WESTERDIJK FUNGAL BIO DIVERSITY INSTITUTE

# 101 *Dothideomycetes* genomes: A test case for predicting lifestyles and emergence of pathogens

S. Haridas[1], R. Albert[1,2], M. Binder[3], J. Bloem[3], K. LaButti[1], A. Salamov[1], B. Andreopoulos[1], S.E. Baker[4], K. Barry[1], G. Bills[5], B.H. Bluhm[6], C. Cannon[7], R. Castanera[1,8,20], D.E. Culley[4], C. Daum[1], D. Ezra[9], J.B. González[10], B. Henrissat[11,12,13], A. Kuo[1], C. Liang[14,21], A. Lipzen[1], F. Lutzoni[15], J. Magnuson[4], S.J. Mondo[1,16], M. Nolan[1], R.A. Ohm[1,17], J. Pangilinan[1], H.-J. Park[10], L. Ramírez[8], M. Alfaro[8], H. Sun[1], A. Tritt[1], Y. Yoshinaga[1], L.-H. Zwiers[3], B.G. Turgeon[10], S.B. Goodwin[18], J.W. Spatafora[19], P.W. Crous[3,17*], and I.V. Grigoriev[1,2*]

[1]*US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA;* [2]*Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA, USA;* [3]*Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands;* [4]*Functional and Systems Biology Group, Environmental Molecular Sciences Division, Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, Washington, USA;* [5]*University of Texas Health Science Center, Houston, TX, USA;* [6]*University of Arkansas, Fayelletville, AR, USA;* [7]*Texas Tech University, Lubbock, TX, USA;* [8]*Institute for Multidisciplinary Research in Applied Biology (IMAB-UPNA), Universidad Pública de Navarra, Pamplona, Navarra, Spain;* [9]*Agricultural Research Organization, Volcani Center, Rishon LeTsiyon, Israel;* [10]*Section of Plant Pathology and Plant-Microbe Biology, School of Integrative Plant Science, Cornell University, Ithaca, NY, USA;* [11]*CNRS, Aix-Marseille Université, Marseille, France;* [12]*INRA, Marseille, France;* [13]*Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia;* [14]*College of Agronomy and Plant Protection, Qingdao Agricultural University, China;* [15]*Department of Biology, Duke University, Durham, NC, USA;* [16]*Bioagricultural Science and Pest Management Department, Colorado State University, Fort Collins, CO, USA;* [17]*Microbiology, Department of Biology, Faculty of Science, Utrecht University, Utrecht, The Netherlands;* [18]*U.S. Department of Agriculture-Agricultural Research Service, 915 W. State Street, West Lafayette, IN, USA;* [19]*Department of Botany & Plant Pathology, Oregon State University, Oregon State University, Corvallis, OR, USA*

*Correspondence: P.W. Crous, p.crous@wi.knaw.nl; I.V. Grigoriev, ivgrigoriev@lbl.gov

[20]Current Address: Centre for Research in Agricultural Genomics CSIC-IRTA-UAB-UB, Campus UAB, Edifici CRAG, Bellaterra, 08193 Barcelona, Spain.

[21]Current Address: College of Plant Health and Medicine, Qingdao Agricultural University, Qingdao, Shandong, China.

**Abstract:** *Dothideomycetes* is the largest class of kingdom Fungi and comprises an incredible diversity of lifestyles, many of which have evolved multiple times. Plant pathogens represent a major ecological niche of the class *Dothideomycetes* and they are known to infect most major food crops and feedstocks for biomass and biofuel production. Studying the ecology and evolution of *Dothideomycetes* has significant implications for our fundamental understanding of fungal evolution, their adaptation to stress and host specificity, and practical implications with regard to the effects of climate change and on the food, feed, and livestock elements of the agro-economy. In this study, we present the first large-scale, whole-genome comparison of 101 *Dothideomycetes* introducing 55 newly sequenced species. The availability of whole-genome data produced a high-confidence phylogeny leading to reclassification of 25 organisms, provided a clearer picture of the relationships among the various families, and indicated that pathogenicity evolved multiple times within this class. We also identified gene family expansions and contractions across the *Dothideomycetes* phylogeny linked to ecological niches providing insights into genome evolution and adaptation across this group. Using machine-learning methods we classified fungi into lifestyle classes with >95 % accuracy and identified a small number of gene families that positively correlated with these distinctions. This can become a valuable tool for genome-based prediction of species lifestyle, especially for rarely seen and poorly studied species.

**Key words:** Fungal evolution, Genome-based prediction, Machine-learning, New taxa.

**Taxonomic novelties: New orders:** *Aulographales* Crous, Spatafora, Haridas & Grigoriev, *Coniosporiales* Crous, Spatafora, Haridas & Grigoriev, *Eremomycetales* Crous, Spatafora, Haridas & Grigoriev, *Lineolatales* Crous, Spatafora, Haridas & Grigoriev; **New families:** *Coniosporiaceae* Crous, Spatafora, Haridas & Grigoriev, *Lineolataceae* Crous, Spatafora, Haridas & Grigoriev, *Rhizodiscinaceae* Crous, Spatafora, Haridas & Grigoriev.

Available online 1 February 2020; https://doi.org/10.1016/j.simyco.2020.01.003.

## INTRODUCTION

*Dothideomycetes* represent the largest and one of the most important classes of ascomycete fungi, currently encompassing more than 23 orders, 110 families, 1 261 genera and 19 000 species (Wijayawardene *et al.* 2017). Numerous species remain undescribed, however, and many well-known taxa actually represent species complexes waiting to be resolved, suggesting that there are likely still thousands of species to be discovered (Lücking & Hawksworth 2018). *Dothideomycetes* are cosmopolitan fungi of diverse lifestyles, which can associate with a wide range of hosts/substrates. The class *Dothideomycetes* is well supported as sister to *Arthroniomycetes*, a class of 1 500-plus lichenized species (Wijayawardene *et al.* 2017) within the subphylum *Pezizomycotina*, which includes most filamentous ascomycetes. *Dothideomycetes* includes two subclasses, *Pleosporomycetidae* and *Dothideomycetidae*, based on phylogeny and the presence or absence of pseudoparaphyses (Schoch *et al.* 2006, 2009). Members of the order *Capnodiales* (*Dothideomycetidae*) seem to have adapted to (hemi-)biotrophy, while *Pleosporales* (*Pleosporomycetidae*) species tend to have adapted to necrotrophy. A previous study of the *Dothideomycetes* by Ohm *et al.* (2012) showed that *Capnodiales* have fewer genes involved in carbohydrate degradation, proteolysis and secondary metabolism. Due to limited sampling, however, it is not known if this holds true for all orders in each subclass, or if further evolutionary specialization has taken place to coincide with their differing ecologies.

Peer review under responsibility of Westerdijk Fungal Biodiversity Institute.





Studies in Mycology

Members of *Dothideomycetes* are known to infect most major crops of mono- and dicotyledons, including species that serve as staple food crops, biomass, and biofuel production (Ohm *et al.* 2012, Condon *et al.* 2013). Well-known examples include southern corn leaf blight caused by *Bipolaris maydis* (Turgeon & Baker 2007, Marin-Felix *et al.* 2017), the Sigatoka complex of banana by *Pseudocercospora* spp. (Chang *et al.* 2016), and multiple wheat pathogens including *Parastagonospora nodorum* (Quaedvlieg *et al.* 2013), *Zymoseptoria tritici* (Stukenbrock *et al.* 2012), and *Pyrenophora tritici-repentis* (Ellwood *et al.* 2012). Diseases of woody hosts caused by *Dothideomycetes* include Sphaerulina canker of poplar by *Sphaerulina* spp. (Foster *et al.* 2015), stem cankers by species of *Botryosphaeria*, *Diplodia* and *Neofusicoccum* (Phillips *et al.* 2013), apple scab by *Venturia inaequalis* (Marin-Felix *et al.* 2017, Citrus Black Spot by *Phyllosticta citricarpa* (Guarnaccia *et al.* 2017), and Teratosphaeria leaf blight of *Eucalyptus* by *Teratosphaeria* spp. (Crous *et al.* 2019), to name a few.

In addition to numerous plant-pathogenic species, *Dothideomycetes* also include an incredible diversity of non-plant-pathogenic life styles, many of which have evolved multiple times. Members of the class are distributed globally, and have important roles as foliar endophytes (*Alternaria*, *Botryosphaeria, Plenodomus, Phyllosticta*), ectophytes (*Phaeothecoidiella, Schizothyrium, Uwebraunia*), and epiphytes (sooty molds) (*Capnodium*, *Scorias*). They can grow on other fungi as mycoparasites (*Dissoconium, Pseudoveronaea*), on scale insects (*Myriangium*), tolerate extreme climatic conditions, and on or in a diversity of substrates, from soil to rocks (*Cryomyces, Coniosporium, Friedmanniomyces*). The rock-inhabiting fungi are found in several unrelated lineages in *Dothideomycetes*, and could represent an ancestral trait (Ruibal *et al.* 2009). Several genera are also found in fresh water (*Jahnula, Massiosphaeria*) or salt water (*Halomassarina, Morosphaeriaceae*) environments, where they occur in the intertidal zones. A small number of species are lichenised and *Dothideomycetes* appears to have the highest number of transitions to the lichenised lifestyle within a class (Nelsen *et al.* 2009), as it occurs in three orders. *Dothideomycetes* also harbour several examples of mycoparasitism of lichens, especially in the *Phoma* complex (Lawrey *et al.* 2012, Valenzuela-Lopez *et al.* 2018).

Some taxa have been shown to be mycorrhizal, such as *Cenococcum geophilum* (Peter *et al.* 2016), which is one of the more commonly encountered ectomycorrhizae, especially in more xeric conditions. *Dothideomycetes* saprobes are commonly found in plant litter, or in more extreme ecological niches. For example, the black yeasts are highly tolerant to extreme temperatures, solar radiation and desiccation. Some species are also tolerant to ethanol vapour, such as *Baudoinia* spp. that grow in the vicinity of whiskey distilleries, or *Zasmidium cellare*, growing in wine cellars (Goodwin *et al.* 2016, Videira *et al.* 2017). Several members are also prominent pathogens of humans and animals, such as *Hortaea werneckii* causing *tinea nigra* of skin, *Piedraia hortae* on hair (De Hoog *et al.* 2000), or are allergenic, such as species of *Alternaria* and *Cladosporium* (Bensch *et al.* 2018).

