

Functional gene content of fungi in a pine needle disease system largely reflects their phylogeny

Gabriel A. A. Silva^{1*}, gabriel.silva@auburn.edu, ORCID 0000-0002-0026-8643

Temitope R. Folorunso¹, trf0019@auburn.edu, ORCID 0000-0003-1276-692X

Lori G. Eckhardt¹, eckhalg@auburn.edu, ORCID 0000-0002-3270-8199

Janna R. Willoughby¹, jwilloughby@auburn.edu, ORCID 0000-0002-0176-1878

1. College of Forestry, Wildlife and Environment, Auburn University, Auburn, AL, USA

* Corresponding author

Abstract

Background: Fungi recovered from diseased host tissue can differ substantially in functional gene content, but interpreting that variation requires knowing how it relates to evolutionary history.

Results: We assembled and annotated genomes for 67 fungi cultured from pine needles with symptoms of brown spot needle blight and compared them with 13 reference genomes, including established pine pathogens. The isolates spanned four classes of Ascomycota. Functional gene content tracked phylogeny. Taxonomic class explained 21.4% of variation in carbohydrate-active enzyme family composition, and phylogenetic distance correlated with functional dissimilarity in every repertoire examined, most strongly for carbohydrate-active enzymes and least for predicted effectors. Comparisons pooling isolates and references across classes produced apparent lifestyle differences that did not persist within Dothideomycetes. Within that class, the needle pathogens *Lecanosticta acicola* and *Dothistroma septosporum* fell at the low end of the isolate distribution for cellulose- and pectin-associated enzyme genes, a reduction that persisted when enzyme families shared among substrate categories were excluded. Their biosynthetic gene cluster counts, secreted enzyme proportions, and predicted effector content fell within the isolate distributions.

Conclusions: These comparisons show that the positions of known pathogens depend on the functional repertoire examined and its evolutionary context. Because the ecological roles of the cultured fungi were not determined, their gene content cannot directly establish which contributed to disease. Outside plant cell wall degradation, the needle pathogens were not distinguishable from co-occurring fungi by gene content, and these genomes provide a comparative foundation for testing which fungi contribute to disease, and which functions are expressed during pine colonization.

Keywords: comparative genomics; fungal phylogenomics; carbohydrate-active enzymes; plant cell wall degrading enzymes; biosynthetic gene clusters; secretome; Dothideomycetes; brown spot needle blight; fungal lifestyle

Background

Fungal genomes contain many of the genes used to obtain nutrients, interact with hosts, and compete with other microorganisms. Comparative studies use these repertoires to investigate functional differences among fungi, but gene content also reflects evolutionary history: relatives inherit much of the same machinery even as individual lineages gain, lose, or modify genes. In a comparison of 18 Dothideomycete genomes, CAZyme family profiles largely followed phylogenetic relationships, although some fungi departed from that pattern [1]. It remains less clear whether evolutionary relatedness corresponds equally strongly to similarity in degradative enzymes, secreted proteins, and secondary metabolite gene clusters. Determining how these relationships differ among functions is essential for interpreting what makes a particular fungus distinctive.

Brown spot needle blight (BSNB), associated with *Lecanosticta acicola*, is one of the most concerning diseases of pines, whose lesions can progress to needle necrosis and premature defoliation, reducing growth after severe or repeated infection [2,3]. Needle lesions associated with BSNB overlap substantially with those caused by other needle pathogens and by abiotic stress, so diagnosis in the field rests largely on symptom appearance [4]. Surveys of affected *Pinus taeda* have also detected diverse fungal assemblages whose community composition varied with symptom presentation [5]. This makes symptomatic needles an important setting in which to examine the broader fungal community present on a diseased host, while recognizing that symptom appearance alone does not establish the ecological role of any individual fungus

recovered. Culturing fungi from such needles, therefore, brings together genomes from lineages encountered in the same host tissue, while leaving each isolate's ecological role unresolved. On top of that, comparing these genomes through a common annotation and analysis pipeline allows us to ask how much of their functional variation occurs across major lineages and how much remains among fungi within those lineages.

Established pine pathogens provide a useful comparison for fungi cultured from symptomatic needles because several predicted gene repertoires plausibly contribute to host interaction. Carbohydrate-active enzymes (CAZymes) include gene families associated with degradation of cellulose, hemicellulose, and pectin in plant cell walls, as well as families that act on chitin and glucans in fungal cell walls [6–8]. Plant-cell-wall-degrading repertoires vary among fungi including pectin-degrading families such as GH28 and the pectate and pectin lyases [6,9]. Secreted proteins, including some effectors, can act at the host interface and suppress of immune recognition [9], while biosynthetic gene clusters can, encode phytotoxins and compounds involved in competition with other fungi [10,11]. Comparing these repertoires across the recovered fungi and established pathogens allows us to ask whether the pathogens are unusual in particular predicted functions, while interpreting those positions in light of their evolutionary relationships.

In essence, fungi recovered from the same host tissue may differ substantially in the genes they carry, but it remains unclear how much of that variation reflects shared evolutionary history and how much occurs among fungi within the same major lineage. We generated annotated genome assemblies for fungi isolated from *Pinus* needles displaying characteristic BSNB symptoms,

collected across the southeastern United States, and compared their phylogenetic relationships and functional repertoires with those of reference fungi. We hypothesized that broad functional composition would reflect evolutionary relatedness, but that the strength of this relationship would differ and carbohydrate-active enzymes, predicted secretomes and effectors, and secondary metabolite gene clusters. We addressed three specific questions: 1. How are the recovered isolates distributed across the fungal phylogeny? 2. How strongly does evolutionary relatedness correspond to similarity in different functional repertoires? And 3. where do established pine pathogens fall within the functional variation observed among the isolates? This comparison allowed us to determine whether functional gene content differs mainly among major fungal lineages or also varies substantially among fungi within the same taxonomic class.

Results

Genome recovery and phylogenetic placement

We sequenced and assembled 176 fungal isolates cultured from pine needles with brown spot needle blight symptoms. Following exclusion of assemblies with BUSCO completeness below 70% or duplication at or above 15%, 67 isolates were retained for comparative analysis (Supplemental Figure 1). Retained assemblies had a median genome size of 33.1 Mb (range 22.2 to 53.0 Mb) across a median of 742 contigs (range 14 to 3,592) with a median contig N50 of 101 kb (range 14 kb to 4.9 Mb). BUSCO completeness of retained assemblies was high (median 94.9%, range 70.2 to 99.7%). Structural annotation predicted a median of 11,182 genes per genome (range 7,791 to 17,662), of which a median of 452 (range 256 to 772) carried at least one CAZy family assignment.

Phylogenetic placement on a BUSCO-based maximum-likelihood phylogeny assigned the 67 retained isolates to four classes of Ascomycota: Dothideomycetes (n = 31), Eurotiomycetes (n = 24), Sordariomycetes (n = 11) and Leotiomyces (n = 1) (Figure 1). Comparison of patristic distance supported provisional assignments to species level for 12 isolates, genus level for 32, family level for 22, and a single isolate to only to order. All 67 samples placement had ultrafast bootstrap support of 100 for the clade uniting them with their nearest reference. A single isolate, BS01_360_046-1, remained distant enough from other reference taxa that we reported here only to order level as *Pleosporales* sp.

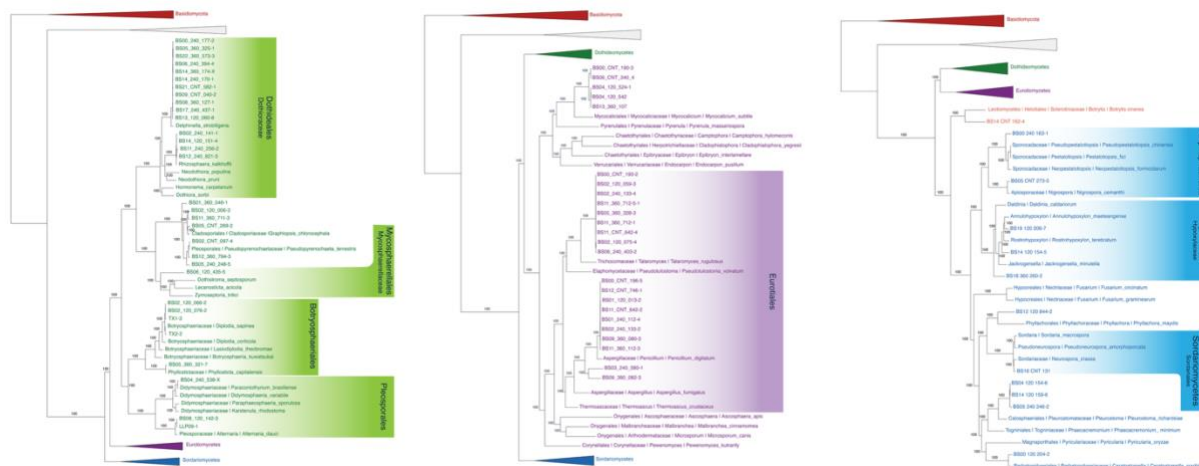


Figure 1. Maximum-likelihood phylogeny of 67 isolates cultured from needles with brown spot needle blight symptoms and 62 reference genomes. The same tree is shown in three panels with different clades collapsed: Dothideomycetes (A), Eurotiomycetes (B), and Sordariomycetes with the single Leotiomyces isolate (C). Labels are coloured by taxonomic class (green = Dothideomycetes, purple = Eurotiomycetes, blue = Sordariomycetes, orange = Leotiomyces, red = Basidiomycota). Reference taxa are labelled by order, family, genus and species, and isolates by sample identifier. Shaded boxes in (A) mark orders and families. Numbers at nodes are ultrafast bootstrap support values from 1,000 replicates, shown only on nodes with references.

Several placements were notable (Figure 1). Four isolates fell at distances consistent with genus-level distance of *Diplodia sapinea*, and four near *Rhizosphaera*, a genus that includes the needle-cast pathogen *R. kalkhoffii*. Eleven isolates formed a well-supported clade within Delphinella,

while nine isolates formed a distinct clade adjacent to *Talaromyces*, and five adjacent to *Mycocalicium*, in both cases at distances exceeding the confamilial median and with UFBoot 100, consistent with lineages not represented among described reference taxa. These are treated under open nomenclature (Table S2). Notably, only one isolate, BS06_120_435-5, was placed in the Mycosphaerellaceae family, the one containing *L. acicola* and *Dothistroma septosporum*.

Evolutionary structure of functional gene content

We first examined whether CAZy family composition differed among major fungal lineages. Principal component analysis of Hellinger-transformed profiles from all 80 genomes recovered 19.7% and 12.0% of variance on the first two axes (Figure 2). PC1 was correlated with total CAZyme content ($r = -0.79$; $r^2 = 0.63$) and separated yeast-like from filamentous fungi rather than distinguishing endophytes from pathogens. Among the 77 filamentous genomes, taxonomic class explained 21.4% of variation in CAZy family composition (PERMANOVA, $p < 0.001$). Multivariate dispersion did not differ significantly among classes (betadisper, $F = 1.91$, $p = 0.135$). When the analysis was restricted to the 67 study isolates, taxonomic class explained 23.1% of variance (PERMANOVA, $p < 0.001$). Phylogenetic relatedness also correlated strongly with CAZy family composition (Mantel test, $r = 0.71$, $p < 0.001$).

