iScience

CellPress
OPEN ACCESS

Article

Evolution of zygomycete secretomes and the origins of terrestrial fungal ecologies

Ying Chang,[1,2,14,*] Yan Wang,[3,4,5] Stephen Mondo,[6] Steven Ahrendt,[6] William Andreopoulos,[6] Kerrie Barry,[6] Jeff Beard,[1] Gerald L. Benny,[7] Sabrina Blankenship,[1] Gregory Bonito,[8] Christina Cuomo,[9] Alessandro Desiro,[8] Kyle A. Gervers,[1] Hope Hundley,[6] Alan Kuo,[6] Kurt LaButti,[6] B. Franz Lang,[10] Anna Lipzen,[6] Kerry O'Donnell,[11] Jasmyn Pangilinan,[6] Nicole Reynolds,[7] Laura Sandor,[6] Matthew E. Smith,[7] Adrian Tsang,[12] Igor V. Grigoriev,[6,13] Jason E. Stajich,[3] and Joseph W. Spatafora[1]

SUMMARY

Fungi survive in diverse ecological niches by secreting proteins and other molecules into the environment to acquire food and interact with various biotic and abiotic stressors. Fungal secretome content is, therefore, believed to be tightly linked to fungal ecologies. We sampled 132 genomes from the early-diverging terrestrial fungal lineage zygomycetes (Mucoromycota and Zoopagomycota) and characterized their secretome composition. Our analyses revealed that phylogeny played an important role in shaping the secretome composition of zygomycete fungi with trophic mode contributing a smaller amount. Reconstruction of the evolution of secreted digestive enzymes revealed lineage-specific expansions, indicating that Mucoromycota and Zoopagomycota followed different trajectories early in their evolutionary history. We identified the presence of multiple pathogenicity-related proteins in the lineages known as saprotrophs, suggesting that either the ecologies of these fungi are incompletely known, and/or that these pathogenicity-related proteins have important functions associated with saprotrophic ecologies, both of which invite further investigation.

INTRODUCTION

Trophic modes and ecological guilds are some of the most defining traits of fungi. Ecologists typically classify fungi into trophic modes based on their association and outcome with hosts and substrates (Nguyen et al., 2016). Symbiotrophs associate with living hosts forming beneficial relationships with both host and fungus, such as mycorrhizae and lichens. Pathotrophs cause disease and ultimately harm their hosts, ranging from biotrophy to necrotrophy; whereas, saprotrophs degrade non-living organic matter, playing an essential role in nutrient cycling. Trophic modes are further divided into ecological guilds, such as mycorrhizae, plant pathogens, animal pathogens, and mycoparasites, which provide greater specificity with respect to host/substrate affiliation and nature of association. Although these categories may be imprecise in some situations, and some fungi may occupy more than one trophic mode or ecological guild, the ecology of a particular fungus is largely based on its ability to secrete biologically active molecules into the environment. These molecules define the heterotrophic nutritional mode of a fungus through its ability to interact with other organisms, decompose substrates comprising complex biopolymers, or obtain nutrients from living hosts through antagonistic, beneficial, or not-yet-defined strategies. As such, fungi are what they secrete.

The importance of fungal secretome has been analyzed in numerous ecological guilds of the fleshy, sporocarp-producing Dikarya. Mycorrhizae are characterized by the reduction of enzymes associated with the breakdown of plant cell walls, and the production of small secreted proteins and effectors that interact with specific plant hormone pathways (Plett and Martin, 2015; Romero et al., 2021). The combination of these features is hypothesized to reduce or inhibit the host defense response, allowing the fungus-plant symbiosis to establish and function. Wood decay fungi possess an arsenal of degradative enzymes acting on different constituents of plant cell walls, and the suite of enzymes employed by a given fungus will determine what plant cell walls biopolymers (e.g., cellulose, lignin, and so forth) will be degraded (Floudas et al., 2012). Mycorrhizae and wood decay represent relatively well-studied ecological guilds in extensively

[1]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA

[2]Division of Science, Yale-NUS College, Singapore 138527, Singapore

[3]Department of Microbiology and Plant Pathology, Institute for Integrative Genome Biology, University of California, Riverside, Riverside, CA 92521, USA

[4]Department of Biological Sciences, University of Toronto Scarborough, Toronto, ON M1C 1A4, Canada

[5]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON M5S 3B2, Canada

[6]US Department of Energy (DOE) Joint Genome Institute (JGI), Lawrence Berkeley National Lab, Berkeley, CA 94720, USA

[7]Department of Plant Pathology, University of Florida, Gainesville, FL 32611, USA

[8]Plant Soil and Microbial Sciences, Michigan State University, East Lansing, MI 48824, USA

[9]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge MA 02142, USA

[10]Robert Cedergren Centre for Bioinformatics and Genomics, Département de Biochimie, Université de Montréal, Montreal, QC, Canada

[11]National Center for Agricultural Utilization Research, US Department of Agriculture, Agricultural Research Service, Peoria, IL 61604, USA

Continued

characterized systems of Dikarya. The majority of fungal species, however, are poorly studied and their function in nature is characterized by limited observations or extrapolations from related species. In this context, the analysis of fungal secretomes represents an opportunity to predict the potential ecologies of many fungi whose function(s) in ecosystems are difficult to observe and characterize.

Zygomycete fungi comprise the first diverging lineages of Kingdom Fungi to emerge after the loss of the flagellum and the origin of the filamentous fungal thallus (Kiss et al., 2019). The two phylum level clades, Mucoromycota and Zoopagomycota, have been resolved as a paraphyletic grade (Chang et al., 2021; Spatafora et al., 2016) or sister groups (Li et al., 2021), depending on taxon sampling and the molecular markers employed. Zygomycetes are ubiquitous in nature and with diverse ecologies. Zoopagomycota comprises pathogens of insects and herptile gut fungi of Entomophthoromycotina; symbionts of aquatic arthropods and mycoparasites of Kickxellomycotina; and mycoparasites and parasites of amoebae, rotifers, and nematodes of Zoopagomycotina. Based on our current understanding of their ecologies, Zoopagomycota is uniquely devoid of plant associations. However, some species can be maintained on simple media consisting of plant and fungal carbohydrates, consistent with a potential for saprotrophy. Conversely, the dominant ecologies of Mucoromycota are plant-associated and include arbuscular mycorrhizae of Glomeromycotina; root endophytes and chitinolytic soil saprotrophs of Mortierellomycotina; and saprotrophs of plant material, mycoparasites, ectomycorrhizae, plant pathogens and opportunistic animal pathogens of Mucoromycotina.

This diversification of ecologies occurred early in the evolution of nonflagellated, terrestrial fungi, and set these two fungal phyla on different evolutionary trajectories. Unfortunately, many of these fungi remain understudied owing to their cryptic nature, small number of economically important species, and a lack of genetic tools. Recently, however, genome sequencing efforts have significantly increased the phylogenetic diversity of zygomycete genomes. This windfall of data provides the unique opportunity to explore the ecological potentials of zygomycete fungi through analysis of their secretomes. Here we report the analyses of the secretome of Kingdom Fungi (Table S1) with a focus on digestive enzymes, putative small secreted proteins (SSPs), and other proteins hypothesized to function in the environment. Our sampling emphasizes the understudied, non-Dikarya phyla Mucoromycota and Zoopagomycota, and testing hypotheses about the evolution of fungal trophic modes and ecological guilds of lineages associated with the origins of terrestrial fungi.

RESULTS

Phylogeny

Zygomycete fungi were resolved as paraphyletic with Zoopagomycota sister to the Mucoromycota-Dikarya clade (Figure S1). The monophyly of each zygomycete subphyla (Glomeromycotina, Mortierellomycotina, Mucoromycotina, Entomophthoromycotina, Kickxellomycotina, and Zoopagomycotina) was well supported. The "rogue" genus *Basidiobolus* (Chang et al., 2021; Li et al., 2021) was resolved as part of Entomophthoromycotina and thus included in the subphylum in the subsequent analyses.

Overview of fungal secretomes

On average, 5.7% of the total number predicted proteins (referred as proteome later in discussion) were secreted to the extracellular environment, with the lowest being 1.1% (*Catenaria*, Blastocladiomycota) and the highest being 12.0% (*Acaulopage*, Zoopagomycotina, Zoopagomycota) (Table S2). Overall, Zoopagomycota had a higher relative secretome abundance than other phyla (Table S2 and Figure S2), with an average of 8.5% (Table S2 and Figure S2). The average secretome:proteome ratio for Mucoromycota was 4.7% and varied across different subphyla (2.8% for Glomeromycotina, 4.7% for Mucoromycotina, and 6.9% for Mortierellomycotina). Secretome composition comprised on average 25% digestive enzymes, 46% SSPs, and 29% OTHER proteins (Figure S2). Glomeromycotina stood out by having the highest SSP content (67%) and the lowest digestive enzyme content (11%). Among zygomycetes, there was a strong negative correlation between the relative abundance of digestive enzymes and the relative abundance of SSPs (Figure 1).

