A genomic perspective on fungal diversity and evolution

Stephen J. Mondo[1,2,3] and Igor V. Grigoriev[1,2,4,†]

[1]U.S. Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

[2]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

[3]Department of Agricultural Biology, Colorado State University, Fort Collins, CO 80521, USA

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

[†]e-mail: ivgrigoriev@lbl.gov

Abstract

Originating from aquatic unicellular ancestors, over the course of ~1 billion years the fungi have evolved to occupy nearly all aerobic environments on the planet, diversified into millions of different 'species' and have evolved complex multicellular structures. Their relatively small, simple genomes have facilitated massive-scale sequencing and allowed us to explore genome evolution across an ancient eukaryotic kingdom. With now thousands of genomes available from diverse lineages, in this review we will discuss insights into fungal biology and evolution gleaned through use of genomics and other multi-omics approaches. Using published genomes available through GenBank and the Joint Genome Institute's MycoCosm platform, we generated kingdom-wide phylogenies which we use to highlight how fungal genomes have changed over time. With this phylogeny as a guide, we also discuss major evolutionary transitions that occurred across the kingdom. Unfortunately, while progress has been made, these efforts are hampered by biases in genome representation and limited characterization of gene functions. Here we discuss these challenges and possible future directions to address them including initiatives to characterize conserved genes of unknown function and scale up sequencing toward 10,000 annotated fungal genomes.

Introduction

With the origin of eukaryotes estimated at ~1.84 billion years[1], the fungi represent one of the oldest kingdoms of eukaryotes[2]. Over ~1 billion years[3,4], fungi have diversified into millions of

different species[5] split across multiple phyla, themselves often containing more diversity than entire other eukaryotic kingdoms. For example, sequence divergence in the budding yeast subphylum alone (Saccharomycotina) exceeds that found between humans and roundworms, or across all Viridiplantae[6]. This pattern persists even at lower taxonomic scales, which can harbor an impressively diverse catalog of species. Examples include genera such as *Aspergillus* and *Penicillium* (both producers of diverse, pharmaceutically-relevant natural products[7]) estimated at ~82 and 74 million years old, respectively[8], among others[9]. For a sense of scale, the genus *Homo*, containing us humans, emerged somewhere around 2.7 to 3 million years ago[10].

As expected given its age, the fungal kingdom hosts an incredible amount of phenotypic and phylogenetic diversity. Starting at the base of the fungal tree and only relatively recently recognized as fungi[11], the earliest-diverging branch is the Microsporidia, obligate intracellular parasites with some of the smallest eukaryotic genomes (2 - 10 Mb). Microsporidia harbor a harpoon-like polar filament to infect their animal (primarily insect) hosts (Figure 1). They also lack chitinous cell walls and mitochondria (harboring a mitosome instead[11]). Also at the root, the Cryptomycota (also known as Rozellomycota) are poorly explored but incredibly diverse[12]. Generally thought to be obligate parasites in aquatic and soil environments, Cryptomycota share some characteristics of Microsporidia (including highly reduced genomes[13]) but have flagella, mitochondria, chitinous cell walls for sporangia (although chitin was not detected in environmental samples[14]), and no polar filaments[15]. Some of these characteristics are also shared with aphelids, another closely related clade to Cryptomycota[16,17]. However, some phylogenies place Cryptomycota and Microsporidia as sister to the Fungi and more closely related to amoeba[15,18], calling for additional data to resolve these relationships.

Many of the earliest diverging fungi are primarily aquatic, e.g. Chytridiomycota and Blastocladiomycota, producing motile zoospores bearing flagella[19]. These taxa are poorly understood, but are generally thought to be degraders of cellulose and chitin[20,21] or pathogens, frequently of algae but also causing diseases in animals (such as chytridiomycosis in amphibians[22]).