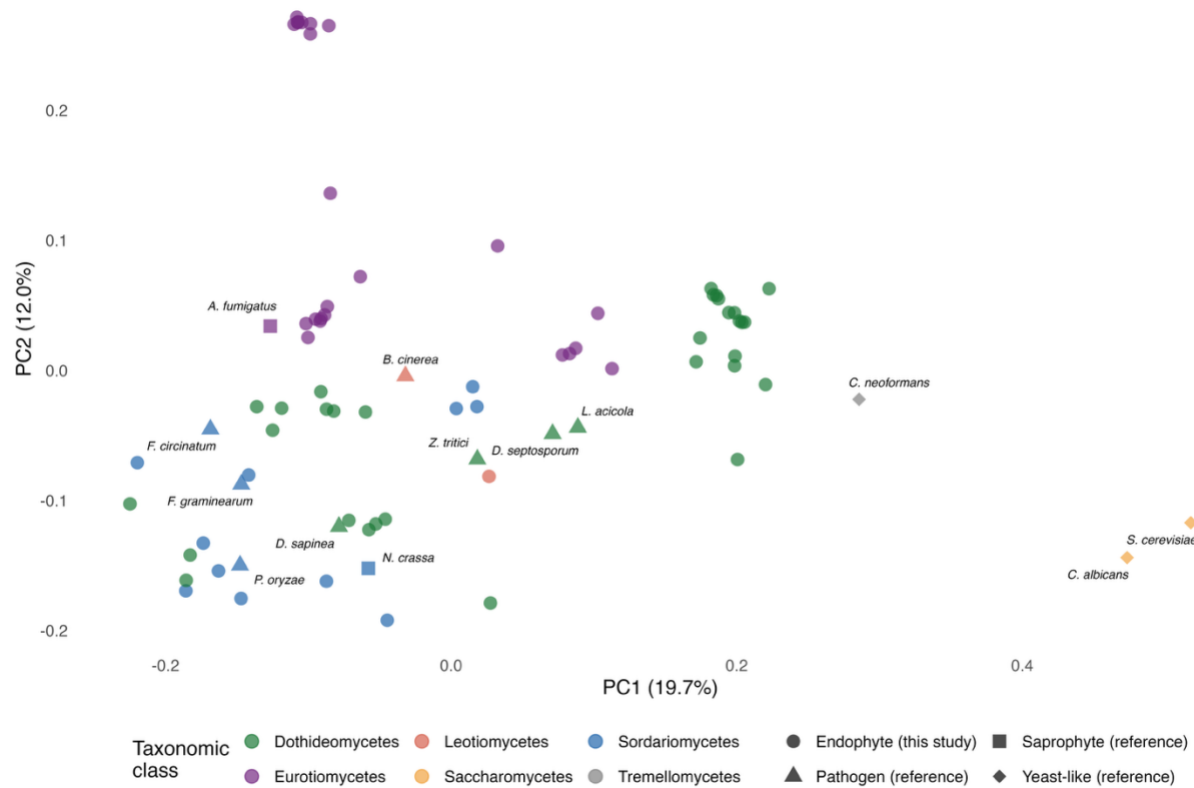


Figure 2. Principal component analysis of CAZy family composition across 80 fungal genomes, including the three yeast-like taxa. Family counts per 1,000 predicted genes were Hellinger transformed before ordination. Points are coloured by taxonomic class and shaped by ecological role, with reference genomes labelled. The first two axes recover 19.7% and 12.0% of total variance. PC1 correlates strongly with total CAZyme content ($r = -0.79$) and separates yeast-like from filamentous fungi. Statistical tests of composition excluded the yeast-like taxa ($n = 77$).

To test whether the taxonomic structure extended beyond CAZymes families, we combined CAZy family, COG category and biosynthetic gene cluster profiles into a block-weighted ordination of the 77 genomes with data for all three repertoires. The first two principal components explained 25.5% and 11.8% of variation, respectively (Figure 3). CAZy, COG, and biosynthetic gene cluster profiles contributed 33.1%, 34.7%, and 32.1% of the squared loadings on PC1, respectively. Functional composition in the combined matrix differed among taxonomic classes (PERMANOVA, $R^2 = 23.3\%$, $n = 77$, $p < 0.001$), while multivariate dispersion did not differ significantly among classes (betadisper, $F = 1.29$, $p = 0.285$). In separate

162 analyses, taxonomic class explained 31% of variation in COG category composition
163 (PERMANOVA, $p < 0.001$), compared with 21.4% for CAZy family composition and 22.6% for
164 biosynthetic gene cluster type composition (both PERMANOVA, $p < 0.001$). Phylogenetic
165 relatedness correlated with functional composition in all three individual repertoires and in the
166 combined profile (Mantel tests: CAZy $r = 0.71$, COG $r = 0.42$, BGC $r = 0.56$, combined $r = 0.68$,
167 all $p < 0.001$). Therefore, the combined signal reflected a blend of its three components,
168 weighted toward the two blocks (CAZy, BGC) most strongly structured by ancestry.

169 Furthermore, lifestyle was also significant in the combined matrix (PERMANOVA, $p < 0.05$),
170 although dispersion differed among lifestyle groups (betadisper, $p < 0.05$), so this result is not
171 interpreted independently. Analyzing each block separately attributed the signal to COG
172 category composition ($p < 0.05$), for which dispersion did not differ among lifestyle groups and
173 which remained significant when class was fitted first ($p < 0.05$) (Supplemental Figure 2).

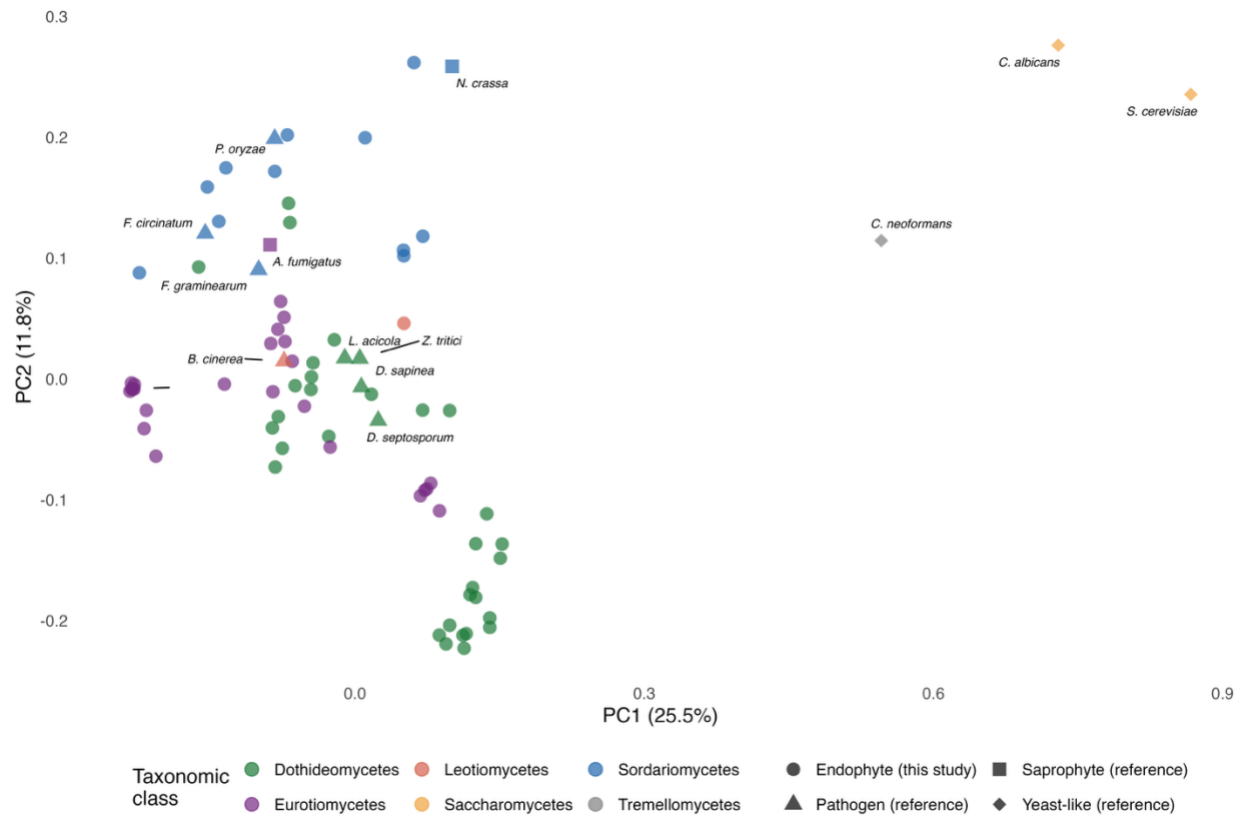


Figure 3. Block-weighted principal component analysis of combined functional profiles across the 80 genomes with data for all three repertoires. CAZy family, COG category and biosynthetic gene cluster type profiles were Hellinger transformed separately and weighted by the inverse of each block's first singular value before concatenation, so that no block dominated through its number of variables. The three blocks contribute near-equally to PC1 (33.1%, 34.7% and 32.1% of squared loadings for CAZy, COG and biosynthetic gene clusters, respectively). The first two axes recover 25.5% and 11.8% of variance.

Biosynthetic gene cluster counts differed by taxonomic class (Kruskal-Wallis, $\chi^2 = 20.30$, $df = 2$, $p < 0.001$; Figure 4). Eurotiomycetes carried a mean of 71.6 clusters per genome, more than twice as many as Dothideomycetes (31.3), with Sordariomycetes intermediate (53.1). This difference was driven by Eurotiomycetes, which differed significantly from both other classes (both $p < 0.05$), while Dothideomycetes and Sordariomycetes did not differ from each other after correction. In contrast to their low plant-cell-wall-degrading enzyme counts, the needle specialists fell within the range of BGC counts observed among the Dothideomycete study

isolates (median 30, range 20-48): *L. acicola* carried 42 predicted regions and *D. septosporum* carried 31 (Figure 4). *D. sapinea* and *Z. tritici* also fell within this range.

Because contig-edge truncation could in principle split a single cluster into multiple partial detections, we tested whether assembly fragmentation impacts the cluster counts above. Across the empirical isolates, the percentage of truncated regions correlated strongly with assembly contiguity (Spearman, contigs: $\rho = +0.931$, $p < 0.001$; N50: $\rho = -0.973$, $p < 0.001$), but total cluster count did not (contigs: $\rho = -0.025$, $p = 0.84$; N50: $\rho = +0.125$, $p = 0.32$). Grouping isolates into contiguity tertiles made this concrete: mean truncation rose from 17.7% in the most contiguous third to 89.5% in the least, while mean cluster count moved only from 46.6 to 41.6. Assembly fragmentation therefore substantially affected the completeness of predicted regions, although we did not detect a corresponding association with their total count, and we report total regions, including those truncated at contig edges.