Understanding the ecology and evolutionary biology of *Dothideomycetes* is therefore fundamental to interpreting fungal evolution and adaptation to abiotic and biotic stress and host specificity. With a phylogenetically interspersed mix of several lifestyles, including a large number of plant pathogens responsible for economic losses, a major outstanding question is whether there are identifiable genomic signatures that correlate with diverse dothideomycete ecological niches. Here, we test this hypothesis by first generating a robust phylogeny (which additionally allowed us to reclassify 25 *Dothideomycetes*) and predicting the ancestral state of *Dothideomycetes*, then explore relationships between genomic content and lifestyle. Using this approach, we were able to identify several gene family expansions and contractions consistent with changes in lifestyle across the *Dothideomycetes*. Furthermore, our large data set enabled the use of machine learning techniques and demonstrate that high quality classification of species can be achieved using a small number of easily identifiable genes. With increasing genomic data becoming available, the ideas presented here suggest that machine-learning techniques can be used to gain insights into ecological speciation and provides clues to several genes that may play a role in lifestyle switches across the fungal tree of life.

## MATERIALS AND METHODS

### Genome sequencing, assembly and annotation

We used 101 *Dothideomycetes* genomes and eight other *Pezizomycotina* genomes as outgroups representing the *Pezizomycetes*, *Eurotiomycetes*, *Leotiomycetes* and *Sordariomycetes* (Supplementary Table 1). The 55 newly sequenced *Dothideomycetes* in this study were assembled and annotated between January 2012 and November 2016. The rapid progress in sequencing and assembly techniques during this period is reflected in the various techniques used to assemble these genomes. Two primary sequencing technologies were used, namely Illumina for earlier genomes and Pacific Biosciences more recently. For the oldest, *Cercospora zeae-maydis*, the consensus assembly was from a mixture of Roche (454) pyrosequencing, Sanger fosmids, and Illumina data using Velvet (Zerbino & Birney 2008). The Illumina-based assemblies fall into two main classes: 1) minimal draft assemblies from single 270-bp insert size libraries using 2 × 150-bp paired-end reads assembled first with Velvet to create an in-silico long mate-pair library with insert sizes of 3 000 +/- 300 bp, which was then assembled together with the original Illumina library using AllPaths-LG (Gnerre *et al.* 2011); and 2) standard draft Illumina assembles where 2 × 100-bp paired-end reads from a 4-kb-long mate-pair library was used in addition to the reads from the 270-bp insert size libraries and assembled using AllPaths-LG. For PacBio sequencing, filtered subread data were assembled together with Falcon (https://github.com/PacificBiosciences/FALCON), improved with finisherSC (Lam *et al.* 2015), and polished with Quiver (https://github.com/PacificBiosciences/GenomicConsensus). A summary of technologies used and assembly methods is shown in Supplementary Table 5. Genomes were annotated using the JGI Annotation pipeline and made available via the JGI fungal genome portal MycoCosm (Grigoriev *et al.* 2014) https://mycocosm.jgi.doe.gov. An overview of the JGI Fungal Annotation Pipeline is available at https://mycocosm.jgi.doe.gov/programs/fungi/FungalGenomeAnnotationSOP.pdf. Functional annotation was performed as in the overview and has also been previously described (Ohm *et al.* 2012). Previously published genomes were functionally re-annotated using the JGI annotation pipeline, but no changes to gene models were made.

## Phylogenetic analyses

The phylogenetic tree is based on 738 single-copy orthologs identified in all 109 genomes (101 *Dothideomycetes* + eight outgroup) using OrthoMCL v. 2.0.9 (Li *et al.* 2003). The genes were aligned using Mafft v. 7.123b (Katoh 2002, Li *et al.* 2003) using the –auto option, concatenated and filtered using Gblocks v. 0.91b (Castresana 2000) using block options -b4 = 5 -b5 = h. To determine the optimal substitution model for tree building, we used PartitionFinder v2.1.1 (Lanfear *et al.* 2017) and analysed the alignment in blocks of 300, 500 and 1 000 bp with options –no-ml-tree and –rcluster-max 100. In all cases, we obtained aamodelpr = fixed(wag) and rates = invgamma as optimal parameters. The alignment was used as input to RAxML v. 8.2.2 (Stamatakis 2014) under the GAMMA model of rate heterogeneity, ML estimate of alpha-parameter, and WAG substitution matrix.

To infer ancestral ecological character states, ecologies of each species was coded based on known ecologies states and using FUNGuild (Nguyen *et al.* 2016). Known ecological states were coded either as multistate: acidophile, aquatic, human-associated, insect pathogen, lichen, mycoparasite, mycorrhiza, plant pathogen, and saprobe, or as binary: plant pathogen and non-plant pathogen. Ancestral ecological character states were reconstructed in Mesquite (http://www.mesquiteproject.org). Multistate ecological character states were analysed using an unordered parsimony model, while binary ecological character states were analysed using the Mk1 (Markov k-state 1 parameter model) likelihood reconstruction model. Similar analysis was also performed using BayesTraits analysis using maximum likelihood (http://www.evolution.rdg.ac.uk/SoftwareMain.html).

## Enrichment analysis

To identify functional annotation differences among the various lifestyles in the *Dothideomycetes*, the two-tailed Fisher exact test for functional annotation was performed using the scipy.stats package in Python with a p-value threshold of 0.05. In order to further limit false positives, annotations with fewer than 100 genes in background distribution were ignored.

## Machine learning

We used the Support Vector Machine method as implemented in Python's scikit-learn library (http://scikit-learn.org) to identify the most informative features (*e.g.*, OrthoMCL gene clusters, PFAM annotations) for differentiating between pathogen and saprobe genomes. The dataset was cleaned using the nearZeroVar function in R to reduce noise in the analysis. We used a subset of known pathogens and saprobes (Supplementary Table 4) to train and identify features that were differentially distributed between the pathogens and saprobes. Features were ranked on ability to predict using C-Support Vector Classification within scikit-learn (sklearn.svm.SVC). Strength of prediction was validated using sklearn.cross_validation and jackknife, and best features were selected using sklearn.feature_selection.SelectKBest. Features that were highly correlated (>0.8) to lifestyle were used in all combinations to improve prediction accuracy on the training set. The best-performing combinations of features were used on the entire study set to assess the accuracy of predictions.

# RESULTS AND DISCUSSION

## Whole-genome data enables reclassification of 25 species

Our study set included 55 newly sequenced *Dothideomycetes* genomes and 45 previously published genomes (Supplementary Table 1) in this class. Genomes were sequenced and assembled as described in Methods, then annotated using the JGI Annotation pipeline (Grigoriev *et al.* 2014). Using these 101 *Dothideomycetes* and 8 outgroups, we were able to generate a high confidence whole-genome-based phylogenetic tree (Fig. 1). The Gblocks filtered alignment of 738 single-copy orthologs consisted of 238 889 unique sites. The species tree provided strong support for the two subclasses, *Pleosporomycetidae* and *Dothideomycetidae*, proposed by Schoch *et al.* (2006) to accommodate the pseudoparaphysate *Pleosporales* (order *Pleosporales*) and the aparaphysate *Dothideales* (orders *Dothideales, Capnodiales* and *Myriangiales*) (Fig. 1).

In the detailed phylogenetic analyses (Figs 2 and 3), the *Pleosporomycetidae* is further expanded to include *Asterinales, Coniosporiales, Eremomycetales, Hysteriales, Lineolatales, Lizoniales, Microthyriales, Mytilinidiales, Patellariales, Phaeotrichales,* and *Venturiales*, while the *Trypetheliales* is added to the *Dothideomycetidae*, with all branches having 99 % or higher bootstrap support. The original division proposed by Schoch *et al.* (2006) still applies, with the newly placed orders listed above exhibiting these characters, except for the *Pleosporomycetidae* genera *Aulographum* (*Asterinales*), *Eremomyces* (*Eremomycetales*), and *Microthyrium* (*Microthyriales*) that lack pseudoparaphyses. The latter examples are, however, embedded in a subclade exhibiting pseudoparaphyses and represent secondary losses. All branches within the *Dothideomycetidae* tree had 100 % bootstrap support (Fig. 3) and indicated that *Acidomyces richmondensis* and *Hortaea acidophila* both belong in the family *Teratosphaeriaceae* of the *Capnodiales*. Based on this phylogenetic analysis, we have proposed new classifications for 25 species (Supplementary Table 2, Supplementary text) including four new orders (*Aulographales, Coniosporiales, Eremomycetales* and *Lineolatales*) and three new families (*Coniosporiaceae, Lineolataceae* and *Rhizodiscinaceae*).

## Saprobes show larger genomes and proteomes than plant pathogens

The *Dothideomycetes* genomes varied more than tenfold in size (from <17 Mbp to >177 Mbp) and encoded from 7 572 to 21 730 protein-coding genes, with neither genome size nor gene count correlated with the lifestyle (Supplementary Figs S1–S3) or phylogenetic position of the species (Fig. 4). However, pathogens generally showed smaller genomes as seen previously, including the recently sequenced species of *Geosmithia* where pathogenic species had smaller genomes than non-pathogenic *Geosmithia* species (Schuelke *et al.* 2017). *Piedraia hortae*, a non-pathogenic opportunist that infects human hair causing black hardened nodules, had the smallest genome assembly at 16.95 Mbp while the two root-associated fungi *Zopfia rhizophila*, an opportunistic pathogen, and *Cenococcum geophilum*, an ectomycorrhizal species, had the largest genomes at 152.8 Mbp and 177.6 Mbp, respectively. Repeat content also varied greatly