From motile unicellular ancestors bearing flagella, the remaining fungal phyla have lost this trait (except *Olpidium*[23]), instead growing filamentously, as unicellular yeasts, or both (dimorphic). At the base of non-flagellated fungi, individuals occupy a range of primarily non-plant-associated niches. This includes members of the Zoopagomycota which are often saprobic or associate with animal hosts (primarily insects), either as mutualists or pathogens, or are mycoparasites. Evolution of the remaining phyla is heavily tied to the diversification of land plants[24], where they have developed extensive mutualistic, pathogenic and saprobic associations. Of these phyla, the Mucoromycota are the least understood, but have important impacts on the bioeconomy, agriculture, plant health and animal disease[25–27]. The others, Ascomycota and Basidiomycota, comprise the Dikarya subkingdom. This hugely successful group with diverse lifestyles includes many mutualists, saprobes and, on occasion, pathogens. While rare, devastating pathogens of plants (e.g. soybean and other rusts[28–30]) and animals (e.g. causing candidiasis[31] and aspergillosis[32]) are also found in the Dikarya. This subkingdom also displays an impressive range of morphological complexity, from simple yeast-like morphologies to complex fruiting bodies like mushrooms[33]. Given how comparatively well studied Dikarya are, researchers often combine non-Dikarya fungi using a single term of ‘Early Diverging Fungi’ (EDF), although terminology here is still under debate[34].

As a hyperdiverse kingdom, the Fungi occupy nearly all aerobic environments (and some anaerobic ones[35]) across the planet, extreme and otherwise, and their capabilities and lifestyles are plentiful[36,37]. Some traits are ancient and conserved such as osmotrophy[38] (a feeding

mechanism whereby fungi absorb dissolved nutrients via osmosis) while others are highly plastic and evolved convergently multiple times[33], including dimorphism (the ability to shift between yeast or filamentous growth[39]), mycorrhization, and lichenization (both mutualistic associations with very divergent photosynthetic partners[40,41]). Yet others, like the loss of flagella and abilities to decay wood appear to have occurred at fixed timepoints[23,42].

From decaying wood to rotting fruit and Antarctic dry valleys to rainforests, fungi inhabit many niches[36] and can produce an incredible array of enzymes and secondary metabolites[43] (of great economic value). They are also important pathogens of plants, animals, insects and other fungi[44], making them both agricultural/health problems and candidate biocontrol agents. Towards understanding these capabilities and their evolution, genomics has emerged as a powerful tool linking genes to traits, metabolic capabilities, and lifestyles. For example, expansions of Carbohydrate Active enZYmes (CAZymes) are often associated with plant biomass decomposition (e.g. white rot fungi) and contractions in biotrophic plant associations (e.g. mycorrhizae[40,45]). Additionally, mycorrhizal fungi, which associate with plant roots and provide their host with nutrients (particularly phosphorus and nitrogen) in return for sugar[26], appear to repeatedly evolve similar genomic profiles through convergent evolution[45]. In contrast, other symbioses such as the *Termitomyces*-termite association - where termites cultivate *Termitomyces* (Agaricomycotina) as a food source - appear to contain genome features that predispose them to domestication[46]. Taking advantage of such genome features, new machine learning and other techniques have been successfully applied to predict lifestyles and niche partitioning in diverse fungi[47–51].

Through the course of evolution, fungal genome characteristics have also changed substantially. While as a kingdom they consistently display small (< 50 megabases), repeat-poor genomes, some clades have taken this to the extremes. For example, lineages like the Saccharomycotina and early diverging Microsporidia have genomes sometimes smaller than bacteria[52,53], while others like the rust fungi (Puccinomycotina[28]), arbuscular mycorrhizal fungi (AMF[54]), and other EDF[55] can have genomes approaching or exceeding 1 gigabase in size.

We also see a diversity of ‘ploidy’ states in fungi. While genomics has revealed that early fungal evolution was dominated by a diploid state[56], many fungi exist vegetatively as haploids. For example, filamentous ascomycetes produce clonal, haploid asexual spores in high numbers with limited shelf-life. Sexual reproductive structures produced by compatible mates in some cases are more long-lived and recalcitrant to environmental stressors compared to asexual spores[57,58]. However, yeasts and non-Dikarya fungi are frequently found as proper diploids[48,56], while members of the Basidiomycota (especially Agaricomycotina) persist predominantly as ‘dikaryons’, where two haploid nuclei coexist vegetatively in the same cell with important evolutionary consequences[59]. Divergence between dikaryotic nuclei can sometimes be exceptionally high, for example in *Schizophyllum commune* where compatible strains can exhibit a remarkable 0.20 synonymous site diversity within a population[60,61]. Genomic approaches have also revealed further nuance in the form of aneuploidy, hybridization, and large-scale segmental duplications which can also be a driver of certain traits[62–64].