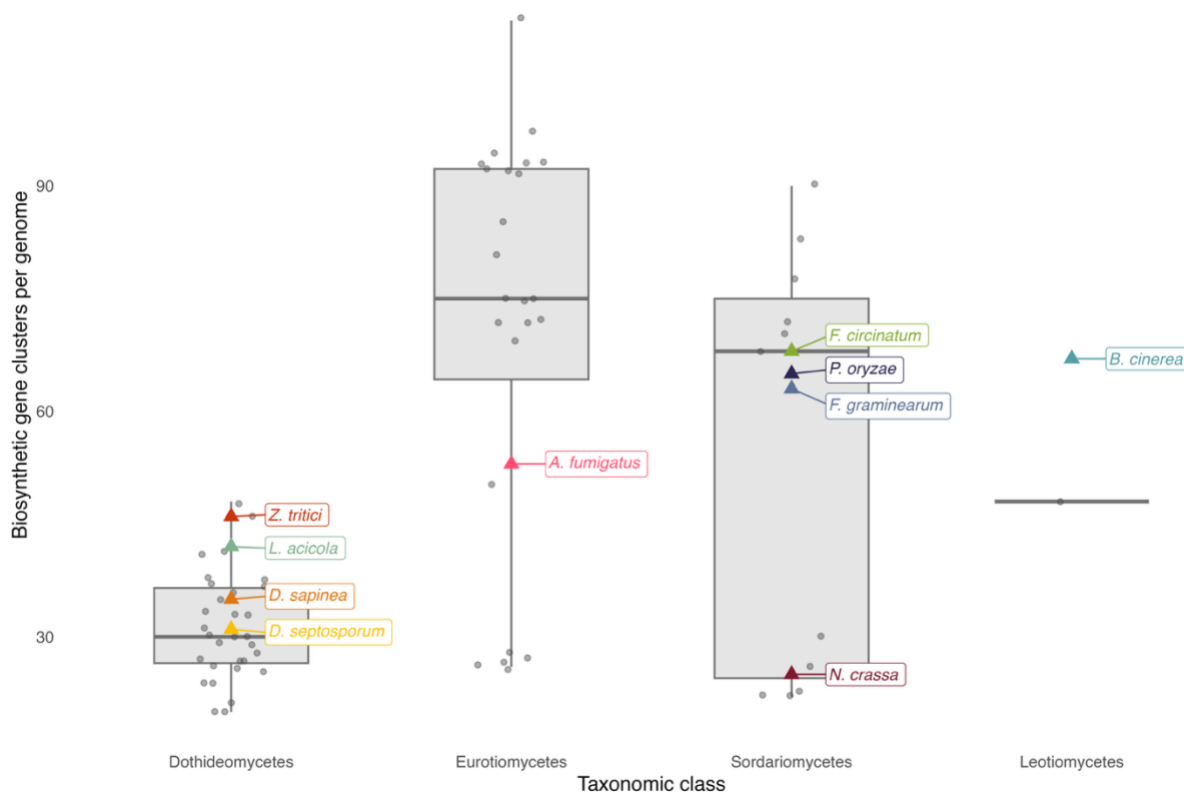


Figure 4. Biosynthetic gene cluster content by taxonomic class. Boxplots show the study isolates (Dothideomycetes n = 31, Eurotiomycetes n = 24, Sordariomycetes n = 11, Leotiomyces n = 1), with reference genomes overlaid as labelled triangles. Counts are total antiSMASH regions per genome, including regions truncated at contig edges. Cluster content differed among the three well-represented classes (Kruskal-Wallis, $p < 0.001$), driven by the higher content of Eurotiomycetes. Within Dothideomycetes, *L. acicola* (42 regions) and *D. septosporum* (31) fall within the isolate range (20 to 48).

The proportion of CAZymes predicted to be secreted was similar across taxonomic classes: 42.3% in Dothideomycetes, 45% in Eurotiomycetes, 43.5% in Leotiomyces, and 41.8% in Sordariomycetes (Figure 5; Kruskal-Wallis, $\chi^2 = 2.14$, $df = 2$, $p > 0.05$, with Leotiomyces excluded as a single isolate). Unlike overall CAZy family composition, the proportion of CAZymes predicted to be secreted therefore showed little variation among classes, although Leotiomyces was represented by only one isolate. Neither secretome size nor effector density differed among classes (Kruskal-Wallis, $\chi^2 = 3.80$ and 3.64 , $df = 2$, both $p > 0.05$), and secretome

size and effector count were nearly collinear across the collection (Spearman $\rho = 0.95$). Reference genomes nevertheless varied substantially in both measures. *Pyricularia oryzae* had 125 predicted secreted proteins and 63 predicted effectors per 1,000 genes, roughly 1.5 and 1.7 times the next highest pathogen reference (*Fusarium graminearum*, 85.5 and 36.8). Because the pathogen references were unevenly distributed among classes, this genome had a large influence on comparisons that pooled references across lineages.

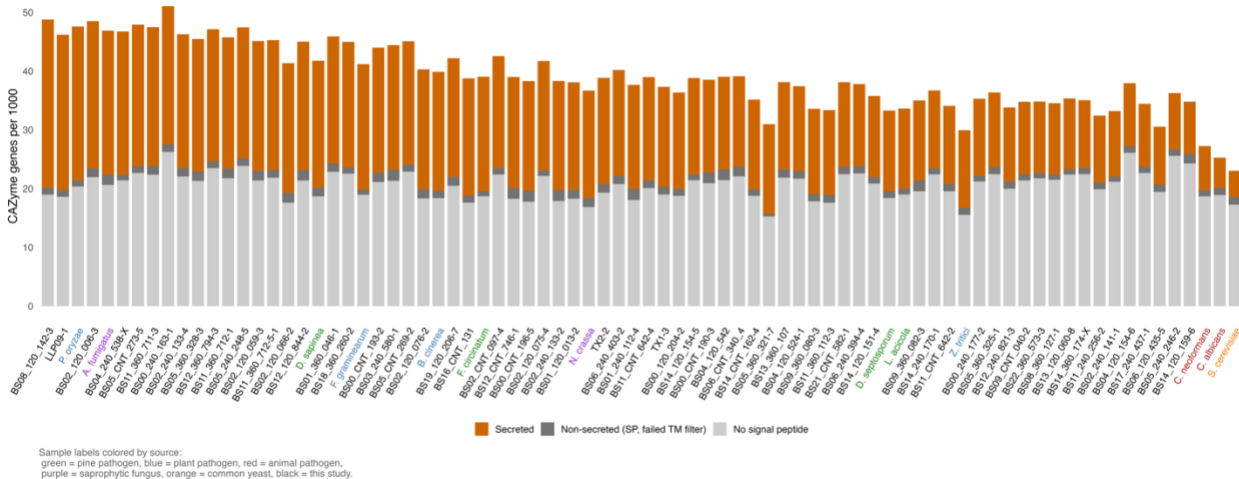


Figure 5. Secreted fraction of the CAZyme repertoire across 80 genomes. Bars show CAZyme genes per 1,000 predicted genes, divided into genes with a signal peptide and no predicted membrane-spanning segment (secreted), genes with a signal peptide and a predicted membrane-spanning segment, and genes without a predicted signal peptide. Signal peptides were predicted with SignalP and the transmembrane filter was applied with DeepTMHMM2. Genes without a signal peptide were not passed to the transmembrane filter and are shown separately. Sample labels are coloured by source. The mean secreted share among study isolates was 42.3% in Dothideomycetes, 45.0% in Eurotiomycetes, 43.5% in Leotiomycetes (one isolate) and 41.8% in Sordariomycetes, and did not differ among the three well-represented classes.

Functional placement of established pine pathogens

To place the needle pathogens *L. acicola* and *D. septosporum* in a narrower taxonomic context, we compared them with the 31 study isolates assigned to Dothideomycetes, the class that also contains the reference pathogens *D. sapinea* and *Z. tritici* (30 isolates for the EffectorP-derived

measures, because one isolate lacked EffectorP output). This restriction reduces the breadth of the comparison but does not fully account for phylogenetic relatedness within the class, since only one isolate falls in the Mycosphaerellales, the order containing *L. acicola* and *D. septosporum*. Mean total plant-cell-wall-degrading gene content, counting each gene once, differed little among the three well-represented study-isolate classes (14.9 genes per 1,000 predicted genes in Dothideomycetes, 15.0 in Eurotiomycetes and 16.8 in Sordariomycetes), although individual genomes varied within classes (ranges 9.3 to 22.0, 9.7 to 19.1 and 13.2 to 22.6, respectively).

Within Dothideomycetes, *L. acicola* and *D. septosporum* occupied the low end of the cellulose-associated, pectin-associated and total distributions (Figure 6, Supplemental Figure 3). They carried 2.32 and 2.85 cellulose-associated genes per 1,000 predicted genes, respectively, compared with a mean of 4.71 among the Dothideomycete isolates (range 2.86 to 7.16). No Dothideomycete isolate had a lower value than either pathogen, although *D. septosporum* fell only marginally below the lowest isolate. For pectin-associated genes (2.32 and 1.99, against a mean of 3.90 and a range of 1.31 to 5.51), only one isolate had a lower value than either pathogen (3.2nd percentile). For total genes (10.7 and 10.3, against a mean of 14.9 and a range of 9.3 to 22.0), two isolates and one isolate fell below *L. acicola* and *D. septosporum*, respectively (6.5th and 3.2nd percentiles). After Benjamini-Hochberg adjustment across the four traits tested, these reductions were significant for cellulose-associated, pectin-associated and total genes (Wilcoxon, $p < 0.05$ for each), whereas hemicellulose-associated counts were not reduced (35.5th and 29.0th percentiles). *D. sapinea*, in contrast, carried a larger degradative repertoire (17.7 total genes per 1,000, 67.7th percentile) that overlapped the isolate distribution, although its pectin-associated count (5.56) slightly exceeded the isolate maximum. *Z. tritici* was as low as

the needle specialists (10.2 total, 2.75 cellulose-associated and 1.80 pectin-associated genes per 1,000).

Because some CAZy families were assigned to more than one substrate category in our activity map, we repeated these comparisons counting only genes carrying at least one family assigned to a single category. Under this restriction, the reductions in cellulose-associated and pectin-associated genes persisted in both needle specialists (cellulose, 1.45 and 1.61 genes per 1,000 predicted genes against an isolate mean of 2.49 and a range of 1.43 to 3.67, pectin, 1.84 and 1.71 against a mean of 3.31 and a range of 0.95 to 4.89, adjusted $p < 0.05$). The separation was narrower than with all assigned families, since one isolate now fell below each pathogen. Total plant-cell-wall-degrading content was no longer reduced under this restriction and hemicellulose-associated counts were similar to the isolates, indicating that the reduction is specific to cellulose- and pectin-associated genes.

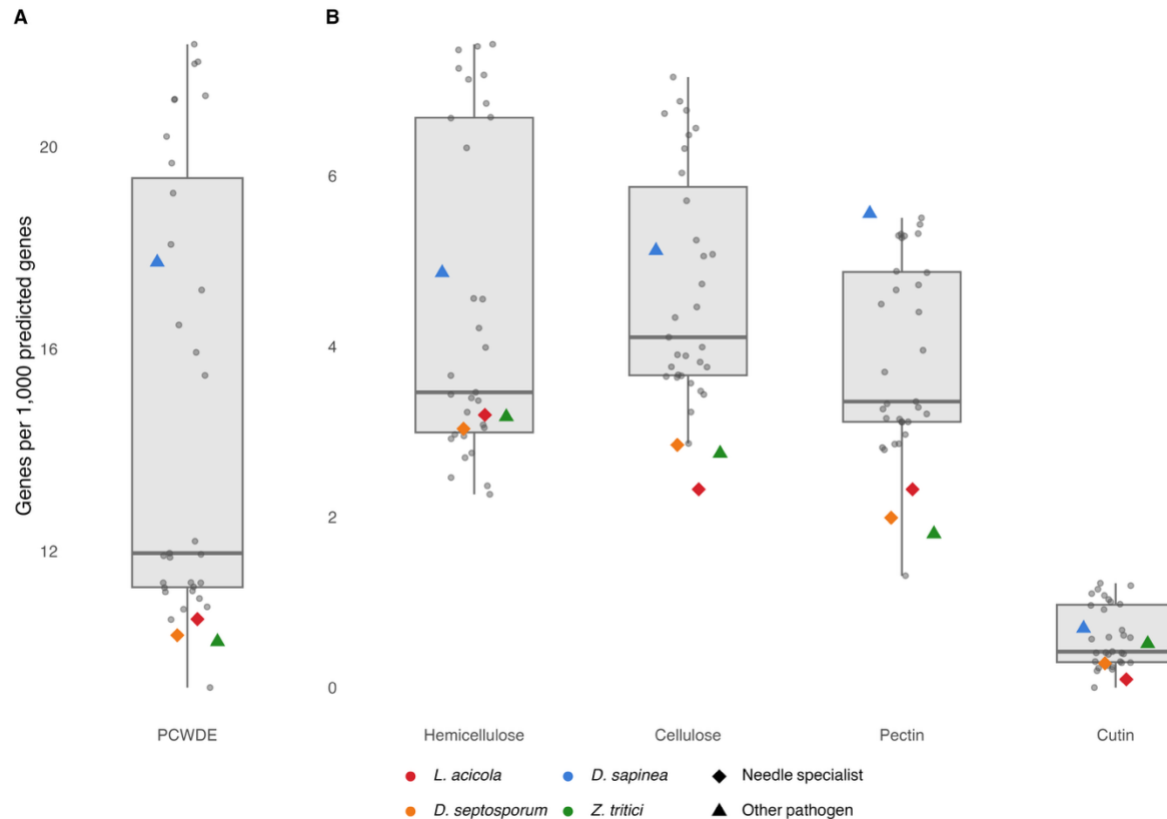


Figure 6. Plant-cell-wall-degrading (PCWDE) gene content of the needle pathogens relative to Dothideomycete study isolates. Panel A displays the total PCWDE content and Panel B displays divided by category. Comparisons are restricted to Dothideomycetes (boxplots and grey points, $n = 31$) to limit the influence of ancestry. Reference pathogens are overlaid, with the needle specialists *L. acicola* and *D. septosporum* as diamonds and the other Dothideomycete pathogens *D. sapinea* and *Z. tritici* as triangles. Values are genes per 1,000 predicted genes, and the total counts each gene once. Both needle specialists fall at or below the 7th percentile of the isolate distribution for cellulose-associated, pectin-associated and total genes (Wilcoxon, Benjamini-Hochberg adjusted $p < 0.05$ for each). Hemicellulose-associated counts were not reduced.