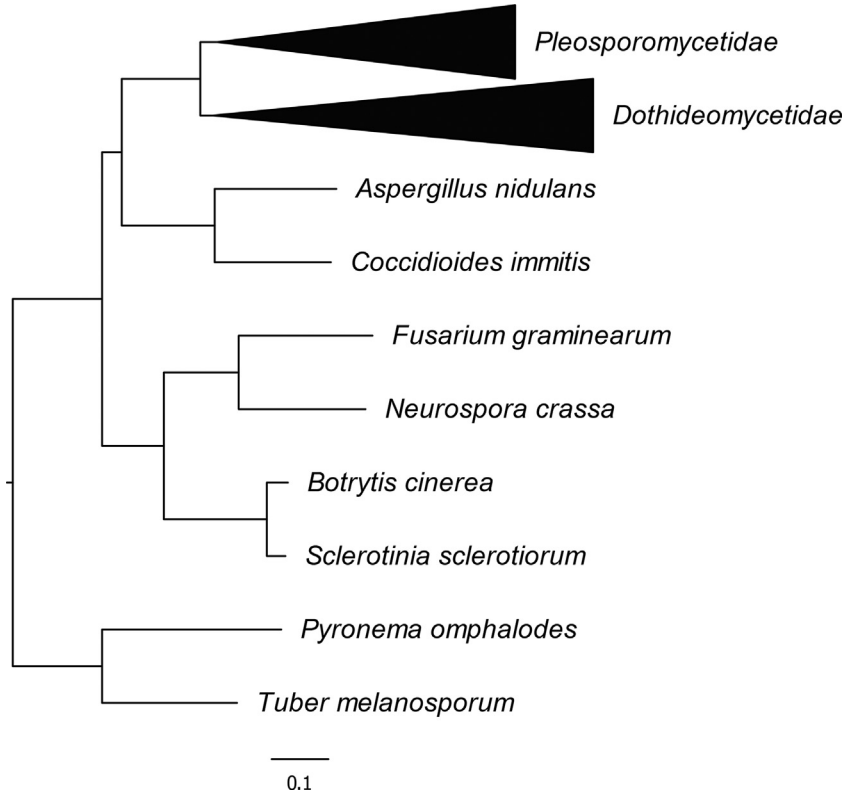


**Fig. 1.** A phylogenetic tree of the *Dothideomycetes* and eight outgroups using whole-genome data. The two sub-classes in the *Dothideomycetes* (*Pleosporomycetidae* and the *Dothideomycetidae*) are well resolved and expanded in Figs 2 and 3.

between the genomes, but this is influenced by the different sequencing and assembly technologies. As expected, smaller genomes had fewer repeats compared to larger assembles ($R^2$ = 0.89). The two largest assemblies were *C. geophilum* and *Z. rhizophila*, which consisted of 65.4 % and 49.6 % repeat content, respectively. This made the non-repeat genome space of *Z. rhizophila* (77.1 Mbp) slightly larger than that of *C. geophilum* (61.5 Mbp). The median assembly was just under 36.5 Mbp or about 33.1 Mbp without repeats. The sizes of the non-repeat genome spaces also did not show any correlation related to lifestyle (Supplementary Fig. S2) or phylogenetic position.

The number of predicted genes was related to non-repeat genome size ($R^2$ = 0.69), but not to functional annotation, lifestyle or phylogenetic position, although saprobes generally had a slightly higher gene count than pathogens (Supplementary Fig. S3). *Piedraia hortae*, which had the smallest genome, also had the fewest genes at 7 572, and was the only one in our study group which had fewer than 10 000 genes. The average *Dothideomycetes* genome had about 12 750 genes of which about 2 100 had homologs in all *Dothideomycetes* (core genes) in our analysis and an additional 8 700 had homologs in at least one other *Dothideomycetes*. Even at our study depth, we were able to identify about 1 950 gene clusters unique to each newly sequenced species (Fig. 5). As expected, the vast majority of core genes had some functional annotation, with over 91 % encoding a PFAM domain. On the other hand, only 8.6 % of "unique genes" encoded a PFAM domain showing the vast gap in our knowledge about key genes that could account for species-specific adaptations. While the average unique gene was smaller than the average core gene (394 *vs* 462 codons), it is well above the 100–150 aa threshold below which gene prediction is often fraught with errors (Haridas *et al.* 2018).

We annotated the gene models to identify PFAMs, transporters, proteases, KEGG Orthology (KO), Carbohydrate-Active Enzymes (CAZy), the secretome and small secreted proteins. We were able to assign some form of functional annotation to about 79 % (average; 67 % median) of the genes in these *Dothideomycetes* genomes (Fig. 4C). The most common was PFAM assignment in over 73 % (average; 62 % median) of the genes. As expected, the number of genes with PFAM annotations was highly correlated with the number of genes ($R^2$ = 0.85). Other functional annotations are also correlated with the number of genes (proteases $R^2$ = 0.66, transporters $R^2$ = 0.67 and secretome $R^2$ = 0.68). In general, saprobes had a higher number of genes, and consequently proportionally higher numbers of PFAMs, proteases and transporters. However, small, secreted cysteine-rich proteins ($R^2$ = 0.46) showed several outliers with the plant pathogen *Venturia inaequalis*, the causal agent of apple scab disease, and *Zymoseptoria ardabiliae*, a hemibiotrophic pathogen of certain grasses, showing the highest relative numbers. We also found that saprobes had a proportionally higher number of CAZy enzymes compared to pathogens. Ohm *et al.* (2012) reported a reduction in genes involved in carbohydrate degeneration in the *Capnodiales*, but this group is primarily composed of plant pathogens even at our increased sampling of 16 *Capnodiales* species compared to seven in Ohm *et al.* (2012). Ancestral state reconstruction (see below) suggests that the switch to parasitism occurred early in the *Capnodiales* (Supplementary Fig. S4, S5) and the reduction in CAZy enzymes may be concurrent with this switch.

## The ancestral dothideomycete was a saprophyte

Parsimony reconstruction of multistate character coding using Mesquite inferred the ancestral ecological character state of

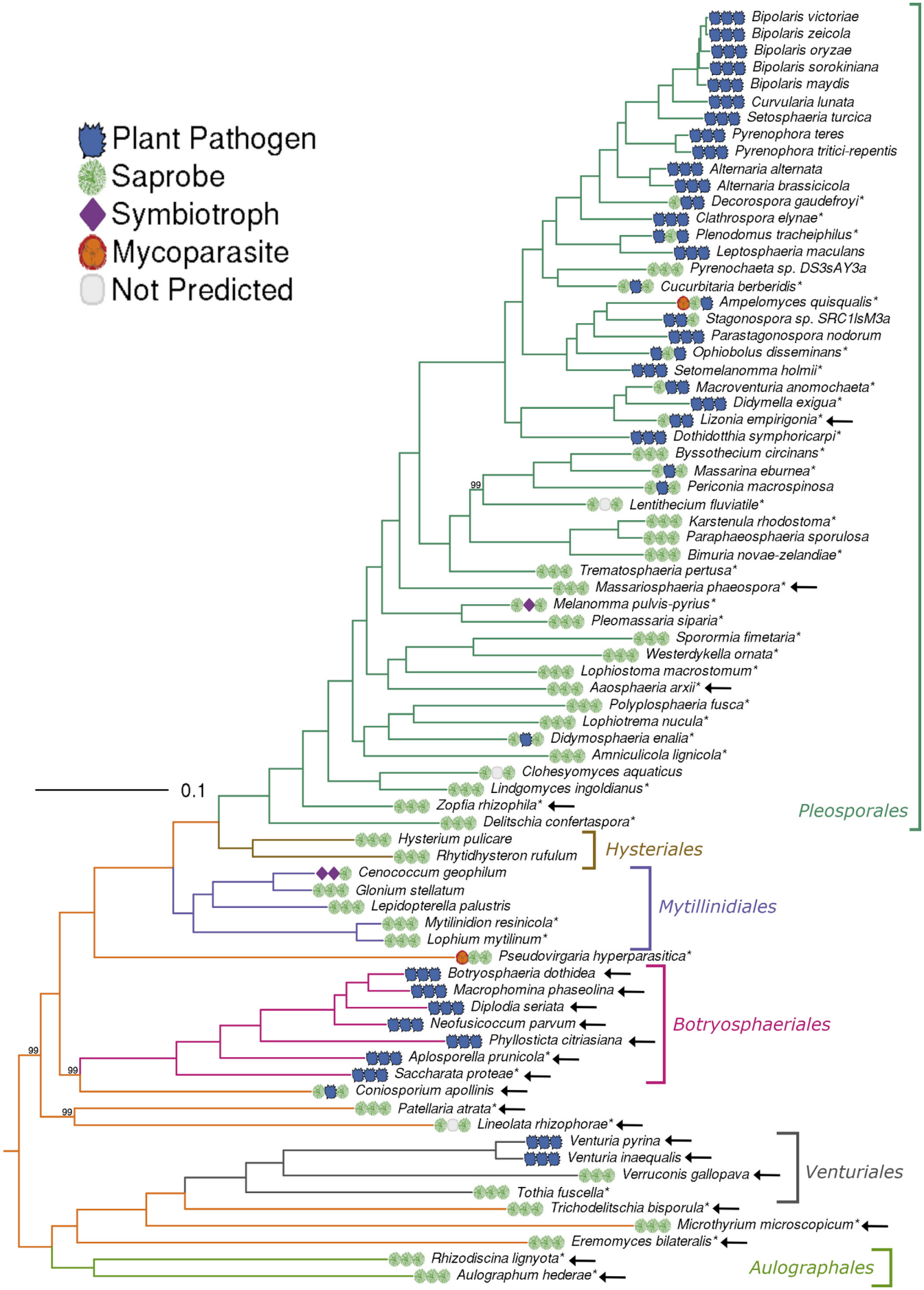


**Fig. 2.** A phylogenetic tree of 76 species from the *Pleosporomycetidae* used in this study. All bootstrap values are 100 % except for those shown. Well resolved orders with multiple species are indicated by brackets on the right. The three icons left of species names represent lifestyle classification based on organism data, FunGuild classification and Machine Learning predictions, respectively. *Represents newly sequenced species for this study. Arrows point to newly reclassified species.

