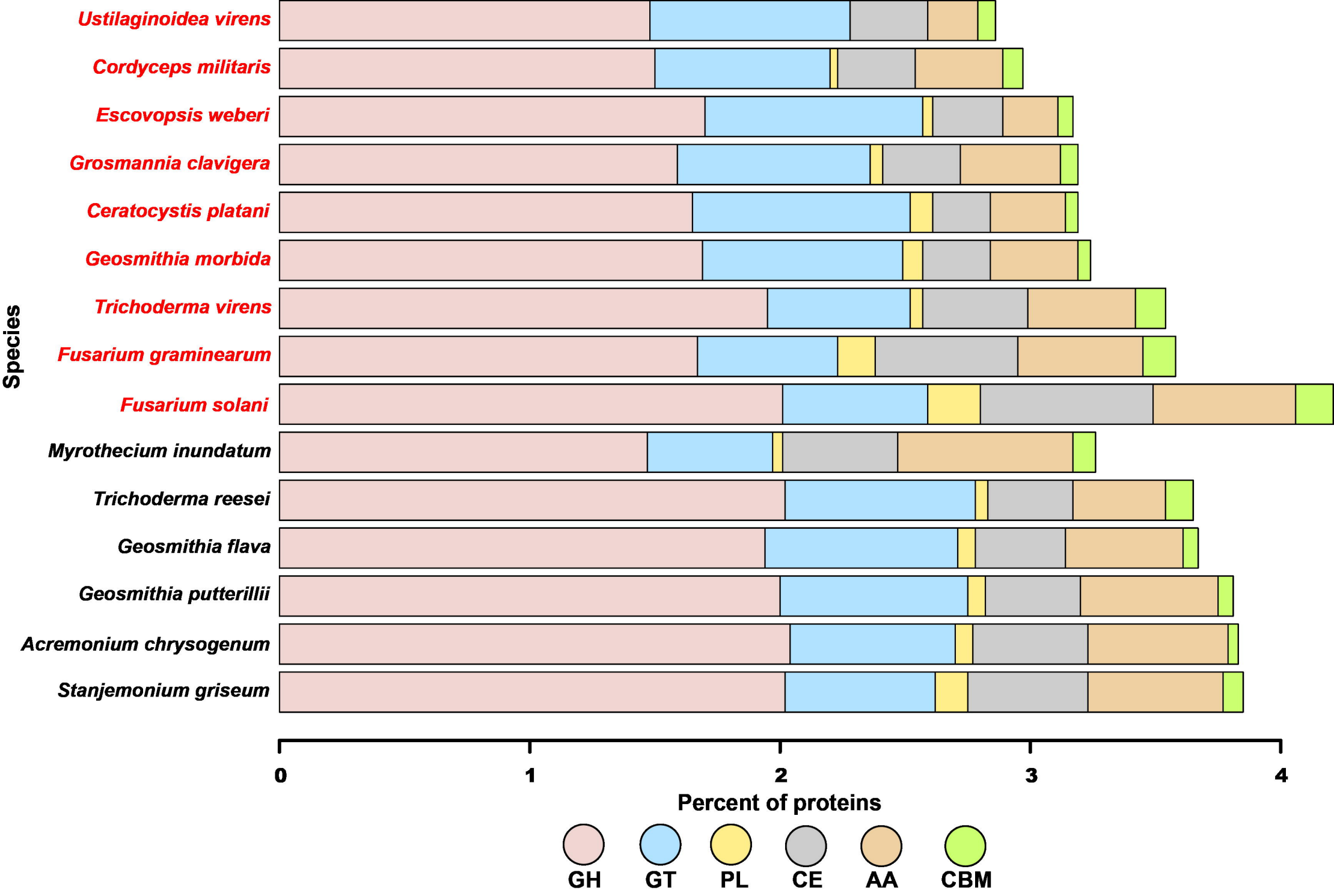
Table 5 Repetitive elements profile of *Geosmithia* species generated with RepeatMasker v4.0.5.