Discussion

The fungi recovered from symptomatic pine needles spanned four classes of Ascomycota, but their functional differences were not distributed randomly across that phylogeny. Taxonomic class accounted for 21.4% of variation in CAZy family composition among filamentous genomes, and class differences were also evident in combined CAZy, COG, and biosynthetic gene cluster profiles. Among the study isolates, more closely related genomes also had more

similar functional profiles in every repertoire examined, but the strength of this correspondence differed among repertoires. It was strongest for CAZy family composition (Mantel $r = 0.71$) and the combined profile ($r = 0.68$), intermediate for biosynthetic gene cluster types ($r = 0.56$) and COG categories ($r = 0.42$), and weakest for predicted effector classes ($r = 0.16$). These results suggest that evolutionary history helps organize functional variation in this collection, although to different degrees for different repertoires, while substantial differences remain among fungi within the same major lineage.

Phylogenetic structure also changes how the reference genomes can be interpreted. The study isolates and pathogen references were unevenly distributed among classes, and comparisons that pooled them across lineages produced apparent lifestyle differences that did not persist within Dothideomycetes. For biosynthetic gene cluster counts, the pooled comparison was dominated by 24 Eurotiomycete isolates with no pathogen counterpart among the references, and suggested lower cluster content in pathogens that was not seen within Dothideomycetes. For secretome size and effector density, the apparent excess in pathogens derived largely from a single reference genome, *Pyricularia oryzae*, and removing it alone eliminated the difference in secretome size and weakened that in effector density. Adjusting for class did not fully substitute for matching, because a model can hold class constant but cannot generate an isolate versus pathogen contrast within a class in which only one of the two is represented. The one lifestyle signal that persisted after class adjustment, in COG category composition, was not recovered within Dothideomycetes, although that comparison included only four pathogen genomes. These results suggest that unmatched designs in fungal comparative genomics may be more prone to this kind

of artifact than is commonly recognized, and we accordingly treat within-Dothideomycetes comparisons as the primary evidence of lifestyle association.

Within that matched comparison, the needle specialists carried fewer cellulose-associated and pectin-associated genes than the co-occurring isolates, a reduction that persisted when CAZy families shared among substrate categories were excluded. A similar pattern has been reported previously. de Wit et al. [9] reported that *Dothistroma septosporum* and *Cladosporium fulvum* both carry fewer predicted cellulolytic genes, and fewer carbohydrate-binding module genes than most comparator fungi, a pattern shared with *Zymoseptoria tritici*, which our data placed among the lowest in total plant cell wall degrading content (Figure 6). At a deeper taxonomic level, Ohm et al. [1] found Capnodiales, the lineage containing the Mycosphaerellaceae, to carry markedly reduced cellulose acting CAZymes relative to Pleosporales and Hysteriales. This correspondence qualifies our result in an important way, since part of the depletion we detect may represent an ancestral feature of the Mycosphaerellaceae rather than an adaptation to needle parasitism, and because only one of our isolates falls within the Mycosphaerellales, ancestry and lifestyle cannot be fully separated for these two species even under a matched design. Resolving this would require genomes from non-pathogenic Mycosphaerellales occupying the same tissue, which would distinguish an ancestral reduction from a lifestyle associated one.

Two mechanistic explanations are commonly invoked for reduced degradative capacity in foliar pathogens, and our data are compatible with both. Both needle specialists enter the host through stomata rather than by enzymatic breach of the cuticle, with *Lecanosticta acicola* germ tubes entering through stomatal openings without forming appressoria [12–14] and *Dothistroma*

septosporum colonising the epistomatal chambers [15], such that a fungus avoiding forced entry and growing intercellularly could require less bulk wall degrading capacity. *Zymoseptoria tritici* offers a supporting parallel as a stomatal invader that is likewise cellulose CAZyme reduced [16], although we could not locate studies demonstrating a causal link between penetration mode and enzyme loss, and *Z. tritici* still deploys these enzymes at its necrotrophic transition. Both species also produce phytotoxins, dothistromin in *D. septosporum* [17,18] and the heat stable LA-I and LA-II in *L. acicola* [19,20], raising the possibility that toxin mediated killing reduces dependence on enzymatic degradation, and our observation that biosynthetic gene cluster content is not reduced in either pathogen is compatible with such a substitution. Still, this interpretation is weakened by evidence that dothistromin is dispensable for pathogenicity [21], by the observation that killing a cell with a toxin does not remove its wall, and by the lack of gene cluster level characterization for LA-I and LA-II. Therefore, we present both explanations as consistent with our observations rather than as inferences the present data support.

One possible explanation for the relatively large predicted degradative repertoires of many study isolates is that fungi inhabiting living needles may also use plant material after tissue death. Conifer needle inhabitants frequently remain latent in living tissue and deploy degradative enzymes once the needle matures, as suggested by the lignocellulose-degrading secretome of *Phialocephala scopiformis* grown on pine wood [22] and by metatranscriptomic evidence that needle mycobiota express CAZymes specifically during the necrotrophic phase [23]. Fungi that must decompose a needle after it dies retain the enzymes to do so, whereas specialists feeding on living mesophyll may not. *Diplodia sapinea* follows the same logic, since it carried substantially more plant cell wall degrading enzymes than the other Dothideomycete pathogens, consistent

with its behavior as a latent endophyte that switches to necrotrophy under drought and wounding [24,25]. Our results therefore suggest that a large degradative repertoire is not in itself evidence of benign behavior.

The ecological roles of the study isolates remain unknown. Recovery from symptomatic needle tissue establishes their presence at the time of sampling, but not whether they were contributing to disease, persisting without symptoms, or using dying tissue. Our lifestyle comparisons therefore contrast fungi of undetermined role with references of established pathogenicity, rather than confirmed endophytes with pathogens. This asymmetry makes the comparisons conservative. If some isolates are themselves latent or opportunistic pathogens, their inclusion should reduce rather than create differences between the groups, so the reductions in the needle specialists are unlikely to be an artifact of misclassification. The same reasoning does not strengthen the null results, which remain consistent with either the absence of a difference or insufficient contrast between groups. Placement alone cannot resolve these roles either. Four isolates were placed within genus-level distance of *D. sapinea* and four within *Rhizosphaera*, a genus containing the needle cast pathogen *R. kalkhoffii*, but these placements do not establish pathogenicity in any isolate. Our culturing approach also recovered only part of the diversity detected in culture-independent surveys [5,26] and predicted gene quantities do not directly correlate to enzymatic activity [9], and overlap between the needle specialists and the isolates in several repertoires does not establish functional equivalence, since the matched comparisons rest on two pathogen genomes.

Predicted gene content alone cannot determine which fungi contributed to the symptoms on the needles we sampled. Closely related fungi may use similar repertoires differently during host colonization [27], and genes involved in infection may occur in rapidly evolving genomic regions that broad comparisons do not resolve well [28]. This uncertainty matters in a system where recent brown spot needle blight outbreaks have raised questions about the fungi present in affected pine needles [5], and where native fungi can become damaging under changing conditions [29,30]. The community sampled and the genomes reported here provide a phylogenetic and functional baseline for addressing those questions. Pathogenicity assays can establish which isolates cause disease, while expression studies across stages of colonization can determine when their predicted degradative, secretory, and secondary metabolic repertoires are used. Until then, the boundary between the fungi that cause these symptoms and those that merely accompany it is better treated as open than settled, which itself argues for monitoring whole needle mycobiomes and their responses to stress rather than exclusively screening against a predetermined list of pathogens.

Conclusion

Overall, functional gene content in fungi cultured from symptomatic pine needles tracked phylogeny. Taxonomic class structured carbohydrate-active enzyme, COG, and biosynthetic gene cluster composition, and phylogenetic relatedness corresponded with functional similarity in every repertoire examined, most strongly for carbohydrate-active enzymes and least for predicted effectors. From this perspective, pooled comparisons of isolates and reference pathogens across lineages produced lifestyle differences that dissipated when the assessment was restricted to a single class, indicating that phylogenetically matched comparisons are needed

before gene content is interpreted in terms of lifestyle. Within Dothideomycetes, the needle pathogens *Lecanosticta acicola* and *Dothistroma septosporum* were distinguished from co-occurring fungi by fewer cellulose- and pectin-associated genes, a reduction that may be partly ancestral to the Mycosphaerellaceae, whereas their secondary metabolite, secretory, and effector repertoires fell within the range of the isolates. Because the ecological roles of the cultured fungi were not determined, these genomes cannot identify which contributed to disease, but they provide a phylogenetic and functional baseline for the pathogenicity and expression studies that can.

Materials and Methods

Sample isolation, culturing, and sequencing

Loblolly (*Pinus taeda* L.) and longleaf pine (*P. palustris* Mill.) needles displaying characteristic symptoms of brown spot needle blight, as defined by the European and Mediterranean Plant Protection Organization [31], were collected from sites across Alabama between March and August in 2023. Collected needles had yellow-margined, dark brown center discolorations progressing to dead tips, central spotted zones, and green bases. Needle material was placed in individual paper bags and held on ice until processing. Recorded metadata included host species, collection date, geographic coordinates, and a tree-level disease severity assessment.

Fungi isolation followed the protocol of Folorunso et al. [32]. Briefly, needles were cut into 2-3 cm sections and surface sterilized by sequential submersion in 8% sodium hypochlorite and 70% ethanol for 30 s each, with sterile distilled water rinses between and after each submersion. Sections were plated on malt extract agar and incubated at room temperature for up to 20 days.

430 Emerging colonies were subcultured onto the same medium until pure cultures free of visible
431 contamination were obtained. For DNA extraction, isolates were grown in liquid yeast extract
432 peptone dextrose medium for 10 days. Mycelia were collected by centrifugation and washed with
433 sterile deionized water. Genomic DNA was extracted using a modified cetyltrimethylammonium
434 bromide (CTAB) protocol optimized for fungal tissues rich in secondary metabolites and
435 polysaccharides [32]. No shearing or size-selection step was performed. DNA was quantified
436 with a Quantus fluorometer using the QuantiFluor ONE dsDNA system (Promega, Madison, WI,
437 USA).