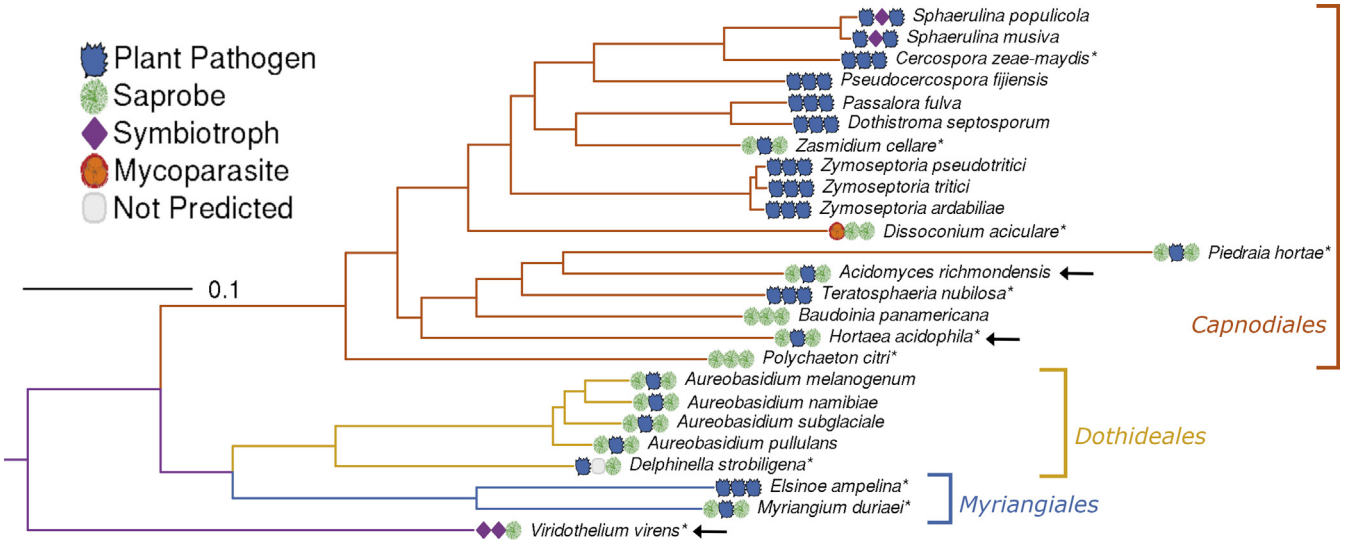


**Fig. 3.** A phylogenetic tree of 25 species from the *Dothideomycetidae* used in this study. All bootstrap support values are 100 percent. Well resolved orders with multiple species are indicated by brackets on the right. The three icons left of species names represent lifestyle classification based on organism data, FunGuild classification and Machine Learning predictions, respectively. *Represents newly sequenced species for this study. Arrows point to newly reclassified species.

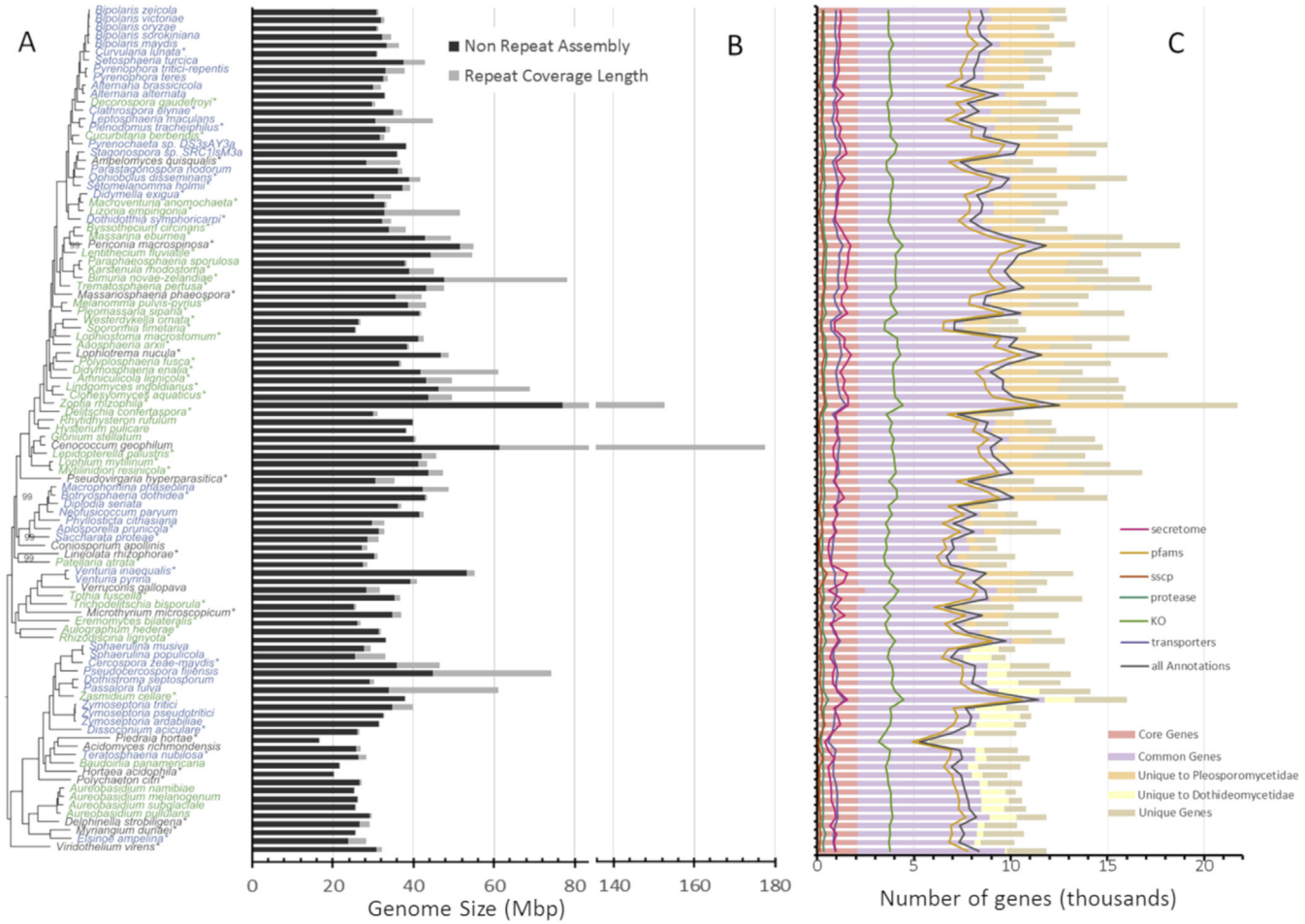


**Fig. 4.** Genome sizes and gene content of 101 *Dothideomycetes*. A. Phylogenetic tree of the *Dothideomycetes* showing plant pathogens in blue and presumed saprobes in green. B. Genome sizes showing proportion of repeat content in these genomes. C. Gene content showing predicted proteome sizes and content. Core genes are found in all *Dothideomycetes* in this study while common genes are found in more than one genome. "Unique to the *Pleosporomycetidae*" and "Unique to the *Dothideomycetidae*" are found in all genomes in the *Pleosporomycetidae* and the *Dothideomycetidae*, respectively.

*Dothideomycetes* as saprobe (Supplementary Fig. S4), and likelihood reconstruction of binary character coding significantly favoured the common ancestor of *Dothideomycetes* as non-plant pathogen (NPP = 0.9987) as compared to plant pathogen (PP = 0.0013) (node 102: Supplementary Fig. S5, Supplementary Table T1). A similar result was obtained using BayesTraits (saprobe: 0.999698, Pathogen: 0.000277, other: 0.000025). Plant pathogenicity is supported as evolving

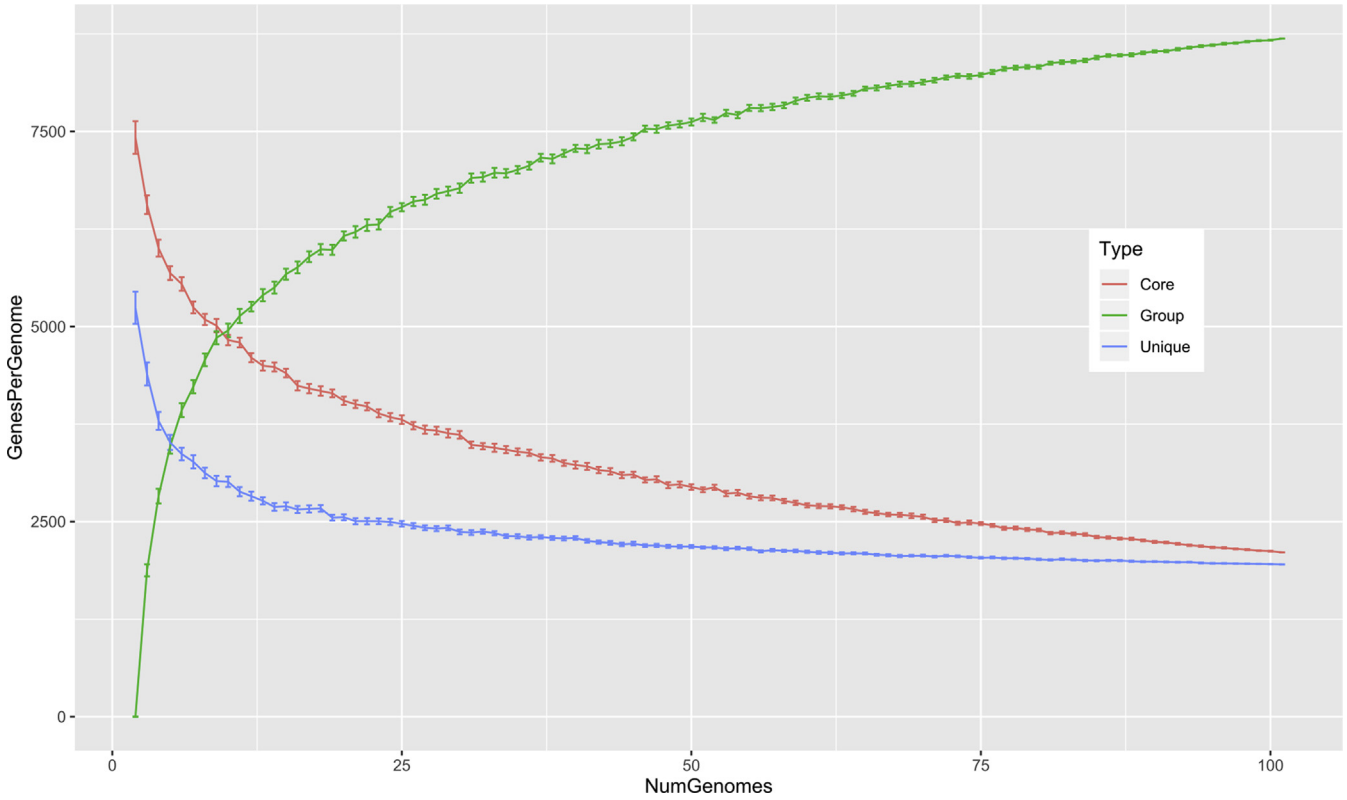


**Fig. 5.** Identifying core, group-specific (common to two or more genomes), and unique genes in OrthoMCL data. Error bars represent 100 random subsamples from the study group.