Relatively small fungal genomes enabled production of good quality fungal genome assemblies using any next generation sequencing technology. Consequently, starting with the genome of *S. cerevisiae*[65] in 1996 followed by a series of other fungal models[66–68], we now see an explosion in available genome resources for diverse fungi (Figure 2). Crucial for large-scale analyses, genomes are also often accessible within the same comparative genomics platforms, including FungiDB[69], Ensembl Fungi[70], the National Center for Biotechnology Information (NCBI) GenBank database[71] and the DOE Joint Genome Institute’s (JGI) MycoCosm[72]. Furthermore,

high throughput genomics centers like the JGI and research consortia such as the Earth BioGenome[73], Darwin Tree of Life (http://darwintreeoflife.org), Y1000+[48] and others have developed standardized workflows to produce high quality, comparable genome assemblies[74] and annotations[72,75], a prerequisite for impactful comparative studies.

Here, we will discuss breakthroughs in our understanding of fungal biology and evolution revealed through genomics and other multi-omic approaches and provide examples of how combining an evolutionary perspective with big data help uncover important biological processes**.** To this end, we used published data from GenBank and MycoCosm (Supplementary Table 1 and viewable interactively via https://mycocosm.jgi.doe.gov/ncbi-combined-tree) to explore sampling gaps across the fungal tree (Figure 3) and reconstructed a phylogeny of ~1440 published fungal genomes (Supplementary Table 2) available through MycoCosm (Figure 4), overlaid with genome statistics to illustrate how fungal genomes differ across clades. We also discuss foreseeable barriers to (and possible solutions for) scaling up genome sequencing efforts, characterizing gene function, and exploring fungi in ecosystems.

Fungal Phylogenomics

A brief history

Originally classified as plants[76], the fungi have experienced a long history of misclassification. This was due to both their microscopic nature and their highly similar morphology during vegetative growth. Prior to DNA sequencing, this meant researchers primarily relied on reproductive structures (when present) to infer taxonomic relationships. This reliance also led to an additional complication: multiple names assigned to a single species, depending on the reproductive stage observed. Many of these challenges were overcome when DNA sequencing and phylogenetics became possible. This allowed us to reconstruct evolutionary relationships through alignment of genes (phylogenetics) and eventually thousands of genes (phylogenomics) across taxa (Box 1). We now know that fungi are more closely related to animals than plants, can resolve relationships amongst fungi at much higher resolution, and can assign a single name to each species[77].

Efforts to build a robust phylogeny of fungi have been underway for decades. Especially challenging has been resolution of taxonomic relationships amongst EDF. A major breakthrough came in 2006 when, as part of the Assembling Fungal Tree of Life project, James et al.[78] generated a 6-gene phylogeny for nearly 200 fungi spanning the entire kingdom. With a well-supported phylogeny in hand, James et al. were able to discern that chytrid fungi, all formerly placed in the single phylum of Chytridiomycota, were not monophyletic[19]. Next, Spatafora et al.[79] used 46 genomes to develop a genome-scale phylogeny which was used as a framework to describe evolutionary relationships across the kingdom, particularly EDF. Moving from the root which consists of motile fungi lacking complex reproductive structures, the chytrids were split into 3 different phyla: Cryptomycota, Chytridiomycota (*sensu stricto*), and Blastocladiomycota. The remaining EDF are predominantly filamentous and were grouped into two phyla, Zoopagomycota and Mucoromycota, each with 3 subphyla, where we start to see the emergence of multicellular structures in the form of sporocarps, e.g *Modicella*[80] and *Endogone*[81] (Figure 1). Next, we see the Dikarya, where complex multicellularity and diverse lifestyles are observed. Recent work expanded this concept to encompass 20 fungal phyla[82]. However, it should be noted that fungal taxonomy remains fluid and under debate. This is especially true for EDF. New data continue to emerge and new classifications are proposed (e.g. Sanchytriomycota[83]). For consistency with genome-based classification and nomenclature in

NCBI's taxonomy database, throughout this review we will use the fungal classification schema described in Spatafora et al., 2016[79]. For additional details on fungal phylogenomics and taxonomy, see James et al., 2020[84] and Li et al., 2021[85].