Phylogenetic vs. ecological signals in secretome variation

Non-metric multidimensional scaling (NMDS) analyses revealed a strong influence of phylogenetic relatedness on the secretome content regardless of whether analyzed as total secretome (Figure 2) or by grouping proteins into digestive enzymes, SSPs, or OTHER (Figure S3). Genomes from Mucoromycota and

[12]Centre for Structural and Functional Genomics, Concordia University, Montréal, QC H4B 1R6, Canada

[13]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA 94720, USA

[14]Lead contact

*Correspondence: ying.chang@yale-nus.edu.sg

https://doi.org/10.1016/j.isci.2022.104840

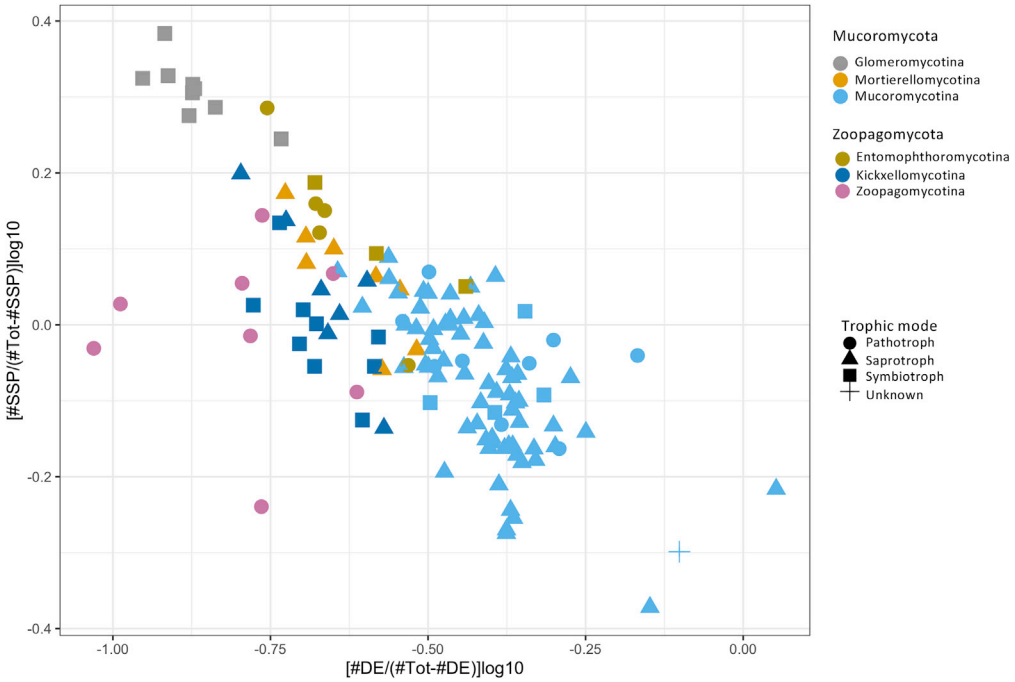


Figure 1. Trade-off between the number of SSPs and digestive enzymes (DEs) in the zygomycete genomes

X-axis: log10[#DE/(#secretome-#DE)]; Y axis: log10[#SSP/(#secretome-#SSP)]. The two variables are negatively correlated (Pearson correlation coefficient r = −0.78, p < 0.01).

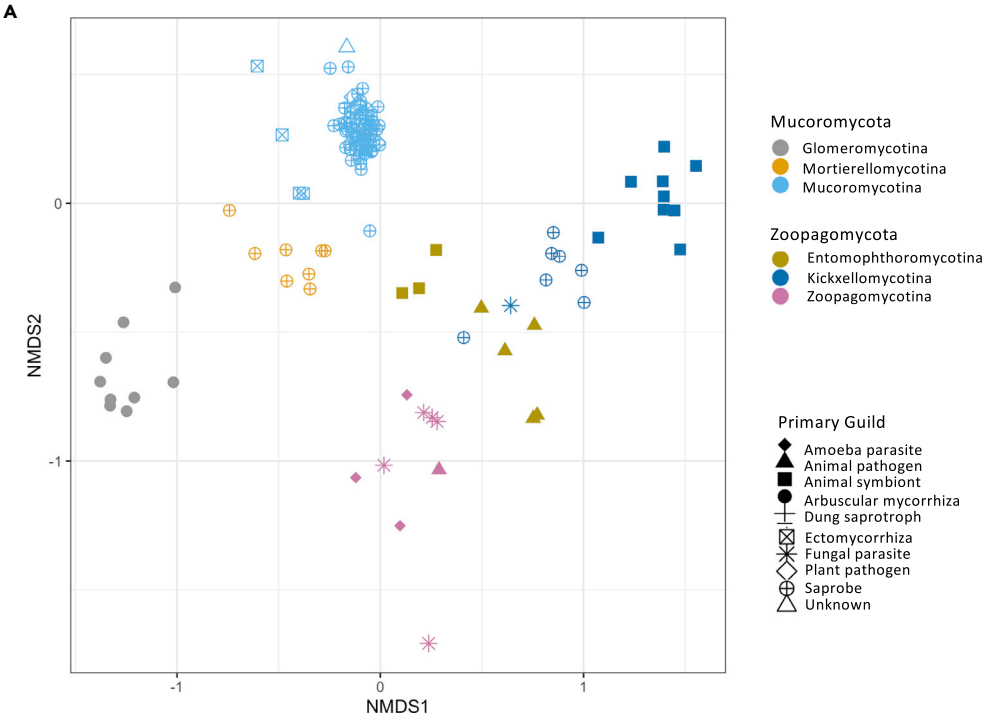
Zoopagomycota were separated from each other in the NMDS space. At the subphylum level, the Mucoromycota secretomes showed similar secretome composition within each subphylum. In contrast, the genomes from each Zoopagomycota subphylum had a more diverse secretome composition and were more loosely grouped in the NMDS plot (Figure 2A). NMDS analyses also revealed patterns associated with taxon primary guild (Figure 2A) among Zoopagomycota genomes. Within Entomophthoromycotina, the three herptile gut fungi (*Basidiobolus*) were clustered apart from the animal pathogenic members. Similarly, the animal symbionts and the saprobes from Kickxellomycotina formed two separate groups. Within Zoopagomycotina, mycoparasites were separated from the animal pathogens. The dual influence from phylogenetic relatedness and trophic ecologies was corroborated by results from Mantel tests. Phylogenetic relatedness was strongly correlated with total secretome content (r = 0.66-0.72, p < 0.01). Ecological guild showed significant but weaker correlation with total secretome content in partial and regular Mantel tests for both primary guild coding (r = 0.41 & 0.56; p < 0.01) and multiple guild coding (r = 0.44 & 0.47; p < 0.01). Similar patterns were recovered for individual protein groups of digestive enzymes, SSPs, and OTHER proteins (Figure 2B).

Secreted digestive enzymes

Peptidases, CAZymes, and lipases on average comprised 8.0%, 6.3%, and 2.9% of the total secreted proteins (Table S3 and Figure S4). On average, anaerobic gut fungi (Neocallimastigomycota, Chytridiomycota) possessed the highest proportions of CAZymes (39.4%) and the highest proportions of lipases (9.9%), and Kickxellomycotina possessed the highest percentage of peptidases (4.8%).

In the NMDS analysis, when enzyme vectors were fit to the ordination plot, 41 enzymes were identified with strong r values (squared correlation coefficient, r > 0.2 & p < 0.001), including 33 cazymes, seven peptidases, and one lipase (Table S4, Figure 3C). Plant cell wall degrading enzymes (PCWDEs) were enriched in Mucoromycotina. In contrast, secretomes of Zoopagomycota, Glomeromycotina, and Mortierellomycotina were enriched in chitinases and chitin-binding modules, including AA11, CBM18, and GH19 (Zoopagomycota only). Chitinase family GH18 was widely distributed across the sampled zygomycete genomes but missing from Glomeromycotina. Chitin-binding CBM5 was commonly found in Mucoromycota but not in Zoopagomycota except for in two of the *Basidiobolus* genomes.

A



B

	Phylogeny			Primary Guild		Multiple Guilds	
	Regular	Partial-PG	Partial-MG	Regular	Partial	Regular	Partial
Total secretome	0.72	0.66	0.71	0.56	0.41	0.47	0.44
Digestive enzymes	0.72	0.67	0.71	0.53	0.40	0.44	0.40
SSPs	0.69	0.62	0.67	0.50	0.35	0.45	0.40
OTHER	0.75	0.69	0.73	0.50	0.35	0.45	0.41

Figure 2. Phylogenetic relatedness demonstrates a strong influence on the secretome composition of zygomycete fungi and that ecological guilds play a relatively weaker role

(A) Ordination plot based on NMDS analysis on secretome composition (k = 2; stress value = 1.3).

(B) R statistics reported from the Mantel tests (regular and partial) on the correlation between secretome composition and phylogenetic relatedness or ecological guild (p value <0.01). Two ecological guild coding schemes were adopted, primary guild and multiple guilds. Partial Mantel tests on the correlation of distance matrices of secretome composition and phylogeny were corrected for the influence of primary guild (PG) or from multiple guilds (MG).

The two peptidase families S01A and A01A had the highest NMDS r scores among all the digestive enzymes ($r = 0.71$ & $p < 0.001$ for both). Serine peptidase S01A was enriched in Kickxellomycotina and Entomophthoromycotina, with average copy numbers of 23 and 20, respectively, whereas this protein was missing or depauperate in the other fungi (average less than one copy per genome). In contrast, aspartic peptidase A01A was enriched in Mucoromycota (average of 17 copies) but found in low copy numbers in Zoopagomycota (average of two copies). Peptidase S10 had an even distribution across most of the genomes, with relatively low copy numbers (present in 122 out of 132 genomes, average of four copies per genome), but was absent from 70% of Zoopagomycotina genomes.