six to eight times within *Dothideomycetes* with at least four major transitions from NPP to PP including in the most recent common ancestor (MRCA) of *Mycosphaerellaceae* of *Dothideomycetidae* (node 158: NPP = 0.1713; PP = 0.8287), and the MRCAs of *Venturia* (node 96: NPP = 0.0033, PP = 0.9967), *Botryosphaeriales* (node 172: NPP = 0.0770, PP = 0.9230), and the *Setomelanomma-Bipolaris* clade (node 67: NPP = 0.0177, PP = 0.9823) of *Pleosporomycetidae*. The increased sampling depth in the *Pleosporales* from nine in Ohm *et al.* (2012) to 49 in this study showed that the transition from saprotrophy occurred late, and is restricted to the *Pleosporineae* (Supplementary Fig. S4, S5). The transition from NPP to PP at the *Setomelanomma-Bipolaris* clade could possibly have occurred one node earlier (MRCA of *Dothidotthia-Bipolaris* clade at node 23), but the parsimony reconstruction was equivocal and the likelihood reconstructions showed increased proportional support for NPP (NPP = 0.1815, PP = 0.8185). Within *Dothideomycetidae*, acidophiles (*Acidomyces richmondensis*, *Hortaea acidophila*) and human-associated (*Piedraia hortae*) species were restricted to a subclade of *Capnodiales*, and lichen (*Viridothelium virens*, *Trypetheliales*) and insect pathogen (*Myriangium duriaei*, *Myriangiales*) each displayed unique origins. Within *Pleosporomycetidae*, mycoparasitic species *Ampelomyces quisqualis* and *Pseudovirgaria hyperparasitica* arose independently; the seven aquatic species sampled represented six independent transitions to aquatic environments, but ectomycorrhizae had a single origin in *Cenococcocum geophilum*. These results are consistent with findings of Schoch *et al.* (2009), who inferred numerous transitions from saprobic life histories to plant pathogens and lichens, and multiple transitions from terrestrial to aquatic habitats.

## Some gene family expansions are correlated with lifestyles

Most of the species in our study set of 101 *Dothideomycetes* were plant pathogens or presumed saprobes, but ecologies can be difficult to code unequivocally for some taxa. The saprobes sampled here also showed a wide variety of ecological niches such as growth in aquatic habitats (*e.g.*, *Amniculicola lignicola*, *Lineolata rhizophorae*), on dung (*e.g.*, *Delitschia confertaspora*, *Eremomyces bilateralis*, *Sporormia fimentaria* and *Trichodelitschia bisporula*), and at relatively high concentrations of alcohol in the air (*Baudoinia panamericana* and *Zasmidium cellare*). Other lifestyles in our study group included mycorrhizal (*Cenococcum geophilum*), human hair pathogen (*Piedraia hortae*), mycoparasites (*Ampelomyces quisqualis* and *Pseudovirgaria hyperparasitica*), and fungal partners of lichen symbiosis (*Viridothelium virens*). As summarized above, the various lifestyles are not grouped phylogenetically, but appear interspersed among different branches of the tree and most likely arose from saprobic ancestors. We used Fisher's exact test with Bonferroni correction to adjust p-values for multiple comparisons and identified gene family expansions and contractions in convergent lifestyle evolution on various branches of the *Dothideomycetes* tree (Supplementary data - Enrichment.xls).

Among the *Dothideomycetes*, *Baudoinia panamericana* and *Zasmidium cellare* are found in wine caves or cellars or proximal to distilleries, environments associated with significant levels of alcohol vapours. These genomes showed a >30 % increase (p value = 0.008) in transport genes related to the Major Facilitator SuperFamily (MFS). The ability to assimilate alcohol from vapours presumably requires transporters, but analysis of their

genomes did not identify any unique genes or gene families. To identify specific MFS transporters that may be involved in alcohol transport, we looked at the distribution of OrthoMCL gene clusters containing MFS genes. *Baudoinia panamericana* and *Z. cellare* did not show an expansion in specific MFS clusters compared to other *Dothideomycetes* in this study. This could suggest that genes in particular MFS families could have been co-opted for alcohol transport, or this ability was independently gained in these two genomes, which would remain lineage-specific, and were therefore omitted in our analysis.

The two animal-associated fungi in our study group, *Piedraia hortae*, which forms dark, hardened nodules (hence piedra, stone) firmly attached to human hair in humid tropical environments, and *Myriangium duriaei*, which is associated with scale insects and is presumably parasitic, showed a 25 % reduction (p-value = 0.001) in MFS and a 16-fold increase (p-value = 0.0) in PF05922 (protease inhibitor). These differences are presumably due to the narrow ecological niche and relatively stable environment inhabited by these two fungi.

An analysis of the genomes of *Pseudovirgaria hyperparasitica* and *Ampelomyces quisqualis* showed a two-fold increase (p-value = 0.0) in PF01636 (Phosphotransferase enzyme family), a family of bacterial antibiotic resistance proteins. These mycoparasites may be using these genes to overcome antibiotics produced by their hosts during infection. However, this class of proteins also showed high numbers in insect-associated fungi like *Tolypocladium inflatum* (Bushley *et al.* 2013) and *Metarhizium robertsii* (Gao *et al.* 2011), and also some plant pathogens such as *Neonectria ditissima* (Gómez-Cortecero *et al.* 2015). In addition, several well-known mycoparasites like *Trichoderma gamsii* (Baroncelli *et al.* 2016), *T. asperellum* and *T. harzianum* (Druzhinina *et al.* 2018) do not show an increase in PF01636, suggesting that they either use alternative strategies or the lack of sufficient sampling in our study set introduced a statistical false positive in this case. Our analysis also showed PF01636 enrichment in aquatic and saprophytic lineages, suggesting that this family has diverse functions and testing individual genes using experimental approaches for roles in mycoparasitism might yield valuable insights.

## Several gene families are reduced in plant pathogens compared to saprobes

Plant pathogens and saprobes were the largest group of lifestyles in our study set. They split evenly into 42 pathogens and 42 presumed saprobes, with other lifestyles making up the rest of the study set. Among the functionally annotated genes (pfams, CAZy, etc), none were unique to either the pathogens or saprobes. Because about a third of the genes in these genomes have no functional annotation, we looked for OrthoMCL gene clusters that were unique to either the pathogens or saprobes. A cluster consisting of 35 genes from 26 pathogen genomes with no representative from other lifestyle classes was identified (Supplementary Table 6). This group consisted of small (<200 aa) proteins, most with signal peptides, but no other functional annotation. We deleted one of these genes in *Bipolaris maydis* strain C4, where only one copy was identified, but the virulence phenotype on the maize host was unaltered when compared to inoculation with the wild type strain (Supplementary Fig. S7). Although we would not expect a single gene to have a large effect on pathogenicity of diverse species to their hosts, a gene conserved only in pathogens likely plays a quantitative role in pathogenicity and additional higher resolution virulence assays in a range of species may be warranted.

Fisher's exact test identified several gene families that were expanded or contracted in pathogens and saprobes. To reduce errors due to misclassification of species into various lifestyles, we used a curated list of 39 well-known saprobes and 39 plant pathogens (removing additional ambiguous lifestyle species) to identify enrichment of gene families in these two groups. We used a cutoff of $p < 0.005$ and at least 100 genes in the test subset to identify 14 GO annotations and 28 PFAM annotations that showed different enrichment between the saprobes and pathogens. Only one PFAM (PF01328, Peroxidase) and one GO (0032440, 2-alkenal reductase activity) were enriched in pathogens compared to saprobes. While we observe a strong association between presence of these functional annotations and lifestyle, further experimental work is necessary to elucidate their potential roles in pathogenicity.

Among the PFAM domains that were reduced in pathogens were several tetratricopeptide repeat (TPR) domains (Supplementary Fig. S6). The TPR structural motif was originally identified in the fission yeast *Schizosaccharomyces pombe* and appears in all branches of life in various unrelated protein families; it is linked to virulence in some human pathogens like *Candida albicans* (Kaneko *et al.* 2006) and *Porphyromonas gingivalis* (Kondo *et al.* 2010). The TPR motif has a role in protein-protein interactions and is found in proteins involved in the anaphase-promoting complex, NADPH oxidases, hsp90-binding proteins, transcription factors, protein kinase inhibitors, and peroxisomal and mitochondrial import proteins, among others (Das *et al.* 1998). TPR proteins are also part of the plant hormone signal-transduction pathway (Schapire *et al.* 2006). Our dataset showed that fungal saprobes had more than double the number of TPR proteins compared to plant pathogens. Much of this expansion was driven by root-associated saprobes (TPE 3, 4, 7, 8, 10 and 16), but TPR 12 was expanded (68 %) in the pathogens compared to saprobes (p-value = 0.0). It is interesting to speculate whether this reduction is due to a decrease in signalling-related TPR proteins in pathogenic fungi, where they may be affected by plant hormone signalling induced by the fungal attack. Other interesting groups of PFAMs depleted in plant pathogens are PF17111 (Fungal N-terminal domain of STAND proteins), PF05729 (NACHT domain) and PF14479 (involved in prion-inhibition and propagation). These domains have a role in heterokaryon incompatibility (Paoletti & Saupe 2009, Greenwald *et al.* 2010, Daskalov *et al.* 2012). In *Colletotrichum lindemuthianum*, fusions between incompatible strains during colony initiation survived and formed heterokaryotic colonies from conidial anastomosis tubes (Ishikawa *et al.* 2012). The authors hypothesized that such fusions allow asexual fungi to increase their genetic diversity and acquire new pathogenic traits. The reduction in PFAMs related to heterokaryon incompatibility suggests that other strategies to reduce heterokaryon incompatibility may exist in pathogenic fungi that may allow for an increase in fitness related to pathogenic traits.

Similarly, among the GO annotations, protein kinases (GO:0004672), and transcription factors (GO:0003700) were under-represented in pathogens compared to saprobes. In addition, GO:0006808 (regulation of nitrogen utilization), and GO:0008270 (Zn-binding) also showed a reduction in pathogens compared to saprobes. The availability of zinc and nitrogen is a

critical factor in infection by fungi (Staats *et al.* 2013, Mur *et al.* 2017), which can be explained in part by the reduction of these genes in pathogens compared to saprobes. Moreover, several hosts have developed nutritional immunity strategies by withholding essential micronutrients from invading microbes. In such a scenario, an increase in uptake efficiency may be preferable to an increase in genes related to assimilation. An evolved reliance on uptake of plant amino acids would be energy sparing for the pathogens given the complexity and energy demands of amino acid biosynthesis pathways.