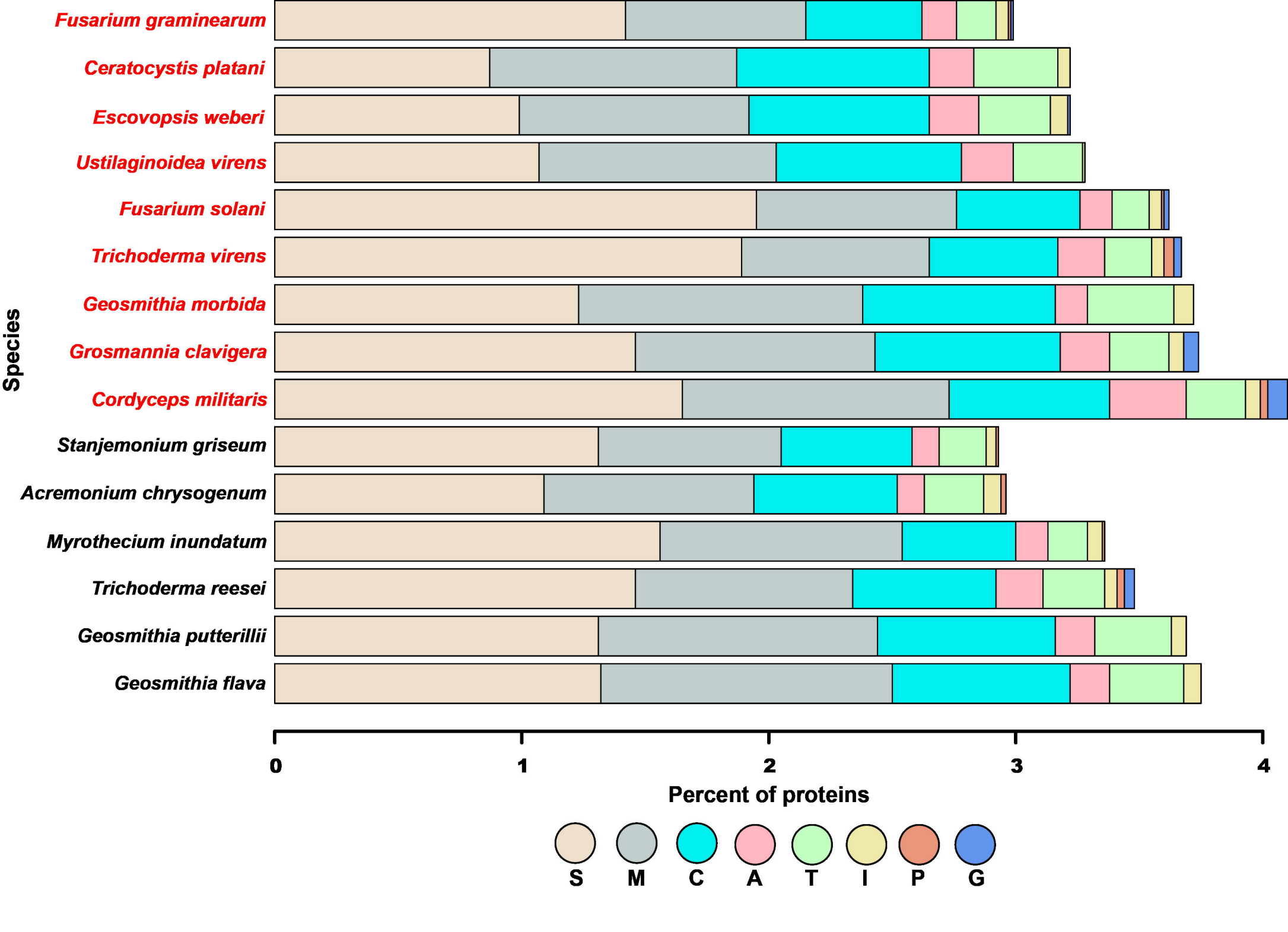
Table 6 Functional analyses of genes under positive selection in *Geosmithia morbida* detected by the branch-site model in PAML 4.8. The gene number corresponds to the sequence ID in the *G. morbida* protein file available at DRYAD.