Species/gene tree reconciliation analysis by Notung mapped ancient diversifications of PCWDEs, chitinases, and peptidases on the branches leading to the most recent common ancestor (MRCA) of the non-flagellated terrestrial fungi and the chytrids, and the MRCA of all non-flagellated terrestrial fungi (Figure 3A). The MRCA of Mucoromycota was characterized by losses of PCWDEs and peptidases with the

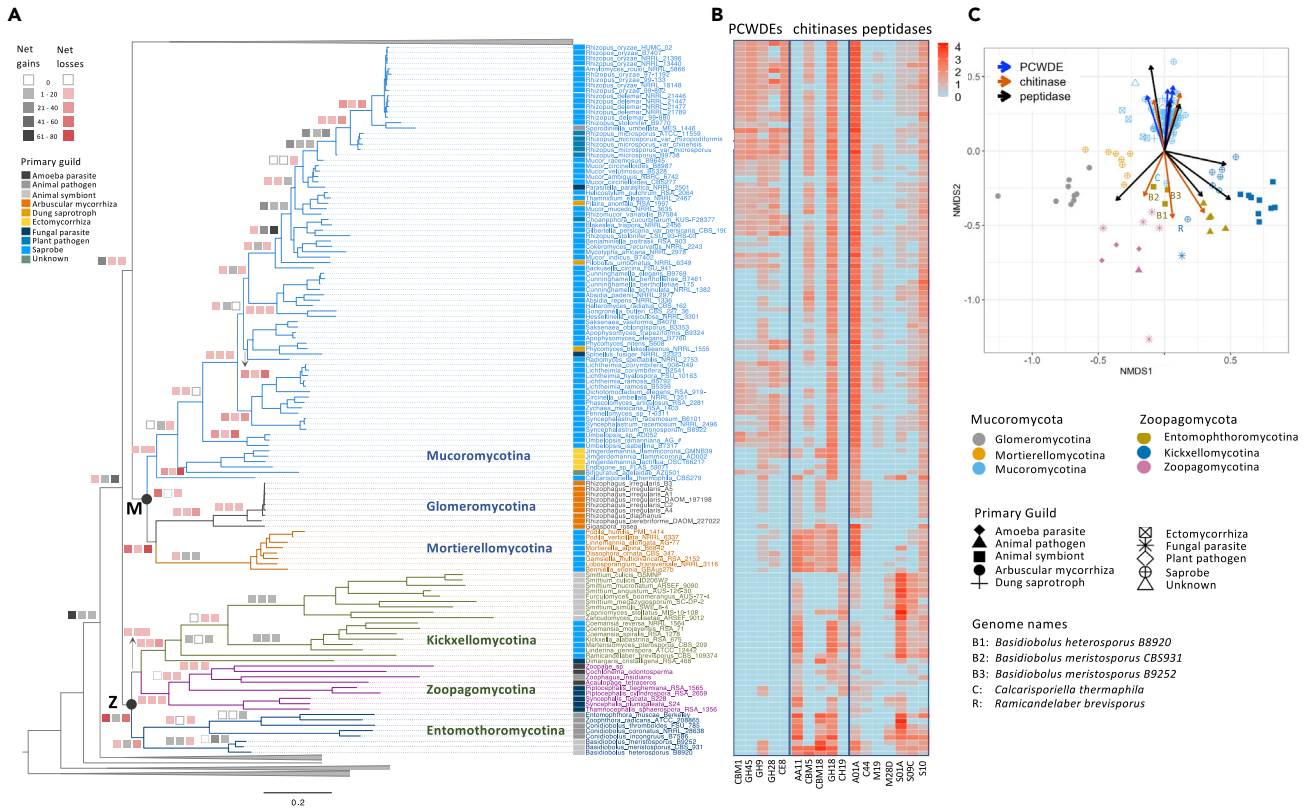


Figure 3. Digestive enzyme composition and evolution across sampled zygomycete genomes

(A) Enzyme evolution was reconstructed using NOTUNG based on the RAxML species tree and best enzyme trees with bootstrap support. Bars above internodes are the total number of net gains (light grey to black) or net losses (pink to red) for cazymes, peptidases, and lipases. Colored bars by taxon names represent the primary guild of the taxon.

(B) Heatmap of raw counts of PCWDEs, chitinases, and peptidases identified as strong vectors in NMDS analysis ($r > 0.2$).

(C) Plot of NMDS analysis of digestive enzyme composition across sampled genomes. Each genome is color coded by its subphylum placement with the shape representing its primary guild. Arrows represent the strong vectors of PCWDEs, chitinases, and peptidases ($r > 0.2$).

diversification of all three classes of enzymes within Mucorales (corresponding approximately to Mucoraceae) and independent expansions of chitinases and peptidases in the MRCA of Mortierellomycotina. The MRCA of Zoopagomycota was characterized by diversification of chitinases only, followed by multiple independent expansions of PCWDEs, chitinases, and peptidases within the subphyla Entomophthoromycotina and Kickxellomycotina.

Small secreted proteins in zygomycete fungal genomes

A total of 47,494 SSPs were identified across the 132 zygomycete genomes. They formed 3,144 non-singleton orthologous groups, including 35,134 SSPs (74%). 97.5% of the OGs were shared by two or more species (3,066). Among them, over two-thirds (2,222) did not show sequence similarity to known pfam proteins.

2 × 2 clustering analyses of the SSP non-singleton OGs identified the enrichment of SSPs shared by closely related taxa with a similar ecological guild (Figure 4), but convergent ecologies across different subphyla did not share enriched orthologous OGs. No indicator OGs (i.e., the proteins OGs that are found mainly associated with particular taxa) were identified at the phylum level. At the subphylum level, Entomophthoromycotina, Glomeromycotina, and Mortierellomycotina had over 180 SSP-OGs identified as indicator OGs, while the other three subphyla had less than 70 (Figure S5). The majority of indicator OGs were found with high specificity A values (Figure S5) but lower fidelity scores (B values), indicating that most of the indicator SSP OGs were specific to, or enriched in a particular subphylum, but these OGs had limited distribution within this subphylum (Figure S5).

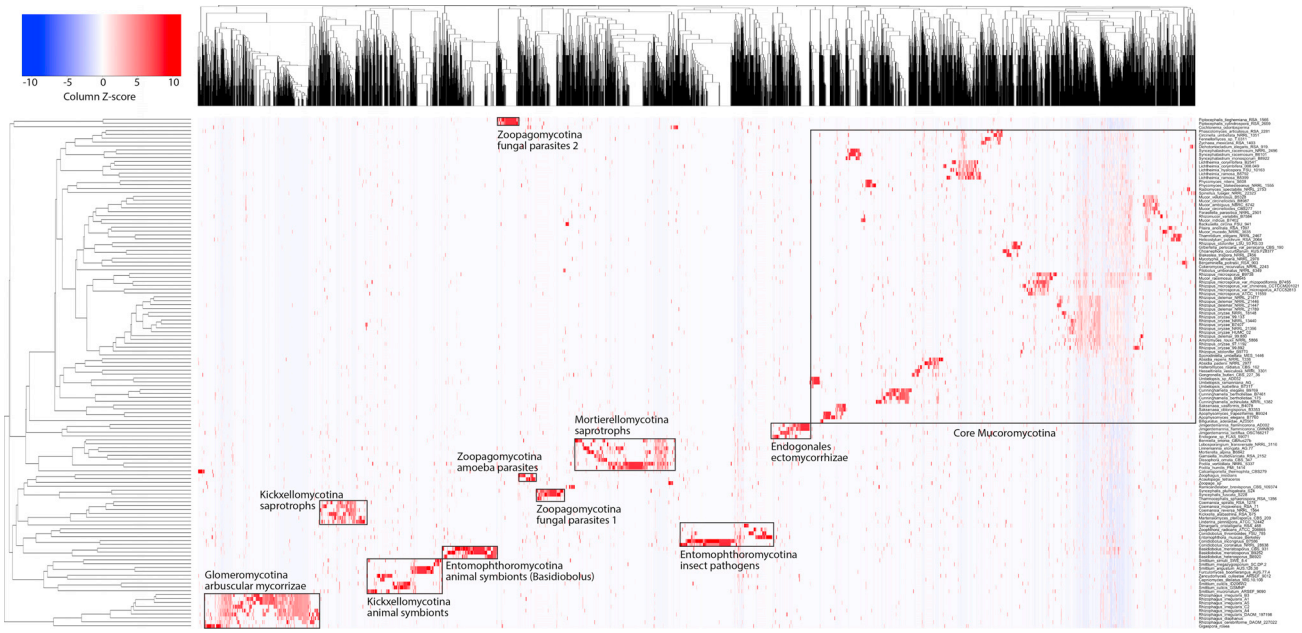


Figure 4. 2 × 2 hierarchical clustering analyses of SSP composition (*hclust* function in *R*; average linkage and the Bray-Curtis dissimilarity matrix)
The result was visualized as a heatmap using the *heatmap.2* function in ggplot.

For Glomeromycotina, many of the indicator OGs with pfam hits were related to pathogenesis and/or defense (Table S5). Similarly, many of the indicator OGs for Entomophthoromycotina and Mortierellomycotina showed sequence similarity to proteins involved in defense (Table S5). In addition, some indicator proteins for Entomophthoromycotina and Mortierellomycotina were related to nutrient acquisition (Table S5). Though not identified as an indicator protein, LysM (PF01476) was found enriched in Mucoromycotina, Mortierellomycotina, and *Basidiobolus* (Entomophthoromycotina, Figure S6). When taxa were grouped by their primary guild, none or a small number of indicator OGs were identified for each category, except for the arbuscular mycorrhizal fungi which were present only in Glomeromycotina (Figures S5C and S5D).

OTHER proteins in zygomycete fungal genomes

A total of 22,712 OTHER proteins were identified across the 132 zygomycete genomes. They formed 1,246 non-singleton OGs, including 20,134 proteins (88.6%). Among the non-singleton OGs, 1,213 (97.35%) OGs were shared by two or more species. Approximately half of the OTHER-OGs (657) had hits to known protein families in the Pfam database.

In the IndicSpecies analysis, six OTHER-OGs were identified as indicator OGs for Zoopagomycota. Three of them showed sequence similarity to Pfam proteins (PF00264—a common central domain of tyrosinase, PF00188—a cysteine-rich secretory protein family, and PF03009—a diester phosphodiesterase family). In contrast, no indicator OTHER-OGs were identified for Mucoromycota. At the subphylum level, Mortierellomycotina had the highest number of indicator OGs (109), followed by Entomophthoromycotina (68), Glomeromycotina (39), Kickxellomycotina (35), Zoopagomycotina (23), and Mucoromycotina (12) (Figure S7). Many of these indicator OTHER-OGs were associated with interactions with other organisms or dealing with environmental stress (Table S6), including *cot*H proteins (PF08757) that are known to be involved in mucormycosis (Figure S6)

Glycosylphosphatidylinositol-anchored cell wall proteins

The three secretome subgroups each contained approximately 10% glycosylphosphatidylinositol (GPI)-anchored proteins. Some protein families had higher proportions of GPI-anchored members, including GH72 (79.4%), CE4 (61.1%), PF01522 (polysaccharide deacetylase family; 39.1%), GH9 (37.5%), and PF08757 (*Cot*H domain; 25.9%). CE4 and PF10342 (Ser-Thr-rich GPI-anchored membrane family) had the highest numbers of GPI-anchored proteins (1136 & 997, respectively), followed by PF01522 (242 proteins)

and PF02469 (Fasciclin domain; 211 proteins). InterProScan annotated another 182 proteins as cell wall proteins, among which 174 were HsbA (PF12296). None of the predicted HsbA proteins contained the C-terminal signal sequence for GPI anchors according to NetGPI prediction.