## Machine learning distinguishes saprobes from plant pathogens

Our knowledge of lifestyles and ecologies of many species sequenced in this study are limited and we attempted to predict the major lifestyles based on genome sequencing alone. Therefore, we used FUNGuild (Nguyen *et al.* 2016) to predict lifestyles of poorly studied taxa and compared this with our current understanding (Figs 2 and 3). In several cases, we found a discrepancy between what we expected and what was predicted by FUNGuild. For example, we sequenced *Plenodomus tracheiphilus* and *Ophiobolus disseminans*, which are known plant pathogens, but FUNGuild classifies them as probable saprobes. On the other hand, we assumed *Massarina eburnea* and *Periconia macrospinosa* are saprobic and endophytic respectively based on their known lifestyles, but are classified as pathogenic by FUNGuild (Supplementary Table 3). Since FUNGuild results are based on database queries, the prediction accuracy is affected by the quality of data present in the database. To reconcile these discrepancies, we used machine-learning methods to identify hidden correlations between lifestyle and gene content. Large data sets are critical to machine learning, and only plant pathogens and saprobes had sufficient sampling depth in our dataset to be adequately separated using these methods. We selected a training set of 34 well known pathogens and 24 saprobes across the tree to identify genes that could discriminate between pathogens and saprobes in this study. Ten PFAM annotations provided accuracy >0.8 in their ability to separate pathogens from saprobes in the machine learning analysis. The best PFAM discriminator was PF13350 (Protein tyrosine phosphatase), which showed an average of three versus four gene copies in pathogens versus saprobes. Other PFAMs showing reduction between pathogens and saprobes included PF06544 (DUF1115), PF00017 (SH2 domain), PF11700 (Vacuole effluxer Atg22 like), PF07522 (DNA repair metallo-beta-lactamase), PF05057 (DUF676), PF03537 (Glycoside hydrolase), PF01183 (Glycoside hydrolase family 25), PF00413 (Matrixin), and PF00076 (RNA recognition motif) (Supplementary data - Enrichment.xls). Of these, enrichment analysis using Fisher's Exact Test had identified PF05057 ($p = 0.03$) and PF03537 ($p = 0.0$) as reduced in pathogens.

As seen in Fig. 4C, over a third of the genes in the *Dothideomycetes* genomes do not have functional annotation. On the other hand, OrthoMCL gene clusters (genes with orthologs) encompass 85 % of the gene space in these genomes (Fig. 5). Using the OrthoMCL gene clusters dataset (which ignores lineage-specific genes), 27 OrthoMCL gene clusters (presumed orthologs) produced accuracy scores >0.8 in their ability to distinguish between saprobes and pathogens. The best discriminator among the OrthoMCL gene clusters was cluster #9837 (C009837 in Supplementary Table 4) which was found in 80 % of the pathogens and less than 5 % of saprobes and produced a prediction accuracy of 0.89. This cluster consisted of small (<200 aa) proteins with no other functional annotation. To improve prediction accuracy, we used all 27 clusters in all possible combinations. Our tests indicated that high prediction accuracy (>95 %) could be achieved by using a small handful of gene clusters (Supplementary Fig. S8).

Among the 50 highest-ranked combinations, six clusters, none of which had any functional annotation, appeared more than 30 times. We used these six clusters to classify the *Dothideomycetes* into pathogens and saprobes as shown in Figs 2, 3 and 6. These six orthologous gene clusters were able to differentiate between pathogens and saprobes with >95 % accuracy in our study set. While the vectors underlying machine-learning remain obscure within the algorithm, the data shows that two of these gene clusters are primarily found in pathogens (PRG, pathogen related genes) while the other four are mostly in the presumed saprobes (SRG, saprobe related genes) (Supplementary Table 4). The data suggest that saprobes have SRG at least two times the number of PRG except for *Macroventuria anomochaeta* and *Stagonospora* sp. SRC1lsM3a. *Macroventuria anomochaeta* has both the PRG and none of the SRG and is strongly predicted to be a pathogen (Macan1 in Fig. 6). The sexual morph of *Macroventuria anomochaeta* was isolated from decayed canvas in a desert in South Africa and its ecology is presumed to be saprotrophic, but the underlying enzymatic mechanisms are unknown. *Stagonospora* sp. SRC1lsM3a (Stasp1 in Fig. 6) has two of the SRG and one PRG and is predicted to be a saprobe. While several *Stagonospora* species are plant pathogens, *Stagonospora* sp. SRC1lsM3a was isolated and studied for its role in the remediation of metal-polluted environments due to the oxidation of manganese compounds. Species-level determination of *Stagonospora* sp. SRC1lsM3a in phylogenetic analysis have been unsuccessful, and the absence of spores or reproductive structures in culture prevents morphological identification (Zeiner *et al.* 2016) (https://mycocosm.jgi.doe.gov/Stasp1). Additional genomic data from closely related species are required to resolve this.

Another organism isolated and studied for its ability to oxidize Mn compounds, *Pyrenochaeta* sp. DS3sAY3a (Pyrsp1), had two PRG and two SRG and therefore has a near equal probability of being classified as a pathogen or saprobe. This ratio (2PRG : 2SRG) is also seen in *Ophiobolus disseminans* (Ophdi1), which was isolated from the highly poisonous plant dog's mercury (*Mercurialis perennis*) in Sweden, but whose ecology remains unknown. Five species in our *Dothideomycetes* study set did not have any of these six gene clusters in their genomes. This included *Dissoconium aciculare*, a hyperparasite of powdery mildew, the alcohol vapour-loving sooty mould *Baudoinia panamericana*, the extremophile *Hortaea acidophila*, the aquatic saprophyte *Lineolata rhizophorae,* and the human hair pathogen *Piedraia hortae*. We generated MEME profiles from the MAFFT alignments of these orthoMCL clusters using the MEME suite (Bailey *et al.* 2009) and scanned 374 previously published fungal genomes (including 47 *Dothideomycetes*) using MAST (part of the MEME suite) for genes matching these profiles in their proteomes. We were able to identify homologs for one or more of these MEME profiles in 134 proteomes, all in the *Pezizomycotina*. Like in the *Dothideomycetes*, the presence or absence of the six gene clusters in these 134 proteomes was able to

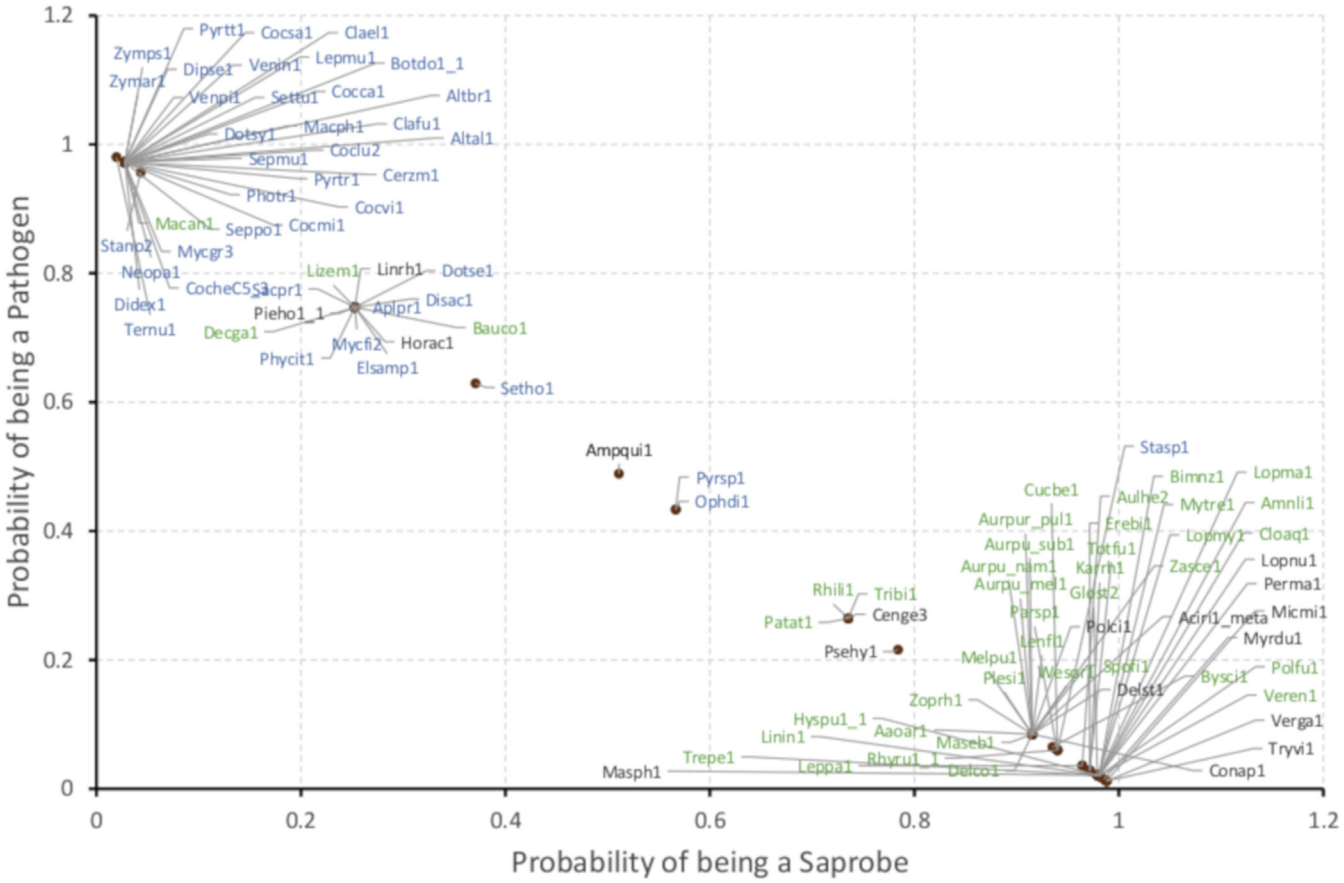


**Fig. 6.** Support Vector Machine (SVM)-based prediction of lifestyle based on 6 gene clusters showed a >95 % accuracy in correctly predicting plant pathogens (blue) vs saprobes (green). (Other lifestyles are indicated in black). Some poorly studied saprobes are predicted to be pathogens suggesting that these may be weak pathogens or have recently diverged from pathogenic ancestors. Short names used are shown in Supplementary Table 4.

differentiate the pathogens and saprobes with >95 % accuracy. This can become a valuable tool to predict lifestyles for rarely seen and poorly studied species beyond *Dothideomycetes*, across Pezizomycotina as our efforts to understand the breadth of the fungal tree of life continues.