Highly diverse early-diverging lineages

Unlike the Dikarya, -omics sampling of EDF remains sparse. This gap is reflected by the number of genomes present in public resources such as MycoCosm[72] where, as of 7/1/2024, Dikarya genomes outnumber EDF almost ten to one (2324 Dikarya versus 279 EDF). While this ratio roughly matches the number of described 'species' in each of these groups[82], we expect that the undersampled EDF harbors much greater richness than described today. Fortunately, interest in genomics of EDF is growing due to a variety of interesting characteristics present in these lineages. For example, some of the most powerful degraders of plant biomass are members of Neocallimastigomycetes. These unusual fungi are found in anaerobic gut environments and lack mitochondria but appear to harbor up to seven times more plant biomass degrading enzymes compared to the most powerful aerobic plant biomass decomposers in Agaricomycetes[14]. EDF also contains arbuscular mycorrhizal fungi from the Glomeromycotina which associate with the vast majority of land plants and are vital to plant health[66]. Other attractive areas under investigation in EDF include i) chytrid algal and plant pathogens[87,88], ii) bacterial endosymbionts altering fungal reproduction and lipid profiles (of relevance to biofuel production)[89,90], and iii) widespread use of non-canonical DNA modifications to regulate gene expression[91].

It is also clear that EDF phylogenetically dominate, where they constitute the majority of the kingdom[82]. At the same time, the limited exploration of these taxa has minimized their visibility within the larger scientific community, creating an artificial separation between these and the better studied Dikarya[34]. Thankfully, resolution here continues to increase with light coverage sequencing of ~1000 EDF genomes[92,93] and single cell sequencing[94] which could further reorganize EDF taxonomy.

Sampling gaps

With an estimated 2.2 to 6.2 million species[5,95], the fungal kingdom is ancient and home to an incredible amount of phenotypic diversity. Unfortunately, with only a few hundred thousand characterized (e.g., approximately 200,000 named fungal taxa in the NCBI taxonomy database) and even fewer available via culture collections (e.g., around 20,000 species available through CBS at the Westerdijk Institute, one of the largest fungal culture collections), we have just barely scratched the surface of potential traits and innovations within this kingdom. One crucial direction towards a better understanding of this diversity is through genome sequencing to describe the genes and features associated with those traits. Such analyses can reveal key genomic features associated with lifestyles, like what was discovered about the origin of the AM symbiosis through interrogation of *Geosiphon pyriformis*[96] - a rare lineage within the Glomeromycotina that forms a nitrogen-fixing symbiosis with *Nostoc* cyanobacteria.

Despite thousands of fungal genome assemblies available from NCBI or MycoCosm (Figure S1), available annotated genomes - especially published and freely accessible ones - for diverse species are surprisingly limited (Figure 2; Figure S1) meaning our genomic representation of diversity across the fungal kingdom is far from comprehensive. Our sampling is also highly skewed, with deep representation for certain clades, particularly surrounding model organisms like *S. cerevisiae*[97] likely leaving others vastly undersampled or lacking

genomic representatives altogether (e.g. Bartheletiomycetes and Entorrhizomycetes). More comprehensive sequencing of such taxa could present huge opportunities for new discoveries. For example, increased sequencing depth in the Neocallimastigomycetes has unearthed a treasure trove of CAZymes which can be used for improved biofuel production[20,98]. Similarly, sequencing of 100 Xylariales[43], 100 *Penicillium*[99], and 300 *Aspergillus*[100,101] genomes revealed hundreds of novel natural products which may yield valuable future pharmaceuticals. To identify parts of the tree likely harboring more novel information (in the form of lineage-specific genes) that are worthy of deeper sequencing, we reconstructed the phylogeny of fungi using all published annotated reference genomes available through JGI and/or NCBI as of August 23rd, 2024 (~2000 genomes; Supplementary Table 1) and explored conservation of genes using mmseqs2[102] to cluster genes into families (Figure 3). To observe the right of primary data producers to publish on their genomes first[103,104], we only included genomes with associated PubMed IDs.

We see that overall, most genes in fungal genomes are conserved across the kingdom, except for Microsporidia[105] and several clades with enlarged gene sets (e.g. Neocallimastigomycetes, Glomeromycotina, Pucciniales, and Pezizomycetes). Additionally, many clades are rich in lineage-specific genes. Through examining abundance of lineage-specific genes under a given node, normalized by node length (distance from node to furthest tip; see Supplementary Information), we find that many groups could harbor more novelty than is currently represented in MycoCosm or GenBank (darker sectors on Figure 3). This includes most EDF, some understudied Basidiomycota where we still only have single genome representatives for many orders[106], and several ascomycete clades (e.g., Orbiliomycetes, Pezizomycetes and Lecanoromycetes). Mushroom-forming fungi (Agaricomycetes) also have high levels of species-specific genes, suggesting that there is a lot more to discover in this subphylum. Figure 3 also shows a need for deeper sequencing in Saccharomycotina, a challenge certainly now addressed with the recent publication of >1000 Saccharomycotina genomes[48] but not yet incorporated into MycoCosm or GenBank. Additionally, we note that for ~4000 fungal taxa only assemblies but not annotations were deposited to NCBI (Figure S1), diminishing the usability of these datasets.