DISCUSSION

Phylogeny is the major driver of secretome content

Earlier studies of Dikarya found that secretome size was directly related to phylogeny but not to trophic modes (e.g., Krijger et al., 2014; Stolzer et al., 2012), whereas Lowe and Howlett (2012) reported a correlation between trophic ecology and secretome size for some fungi. Alfaro et al. (2014) argued that lifestyle shapes secretome composition, rather than secretome size. Here, we found that zygomycete secretome composition is primarily determined by phylogeny, with trophic mode and ecological guild playing an important secondary role in the diversification of secretome content (Figures 2, 5 and S3).

Mantel tests suggest that phylogenetic relatedness strongly affects the secretome composition and that trophic ecology played a minor role (Figure 2). This is consistent with the adaptive potential of a fungus being largely determined by the genetic background inherited from its ancestors, which results in fewer shared homologies in secretome content among fungi with similar trophic ecologies but with different phylogenetic backgrounds. Although fungi with the same trophic ecology may utilize common nutrient resources, and encounter similar environmental stresses, they have developed independent, non-orthologous enzymatic and molecular mechanisms to do so.

Early diversification of nutritional modes with the emergence of nonflagellated, terrestrial fungi

Gene tree - species tree reconciliation analyses support early diversifications of digestive enzymes prior to the loss of the flagellum and origin of nonflagellated fungi (Figure 3). Reconstruction of digestive enzyme evolution concurrent with the emergence of nonflagellated, terrestrial fungi supports an early split in nutritional modes between Mucoromycota and Zoopagomycota. The majority of Mucoromycota taxa are associated with plants or plant-derived substrates, whereas Zoopagomycota taxa are mainly associated with animals or other fungi. PCWDEs are enriched in Mucoromycotina taxa, consistent with ecologies dependent on the breakdown of plant cell wall carbohydrates (Figure 2 and Table S7). In contrast, Zoopagomycota taxa possess more chitinases and peptidases, consistent with animal-and/or fungal-associated ecologies and the utilization of chitin and proteins as nutritional sources.

Rather than being phylum-level secretome synapomorphies, however, the emergence of these two ecological trajectories is characterized by ancient losses and differential retention of proteins in the MRCAs of Mucoromycota and Zoopagomycota, followed by more derived, lineage-specific secretome diversifications. For example, even though both lineages underwent losses of PCWDEs in their early evolutionary history, the MRCA of Zoopagomycota was estimated to possess only 18 copies of PCWDEs and few or no pectinases, while the MRCA of Mucoromycota is hypothesized to have retained more PCWDEs (75 copies), targeting a diverse array of plant-derived polysaccharides (Table S7). Subsequent diversifications of extant ecologies are the product of more derived expansions of enzyme families within subphyla, as evidenced in the increase in PCWDEs, chitinases, and peptidases within Mucorales, chitinases, and peptidases in Mortierellomycotina, and independent expansions of chitinases and peptidases within the subphyla Entomophthoromycotina and Kickxellomycotina (Figure 3).

These results are consistent with the hypothesis that Mucoromycota fungi were associated with streptophytes early in their evolutionary history. The earliest Zoopagomycota fungi, on the other hand, adapted to animal-based and fungal-based ecologies and utilized nutrient sources other than biopolymers derived from streptophyte cell walls. As such, early diversification events in Kingdom Fungi involved two different phylogenetic trajectories defined by host association and substrate utilization.

Evolution of fungal cell wall masking LysM domain

LysM effectors are small effector proteins with one or more Lysin motifs (LysM; PF01476). LysM motifs are typically between 44 and 65 amino acids long and bind to peptidoglycan, chitin, and their derivatives (Dubey et al., 2020). Plants detect the presence of fungi via cell surface receptors containing LysM domains to recognize the potential fungal symbiont or pathogens and start subsequent responses leading either to

the establishment of symbiosis or to defense response against pathogens (Hu et al., 2021). Fungi, on the other hand, also use LysM effectors for a diversity of functions. By binding to chitin in the fungal cell walls, LysM effector helps the fungi to avoid detection by the plant immune system and to protect fungi against mycoparasitism (Kombrink and Thomma, 2013). LysM effectors can also immobilize bacterial competitors through binding of bacterial peptidoglycan (Kombrink and Thomma, 2013). LysM effector domains are enriched in Mucoromycota but with poor representation in Zoopagomycota with the exception of *Basidiobolus* (Figure S6). This striking difference in the distribution of LysM effector domains positively correlates with Mucoromycota ecologies being primarily associated primarily with plants as symbionts (e.g., mycorrhizae and root endophytes) and saprotrophs, while Zoopagomycota are defined by animal and fungi-associated ecologies.

A parallel pattern of enrichment is observed for the CE4/PF015522 domain, which functions in cell-wall remodeling and the deacetylation of chitin to chitosan, consistent with chitosan being an abundant cell wall carbohydrate of Mucoromycota (Figure S6). But this domain is also involved in protecting fungi from attack by plant hydrolases by de-N-acetylating chitin, and avoidance of the plant immune system (Toro and Brachmann, 2016).

Zoopagomycota with animal-associated ecologies have higher secretome:proteome ratios

Secretome:proteome ratios of the zygomycete genomes in this study are comparable to those of Dikarya fungi (e.g., Lo Presti et al., 2015; Toro and Brachmann, 2016; Figure S2). Glomeromycotina stands out in having lower secretome:proteome ratios than the other zygomycete subphyla, consistent with the findings of previous studies (Kamel et al., 2017; Miyauchi et al., 2020; Tisserant et al., 2013). Zoopagomycota, especially the subphyla Zoopagomycotina and Entomophthoromycotina, has a higher relative secretome content compared to other fungi. Many members of Zoopagomycota are associated with animals, as either pathogens or symbionts. The high relative secretome content in these taxa is different from what has been reported for animal pathogens from Dikarya, which generally have low secretome:proteome ratios (Krijger et al., 2014; Lowe and Howlett, 2012). It should be noted that the majority of the Dikarya animal pathogens studied so far are mostly from Eurotiomycetes and are opportunistic pathogens of mammals, which only represents a small portion of the animal pathogen diversity in Dikarya.

Entomophthoromycosis fungi are best known as insect pathogens, such as *Entomophthora muscae*, a pathogen of the common house fly (Gryganskyi et al., 2017). These insect pathogens infect their hosts through spores, grow vegetatively within the host, ultimately cause harm and death, and then sporulate on the insect exoskeleton to infect other hosts. Many species within the subphylum exhibit complex life histories, however, which involve cycling between complex animal associations characterized as symbiotrophs (e.g., insect dispersed, herptile gut symbionts of *Basidiobolus*) and pathotrophs (e.g., insect pathogens of *Conidiobolus*), and saprotrophic life history stages associated with soil and plant litter environments. Furthermore, both *Basidiobolus* and *Conidiobolus* can function as opportunistic pathogens of humans and other animals (Vilela and Mendoza, 2018). In contrast, the sampled Zoopagomycotina fungi are mycoparasites, amoeba parasites, and predators of amoeba or rotifers. Although little is known about the cellular interactions between Zoopagomycota fungi and their animal or fungal hosts, the high relative abundance of the Zoopagomycota secretome is consistent with ecologies characterized by complex fungal-host interactions and nutrient acquisition strategies.

Secretome content of GPI-anchored proteins involved in maintaining cell integrity and interacting with the environment

A portion of the predicted secreted proteins are not released as free extracellular proteins, but are bound to the cell wall or plasma membrane. A proportion of these proteins are attached to the cell walls or plasma membrane by a glycerophospholipid anchor (GPI anchor) and extend to the extracellular side of the fungal cell. These proteins play important roles in retaining the integrity of fungal cells, in protecting fungi from biotic and abiotic environmental stresses, and are also related to pathogenicity (Garcia-Rubio et al., 2020; Gow et al., 2017; Vogt et al., 2020).

An average of 10% of the secretome is predicted to be GPI-anchored. Pfam annotation of the GPI-anchored proteins is consistent with the aforementioned functions of the fungal cell wall proteins. The most common GPI-anchored proteins identified in this study, Ser-Thr-rich GPI-anchored proteins (PF10342), are known to function in cell wall organization (Samalova et al., 2020). GH72, CE4, and other

chitin deacetylases from PF01522 are also involved in the remodeling of the fungal cell walls (Toro and Brachmann, 2016). All these proteins are also involved in pathogenesis, and/or protection of pathogenic fungi from attack by their host plants (Samalova et al., 2017). In addition, we recovered several other pathogenesis-related protein families, including fasciclin proteins (PF02469) and cotH proteins (PF08757). Both proteins were described as surface proteins with functions in cell-to-cell interactions. Fasciclin proteins mediate interactions between cell exterior and cell surface and they function in pathogenesis and/or symbiosis in fungi (Seifert, 2018; Wang et al., 2018). CotH proteins play key roles in the invasion of human endothelial cells during mucormycosis by adhering the fungal cell walls to the endothelial cell membrane (Baldin and Ibrahim, 2017). Both of these pathogenesis-related proteins are present in many non-pathogenic fungi sampled in this study (Figure S6). Importantly, most of the predicted fasciclin proteins (89%) and cotH proteins (75%) are not GPI-anchored, consistent with these proteins likely having functions other than pathogenesis and are consistent with a co-opted function in opportunistic pathogens.

The trade-off between small secreted proteins and digestive enzymes

SSPs and digestive enzymes displayed a strong, negative logarithmic relationship with respect to secretome content (Figure 1). Taxa with abundant digestive enzymes possessed relatively few SSPs, and vice-versa. The highest SSP:DE ratios were predictably found in symbiotrophs of arbuscular mycorrhizal Glomeromycotina, consistent with previous observations and SSP importance in the mycorrhizal symbiosis (Miyauchi et al., 2020). Likewise, the lowest SSP:DE ratios consisted of saprotrophic species of Mucoromycotina, in line with known ecologies of soil saprobes. Species characterized by more intermediate SSP:DE values exhibited a range of trophic modes and ecologies, and while symbiotrophs and pathotrophs possessed higher SSP:DE ratios on average, numerous exceptions were observed.