438

439 Long-read libraries were prepared with the Native Barcoding Kit SQK-NBD114.24 (Oxford
440 Nanopore Technologies, Oxford, UK) following the manufacturer's standard protocol. Libraries
441 were sequenced on R10.4.1 flow cells (FLO-MIN114) using a MinION Mk1C or Mk1D for up
442 to 72 h, with 6 samples multiplexed per flow cell. Raw signal data were basecalled with Dorado
443 v0.9.1 (Oxford Nanopore Technologies) using the super-accuracy model v5.0.0. Reads were
444 filtered with Chopper v0.11.0 [33] to remove reads shorter than 1,000 bp (*--minlength 1000*); no
445 fixed number of bases was trimmed from read ends. Read quality and yield were summarized
446 with NanoPlot [33]. Genomes were assembled with Flye [34] in *--nano-hq* mode with *--min-*
447 *overlap 1000*, followed by two rounds of polishing with Medaka[35]. Medaka polished
448 assemblies were subsequently scaffolded by LINKS v2.0.1 [36].

449

450 Assembly contiguity and size were evaluated with QUAST v5.3.0 [37]. Gene-space
451 completeness was assessed with BUSCO v6.1.0 [38] against the fungi_odb12 lineage dataset.

Genome assemblies with a BUSCO completeness below 70% or a duplication rate exceeding 15 were excluded from downstream comparative analyses.

Short-read data were available for a subset of isolates, and was used to assess consensus accuracy by reference-free k-mer analysis using Merqury v1.4.1[39], rather than used for assembly or polishing. Hence all genomes received identical treatment and downstream comparisons of gene content were not confounded by differences in polishing. This approach is supported by the accuracy of R10.4.1 super-accuracy long-read-only assemblies [40] and by benchmarking of this workflow on fungal genomes [41], which showed that structure and gene content can be accurately obtained even at lower long-reads depth.

Phylogenomic inference and taxonomic placement

Because the isolates were not identified beyond colony morphology prior to sequencing, candidate reference genomes were selected using genome similarity searches. A representative genome assembly was downloaded from NCBI for each fungal genus, preferring reference assemblies where available. These were compiled into a skani database (*--slow* mode), and each isolate assembly was queried against it with skani v0.3.2 [42] in search mode. Because skani reports average nucleotide identity only above an approximate 80% detection threshold, this step was used to identify candidate close relatives and to define a tractable reference set for phylogenomic analysis, not to assign taxonomy directly. Isolates returning one or more matches had the matched reference assemblies taken for phylogenomic analysis. Since absence of a skani match indicates only that no reference in the database fell within skani's detection range rather

than absence of a relative, the isolates returning no match above threshold were retained in the analysis and closely related taxa were included for taxonomical refinement where possible.

Phylogenomic placement was performed by running BUSCO against the fungi_odb12 lineage dataset for all isolates and reference genomes and the results processed by the BUSCO phylogenomics pipeline [43]. The pipeline identified loci that were complete and single copy in at least 80% (*-psc 80*) of individuals, align each locus with MAFFT [44,45], trim each alignment with trimAl [46], and concatenate the alignments into a supermatrix. A maximum-likelihood phylogeny was inferred from the supermatrix with IQ-TREE 3.1.4 [47], using the best-fit substitution model selected by ModelFinder [48] (Q.YEAST+F+I+R10) and 1000 ultrafast bootstrap replicates [49]. The tree was rooted with Basidiomycota as outgroup and analyzed in R using ape [50] and phangorn [51]. Patristic distances were computed between all pairs of tips, and each isolate was assigned the lineage of its nearest reference genome by patristic distance.

To determine the finest taxonomic rank that could be assigned with confidence, distance thresholds were calibrated empirically from the reference genomes themselves. For every pair of references, we recorded the lowest rank they shared and took the median patristic distance for pairs sharing genus, family, order and class. An isolate was then assigned to the rank whose median exceeded its distance to the nearest reference. Isolates within a conservative threshold of 0.020 of a reference were considered species-level candidates. Support for each placement was taken as the UFBoot value of the smallest clade containing the isolate and at least one reference genome.

Taxonomic assignments were based on phylogenetic placement and comparison of patristic distances with those among reference taxa of established classification. Isolates were only assigned to the rank best supported by those comparisons and otherwise retained under open nomenclature.

Genome annotation

Publicly available genome assemblies were retrieved from NCBI for phylogenetic and functional comparisons (Table S2). The reference panel comprised pine needle and shoot pathogens (*Lecanosticta acicola*, *Dothistroma septosporum*, *Diplodia sapinea*, *Fusarium circinatum*), broad host-range plant pathogens (*Zymoseptoria tritici*, *Botrytis cinerea*, *Fusarium graminearum*, *Pyricularia oryzae*), saprotrophic taxa (*Neurospora crassa*, *Saccharomyces cerevisiae*) and animal pathogens as outgroups (*Aspergillus fumigatus*, *Cryptococcus neoformans*, *Candida albicans*). All reference genomes were re-annotated with the pipeline described below to ensure that comparisons of gene content reflected biological differences rather than differences in annotation method.

Genomes were annotated with Funannotate v1.8.16 [52]. Repetitive contigs were removed and remaining contigs sorted and soft-masked, after which genes were predicted ab initio with AUGUSTUS [53] and GeneMark-ES [54] and reconciled with EVidenceModeler [55]. Transfer RNAs were predicted with tRNAscan-SE [56]. Functional annotation was assigned using InterProScan5 [57], eggNOG-mapper [58], Pfam-A [59], MEROPS [60], UniProt/Swiss-Prot [61], and the full NCBI fungal transcript and protein sets.

Carbohydrate-active enzymes (CAZymes) were identified with dbCAN3 [62] through hmmscan against dbCAN HMMs with Funannotate's default parameters. Family assignments were parsed from the Funannotate GFF3 output using a custom Python script and grouped into substrate-based activity categories following published family-to-substrate assignments[6–8]. Categories distinguished plant-cell-wall-degrading activities (cellulose, hemicellulose, mannan, pectin) from fungal-cell-wall-active activities (chitin, β - and α -glucan), with additional categories for cuticle esterases, oxidative polysaccharide cleavage, auxiliary redox activities, and storage polysaccharides. Because several CAZy families act on more than one substrate, families mapping to multiple categories were retained in each and are identified as such; analyses were additionally repeated using only substrate-diagnostic families. Gene counts per category were de-duplicated at the gene level, so that a protein carrying multiple families assigned to the same category. Counts were expressed per 1,000 predicted genes to control for variation in gene-model number among assemblies.

Secreted proteins were predicted from the Funannotate-derived proteome of each genome, with all reference genomes processed identically. Signal peptides were identified with SignalP v6.0 [63] in eukarya mode using the fast model. Proteins carrying a predicted signal peptide were retained, and the corresponding mature sequences, with the predicted signal peptide removed, were submitted to DeepTMHMM2 [64] for transmembrane topology prediction.

Transmembrane assignments were parsed from the DeepTMHMM2 output using named topology segments and proteins containing one or more membrane-spanning segments, defined as transmembrane alpha helices (TMhelix) or transmembrane beta strands (Beta sheet), were

excluded. Membrane-associated but non-spanning elements predicted by DeepTMHMM2, namely interfacial helices and re-entrant regions, were not treated as disqualifying, because these associate with or penetrate the bilayer without traversing it and therefore do not preclude secretion; interfacial helices in particular are reported in approximately one quarter of alpha-helical transmembrane proteins DeepTMHMM2 [64], so treating them as transmembrane domains would remove a substantial number of legitimately secreted proteins.

The resulting predicted secretomes were classified with EffectorP-fungi 3.0 [65], which assigns secreted proteins as apoplastic effectors, cytoplasmic effectors, both, or non-effectors. Secretome size was expressed as an absolute count and as a percentage of the predicted proteome, and effector candidates were additionally normalized per 1,000 predicted genes to control for variation in gene-model number among assemblies. Biosynthetic gene clusters were identified with fungal antiSMASH 8.0.4 [66]. Clusters were classified by backbone type (NRPS, PKS, PKS-NRPS hybrid, terpene, and other) and counted per genome.

Statistical analyses and visualization

To evaluate how evolutionary history relates to functional gene content, we analyzed CAZyme families, COG categories, and biosynthetic gene cluster types as multivariate profiles, and summarized CAZyme activity categories, cluster counts, predicted secretomes, and effector repertoires as genome-level traits. We first evaluated variation within and among major fungal lineages using ordination and comparisons among taxonomic classes. We then examined the correspondence between phylogenetic relatedness and each functional profile using Mantel tests (Pearson correlation, 10k permutations) between pairwise patristic distances from the maximum-

likelihood phylogeny and the Euclidean distances among Hellinger-transformed profiles used for PERMANOVA (CAZyme family, COG category, biosynthetic gene cluster type, effector class composition and the combined profile). Because patristic distances were available for study isolates only, these tests included 67 isolates (66 for effector composition). Finally, we placed established pathogen genomes within the functional distributions of the sequenced isolates, with comparisons involving the needle pathogens restricted to Dothideomycetes.

For analyses of functional composition, gene or cluster counts within each genome were converted to relative abundances and Hellinger-transformed (square-root of relative abundance [67]). CAZyme family, COG category, and biosynthetic gene cluster type profiles were analyzed separately. Principal component analysis (PCA) was used to visualize CAZy family composition, COG and biosynthetic gene cluster ordinations. To examine broader functional composition, the three transformed profiles were weighted separately by the inverse of each block's first singular value, concatenated, and also analyzed by PCA. This weighting reduced differences in the contributions of the three data types to the combined ordination.

Differences in functional composition among taxonomic classes and lifestyles were tested with PERMANOVA [68] on Euclidean distances among Hellinger-transformed profiles, using *adonis2()* in the vegan package v2.8 [69] with 10k permutations. Class and lifestyle were tested separately and sequentially, with class entered first so that lifestyle was evaluated after accounting for class. Homogeneity of multivariate dispersion was tested with *betadisper()* for each grouping factor, and significant heterogeneity was taken to limit the interpretation of the corresponding PERMANOVA. To compare lifestyles without the confound of ancestry, CAZy

and COG profiles were also analyzed within Dothideomycetes alone (31 isolates and 4 pathogen references). Yeast-like taxa (*Candida albicans*, *Saccharomyces cerevisiae* and *Cryptococcus neoformans*) were retained in the ordinations but excluded from these tests.

For genome-level traits, CAZyme genes were grouped into the substrate-based activity categories described above and expressed per 1,000 predicted genes. Secretome size and predicted effector counts were likewise expressed per 1,000 predicted genes, and the effector fraction of the secretome was calculated as the proportion of transmembrane-filtered secreted proteins classified as effectors. The secreted fraction of the CAZyme repertoire was expressed as the percentage of CAZyme genes carrying a predicted signal peptide and passing transmembrane filtering. Biosynthetic gene clusters were counted as antiSMASH regions per genome, including regions truncated at contig edges. Total cluster count, effector density, secretome size and the secreted fraction of the CAZyme repertoire were compared among the three well-represented isolate classes using Kruskal-Wallis tests, with pairwise Wilcoxon rank-sum tests (continuity correction, Benjamini-Hochberg adjustment) where reported. Leotiomycetes, represented by a single isolate, were excluded from these tests. Associations between assembly contiguity (contig count and N50) and the percentage of truncated cluster regions or total cluster count, and between secretome size and effector count, were assessed with Spearman correlations. Differences among classes in the plant-cell-wall-degrading activity categories were summarized descriptively.