## CONCLUSIONS

This study represents the largest genome comparison of fungal plant pathogens and introduces the genomes of 55 newly sequenced *Dothideomycetes*. The availability of whole-genome data also produced high-confidence phylogenetic trees with which we have been able to reclassify 25 organisms including the clarification of phylogenetic placement of several species that were previously considered to be "*incertae sedis*" and the addition of new taxonomic units. Ancestral character state analyses support a terrestrial saprobic lifestyle as being ancestral with subsequent derivations of plant pathogen, lichen, mycorrhiza, mycoparasite, aquatic and extremophile. Genomic traits were identified as enriched or reduced for selected ecologies, but often these were unannotated orthologous clusters or PFAM domains with broad high-level annotations, and subsequent experimentation is needed to test potential function related to ecology. The identification of gene family expansion/contraction is just one piece of the puzzle, with additional factors like changes in gene expression regulation, SNPs, *etc*. also playing important roles in determining the proteome and consequently the ecological adaptations of a particular species. Growth specific gene regulation is likely a key factor in understanding the saprophyte-pathogen continuum of several species and closely related lineages. Nevertheless, this large-scale comparison combined with machine learning was able to identify six gene clusters, most of which contain single-copy genes that can classify plant pathogens and saprobes in *Dothideomycetes* with greater than 95 % accuracy and we are encouraged by the ability of machine learning to differentiate pathogen vs saprobe based on genomic features alone. With increasing genomic data becoming available, a fundamental requirement for machine learning techniques, we can envision the use of these methods to identify genes and gene families driving other phenotypic changes, for example identifying genes related to fungal morphology (filamentous vs yeast vs dimorphic growth). This study encompassed 48 of the 110 known families in *Dothideomycetes*, but represents an important step in our efforts to explore fungal diversity and identify sets of genes determining fungal lifestyle.

## DATA AVAILABILITY

Genome data is available on the JGI MycoCosm portal using the DBID shown in Supplementary Table 4 in the form https://mycocosm.jgi.doe.gov/DBID; eg: https://mycocosm.jgi.doe.gov/Aaoar1 for *Aaosphaeria arxii*. All *Dothideomycetes* genomes available on MycoCosm can be seen at https://mycocosm.jgi.doe.gov/Dothideomycetes. OrthoMCL Clustering run data for 101 Dothideomycetes is available at https://mycocosm.jgi.doe.gov/clm/run/101-Dothideomycetes.3460. Data for newly sequenced genomes has been submitted to GenBank and accession numbers are available in Supplementary Table 5.

## ACKNOWLEDGEMENTS

The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The U.S. Government retains, and the publisher, by accepting the article for publication, acknowledges, that the U.S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for U.S. Government purposes. R.A. and I.V.G. were supported by the NSF grants DEB-1354625 and IOS-1456958. We thank Daniele Armaleo who provided materials for sequencing. Sequencing of *Dothideomycetes* genomes were conducted as part of the 1000 Fungal Genomes Project supported by the JGI Community Sequencing Project #662.

## SUPPLEMENTARY TEXT

### Taxonomy

***Coniosporiales*** Crous, Spatafora, Haridas & Grigoriev, ***ord. nov.*** MycoBank MB831889.

*Etymology*: Name reflects the genus *Coniosporium.*

*Conidiomata* sporodochial, pulvinate, punctiform, olivaceous to black. *Mycelium* immersed. *Conidiophores* arising from a stroma, macronematous straight to flexuous, densely aggregated, straight to flexuous, unbranched, smooth, rugulose to verruculose. *Conidiogenous cells* integrated, terminal, subcylindrical, fragmenting. *Conidia* catenate, schizogenous, dry, ellipsoid, oblong, obovoid, pyriform to subspherical, pale to dark brown, muriformly septate, smooth to verruculose (adapted from Ellis 1971).

***Coniosporiaceae*** Crous, Spatafora, Haridas & Grigoriev, ***fam. nov.*** MycoBank MB831890.

*Etymology*: Name reflects the genus *Coniosporium.*

*Conidiomata* sporodochial, pulvinate, punctiform, olivaceous to black. *Mycelium* immersed. *Conidiophores* arising from a stroma, macronematous straight to flexuous, densely aggregated, straight to flexuous, unbranched, smooth, rugulose to verruculose. *Conidiogenous cells* integrated, terminal, subcylindrical, fragmenting. *Conidia* catenate, schizogenous, dry, ellipsoid, oblong, obovoid, pyriform to subspherical, pale to dark brown, muriformly septate, smooth to verruculose (adapted from Ellis 1971).

*Type genus*: *Coniosporium* Link, Mag. Gesell. naturf. Freunde, Berlin 3: 8. 1809.
*Type species*: *Coniosporium olivaceum* Link, Mag. Gesell. naturf. Freunde, Berlin 3: 8. 1809.
*Notes*: The genus *Coniosporium* is polyphyletic (see Li *et al.* 2008), and *C. apollinis* (see Sterflinger *et al*. 1997) possibly represents an undescribed genus. However, there is presently no DNA sequence of the type, *Coniosporium olivaceum*, and the latter needs to be recollected (from *Pinus maritima* in the Mediterranean area) before this can be resolved.

***Eremomycetales*** Crous, Spatafora, Haridas & Grigoriev, ***ord. nov.*** MycoBank MB831891.

*Etymology*: Name reflects the genus *Eremomyces.*

*Mycelium* hyaline or brown. *Ascocarps* subglobose to globose, non-ostiolate, glabrous to setose, dark brown to black, with a single cavity. *Asci* irregularly disposed, subglobose to clavate, evanescent. *Ascospores* aseptate, hyaline, lacking germ pores. No asexual morph observed (adapted from Malloch & Cain 1971).

***Eremomycetaceae*** Malloch & Cain, Canad. J. Bot. 49: 847. 1971.
*Type genus*: *Eremomyces* Malloch & Cain, Canad. J. Bot. 49: 847. 1971.
*Type species*: *Eremomyces bilateralis* Malloch & Cain, Canad. J. Bot. 49: 849. 1971.

***Lineolatales*** Crous, Spatafora, Haridas & Grigoriev, ***ord. nov.*** MycoBank MB831892.

*Etymology*: Name reflects the genus *Lineolata.*

*Ascomata* obpyriform, immersed to superficial, ostiolate, papillate, subcarbonaceous to subcoriaceous, periphysate, and dark brown to black, pseudoparaphyses trabeculate, in a gelatinous matrix. *Asci* 8-spored, cylindrical, short pedunculate, non-amyloid, with multi-layered refractive ring, fissitunicate. *Ascospores* uniseriate, 1-septate, ellipsoidal, brown with surface sculpturing.

***Lineolataceae*** Crous, Spatafora, Haridas & Grigoriev, ***fam. nov.*** MycoBank MB831893.

*Etymology*: Name reflects the genus *Lineolata.*

*Ascomata* obpyriform, immersed to superficial, ostiolate, papillate, subcarbonaceous to subcoriaceous, periphysate, and dark brown to black, pseudoparaphyses trabeculate, in a gelatinous matrix. *Asci* 8-spored, cylindrical, short pedunculate, non-amyloid, with multi-layered refractive ring, fissitunicate. *Ascospores* uniseriate, 1-septate, ellipsoidal, brown with surface sculpturing (adapted from Kohlmeyer & Volkmann-Kohlmeyer 1990).

*Type genus*: *Lineolata* Kohlm. & Volkm.-Kohlm., Mycol. Res. 94: 687. 1990.
*Type species*: *Lineolata rhizophorae* (Kohlm. & E. Kohlm.) Kohlm. & Volkm.-Kohlm., Mycol. Res. 94: 688. 1990.

***Aulographales*** Crous, Spatafora, Haridas & Grigoriev, ***ord. nov.*** MycoBank MB831896.

*Etymology*: Name reflects the genus *Aulographum.*

Saprobic on leaves and wood. *Mycelium* superficial and internal, sparse. *Ascomata* apothecial to thyrothecial, globose to ellipsoidal, simple or branched, scattered to loosely clustered, superficial, black. *Hamathecium* dissolving early, or of branched, hyaline, septate, anastomosing pseudoparaphyses. *Asci* 8-spored, bitunicate, clavate or globose to subglobose, with or without pedicel. *Ascospores* irregularly arranged, obovoid to narrowly clavate, hyaline to brown, constricted at the septum, smooth-walled, with or without remnants of mucilage.
*Families included in order*: *Aulographaceae* and *Rhizodiscinaceae*.

***Aulographaceae*** Luttr. ex P.M. Kirk *et al*., In: Kirk *et al*., Ainsworth & Bisby's Dictionary of the Fungi, Edn 9 (Wallingford): ix. 2001.
*Type genus*: *Aulographum* Lib., Pl. crypt. Arduenna, fasc. (Liège) 3(nos 201–300): no. 272. 1834.
*Type species*: *Aulographum hederae* Lib., Pl. crypt. Arduenna, fasc. (Liège) 3(nos 201–300): no. 272. 1834.

***Rhizodiscinaceae*** Crous, Spatafora, Haridas & Grigoriev, ***fam. nov.*** MycoBank MB831897.

*Etymology*: Name reflects the genus *Rhizodiscina.*

*Ascomata* apothecial, superficial, initially closed, exposing a black hymenium at maturity, convex, anchored at base. Exciple pseudoparenchymatous, with black outer layers; inner layers of dense textura angularis, dark brown. *Hamathecium* of branched, hyaline, septate pseudoparaphyses, slightly swollen on the apex, anastomosing, forming a pale brown epithecium above or at the level of asci. *Asci* 8-spored, bitunicate, clavate, long pedicellate, with prominent ocular chamber. *Ascospores* irregularly arranged, obovoid to oblong, brown, smooth-walled, constricted at the septum.