Compared to the Dikarya, zygomycete genomes generally have higher proportions of SSPs and relatively lower proportions of digestive enzymes in their secretomes (Figure S2). In addition, a high proportion of zygomycete SSPs is predicted to be putative effector proteins (Figure S8), regardless of their subphylum, trophic mode, or ecological guild. Effector proteins are best characterized in pathogenic and symbiotic fungi, and function in the interaction and colonization of host cells (Lo Presti et al., 2015). Saprotrophic zygomycetes maintain a wide array of effector-like proteins, suggesting that either effector proteins play important roles in their survival in the environment away from host cells, or that their ecologies are more complex in nature than currently appreciated, or both. Together with previous reports on effectors shared by pathogenic fungi and their saprotrophic relatives (Baldin and Ibrahim, 2017; Samalova et al., 2017; Seifert, 2018; Wang et al., 2018), our study lends support to the hypothesis that effectors play dual functions in both pathogen and symbiont fitness and survival of saprotrophs (Kim et al., 2016; Morris et al., 2009; Stergiopoulos et al., 2012).

Diversification of small secreted proteins and divergent and convergent ecologies

SSPs, including effectors, function in symbiotrophic and pathotrophic ecologies through regulating biochemical and cellular processes associated with host defense mechanisms (Feldman et al., 2020). Our analyses revealed abundant SSPs and putative effectors for zygomycete fungi as well as three main patterns of SSP distribution. First, we found that related species (either the members of the same genus or family) exhibited strong divergence and shared only a limited set of SSP orthologous clusters. This finding is consistent with a conserved core function among closely related taxa that possess common ecologies. Second, species classified within a subphylum, but with different trophic modes and ecological guilds, do not share SSP OGs. Instead, each ecological group was characterized by a unique set of SSPs. For example, animal pathogens and soil saprotrophs of Kickxellomycotina, and herptile gut symbionts and insect pathogens of Entomophthoromycotina, possess divergent sets of SSPs consistent with their phylogeny (Figure 4). Third, convergent ecologies do not share orthologous SSPs. Instead, they possess a unique repertoire of SSP OGs, a finding consistent with non-orthologous solutions to common ecological guilds.

Concluding remarks

These data and analyses revealed that the secretome composition of zygomycete fungi is primarily influenced by phylogeny with trophic mode and ecological guild accounting for a smaller impact. The early evolution of terrestrial fungi secretome content is strongly influenced by adaptation to plant versus non-plant hosts and substrates. This diversification is characterized by ancient losses of enzymatic diversity in the MRCAs and more derived lineage-specific expansions, that resulted in the

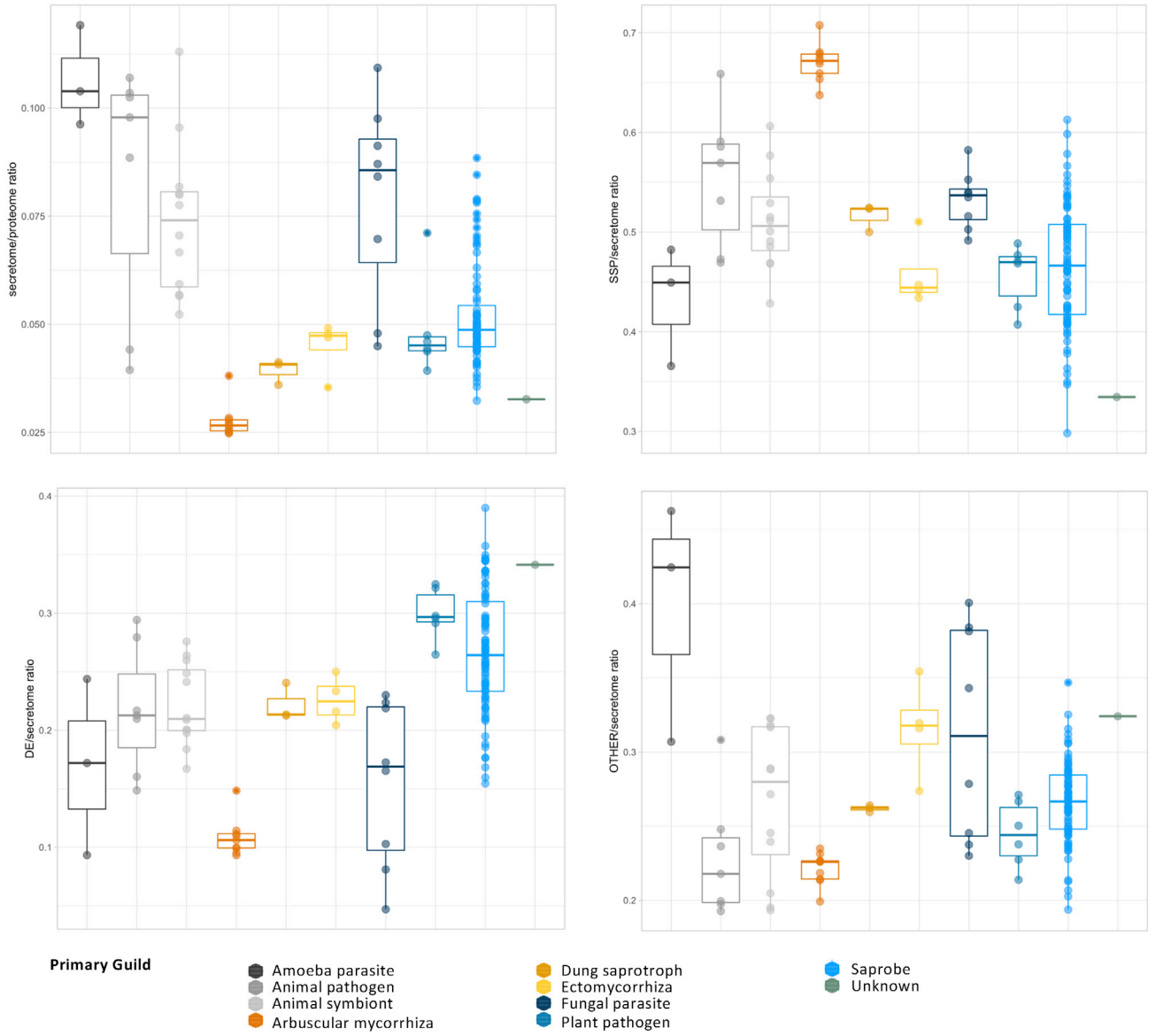


Figure 5. Box plots with data points showing zygomycete secretome composition across each primary guild

diversification of PCWDE in Mucoromycotina, chitinases, and peptidases in Mortierellomycotina, and independent expansions of chitinases and proteases within Zoopagomycota. Regardless of ecology, there is a trade-off between SSPs and digestive enzymes, suggesting that either the ecologies of these organisms are incompletely known, or that SSPs are involved in important functions associated with saprotrophic ecologies, or both.

Limitations of the study

With the most extensive genome sampling on zygomycetes up to date, this work systematically investigated how genealogy, genome/secretome composition, and fungal lifestyle were interconnected with each other. Although we took similarity-based approaches to annotate the secretome, we noted there were potential disparities between sequence similarity and function similarity for some proteins. In addition, just like any other large-scale studies, the application of the same threshold for different proteins from different fungal lineages (e.g., the same e-value cut-offs for all peptidases) might also

introduce inaccuracies in the secretome annotation. Although the life-style annotation of the sampled fungi was based on the authors' best knowledge and cross-validated by mycologists with expertise in different fungal groups, what fungi do in the natural environment is challenging to study. Hence the life-style annotation included in this study is likely a simplified and partial reflection of their real ecological functions.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Material availability
  - Data and code availability
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
- METHOD DETAILS
  - Taxon sampling, trophic mode coding, and genome sequencing, assembly and annotation
  - Phylogenetic reconstruction
  - Secretome retrieval and characterization
  - Comparative analysis on zygomycete secretomes
  - Reconstruction of digestive enzyme evolution

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.104840.

ACKNOWLEDGMENTS

The authors would like to thank Drs. M. Catherine Aime, David Culley, William J. Davies, Gunther Doehlemann, Jon Magnuson, and Timothy Y. James for generously sharing their unpublished genome data with us. We are grateful to Dr. Bruce P. McCune for the constructive advice and discussion on data analysis. We would like to express our gratitude to the two anonymous reviewers for their effort in reviewing this article and for providing insightful comments. This material is based upon work supported by the National Science Foundation (DEB-1441604 to JWS, DEB-1441715 to JES). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer. 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. This work was partially supported by the National Institutes of Health, National Human Genome Research Institute grant U54HG003067 to the Broad Institute.

AUTHOR CONTRIBUTIONS

Y.C., J.W.S, J.E.S., and Y.W. designed this study. Y.C. and J.W.S. conducted the analysis and wrote the article. Y.C., W.Y., S.B., A.D., J.B., and N.R. generated DNA & RNA materials; K.B., S.M., W.A., H.H., A.K., K.L., A.L., J.*P.*, *and* L.S. conducted genome sequencing, assembly, and annotation; Y.C., J.W.S., J.E.S., Y.W., G.L.B., G.B., C.C., K.A.G., B.F.L., K.O'D., M.W.S., A.T., and I.V.G. discussed and commented on the articles.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

Received: February 6, 2022
Revised: May 9, 2022
Accepted: July 21, 2022
Published: August 19, 2022

REFERENCES

Alfaro, M., Oguiza, J.A., Ramírez, L., and Pisabarro, A.G. (2014). Comparative analysis of secretomes in basidiomycete fungi. J. Proteomics *102*, 28–43. https://doi.org/10.1016/J.JPROT.2014.03.001.

Baldin, C., and Ibrahim, A.S. (2017). Molecular mechanisms of mucormycosis-The bitter and the sweet. PLoS Pathog. *13*, e1006408. https://doi.org/10.1371/journal.ppat.1006408.