Gene number	Function	dn/ds	Transmembrane domain (N)
3078	Takes part in intracellular signaling, protein recruitment to various membranes	2.04	0
2666	Involved in receptor-mediated endocytosis and vesicle trafficking	2.01	0
563	Unclear function	1.94	1
2194	Unknown function	1.94	0
801	Catalyzes the transfer of electrons from ferrocytochrome c to oxygen converting the cytochrome c into water	1.93	1
3944	Involved in methylation and have a wide range of substrate specificity	1.90	5
5058	Involved in ubiquitination of proteins target for degradation	1.90	0
1843	Involved in heat-shock response	1.86	0
521	Involved in damage DNA binding and repair	1.85	0
5111	Involved in receptor-mediated endocytosis and vesicle trafficking	1.84	0
4128	Catalyzes the hydrolysis of esters	1.84	0
923	Hydrolases the peptide bond at the C-terminus of ubiquitin	1.83	1
4405	Involved in transport and metabolism of lipids	1.83	1
3137	Part of proteins with diverse functions such as cell-cycle regulators, signal transducers, transcriptional initiators	1.78	0
4359	Unknown function	1.73	2
5639	Involved in rRNA synthesis	1.67	0
5	Involved in vesicular transport	1.63	0
624	Involved in transfer of glucose molecules that are part of a larger glycosylation machinery	1.62	9
3929	Unknown function but associates with GRAM domain found in glucosyltransferases and other membrane affiliated proteins	1.61	0
1456	Involved in DNA repair and replication	1.59	0
4829	Form cAMP	1.59	0
254	Major ATP transporters	1.59	2
4888	Unknown function	1.54	0
5426	Hydrolyzes nonubiquitinated peptides	1.54	0
5709	Transcription factors	1.50	0
859	May be involved in the timing of nuclear migration	1.50	0

5703	Cleave peptide bonds in other proteins	1.47	6
5255	Heat shock protein involved in induced stress response to ethanol	1.46	3
5704	Regulates gene expression during oxidative stress caused by the host plant	1.46	0
2485	Transfer phosphates	1.39	0
6116	Hydratase and/or isomerase	1.38	0
5266	Breaks down actin, cell membrane deformations	1.34	0
5000	Catalyzes the first step in histidine biosynthesis	1.34	0
3326	Involved in de novo synthesis of nucleotide purine	1.32	0
2142	E2 enzymes that catalyze the binding of activated ubiquitin to the substrate protein. The substrate proteins are targeted for degradation by the proteasome	1.24	0
581	Ribosomal protein	1.17	0
5948	Involved in initiation of transcription	1.14	1
3700	Part of the TOM complex that recognizes and regulates the transport of mitochondrial precursor molecules from the cytosol to the intracellular space of the mitochondrion	1.03	0

dn/ds is the ratio of nonsynonymous substitutions to synonymous changes.


Species
Ustilaginoidea virens
Cordyceps militaris
Escovopsis weberi
Grosmannia clavigera
Ceratocystis platani
Geosmithia morbida
Trichoderma virens
Fusarium graminearum
Fusarium solani
Myrothecium inundatum
Trichoderma reesei
Geosmithia flava
Geosmithia putterillii
Acremonium chrysogenum
Stanjemonium griseum
0
1
2
3
4
Percent of proteins
GH
GT
PL
CE
AA
CBM


Fusarium graminearum
Ceratocystis platani
Escovopsis weberi
Ustilaginoidea virens
Fusarium solani
Trichoderma virens
Geosmithia morbida
Grosmannia clavigera
Cordyceps militaris
Stanjemonium griseum
Acremonium chrysogenum
Myrothecium inundatum
Trichoderma reesei
Geosmithia putterillii
Geosmithia flava
Species
0
1
2
3
4
Percent of proteins
S
M
C
A
T
I
P
G

Botrytis cinerea (42.6 Mbp, 12103)

Oidiodendron maius (46.2 Mbp, 16703)

Leotiomycetes

Xylariales

Eutypa lata (53.8 Mbp, 11685)

Sordariomycetes

Ophiostomatales

Grosmannia clavigera (29.1 Mbp, 8312)

Sordariales

Chaetomium globosum (34.3 Mbp, 11124)

Neurospora crassa (39.2 Mbp, 9839)

Microascales

Ceratocystis platani (29.1 Mbp, 5622)

Hypocreales

Fusarium graminearum (38.1 Mbp, 14164)

Fusarium solani (51.2 Mbp, 15705)

Stanjemonium griseum (31.9 Mbp, 11832)

Acremonium chrysogenum (28.6 Mbp, 8901)

0.06

Geosmithia putterillii (30.0 Mbp, 7086)

Geosmithia flava (29.5 Mbp, 6976)

Geosmithia morbida (26.0 Mbp, 6273)

Myrothecium inundatum (38.8 Mbp, 13553)

Cordyceps militaris (32.2 Mbp, 9651)

Ustilaginoidea virens (30.2 Mbp, 7110)

Escovopsis weberi (26.7 Mbp, 6870)

Trichoderma reesei (33.4 Mbp, 9115)

Trichoderma virens (39.0 Mbp, 12427)