Established pathogen reference genomes were overlaid on the functional ordinations and trait distributions. To examine the position of the needle-associated pathogens *Lecanosticta acicola*

and *Dothistroma septosporum* in a narrower taxonomic context, we compared their genome-level traits with those of the study isolates assigned to Dothideomycetes. For each trait, the two needle specialists were compared with the Dothideomycete study isolates using a two-sided Wilcoxon rank-sum test, with Benjamini-Hochberg adjustment across the four traits tested (cellulose, pectin, hemicellulose and total plant-cell-wall-degrading genes). Percentiles were reported for the plant cell wall degrading activity categories (cellulose, hemicellulose, pectin, cutin and total plant cell wall degrading content), total biosynthetic gene cluster count, the secreted fraction of the CAZyme repertoire, and the EffectorP-derived measures of secretome size, effector density and effector fraction of the secretome. Values and percentiles for the other Dothideomycete pathogen references, *Diplodia sapinea* and *Zymoseptoria tritici*, were reported alongside those of the needle specialists where computed.

All statistical analyses and visualization were performed in R v4.6.1 [70]. Phylogenetic tree was produced with TreeViewer [71], heatmaps with ComplexHeatmap [72,73] and remaining figures with ggplot2 [74]. All analysis scripts are available on GitHub.

Additional Files

Silva_Comparative_Fungi_SuppFigures

File format: DOCX document with images

Title of data: Supplementary Figures

632 Description of data: Document containing the supplementary figures cited in this
633 manuscript and their caption.

634 Silva_Comparative_Fungi_SuppTables

635 File format: DOCX document with tables

636 Title of data: Supplementary Tables

637 Description of data: Document containing the supplementary tables cited in this
638 manuscript and their caption.

639

640 **Declarations**

641 *Ethics approval and consent to participate*

642 Not applicable

643

644 *Consent for publication*

645 Not applicable

646

647 *Availability of data and materials*

648 Raw sequencing reads, genome assemblies and annotations are deposited in the NCBI Sequence

649 Read Archive and GenBank under BioProject PRJNA1474061.

650 Code is available on “github.com/bielasilva/comparative_fungi_pine_needle”.

651

652 *Competing interests*

653 The authors declare that they have no competing interests

654

Funding

This project was supported by funding from the U.S. Forest Service (Award Number 22-CA-11330160-067, L.G.E. and J.R.W.) and in part by the Alabama Agricultural Experiment Station of the U.S. Department of Agriculture National Institute of Food and Agriculture (Hatch projects 1025651 and 7011221, JRW).

Authors' contributions

G.A.A.S. and J.R.W. conceptualized the study and planned the study design. Sample collection, culturing, DNA extraction, and sequencing were done by G.A.A.S., T.R.F., and L.G.E. Genome assembly, annotation, phylogenomic placement, comparative and statistical analyses, and figure preparation were done by G.A.A.S. Supervision and resource acquisition were done by J.R.W. and L.G.E. The original draft was written by G.A.A.S., and subsequent revisions and editing were done by all authors. All authors read and approved the final manuscript.

Acknowledgements

We thank members of the Willoughby Lab for constructive comments on earlier versions of this work. In addition, we are grateful to the many field assistants that worked collecting the original fungal samples used in our sequencing efforts. We also thank Mick Persyn, Samantha Huey and Tess Lindow who assisted with fungal culturing and DNA extractions. This work was completed in part with resources provided by the Auburn University Easley Cluster.

References

1. Ohm RA, Feau N, Henrissat B, Schoch CL, Horwitz BA, Barry KW, et al. Diverse Lifestyles and Strategies of Plant Pathogenesis Encoded in the Genomes of Eighteen

679 Dothideomycetes Fungi. Andrianopoulos A, editor. PLoS Pathog. 2012;8:e1003037.
680 <https://doi.org/10.1371/journal.ppat.1003037>

681 2. Siggers PV. The Brown Spot Needle Blight of Pine Seedlings. 1944;41.
682 <https://doi.org/10.22004/AG.ECON.169629>

683 3. Jewell, Sr. FF. Histopathology of the Brown Spot Fungus on Longleaf Pine Needles.
684 Phytopathology. 1983;73:854. <https://doi.org/10.1094/Phyto-73-854>

685 4. Silva G, Folorunso TR, Paez AM, Lamka GF, Singh S, King J, et al. Biology, Detection, and
686 Management of *Lecanosticta acicola* : Emerging Challenges for the Management of Brown
687 Spot Needle Blight in Pine Forests. Forest Pathology. 2026;56:e70061.
688 <https://doi.org/10.1111/efp.70061>

689 5. Meinecke CD, Niyas AMM, McCarty E, Quesada T, Smith JA, Villari C. Searching the *Pinus*
690 *taeda* Foliar Mycobiome for Emerging Pathogens Among Brown-Spot Needle Blight and
691 Needlecast Outbreaks in the Southeast United States. Phytobiomes Journal. 2024;8:512–
692 28. <https://doi.org/10.1094/PBIOMES-02-24-0013-R>

693 6. Zhao Z, Liu H, Wang C, Xu J-R. Comparative analysis of fungal genomes reveals different
694 plant cell wall degrading capacity in fungi. BMC Genomics. 2013;14:274.
695 <https://doi.org/10.1186/1471-2164-14-274>

696 7. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-
697 active enzymes database (CAZy) in 2013. Nucl Acids Res. 2014;42:D490–5.
698 <https://doi.org/10.1093/nar/gkt1178>

699 8. Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-
700 active enzyme database: functions and literature. Nucleic Acids Research. 2022;50:D571–
701 7. <https://doi.org/10.1093/nar/gkab1045>

702 9. De Wit PJGM, Van Der Burgt A, Ökmen B, Stergiopoulos I, Abd-Elsalam KA, Aerts AL, et al.
703 The Genomes of the Fungal Plant Pathogens *Cladosporium fulvum* and *Dothistroma*
704 *septosporium* Reveal Adaptation to Different Hosts and Lifestyles But Also Signatures of
705 Common Ancestry. Barsh GS, editor. PLoS Genet. 2012;8:e1003088.
706 <https://doi.org/10.1371/journal.pgen.1003088>

707 10. De Jonge R, Peter Van Esse H, Kombrink A, Shinya T, Desaki Y, Bours R, et al. Conserved
708 Fungal LysM Effector Ecp6 Prevents Chitin-Triggered Immunity in Plants. Science.
709 2010;329:953–5. <https://doi.org/10.1126/science.1190859>

710 11. Bradshaw RE. Dothistroma (red-band) needle blight of pines and the dothistromin
711 toxin: a review. Forest Pathology. 2004;34:163–85. <https://doi.org/10.1111/j.1439-0329.2004.00356.x>

- 713 12. van der Nest A, Wingfield MJ, Janoušek J, Barnes I. *Lecanosticta acicola*: A growing
714 threat to expanding global pine forests and plantations. *Molecular Plant Pathology*. 2019;
715 <https://doi.org/10.1111/mpp.12853>
- 716 13. Setliff EC, Patton RF. Germination Behavior of *Scirrhia acicola* Conidia on Pine Needles.
717 *Phytopathology*. 1974;64:1462–4. <https://doi.org/10.1094/Phyto-64-1462>
- 718 14. Kais AG. Environmental Factors Affecting Brown-Spot Infection on Longleaf Pine.
719 *Phytopathology*. 1975;65:1389. <https://doi.org/10.1094/Phyto-65-1389>
- 720 15. Gadgil PD. Infection of *Pinus radiata* needles by *Dothistroma pini*. *New Zealand Journal*
721 *of Botany*. 1967;5:498–503. <https://doi.org/10.1080/0028825X.1967.10428769>
- 722 16. Brunner PC, Torriani SFF, Croll D, Stukenbrock EH, McDonald BA. Coevolution and Life
723 Cycle Specialization of Plant Cell Wall Degrading Enzymes in a Hemibiotrophic Pathogen.
724 *Molecular Biology and Evolution*. 2013;30:1337–47.
725 <https://doi.org/10.1093/molbev/mst041>
- 726 17. Zhang S, Schwelm A, Jin H, Collins LJ, Bradshaw RE. A fragmented aflatoxin-like gene
727 cluster in the forest pathogen *Dothistroma septosporum*. *Fungal Genetics and Biology*.
728 2007;44:1342–54. <https://doi.org/10.1016/j.fgb.2007.06.005>
- 729 18. Bradshaw RE, Slot JC, Moore GG, Chettri P, De Wit PJGM, Ehrlich KC, et al.
730 Fragmentation of an aflatoxin-like gene cluster in a forest pathogen. *New Phytologist*.
731 2013;198:525–35. <https://doi.org/10.1111/nph.12161>
- 732 19. Yang B, Ye J, Bao H, Lui jiKai, Dong Z. Studies on the phytotoxic activities of LA- I and
733 LA- II producing by brown spot needle blight fungus (*Lecanosticta acicola*). *Scientia Silvae*
734 *Sinicae*. Chinese Academy of Forestry; 2002;38:84–8.
- 735 20. Yang B, Ye J, Bao H, Lui jiKai, Dong Z. Separation and purification, structure of LA - I
736 toxin produced by brown spot needle blight fungus (*Lecanosticta acicola*). *Scientia Silvae*
737 *Sinicae*. Beijing: Chinese Academy of Forestry; 2005;41:86–90.
- 738 21. Kabir MS, Ganley RJ, Bradshaw RE. Dothistromin toxin is a virulence factor in
739 dothistroma needle blight of pines. *Plant Pathology*. 2015;64:225–34.
740 <https://doi.org/10.1111/ppa.12229>
- 741 22. Bhatnagar JM, Sabat G, Cullen D. The Foliar Endophyte *Phialocephala scopiformis*
742 DAOMC 229536 Proteome When Grown on Wood Used as the Sole Carbon Source. Stajich
743 JE, editor. *Microbiol Resour Announc*. 2019;8:e01280-18.
744 <https://doi.org/10.1128/MRA.01280-18>
- 745 23. Ata JP, Ibarra Caballero JR, Abdo Z, Mondo SJ, Stewart JE. Transitions of foliar mycobiota
746 community and transcriptome in response to pathogenic conifer needle interactions. *Sci*
747 *Rep*. 2022;12:7832. <https://doi.org/10.1038/s41598-022-11907-0>

748 24. Stanosz GR, Blodgett JT, Smith DR, Kruger EL. Water stress and *Sphaeropsis sapinea* as
749 a latent pathogen of red pine seedlings. *New Phytologist*. 2001;149:531–8.
750 <https://doi.org/10.1046/j.1469-8137.2001.00052.x>

751 25. Blumenstein K, Bußkamp J, Langer GJ, Langer EJ, Terhonen E. The Diplodia Tip Blight
752 Pathogen *Sphaeropsis sapinea* Is the Most Common Fungus in Scots Pines' Mycobiome,
753 Irrespective of Health Status—A Case Study from Germany. *JoF*. 2021;7:607.
754 <https://doi.org/10.3390/jof7080607>

755 26. Mesanza N, Aitken J, Uria A, Iturrutxa E. Fungal Communities in Asymptomatic and
756 Symptomatic Needles of *Pinus* spp. Affected by Pine Needle Diseases. *Microorganisms*.
757 2025;14:88. <https://doi.org/10.3390/microorganisms14010088>

758 27. Muñoz-Barrios A, Sopena-Torres S, Ramos B, López G, Del Hierro I, Díaz-González S, et
759 al. Differential Expression of Fungal Genes Determines the Lifestyle of *Plectosphaerella*
760 Strains During *Arabidopsis thaliana* Colonization. *MPMI*. 2020;33:1299–314.
761 <https://doi.org/10.1094/MPMI-03-20-0057-R>

762 28. Shaw PL, Slippers B, Wingfield BD, Laurent B, Penaud B, Wingfield MJ, et al.
763 Chromosome-level genome assemblies for the latent pine pathogen, *Diplodia sapinea* ,
764 reveal two accessory chromosomes with distinct genomic features and evolutionary
765 dynamics. LaBella A, editor. *G3: Genes, Genomes, Genetics*. 2025;15:jkaf239.
766 <https://doi.org/10.1093/g3journal/jkaf239>