Beaudet, D., Chen, E.C.H., Mathieu, S., Yildirir, G., Ndikumana, S., Dalpé, Y., Séguin, S., Farinelli, L., Stajich, J.E., and Corradi, N. (2018). Ultra-low input transcriptomics reveal the spore functional content and phylogenetic affiliations of poorly studied arbuscular mycorrhizal fungi. DNA Res. *25*, 217–227. https://doi.org/10.1093/dnares/dsx051.

Buchfink, B., Xie, C., and Huson, D.H. (2014). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60.

Cáceres, M. De, and Legendre, P. (2009). Associations between species and groups of sites: indices and statistical inference. Ecology *90*, 3566–3574. https://doi.org/10.1890/08-1823.1.

Chang, Y., Desirò, A., Na, H., Sandor, L., Lipzen, A., Clum, A., Barry, K., Grigoriev, I.V., Martin, F.M., Stajich, J.E., et al. (2019). Phylogenomics of endogonaceae and evolution of mycorrhizas within Mucoromycota. New Phytol. *222*, 511–525. https://doi.org/10.1111/nph.15613.

Chang, Y., Rochon, D., Sekimoto, S., Wang, Y., Chovatia, M., Sandor, L., Salamov, A., Grigoriev, I.V., Stajich, J.E., and Spatafora, J.W. (2021). Genome-scale phylogenetic analyses confirm Olpidium as the closest living zoosporic fungus to the non-flagellated, terrestrial fungi. Sci. Rep. *11*, 3217. https://doi.org/10.1038/s41598-021-82607-4.

Chin, C.-S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods *10*, 563–569. https://doi.org/10.1038/nmeth.2474.

Chin, C.-S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods *13*, 1050–1054. https://doi.org/10.1038/nmeth.4035.

Dubey, M., Vélëz, H., Broberg, M., Jensen, D.F., and Karlsson, M. (2020). LysM proteins regulate fungal development and contribute to hyphal protection and biocontrol traits in clonostachys rosea. Front. Microbiol. *11*, 679. https://doi.org/10.3389/fmicb.2020.00679.

Eddy, S.R. (2011). Accelerated profile HMM searches. PLoS Comput. Biol. *7*, e1002195. https://doi.org/10.1371/journal.pcbi.1002195.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.

Feldman, D., Yarden, O., and Hadar, Y. (2020). Seeking the roles for fungal small-secreted proteins in affecting saprophytic lifestyles. Front. Microbiol. *11*, 455. https://doi.org/10.3389/FMICB.2020.00455.

Floudas, D., Binder, M., Riley, R., Barry, K., Blanchette, R.A., Henrissat, B., Martínez, A.T., Otillar, R., Spatafora, J.W., Yadav, J.S., et al. (2012). The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. Science *336*, 1715–1719. https://doi.org/10.1126/science.1221748.

Garcia-Rubio, R., de Oliveira, H.C., Rivera, J., and Trevijano-Contador, N. (2020). The fungal cell wall: Candida, cryptococcus, and Aspergillus species. Front. Microbiol. *10*, 2993. https://doi.org/10.3389/fmicb.2019.02993.

Gíslason, M.H., Nielsen, H., Almagro Armenteros, J.J., and Johansen, A.R. (2021). Prediction of GPI-Anchored proteins with pointer neural networks. Curr. Res. Biotechnol. *3*, 6–13. https://doi.org/10.1016/J.CRBIOT.2021.01.001.

Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., Sharpe, T., Hall, G., Shea, T.P., Sykes, S., et al. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc. Natl. Acad. Sci. USA *108*, 1513–1518. https://doi.org/10.1073/pnas.1017351108.

Gow, N.A.R., Latge, J.-P., and Munro, C.A. (2017). The fungal cell wall: structure, biosynthesis, and function. Microbiol. Spectr. *5*. FUNK-0035-2016. https://doi.org/10.1128/microbiolspec.FUNK-0035-2016.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652. https://doi.org/10.1038/nbt.1883.

Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., et al. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. *42*, D699–D704. https://doi.org/10.1093/nar/gkt1183.

Gryganskyi, A.P., Mullens, B.A., Gajdeczka, M.T., Rehner, S.A., Vilgalys, R., and Hajek, A.E. (2017). Hijacked: Co-option of host behavior by entomophthoralean fungi. PLoS Pathog. *13*, e1006274. https://doi.org/10.1371/journal.ppat.1006274.

Haas, B.J., Zeng, Q., Pearson, M.D., Cuomo, C.A., and Wortman, J.R. (2011). Approaches to fungal genome annotation. Mycology *2*, 118–141. https://doi.org/10.1080/21501203.2011.606851.

Hu, S.-P., Li, J.-J., Dhar, N., Li, J.-P., Chen, J.-Y., Jian, W., Dai, X.-F., and Yang, X.-Y. (2021). Lysin motif (LysM) proteins: interlinking manipulation of plant immunity and fungi. Int. J. Mol. Sci. *22*, 3114. https://doi.org/10.3390/ijms22063114.

Jaffe, D.B., Butler, J., Gnerre, S., Mauceli, E., Lindblad-Toh, K., Mesirov, J.P., Zody, M.C., and Lander, E.S. (2003). Whole-genome sequence assembly for mammalian genomes: Arachne 2. Genome Res. *13*, 91–96. https://doi.org/10.1101/gr.828403.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

Kamel, L., Tang, N., Malbreil, M., San Clemente, H., Le Marquer, M., Roux, C., and Frei dit Frey, N. (2017). The comparison of expressed candidate secreted proteins from two arbuscular mycorrhizal fungi unravels common and specific molecular tools to invade different host plants. Front. Plant Sci. *8*, 124. https://doi.org/10.3389/fpls.2017.00124.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

Kim, K.-T., Jeon, J., Choi, J., Cheong, K., Song, H., Choi, G., Kang, S., and Lee, Y.-H. (2016). Kingdom-wide analysis of fungal small secreted proteins (SSPs) reveals their potential role in host association. Front. Plant Sci. *7*, 186. https://doi.org/10.3389/fpls.2016.00186.

Kiss, E., Hegedüs, B., Virágh, M., Varga, T., Merényi, Z., Kószó, T., Bálint, B., Prasanna, A.N., Krizsán, K., Kocsubé, S., et al. (2019). Comparative genomics reveals the origin of fungal hyphae and multicellularity. Nat. Commun. *10*, 4080. https://doi.org/10.1038/s41467-019-12085-w.

Kombrink, A., and Thomma, B.P.H.J. (2013). LysM effectors: secreted proteins supporting fungal life. PLoS Pathog. *9*, e1003769. https://doi.org/10.1371/journal.ppat.1003769.

Krijger, J.-J., Thon, M.R., Deising, H.B., and Wirsel, S.G.R. (2014). Compositions of fungal secretomes indicate a greater impact of phylogenetic history than lifestyle adaptation. BMC Genom. *15*, 722. https://doi.org/10.1186/1471-2164-15-722.

Lam, K.-K., LaButti, K., Khalak, A., and Tse, D. (2015). FinisherSC: a repeat-aware tool for upgrading de novo assembly using long reads. Bioinformatics *31*, 3207–3209. https://doi.org/10.1093/bioinformatics/btv280.

Li, Y., Steenwyk, J.L., Chang, Y., Wang, Y., James, T.Y., Stajich, J.E., Spatafora, J.W., Groenewald, M., Dunn, C.W., Hittinger, C.T., et al. (2021). A genome-scale phylogeny of the kingdom Fungi. Curr. Biol. *31*, 1653–1665.e5. https://doi.org/10.1016/J.CUB.2021.01.074.

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., and Kahmann, R. (2015). Fungal effectors and plant susceptibility. Annu. Rev. Plant Biol. *66*, 513–545. https://doi.org/10.1146/annurev-arplant-043014-114623.

Lowe, R.G.T., and Howlett, B.J. (2012). Indifferent, affectionate, or deceitful: lifestyles and secretomes of fungi. PLoS Pathog. *8*, e1002515. https://doi.org/10.1371/journal.ppat.1002515.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. Cancer Res. *27*, 209–220.

Miyauchi, S., Kiss, E., Kuo, A., Drula, E., Kohler, A., Sánchez-García, M., Morin, E., Andreopoulos, B., Barry, K.W., Bonito, G., et al. (2020). Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. Nat. Commun. *11*, 5125. https://doi.org/10.1038/s41467-020-18795-w.

Morris, C.E., Bardin, M., Kinkel, L.L., Moury, B., Nicot, P.C., and Sands, D.C. (2009). Expanding the paradigms of plant pathogen life history and evolution of parasitic fitness beyond agricultural boundaries. PLoS Pathog. *5*, e1000693. https://doi.org/10.1371/journal.ppat.1000693.

Nguyen, N.H., Song, Z., Bates, S.T., Branco, S., Tedersoo, L., Menke, J., Schilling, J.S., and Kennedy, P.G. (2016). FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. Fungal Ecol. *20*, 241–248. https://doi.org/10.1016/J.FUNECO.2015.06.006.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., et al. (2019). Vegan: Community Ecology Package. R Package Version 2.5-6 (The Comprehensive R Archive Network).

Paradis, E., and Schliep, K. (2019). Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526–528. https://doi.org/10.1093/bioinformatics/bty633.

Plett, J.M., and Martin, F. (2015). Reconsidering mutualistic plant–fungal interactions through the lens of effector biology. Curr. Opin. Plant Biol. *26*, 45–50. https://doi.org/10.1016/j.pbi.2015.06.001.

Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A., and Finn, R.D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. *46*, D624–D632. https://doi.org/10.1093/nar/gkx1134.

Romero, F., Cazzato, S., Walder, F., Vogelgsang, S., Bender, S.F., and van der Heijden, M.G.A. (2021). Humidity and high temperature are important for predicting fungal disease outbreaks worldwide. New Phytol. *234*, 1553–1556. https://doi.org/10.1111/nph.17340.

Samalova, M., Carr, P., Bromley, M., Blatzer, M., Moya-Nilges, M., Latgé, J.-P., and Mouyna, I. (2020). GPI anchored proteins in Aspergillus fumigatus and cell wall morphogenesis. Curr. Top. Microbiol. Immunol. 167–186. https://doi.org/10.1007/82_2020_207.