*Type genus*: *Rhizodiscina* Hafellner, Beih. Nova Hedwigia 62: 195. 1979.
*Type species*: *Rhizodiscina lignyota* (Fr.) Hafellner, Beih. Nova Hedwigia 62: 195. 1979.

## APPENDIX A. SUPPLEMENTARY DATA

Supplementary data to this article can be found online at https://doi.org/10.1016/j.simyco.2020.01.003.

## REFERENCES


Bailey TL, Boden M, Buske FA, *et al*. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* **37**: W202–W208.

Baroncelli R, Zapparata A, Piaggeschi G, *et al*. (2016). Draft whole-genome sequence of *Trichoderma gamsii* T6085, a promising biocontrol agent of Fusarium Head Blight on wheat. *Genome Announcements* **4**(1) e01747-15.

Bensch K, Groenewald JZ, Meijer M, *et al*. (2018). *Cladosporium* species in indoor environments. *Studies in Mycology* **89**: 177–301.

Bushley KE, Raja R, Jaiswal P, *et al*. (2013). The genome of *Tolypocladium inflatum*: evolution, organization, and expression of the cyclosporin biosynthetic gene cluster. *PLoS Genetics* **9**: e1003496.

Castresana J (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**: 540–552.

Chang T-C, Salvucci A, Crous PW, *et al*. (2016). Comparative genomics of the Sigatoka Disease complex on banana suggests a link between parallel evolutionary changes in *Pseudocercospora fijiensis* and *Pseudocercospora eumusae* and increased virulence on the banana host. *PLoS Genetics* **12**: e1005904.

Condon BJ, Leng Y, Wu D, *et al*. (2013). Comparative genome structure, secondary metabolite, and effector coding capacity across *Cochliobolus* pathogens. *PLoS Genetics* **9**: e1003233.

Crous PW, Wingfield MJ, Cheewangkoon R, *et al*. (2019). Foliar pathogens of eucalypts. *Studies in Mycology* **94**: 125–298.

Das AK, Cohen PW, Barford D (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *The EMBO Journal* **17**: 1192–1199.

Daskalov A, Paoletti M, Ness F, *et al*. (2012). Genomic clustering and homology between HET-S and the NWD2 STAND protein in various fungal genomes. *PLoS One* **7**: e34854.

De Hoog GS, Guarro J, Gené J, *et al*. (2000). *Atlas of Clinical Fungi*, Second Edition. Centraalbureau voor Schimmelcultures, Utrecht, the Netherlands.

Druzhinina IS, Chenthamara K, Zhang J, *et al*. (2018). Massive lateral transfer of genes encoding plant cell wall-degrading enzymes to the mycoparasitic fungus *Trichoderma* from its plant-associated hosts. *PLoS Genetics* **14**: e1007322.

Ellwood SR, Syme RA, Moffat CS, *et al*. (2012). Evolution of three *Pyrenophora* cereal pathogens: recent divergence, speciation and evolution of non-coding DNA. *Fungal Genetics and Biology* **49**: 825–829.

Foster AJ, Pelletier G, Tanguay P, *et al*. (2015). Transcriptome analysis of Poplar during leaf spot infection with *Sphaerulina* spp. *PLoS One* **10**: e0138162.

Gao Q, Jin K, Ying S-H, *et al*. (2011). Genome sequencing and comparative transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M. acridum*. *PLoS Genetics* **7**: e1001264.

Gnerre S, Maccallum I, Przybylski D, *et al*. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences USA* **108**: 1513–1518.

Gómez-Cortecero A, Harrison RJ, Armitage AD (2015). Draft genome sequence of a European isolate of the apple canker pathogen *Neonectria ditissima*. *Genome Announcements* **3**(6) e01243-15.

Goodwin SB, McCorison CB, Cavaletto JR, *et al*. (2016). The mitochondrial genome of the ethanol-metabolizing, wine cellar mold *Zasmidium cellare* is the smallest for a filamentous ascomycete. *Fungal Biology* **120**: 961–974.

Greenwald CJ, Kasuga T, Glass NL, *et al*. (2010). Temporal and spatial regulation of gene expression during asexual development of *Neurospora crassa*. *Genetics* **186**: 1217–1230.

Grigoriev IV, Nikitin R, Haridas S, *et al*. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research* **42**: D699–D704.

Guarnaccia V, Groenewald JZ, Li H, *et al*. (2017). First report of *Phyllosticta citricarpa* and description of two new species, *P. paracapitalensis* and *P. paracitricarpa*, from citrus in Europe. *Studies in Mycology* **87**: 161–185.

Haridas S, Salamov A, Grigoriev IV (2018). Fungal genome annotation. *Methods in Molecular Biology* **1775**: 171–184.

Ishikawa FH, Souza EA, Shoji J-Y, *et al*. (2012). Heterokaryon incompatibility is suppressed following conidial anastomosis tube fusion in a fungal plant pathogen. *PLoS One* **7**: e31175.

Kaneko A, Umeyama T, Utena-Abe Y, *et al*. (2006). Tcc1p, a novel protein containing the tetratricopeptide repeat motif, interacts with Tup1p to regulate morphological transition and virulence in *Candida albicans*. *Eukaryotic Cell* **5**: 1894–1905.

Katoh K (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**: 3059–3066.

Kondo Y, Ohara N, Sato K, *et al*. (2010). Tetratricopeptide repeat protein-associated proteins contribute to the virulence of *Porphyromonas gingivalis*. *Infection and Immunity* **78**: 2846–2856.

Lam K-K, LaButti K, Khalak A, *et al*. (2015). FinisherSC: a repeat-aware tool for upgrading de novo assembly using long reads. *Bioinformatics* **31**: 3207–3209.

Lanfear R, Frandsen PB, Wright AM, *et al*. (2017). PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* **34**: 772–773.

Lawrey JD, Diederich P, Nelsen MP, *et al*. (2012). Phylogenetic placement of lichenicolous *Phoma* species in the *Phaeosphaeriaceae* (*Pleosporales, Dothideomycetes*). *Fungal Diversity* **55**: 195–213.

Li L, Stoeckert Jr CJJr., Roos DS (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* **13**: 2178–2189.

Lücking R, Hawksworth DL (2018). Formal description of sequence-based voucherless: promises and pitfalls, and how to resolve them. *IMA Fungus* **9**: 143–166.

Marin-Felix Y, Groenewald JZ, Cai L, *et al*. (2017). Genera of phytopathogenic fungi: GOPHY 1. *Studies in Mycology* **86**: 99–216.

Mur LAJ, Simpson C, Kumari A, *et al*. (2017). Moving nitrogen to the centre of plant defence against pathogens. *Annals of Botany* **119**: 703–709.

Nelsen MP, Lücking R, Grube M, *et al*. (2009). Unravelling the phylogenetic relationships of lichenised fungi in *Dothideomyceta*. *Studies in Mycology* **64**: 135–144.

Nguyen NH, Song Z, Bates ST, *et al*. (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology* **20**: 241–248.

Ohm RA, Feau N, Henrissat B, *et al*. (2012). Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. *PLoS Pathogens* **8**: e1003037.

Paoletti M, Saupe SJ (2009). Fungal incompatibility: evolutionary origin in pathogen defense? *Bioessays* **31**: 1201–1210.

Peter M, Kohler A, Ohm RA, *et al*. (2016). Ectomycorrhizal ecology is imprinted in the genome of the dominant symbiotic fungus *Cenococcum geophilum*. *Nature Communications* **7**: 12662.

Phillips AJL, Alves A, Abdollahzadeh J, *et al*. (2013). The *Botryosphaeriaceae*: genera and species known from culture. *Studies in Mycology* **76**: 51–167.

Quaedvlieg W, Verkley GJM, Shin H-D, *et al*. (2013). Sizing up *Septoria*. *Studies in Mycology* **75**: 307–390.

Ruibal C, Gueidan C, Selbmann L, *et al*. (2009). Phylogeny of rock-inhabiting fungi related to *Dothideomycetes*. *Studies in Mycology* **64**: 123–133S7.

Schapire AL, Valpuesta V, Botella MA (2006). TPR proteins in plant hormone signaling. *Plant Signaling & Behavior* **1**: 229–230.

Schoch CL, Crous PW, Groenewald JZ, *et al.* (2009). A class-wide phylogenetic assessment of Dothideomycetes. *Studies in Mycology* **64**: 1–15.

Schoch CL, Shoemaker RA, Seifert KA, *et al.* (2006). A multigene phylogeny of the *Dothideomycetes* using four nuclear loci. *Mycologia* **98**: 1041–1052.

Schuelke TA, Wu G, Westbrook A, *et al.* (2017). Comparative genomics of pathogenic and nonpathogenic beetle-vectored fungi in the genus *Geosmithia*. *Genome Biology and Evolution* **9**: 3312–3327.

Staats CC, Kmetzsch L, Schrank A, *et al.* (2013). Fungal zinc metabolism and its connections to virulence. *Frontiers in Cellular and Infection Microbiology* **3**: 65.

Stamatakis A (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Stukenbrock EH, Quaedvlieg W, Javan-Nikhah M, *et al.* (2012). *Zymoseptoria ardabiliae* and *Z. pseudotritici*, two progenitor species of the septoria tritici leaf blotch fungus *Z. tritici* (synonym: *Mycosphaerella graminicola*). *Mycologia* **104**: 1397–1407.

Turgeon BG, Baker SE (2007). Genetic and genomic dissection of the Cochliobolus heterostrophus Tox1 locus controlling biosynthesis of the polyketide virulence factor T-toxin. *Advances in Genetics* **57**: 219–261.

Valenzuela-Lopez N, Cano-Lira JF, Guarro J, *et al.* (2018). Coelomycetous *Dothideomycetes* with emphasis on the families *Cucurbitariaceae* and *Didymellaceae*. *Studies in Mycology* **90**: 1–69.

Videira SIR, Groenewald JZ, Nakashima C, *et al.* (2017). Mycosphaerellaceae – Chaos or clarity? *Studies in Mycology* **87**: 257–421.

Wijayawardene NN, Hyde KD, Rajeshkumar KC, *et al.* (2017). Notes for genera: *Ascomycota*. *Fungal Diversity* **86**: 1–594.

Zeiner CA, Purvine SO, Zink EM, *et al.* (2016). Comparative analysis of secretome profiles of manganese(II)-oxidizing Ascomycete fungi. *PLoS One* **11**: e0157844.

Zerbino DR, Birney E (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18**: 821–829.