767 29. Burgess TI, Oliva J, Sapsford SJ, Sakalidis ML, Balocchi F, Paap T. Anthropogenic
768 Disturbances and the Emergence of Native Diseases: a Threat to Forest Health. *Curr*
769 *Forestry Rep*. 2022;8:111–23. <https://doi.org/10.1007/s40725-022-00163-0>

770 30. Ghelardini L, Pepori AL, Luchi N, Capretti P, Santini A. Drivers of emerging fungal
771 diseases of forest trees. *Forest Ecology and Management*. 2016;381:235–46.
772 <https://doi.org/10.1016/j.foreco.2016.09.032>

773 31. EPPO. PM 7/46 (3) *Lecanosticta acicola* (formerly *Mycosphaerella dearnessii*),
774 *Dothistroma septosporum* (formerly *Mycosphaerella pini*) and *Dothistroma pini*. EPPO
775 Bulletin. 2015;45:163–82. <https://doi.org/10.1111/epp.12217>

776 32. Folorunso TR, Silva G, Girón ME, Lindow T, Persyn M, Eckhardt L, et al. Optimized
777 protocol for culturing and extracting DNA from fungal isolates associated with brown spot
778 needle blight in pine trees. *PLoS One*. 2025;20:e0337218.
779 <https://doi.org/10.1371/journal.pone.0337218>

780 33. De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read
781 sequencing data. Alkan C, editor. *Bioinformatics*. 2023;39:btad311.
782 <https://doi.org/10.1093/bioinformatics/btad311>

783 34. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using
784 repeat graphs. *Nat Biotechnol.* 2019;37:540–6. <https://doi.org/10.1038/s41587-019-0072-8>

785 35. Medaka [Internet]. Oxford Nanopore Technologies;
786 <https://github.com/nanoporetech/medaka>

787 36. Warren RL, Yang C, Vandervalk BP, Behsaz B, Lagman A, Jones SJM, et al. LINKS:
788 Scalable, alignment-free scaffolding of draft genomes with long reads. *Gigascience.*
789 2015;4:s13742-015-0076–3. <https://doi.org/10.1186/s13742-015-0076-3>

790 37. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome
791 assembly evaluation with QUAST-LG. *Bioinformatics.* 2018;34:i142–50.
792 <https://doi.org/10.1093/bioinformatics/bty266>

793 38. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO Update: Novel and
794 Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring
795 of Eukaryotic, Prokaryotic, and Viral Genomes. Kelley J, editor. *Molecular Biology and*
796 *Evolution.* 2021;38:4647–54. <https://doi.org/10.1093/molbev/msab199>

797 39. Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality,
798 completeness, and phasing assessment for genome assemblies. *Genome Biol.*
799 2020;21:245. <https://doi.org/10.1186/s13059-020-02134-9>

800 40. Sanderson ND, Hopkins KMV, Colpus M, Parker M, Lipworth S, Crook D, et al. Evaluation
801 of the accuracy of bacterial genome reconstruction with Oxford Nanopore R10.4.1 long-
802 read-only sequencing. *Microbial Genomics [Internet].* 2024 [cited 2025 Oct 7];10.
803 <https://doi.org/10.1099/mgen.0.001246>

804 41. Silva GADA, Folorunso TR, Eckhardt LG, Willoughby JR. Quantitative comparison of
805 fungal genome assembly strategies using short and long reads from simulated and
806 empirical sequencing data. *Microbial Genomics [Internet].* 2026 [cited 2026 Sept 8];12.
807 <https://doi.org/10.1099/mgen.0.001824>

808 42. Shaw J, Yu YW. Fast and robust metagenomic sequence comparison through sparse
809 chaining with skani. *Nat Methods.* 2023;20:1661–5. [https://doi.org/10.1038/s41592-023-](https://doi.org/10.1038/s41592-023-02018-3)
810 [02018-3](https://doi.org/10.1038/s41592-023-02018-3)

811 43. McGowan J. BUSCO_Phylogenomics [Internet]. 2026.
812 https://github.com/jamiemcg/BUSCO_phylogenomics

813 44. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7:
814 Improvements in Performance and Usability. *Molecular Biology and Evolution.*
815 2013;30:772–80. <https://doi.org/10.1093/molbev/mst010>

- 816 45. Nakamura T, Yamada KD, Tomii K, Katoh K. Parallelization of MAFFT for large-scale
817 multiple sequence alignments. Hancock J, editor. *Bioinformatics*. 2018;34:2490–2.
818 <https://doi.org/10.1093/bioinformatics/bty121>
- 819 46. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated
820 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25:1972–3.
821 <https://doi.org/10.1093/bioinformatics/btp348>
- 822 47. Wong TKF, Ly-Trong N, Ren H, Demotte P, Baños H, Roger AJ, et al. IQ-TREE 3:
823 phylogenomic inference software using complex evolutionary models. Tamura K, editor.
824 *Molecular Biology and Evolution*. 2026;43:msag117.
825 <https://doi.org/10.1093/molbev/msag117>
- 826 48. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: fast
827 model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14:587–9.
828 <https://doi.org/10.1038/nmeth.4285>
- 829 49. Minh BQ, Nguyen MAT, Von Haeseler A. Ultrafast Approximation for Phylogenetic
830 Bootstrap. *Molecular Biology and Evolution*. 2013;30:1188–95.
831 <https://doi.org/10.1093/molbev/mst024>
- 832 50. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and
833 evolutionary analyses in R. Schwartz R, editor. *Bioinformatics*. 2019;35:526–8.
834 <https://doi.org/10.1093/bioinformatics/bty633>
- 835 51. Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics*. 2011;27:592–3.
836 <https://doi.org/10.1093/bioinformatics/btq706>
- 837 52. Palmer JM, Stajich J. Funannotate v1.8.1: Eukaryotic genome annotation [Internet].
838 Zenodo; 2020 [cited 2024 Jan 26]. <https://doi.org/10.5281/ZENODO.4054262>
- 839 53. Keller O, Kollmar M, Stanke M, Waack S. A novel hybrid gene prediction method
840 employing protein multiple sequence alignments. *Bioinformatics*. 2011;27:757–63.
841 <https://doi.org/10.1093/bioinformatics/btr010>
- 842 54. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. Gene prediction in novel
843 fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res*.
844 2008;18:1979–90. <https://doi.org/10.1101/gr.081612.108>
- 845 55. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic
846 gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced
847 Alignments. *Genome Biol*. 2008;9:R7. <https://doi.org/10.1186/gb-2008-9-1-r7>
- 848 56. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional
849 classification of transfer RNA genes. *Nucleic Acids Research*. 2021;49:9077–96.
850 <https://doi.org/10.1093/nar/gkab688>

851 57. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-
852 scale protein function classification. *Bioinformatics*. 2014;30:1236–40.
853 <https://doi.org/10.1093/bioinformatics/btu031>

854 58. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-
855 mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the
856 Metagenomic Scale. Tamura K, editor. *Molecular Biology and Evolution*. 2021;38:5825–9.
857 <https://doi.org/10.1093/molbev/msab293>

858 59. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al.
859 Pfam: The protein families database in 2021. *Nucleic Acids Research*. 2021;49:D412–9.
860 <https://doi.org/10.1093/nar/gkaa913>

861 60. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. The MEROPS
862 database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison
863 with peptidases in the PANTHER database. *Nucleic Acids Research*. 2018;46:D624–32.
864 <https://doi.org/10.1093/nar/gkx1134>

865 61. The UniProt Consortium, Bateman A, Martin M-J, Orchard S, Magrane M, Ahmad S, et al.
866 UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*.
867 2023;51:D523–31. <https://doi.org/10.1093/nar/gkac1052>

868 62. Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y. dbCAN3: automated carbohydrate-active
869 enzyme and substrate annotation. *Nucleic Acids Research*. 2023;51:W115–21.
870 <https://doi.org/10.1093/nar/gkad328>

871 63. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al.
872 SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat*
873 *Biotechnol*. 2022;40:1023–5. <https://doi.org/10.1038/s41587-021-01156-3>

874 64. Teufel F, Hallgren J, Nielsen H, Krogh A, Tsirigos KD, Winther O. DeepTMHMM2 enables
875 accurate prediction of transmembrane protein topology and subcellular location [Internet].
876 *Bioinformatics*; 2026 [cited 2026 Sept 7]. <https://doi.org/10.64898/2026.08.24.746435>

877 65. Sperschneider J, Dodds PN. EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic
878 Effectors in Fungi and Oomycetes. *MPMI*. 2022;35:146–56. [https://doi.org/10.1094/MPMI-](https://doi.org/10.1094/MPMI-08-21-0201-R)
879 [08-21-0201-R](https://doi.org/10.1094/MPMI-08-21-0201-R)

880 66. Blin K, Shaw S, Vader L, Szenei J, Reitz ZL, Augustijn HE, et al. antiSMASH 8.0: extended
881 gene cluster detection capabilities and analyses of chemistry, enzymology, and regulation.
882 *Nucleic Acids Research*. 2025;53:W32–8. <https://doi.org/10.1093/nar/gkaf334>

883 67. Legendre P, Gallagher ED. Ecologically meaningful transformations for ordination of
884 species data. *Oecologia*. 2001;129:271–80. <https://doi.org/10.1007/s004420100716>

885 68. Anderson MJ. A new method for non-parametric multivariate analysis of variance.
886 Austral Ecology. 2001;26:32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>

887 69. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. vegan:
888 Community Ecology Package [Internet]. 2026 [cited 2026 Sept 13]. p. 2.7-6.
889 <https://doi.org/10.32614/CRAN.package.vegan>

890 70. R Core Team. R: A Language and Environment for Statistical Computing [Internet].
891 Vienna, Austria: R Foundation for Statistical Computing; 2024. <https://www.R-project.org/>

892 71. Bianchini G, Sánchez-Baracaldo P. TREEVIEWER : Flexible, modular software to visualise
893 and manipulate phylogenetic trees. Ecology and Evolution. 2024;14:e10873.
894 <https://doi.org/10.1002/ece3.10873>

895 72. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in
896 multidimensional genomic data. Bioinformatics. 2016;32:2847–9.
897 <https://doi.org/10.1093/bioinformatics/btw313>

898 73. Gu Z. Complex heatmap visualization. iMeta. 2022;1:e43.
899 <https://doi.org/10.1002/imt2.43>

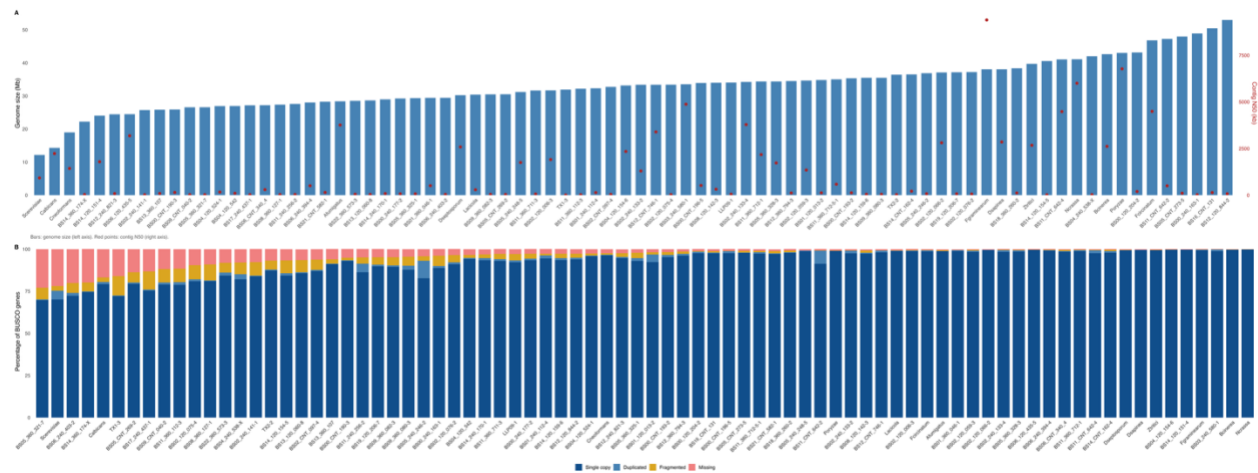
900 74. Wickham H. ggplot2: Elegant graphics for data analysis [Internet]. Springer-Verlag New
901 York; 2016. <https://ggplot2.tidyverse.org>