Samalova, M., Mélida, H., Vilaplana, F., Bulone, V., Soanes, D.M., Talbot, N.J., and Gurr, S.J. (2017). The β-1, 3-glucanosyltransferases (Gels) affect the structure of the rice blast fungal cell wall during appressorium-mediated plant infection. Cell Microbiol. *19*, e12659. https://doi.org/10.1111/cmi.12659.

Seifert, G.J. (2018). Fascinating fasciclins: a surprisingly Widespread family of proteins that mediate interactions between the cell exterior and the cell surface. Int. J. Mol. Sci. *19*, 1628. https://doi.org/10.3390/ijms19061628.

Smouse, P.E., Long, J.C., and Sokal, R.R. (1986). Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Syst. Zool. *35*, 627. https://doi.org/10.2307/2413122.

Spatafora, J.W., Chang, Y., Benny, G.L., Lazarus, K., Smith, M.E., Berbee, M.L., Bonito, G., Corradi, N., Grigoriev, I., Gryganskyi, A., et al. (2016). A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. Mycologia *108*, 1028–1046. https://doi.org/10.3852/16-042.

Sperschneider, J., Dodds, P.N., Gardiner, D.M., Singh, K.B., and Taylor, J.M. (2018). Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. Mol. Plant Pathol. *19*, 2094–2110. https://doi.org/10.1111/mpp.12682.

Stajich, J. (2020). 1KFG/PHYling_HMMs_fungi: PHYling Markers 1*3*. https://doi.org/10.5281/ZENODO.3630031.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

Stergiopoulos, I., Kourmpetis, Y.A.I., Slot, J.C., Bakker, F.T., De Wit, P.J.G.M., and Rokas, A. (2012). In silico characterization and molecular evolutionary analysis of a novel superfamily of fungal effector proteins. Mol. Biol. Evol. *29*, 3371–3384. https://doi.org/10.1093/molbev/mss143.

Stolzer, M., Lai, H., Xu, M., Sathaye, D., Vernot, B., and Durand, D. (2012). Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. Bioinformatics *28*, i409–i415. https://doi.org/10.1093/bioinformatics/bts386.

Tisserant, E., Malbreil, M., Kuo, A., Kohler, A., Symeonidi, A., Balestrini, R., Charron, P., Duensing, N., Frei dit Frey, N., Gianinazzi-Pearson, V., et al. (2013). Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. Proc. Natl. Acad. Sci. USA *110*, 20117–20122. https://doi.org/10.1073/pnas.1313452110.

Toro, K.S., and Brachmann, A. (2016). The effector candidate repertoire of the arbuscular mycorrhizal fungus Rhizophagus clarus. BMC Genom. *17*, 101. https://doi.org/10.1186/s12864-016-2422-y.

Vilela, R., and Mendoza, L. (2018). Human pathogenic entomophthorales. Clin. Microbiol. Rev. *31*. e00014-18. https://doi.org/10.1128/CMR.00014-18.

Vogt, M.S., Schmitz, G.F., Varón Silva, D., Mösch, H.U., and Essen, L.-O. (2020). Structural base for the transfer of GPI-anchored glycoproteins into fungal cell walls. Proc. Natl. Acad. Sci. USA *117*, 22061–22067. https://doi.org/10.1073/pnas.2010661117.

Wang, Y., Stata, M., Wang, W., Stajich, J.E., White, M.M., and Moncalvo, J.-M. (2018). Comparative genomics reveals the core gene toolbox for the fungus-insect symbiosis. mBio *9*. e00636-18. https://doi.org/10.1128/mBio.00636-18.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. *40*, W445–W451. https://doi.org/10.1093/nar/gks479.

Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. *18*, 821–829. https://doi.org/10.1101/gr.074492.107.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Zygomycete fungi and outgroups	This study and previous publications (Table S1)	Table S1
Deposited data		
Fungal genome sequences at GenBank	This study and previous publications (Table S1)	Table S1
HMM files for the 434 markers used for phylogenetic recontruction		https://github.com/1KFG/Phylogenomics_HMMs
Software and algorithms		
AllPathsLG	Gnerre et al. (2011)	http://software.broadinstitute.org/allpaths-lg/blog/?page_id=12
ape 5.5	Paradis and Schliep (2019)	http://ape-package.ird.fr/
Arachne Hybrid Assembler v. Nov. 2010	Jaffe et al. (2003)	http://www.genome.wi.mit.edu/wga
Arrow		https://github.com/PacificBiosciences/GenomicConsensus
BBDuk		https://sourceforge.net/projects/bbmap/
dbCAN database release 4.0	Yin et al. (2012)	http://bcb.unl.edu/dbCAN/
DIAMOND v.0.7.9.58	Buchfink et al. (2014)	https://github.com/bbuchfink/diamond
EffectorP2	Sperschneider et al. (2018)	https://github.com/JanaSperschneider/EffectorP-2.0
Falcon	Chin et al. (2016)	https://github.com/PacificBiosciences/FALCON
FinisherSC	Lam et al. (2015)	https://kakitone.github.io/finishingTool/
HMMSCAN in HMMER v3.1b	Rawlings et al. (2018)	http://hmmer.org
Indicspecies	Cáceres and Legendre (2009)	https://CRAN.R-project.org/package=indicspecies
InterProScan 5	Jones et al. (2014)	https://www.ebi.ac.uk/interpro/download/
LED database release 3.0		http://www.led.uni-stuttgart.de
MAFFT 7.402	Katoh and Standley (2013)	https://mafft.cbrc.jp/alignment/software/
MEROPS peptidase database release 11.0		https://www.ebi.ac.uk/merops/index.shtml
MycoCosm portal	Grigoriev et al. (2014)	https://mycocosm.jgi.doe.gov
NetGPI-1.1 online server	Gíslason et al. (2021)	https://services.healthtech.dtu.dk/service.php?NetGPI-1.1
NOTUNG 2.9	Stolzer et al. (2012)	http://www.cs.cmu.edu/~durand/Lab/Notung/
OrthoFinder 2.27	Emms and Kelly (2019)	https://github.com/davidemms/OrthoFinder
PHYLING	Stajich (2020)	http://github.com/stajichlab/PHYling_unified
Quiver	Chin et al. (2013)	https://www.pacb.com/products-and-services/analytical-software/open-source-tools/
RAxML v.8.0.26	Stamatakis (2014)	https://cme.h-its.org/exelixis/web/software/raxml/
Trinity v. 2.1.1.1 or 2.3.2	Grabherr et al. (2011)	https://github.com/trinityrnaseq/trinityrnaseq
VEGAN	Oksanen et al. (2019)	https://github.com/vegandevs/vegan
Velvet version 1.2.07	Zerbino and Birney (2008)	https://github.com/dzerbino/velvet

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Ying Chang (ying.chang@yale-nus.edu.sg).

Material availability

This study did not generate new unique reagents.

Data and code availability

The accession numbers for the genome assemblies are listed in Table S1.

This paper does not report original code.

Any additional information of the data is available from the lead contact upon request.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

No experimental model organism was used in this study.

METHOD DETAILS

Taxon sampling, trophic mode coding, and genome sequencing, assembly and annotation

We sampled 182 fungal genomes, with representative species from all the subphyla in Kingdom Fungi, including 98 Mucoromycota and 34 Zoopagomycota genomes (Table S1). We included three non-fungal Opisthokont genomes (*Monosiga brevicollis*, *Capsaspora owczarzki* and *Drosophila melanogaster*) as the outgroups for phylogenetic reconstruction.

We coded the sampled zygomycete fungi according to their trophic mode and ecological guild (Tables S2 and S3). The trophic mode categories included symbiotroph, pathotroph, and saprotroph, while the ecological guilds included amoeba parasite, animal pathogen, animal symbiont, arbuscular mycorrhiza, dung saprotroph, ectomycorrhiza, endophyte, mycoparasite, plant pathogen, saprotroph, & 'unknown' (1). We adopted two coding schemes – 1) Each genome coded by its primary trophic mode or ecological guild; 2) a binary coding method with '0' for absence and '1' for presence of a particular ecological guild. In this scheme, we allowed a fungus to have multiple ecological guilds when evidence suggests this possibility.

We newly sequenced 29 genomes in this work (Table S1). The genome of *Podila verticillata* (=*Mortierella verticillata*) NRRL 6337 was sequenced by the Broad Institute using Sanger and Illumina technology. Three libraries were constructed for Sanger sequencing, 4 and 10 kb plasmids and 40 kb Fosmids, and paired-end reads were generated. Illumina sequencing was performed on an Genome Analyzer II to generate 101 base paired reads. Reads were assembled using the Arachne Hybrid Assembler v. Nov. 2010 (Jaffe et al., 2003). The assembly was annotated using the Broad Institute fungal pipeline (Haas et al., 2011).

The remaining 28 genomes were sequenced by the Joint Genome Institution. Out of these 28 genomes, the genomes of *Cunninghamella echinulata* NRRL 1382 and *Dissophora ornata* CBS 347.77 were sequencing using Illumina platform, for which 100 ng of DNA was sheared to 300 bp using the Covaris LE220 and size selected using SPRI beads (Beckman Coulter). The fragments were treated with end-repair, A-tailing, and ligation of Illumina compatible adapters (IDT, Inc) using the KAPA-Illumina library creation kit (KAPA biosystems). qPCR was used to determine the concentration of the libraries and were sequenced on the Illumina Hiseq. The prepared libraries were quantified using KAPA Biosystems' next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified libraries were then prepared for sequencing on the Illumina HiSeq sequencing platform utilizing a TruSeq paired-end cluster kit, v4. Sequencing of the flowcell was performed on the Illumina HiSeq2500 sequencer using HiSeq TruSeq SBS sequencing kits, v4, following a 2 × 150 indexed run recipe.