In cases where genomic representation is good, we largely have focused, large-scale sequencing efforts like the Y1000+ project to thank. Additional examples include sequencing and analysis of 101 Dothideomycetes[47], 99 Sordariomycetes[107], 100 *Penicillium*[99] and 300 *Aspergillus* genomes[100,101,108], which have been helpful in closing knowledge gaps and enabling reclassification of many species. Other ambitious projects such as the 1000 Fungal Genomes Project[72], a community effort to develop reference genomes across the fungal tree of life (https://mycocosm.jgi.doe.gov/1000-fungal-genomes) have played a crucial role in providing a backbone phylogeny which now must be filled out in greater detail to provide comprehensive representation of diversity within clades.

Massive scale sequencing elucidates fungal evolution

Shifts in genome features

Genomics has led to multiple breakthroughs in our understanding of how fungi have evolved. Comparative analysis across 83 eukaryotes revealed that compared to Metazoa, fungal gene repertoires are more similar to the last common ancestor of Opisthokonta and contain more metabolic genes while also showing increased prevalence of Horizontal Gene Transfer (HGT; also discussed below)[109]. In addition to shifting gene landscapes, the genome architecture and content itself can change. See Stajich, 2017[110] for a review on some of these changes across different fungal clades, and Box 2 for a description of mechanisms and implications of fungal genome rearrangements. Genome features and gene content can also vary considerably across fungi (Figure 4; Supplementary Table 2), providing us with a wealth of information about fungal biology and evolution.

For example, in the large and extremely A/T biased genomes of gut-dwelling Neocallimastigomycetes we see elevated numbers of CAZymes due to their powerful plant biomass degradation capabilities[98] and potential to manipulate microbial communities due to their high numbers of biosynthetic gene clusters[111] (BGCs; Figure 4). While Neocallimastigomycetes are an exception amongst EDF, BGCs are typically most enriched in the Pezizomycotina, particularly Sordariomycetes and Eurotiomycetes[43,100]. We also see general trends, where significant genome compaction correlates with reduced gene content and increasing density (e.g. Microsporidia, Ustilaginomycotina and Saccharomycotina[6,106,112]). Interestingly, most of these lineages have proteins of similar length to other fungi, but seem to have further saved on space through kicking out most introns. In contrast, rust genomes in Pucciniomycotina are the largest with huge expansions in repeat content[28] followed by *Gigaspora* in Glomeromycotina[54] and *Zoophthora* in Entomophthoromycotina[56]. See Box 3 for a discussion on repetitive content and its potential implications in fungal genome evolution.

Gene loss drives evolution across a subphylum

Over long timescales, it appears that many fungal clades have undergone substantial refinement of gene content. Such refinement was taken to the extreme in the Saccharomycotina (Ascomycota). The Saccharomycotina is an ancient clade of primarily yeast-like organisms that occupy a range of niches and consume a diversity of sugars[6,48], and includes several major model organisms (e.g. baker yeast *Saccharomyces cerevisiae* and human pathogen *Candida albicans*). In this subphylum, genome refinement was observed through massive subphylum-wide sequencing as part of the Y1000+ project. Using ~330 yeast genomes, Shen et al., 2018[6] coupled evolutionary histories with metabolic capabilities to trace how yeast metabolism has changed over time. This revealed that the common ancestor of yeasts was more metabolically complex than most extant yeasts and that pervasive gene and trait loss gave rise to the Saccharomycotina we see today.

Next, after completing the sequencing of >1000 genomes representing nearly all known Saccharomycotina species, Opulente et al., 2024[48] explored niche breadth across the subphylum, which revealed that intrinsic factors (e.g. gene content) rather than extrinsic ones (e.g. environment) best explain yeast metabolic breadth. They show that while generalists and

specialists tend to maintain a similar repertoire of metabolic genes, interconnectedness between metabolic pathways was greater in generalists compared to specialists. Using machine learning, they found that genes associated with elevated growth rates and metabolic flexibility (e.g. genes involved in ATP production) were often enriched in generalists compared to specialists. Taken together, these findings suggest little downside to being a generalist, raising important questions about how specialists emerge in fungi (e.g. drift versus adaptation).