902

Functional gene content of fungi in a pine needle disease system largely reflects their phylogeny

Gabriel A. A. Silva, Temitope R. Folorunso, Lori G. Eckhardt, Janna R. Willoughby

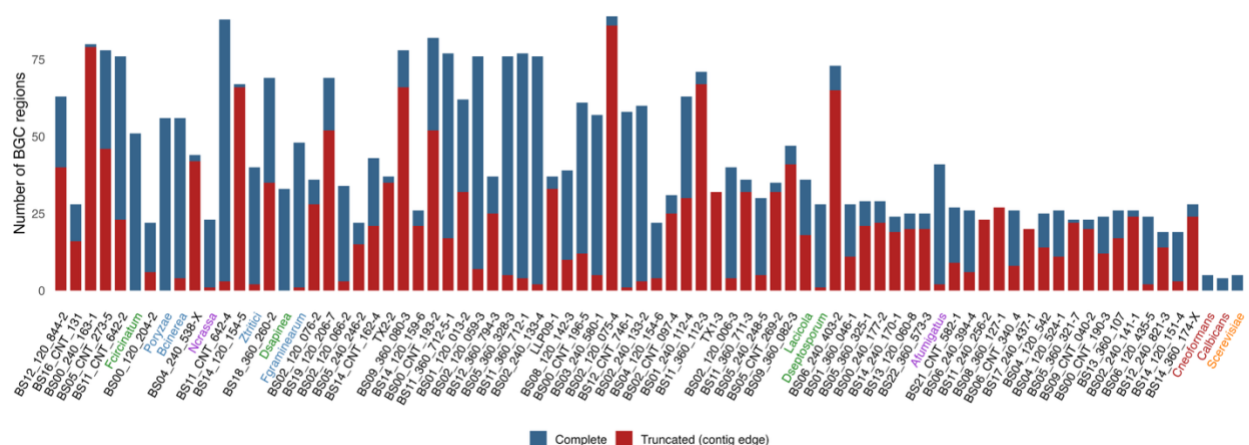
Supplementary Figures



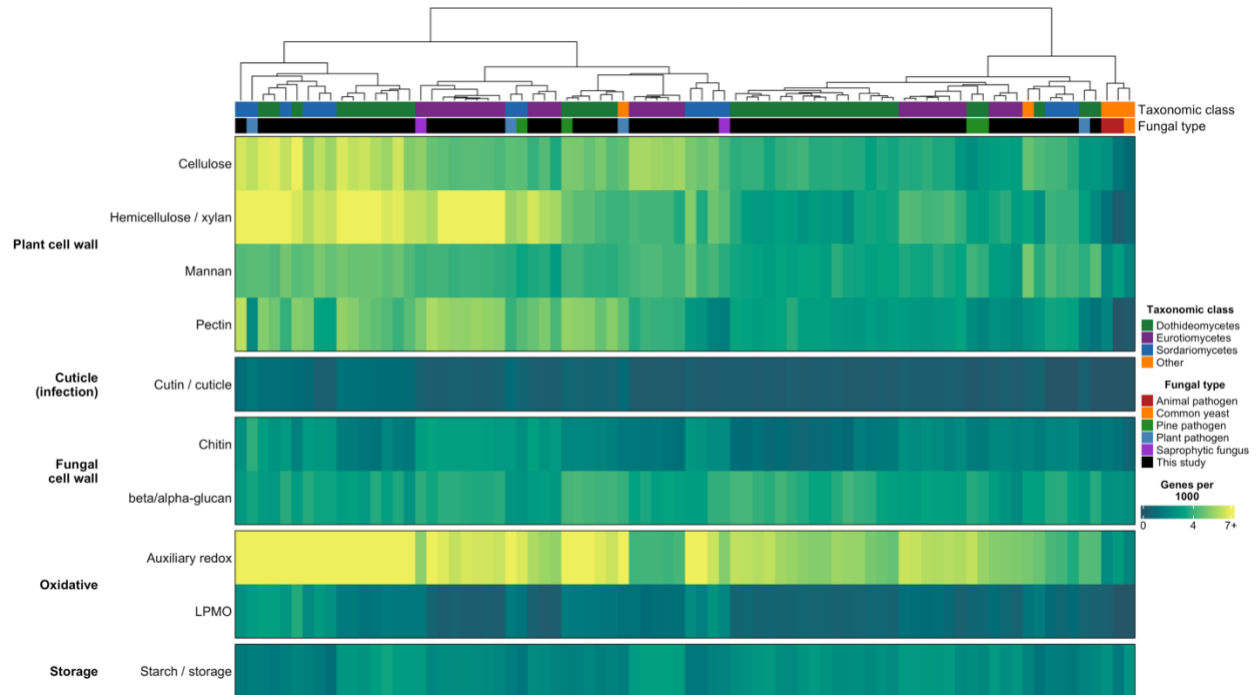
Supplementary Figure 1. Assembly quality of the 67 sequenced isolates and 13 reference genomes. (A) Assembly size in megabases (bars, left axis) and contig N50 in kilobases (red points, right axis), ordered by increasing assembly size. (B) BUSCO completeness against the fungi_odb12 lineage, partitioned into single-copy, duplicated, fragmented and missing genes, with genomes ordered by decreasing percentage of missing genes. All genomes shown met the retention criteria (BUSCO complete $\geq 70\%$ and duplicated $< 15\%$). The two panels use different orderings, so a genome occupies a different position in each.



Supplemental Figure 2. COG functional category composition across the 80 genomes. Rows are COG categories and columns are genomes, clustered by profile similarity. Colour shows the number of genes per 1,000 predicted genes, capped at the 95th percentile of non-zero values to prevent a few high values from flattening the scale. Annotation bars above the heatmap give each genome's taxonomic class and fungal type (study isolate or reference). Classes other than Dothideomycetes, Eurotiomycetes and Sordariomycetes are shown together as “Other”.



Supplementary Figure 3. Biosynthetic gene cluster regions per genome, partitioned into complete regions and regions truncated at contig edges. Among the 67 study isolates, the percentage of truncated regions rose with assembly fragmentation, but total region count did not. Mean truncation increased from 17.7% in the most contiguous third of assemblies to 89.5% in the least contiguous, whereas mean region count changed only from 46.6 to 41.6. Fragmentation therefore changes how many detected regions are flagged as incomplete but not how many are detected.



Supplementary Figure 4. CAZyme activity group composition across the 80 genomes. Rows are the ten substrate-based activity groups, arranged in five blocks: plant cell wall (cellulose, hemicellulose/xylan, mannan, pectin), cuticle (cutin/cuticle), fungal cell wall (chitin, beta/alpha-glucan), oxidative (auxiliary redox, LPMO) and storage (starch/storage). Columns are genomes, clustered by profile similarity. Colour shows genes per 1,000 predicted genes, with the scale capped at 7 and above. Annotation bars above the heatmap give each genome's taxonomic class and fungal type.

Functional gene content of fungi in a pine needle disease system largely reflects their phylogeny

Gabriel A. A. Silva, Temitope R. Folorunso, Lori G. Eckhardt, Janna R. Willoughby

Supplementary Tables

Supplementary Table 1. Reference-free assembly quality estimates from Merqury for the 67 sequenced isolates. K-mer databases were built with Meryl (k = 21) from Illumina short reads for each isolate and compared with the corresponding assembly.

Iolate ID	QV	Error Rate (%)	Unique K-mers	Common K-mers
BS00_120_204-2	32.35	0.0583	769150	42953875
BS00_240_163-1	34.08	0.0391	575411	47782105
BS00_240_165-2	22.37	0.5795	6986072	42371002
BS00_240_177-2	28.62	0.1375	1205685	28870926
BS01_120_013-2	10.22	9.5025	32634107	34181194
BS02_120_076-2	30.98	0.0798	897855	36735580
BS02_240_133-2	37.81	0.0165	170431	33309062
BS02_240_133-4	7.89	16.2463	34087180	34227625
BS04_120_524-1	28.68	0.1356	1104631	26821893
BS05_240_246-2	32.68	0.0539	604269	36447959
BS05_240_248-5	44.50	0.0035	34271	31184914
BS05_360_328-3	36.34	0.0232	246587	34353720
BS05_CNT_273-5	35.06	0.0312	457060	47471375
BS06_120_435-5	38.01	0.0158	119566	24462856
BS06_240_394-4	32.07	0.0622	534004	27972179
BS06_360_361-6	8.86	13.0137	53079	53793
BS08_120_142-3	38.21	0.0151	158316	33860220
BS09_120_061-2	10.88	8.1734	81427	87662
BS09_360_082-3	33.25	0.0474	436605	29953557
BS11_240_256-2	32.08	0.0619	507322	26695377
BS11_CNT_642-2	13.08	4.9259	36881734	46620646
BS11_CNT_642-4	22.28	0.5910	6888537	41039973
BS12_120_844-2	34.25	0.0376	605239	52273967
BS12_240_821-3	34.00	0.0398	297115	24203729
BS14_240_170-1	32.06	0.0622	548711	28733191
BS14_CNT_162-4	36.15	0.0243	271515	36210111
BS21_CNT_582-1	33.75	0.0421	364855	28115896
BS22_360_573-3	29.24	0.1192	1016738	28014712
LLP09-1	32.91	0.0512	529335	33603435
Mean	29.06	1.87		
SD	9.81	4.29		
95CI	1.24	0.54		

9 Supplementary Table 2. Reference genomes included in the functional comparisons. Genome
10 assemblies were obtained from NCBI.

Species	NCBI Accession	Reason	Associated disease or role	Lifestyle in analyses
<i>Lecanosticta acicola</i>	GCA_963674455.1	Pine pathogen	BSNB Pines	Pathogen
<i>Dothistroma septosporum</i>	GCA_002236755.2	Pine pathogen	DNB Pines	Pathogen
<i>Fusarium circinatum</i>	GCA_024047395.1	Pine pathogen	Pine Pitch Canker	Pathogen
<i>Diplodia sapinea</i>	GCA_056582465.1	Pine pathogen	Diplodia Tip Blight	Pathogen
<i>Fusarium graminearum</i>	GCA_020991245.1	Plant pathogen	Fusarium Head Blight	Pathogen
<i>Pyricularia oryzae</i>	GCA_050157985.1	Plant pathogen	Rice Blast	Pathogen
<i>Botrytis cinerea</i>	GCF_000143535.2	Plant pathogen	Gray Mold	Pathogen
<i>Zymoseptoria tritici</i>	GCF_000219625.1	Plant pathogen	Wheat leaf blotch	Pathogen
<i>Aspergillus fumigatus</i>	GCA_040126065.1	Saprophytic fungus	Saprotroph Human pathogen	Saprophyte
<i>Neurospora crassa</i>	GCF_000182925.2	Saprophytic fungus	Saprotroph Model organism	Saprophyte
<i>Candida albicans</i>	GCF_000182965.3	Animal pathogen	Candidiasis	Yeast-like
<i>Cryptococcus neoformans</i>	GCA_056621495.1	Animal pathogen	Cryptococcosis	Yeast-like
<i>Saccharomyces cerevisiae</i>	GCF_000146045.2	Common yeast	Model organism	Yeast-like

11