All raw Illumina sequence data was filtered for artifact/process contamination and initially assembled using VelvetOptimiser version 2.1.7 with Velvet version 1.2.07 (Zerbino and Birney, 2008) using the following parameters; "–s 61 –e 97 –i 4 –t 4, –o "-ins_length 250 -min_contig_lgth 500" ". The resulting assembly was

used to simulate 2 × 100bp 3000 ± 300bp insert long mate-pair library with wgsim version 0.3.1-r13 using "-e 0–1 100–2 100 -r 0 -R 0 -X 0 -d 3000 -s 30". 25× of the simulated long mate-pair was then co-assembled together with 125× of the original Illumina filtered fastq with AllPathsLG (Gnerre et al., 2011) version R49403 to produce the final nuclear assembly.

All other genomes were sequenced using the PacBio platform, for which, 1-6 ug of genomic DNA was sheared to 10kb or >10kb using Covaris g-Tubes. The sheared DNA was treated with exonuclease to remove single-stranded ends and DNA damage repair mix followed by end repair and ligation of blunt adapters using SMRTbell Template Prep Kit 1.0 (Pacific Biosciences). All libraries were then purified with AMPure PB beads. After libraries were prepared, PacBio Sequencing primer was then annealed to the SMRTbell template library and Version P6 sequencing polymerase was bound to them. Then, the prepared SMRTbell template libraries were then sequenced on a Pacific Biosciences RSII sequencer using Version C4 chemistry and 1 × 240 sequencing movie run times. However, for *C. cucurbitarum*, *H. radiatus*, *H. pulchrum* and *C. mojavensis*, sequencing polymerase was bound to them using Sequel Binding kit 2.0 and then the prepared SMRTbell template libraries were sequenced on a Pacific Biosystems' Sequel sequencer using v3 sequencing primer, 1M v2 SMRT cells, and Version 2.1 sequencing chemistry with 1 × 360 & 1 × 600 sequencing movie run times. For PacBio-sequenced genomes, a variety of methods were used for assembly as new tools became available. All genomes were assembled using Falcon (Chin et al., 2016). FinisherSC (Lam et al., 2015), was then used in most cases to improve the assembly, followed by polishing with either Quiver (Chin et al., 2013). or Arrow (https://github.com/PacificBiosciences/GenomicConsensus). See Table S1 for details on tools used for assembly, finishing and polishing.

All transcriptomes were sequenced with Illumina. For all lineages except *D. ornata* CBS 347.77 and *H. pulchrum* RSA 2064, stranded cDNA libraries were generated using the Illumina Truseq Stranded mRNA Library Prep kit. mRNA was purified from 1 ug of total RNA using magnetic beads containing poly-T oligos. mRNA was fragmented and reversed transcribed using random hexamers and SSII (Invitrogen) followed by second strand synthesis. The fragmented cDNA was treated with end-pair, A-tailing, adapter ligation, and 8 cycles of PCR. The prepared libraries were quantified using KAPA Biosystems' next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified libraries were then prepared for sequencing on the Illumina HiSeq sequencing platform utilizing a TruSeq paired-end cluster kit, v4. Sequencing of the flowcell was performed on the Illumina HiSeq2000 sequencer using HiSeq TruSeq SBS sequencing kits, v4, following a 2 × 150 indexed run recipe.

For *D. ornata* CBS 347.77 and *H. pulchrum* RSA 2064, plate-based RNA sample prep was performed on the PerkinElmer Sciclone NGS robotic liquid handling system using Illumina's TruSeq Stranded mRNA HT sample prep kit utilizing poly-A selection of mRNA following the protocol outlined by Illumina in their user guide (https://support.illumina.com/sequencing/sequencing_kits/truseq-stranded-mrna.html). For *D. ornata*, the quantified libraries were sequenced as described earlier but using HiSeq2500. For *H. pulchrum*, sequencing of the flowcell was performed on the Illumina NovaSeq sequencer using NovaSeq XP V1 reagent kits, S4 flowcell, following a 2 × 150 indexed run recipe.

All reads were filtered and trimmed for quality. BBDuk (https://sourceforge.net/projects/bbmap/) was used to remove artifact sequences by kmer matching (kmer = 25), allowing 1 mismatch. Detected artifacts were trimmed from the 3′ end of the reads. RNA spike-in reads, PhiX reads and reads containing any Ns were removed. Quality trimming was performed using the phred trimming method set at Q6. Finally, following trimming, reads under the length threshold were removed (minimum length 25 bases or 1/3 of the original read length - whichever is longer). Filtered reads were assembled *de novo* using Trinity (Grabherr et al., 2011) version 2.1.1.1 or 2.3.2. The assemblies and annotations of newly sequenced genomes are available at the fungal genome portal MycoCosm (https://mycocosm.jgi.doe.gov) (Grigoriev et al., 2014) and in the DDBJ/EMBL/GenBank repository (https://www.ncbi.nlm.nih.gov; Table S1).

Phylogenetic reconstruction

The phylogenetic reconstruction was based on a dataset comprising 434 protein markers that have previously proved useful for higher level phylogenetic analysis of fungi (Beaudet et al., 2018; Chang et al., 2019) (https://github.com/1KFG/Phylogenomics_HMMs) (12). The retrieval and subsequent alignment of the markers were performed using the pipeline PHYLING (http://github.com/stajichlab/PHYling_unified) (Stajich, 2020) with the default settings. Phylogenetic reconstruction was performed on the concatenated

alignment of the 434 markers in RAxML v.8.0.26 (Stamatakis, 2014), with the '-f a' option, 100 bootstrap replicates and the PROTGAMMALG model.

Secretome retrieval and characterization

The secretome identification followed the pipeline from Chang et al. (2019). The only modification was the prediction of proteins with non-permanent residence in the endoplasmic reticulum which was done using Prosite (PS00014; https://prosite.expasy.org/scanprosite/). The CAZyme classes analyzed included auxiliary activities (AAs), carbohydrate esterases (CEs), glycoside hydrolases (GHs), and polysaccharide lyases (PLs), as well as the non-catalytic carbohydrate binding modules (CBMs). Hidden Markov model (HMM) profiles of 259 CAZyme families were downloaded from the dbCAN database release 4.0 (http://bcb.unl.edu/dbCAN/) (Yin et al., 2012) (14) and used to query the predicted proteomes with HMMSCAN in HMMER v3.1b (Eddy, 2011). The results were filtered with an *e*-value cutoff of 1e-15 and an alignment coverage cutoff of 50% of the domain query. The predicted proteome were searched against sequences from 453 peptidase subfamilies from MEROPS peptidase database release 11.0 DIAMOND v.0.7.9.58 (Buchfink et al., 2014) (merops_scan.lib, https://www.ebi.ac.uk/merops/index.shtml) (Rawlings et al., 2018) with an *e*-value cutoff of 1e-10. Characterization of lipases was performed with DIAMOND searches against the 38 lipase superfamilies from the LED database release 3.0 (http://www.led.uni-stuttgart.de) with an *e*-value cutoff of 1e-5. SSPs were defined as secreted proteins shorter than 300 amino acids. SSPs containing one or more catalytic domains of digestive enzymes were considered as digestive enzymes. The remaining secreted proteins, neither digestive enzymes nor SSPs, were arbitrarily grouped together as OTHER proteins. We performed orthologous clustering on zygomycete SSPs and OTHER proteins using OrthoFinder 2.27 (Emms and Kelly, 2019) with default settings.

Functional annotation of SSPs and OTHER proteins was performed using InterProScan 5 (Jones et al., 2014). Putative effectors were predicted using EffectorP2 with default settings (Sperschneider et al., 2018). NetGPI-1.1 online server (https://services.healthtech.dtu.dk/service.php?NetGPI-1.1) (Gíslason et al., 2021) was applied to predict the proteins bound to cell membrane or cell walls via Glycosylphosphatidylinositol-anchors (GPI-anchors).

Comparative analysis on zygomycete secretomes

Non-metric multidimensional scaling (NMDS) analyses were performed using VEGAN (Oksanen et al., 2019) to compare the protein composition among sampled zygomycete secretomes. Bray-Curtis dissimilarity matrices were used. The results were visualized by plotting NMDS coordinates for each genome.

We generated protein and genome clusters with 2 × 2 hierarchical clustering analyses using the hclust function in R, with average linkage clustering and a Bray-Curtis dissimilarity matrix. Results were visualized as heatmaps using heatmap.2 in ggplot.

The R package Indicspecies (Cáceres and Legendre, 2009) was used to identify representative proteins for fungi belonging to the same subphylum or with the same primary ecological guild, using the multipatt function. The indicator values included two factors – specificity factor A estimating the probability of a particular indicator protein belonging to the target fungal group and fidelity factor B denoting the probability of finding the indicator protein in the target fungal group.

Mantel tests (Mantel, 1967) were performed using VEGAN to evaluate the explanatory power of phylogeny and ecological guild on the secretome compositions. The phylogenetic distance matrix was generated based on the best RAxML tree using the cophenetic.phylo function from ape v5.5 (Paradis and Schliep, 2019). Four distance matrices based on Bray-Curtis dissimilarity were calculated for the total secretome, digestive enzymes, SSPs, and OTHER proteins. The ecological guild (primary or multiple) was coded in a binary manner and was used to calculate Gower distance matrices (Table S3). The influence of the phylogenetic relatedness was corrected in the partial Mantel tests (Smouse et al., 1986) on ecological guild and protein compositions by including phylogenetic distance matrix as the dependent matrix. Likewise, the influence of ecological guilds was corrected in the partial Mantel analysis on the phylogenetic distance and the protein compositions. Pearson correlation was used for all the regular and partial Mantel tests with 999 permutations.

Reconstruction of digestive enzyme evolution

The sequences for each digestive enzyme family were aligned using MAFFT 7.402 (Katoh and Standley, 2013) (L-INS-i, maxiterate = 1000). Individual enzyme trees were inferred using RAXML v.8.0.26 (Stamatakis, 2014) (-f a option, 100 bootstrap replicates and a PROTGAMMALG model). The evolution of enzyme families was reconstructed by reconciling the enzyme trees and the best RAxML species tree using NOTUNG 2.9 (Stolzer et al., 2012). To take into account the uncertainty in the enzyme trees, we allowed the branches with <70% ML support to be rearranged to obtain the most parsimonious solution during the reconciliation.