Like Saccharomycotina, as noted above, several large clades exhibit a shift towards small, intron-poor genomes, for example, Taphrinomycotina and Ustilaginomycotina (Figure 4). In these cases, similar routes to genome simplification were also found, including diversification of transcription factors associated with regulation of the yeast-filamentous transition[39]. It is interesting to note that when this transition occurs, yeast (or dimorphic) morphology often dominates (Figure 1). Given these similarities and the findings in Saccharomycotina[6,48], it is tempting to speculate that similar mechanisms of genome reduction are occurring across all these clades. As we sequence more from these understudied clades, comparing across them will be valuable for extracting themes associated with fungal genome reduction and niche specialization.

Horizontal gene transfer

Horizontal Gene Transfer (HGT) is also an important driver of fungal evolution[98,113,114]. As opposed to vertical transmission from parent to offspring, HGT occurs when genetic material is transferred between unrelated individuals. In fungi, HGT has been documented to impact diverse biological functions, from ancient processes like the cell cycle and osmotrophy[38,115,116] to production of natural products through transfer of entire biosynthetic gene clusters (see Box 2) which produce new toxins and expand niche space[111,117]. Particularly impressive, one clade of yeasts acquired an entire bacterial operon for siderophore biosynthesis [118]. Other means of DNA transfer may involve transposable elements known as ‘Starships'[119], which were shown to drive fungal HGT[120] and are discussed further in Box 3.

While rare in some clades[118], HGT can play outsized roles in others, for example, the Neocallimastigomycetes, *Trichoderma* and *Armillaria*[98,113,117,121]. In Neocallimastigomycetes in particular, HGT rates are high, thousands of bacterial genes are present, and HGT plays a central role in adaptation within the anaerobic ruminant gut environment. Specifically, many genes of bacterial origin are involved in breakdown of lignocellulose[98] while BGCs are dominated by inter-fungal HGT[111]. In *Trichoderma*, *Armillaria*, and several other genera, HGT was reported to happen primarily between fungi[113,117]. Large-scale assessment through gene-tree species-tree reconciliation (Box 1) may help resolve the impact of HGT on fungal evolution. HGT from bacterial endosymbionts that inhabit fungal mycelium also likely provided important fungal regulatory machinery[122].

Endosymbiosis

While innovation can come through HGT or other changes within a genome, another means towards metabolic innovation and niche expansion comes through symbiosis. Unlike most other fungal phyla, the Mucoromycota show a predilection to host intimate, long-term associations with bacterial endosymbionts[123]. Specifically, members of all 3 subphyla have established endosymbioses, some which originated long ago[89] and others more recently[124]. As many fungal endosymbionts cannot be cultivated independent of their hosts, genomics has been key to understanding these interactions.

Often mutualistic in nature, these symbiotic associations provide important benefits to their fungal hosts. For example, a bacterium, *Mycetohabitans rhizoxinica*, provides its fungal host *Rhizopus microsporus* - important in food spoilage and a causal agent of Mucoromycosis[125] - with potent secondary metabolites enabling fungal infection of rice[126] and *Candidatus* Glomeribacter gigasporarum enhances presymbiotic growth rates of their arbuscular mycorrhizal (AM) hosts[127]. Potentially associated, endobacteria appear to grant AM fungi with higher respiration activity and ATP compared to fungi without bacteria[128]. This observation aligns with results of genome-wide positive selection analyses in endosymbionts which indicate an enrichment in energy metabolism-related proteins under selection[129].

However, these associations are not without their costs - even seemingly mutualistic associations show evidence of ancestral antagonism[130]. Comparative genomic and transcriptomic analyses within these systems have revealed manipulation of reproductive signaling cascades, alterations of lipid metabolism, and other impacts on host biology[131,132]. Those which impact host reproduction are expected to have important implications on evolution and mutualistic outcomes[133]. While a hijacked reproductive system may pose a problem for the fungus, it is of great advantage to us, especially in a subphylum largely recalcitrant to genetic modification. This feature allowed us to shed light onto the genetics of sexual reproduction in this subphylum[131], revealing common themes across Mucoromycota and Dikarya, as well as potential receptors for phylum-specific mating pheromones (trisporic acid derivatives).

Evolution of fungal multicellularity

Over time, fungi have evolved a variety of diverse reproductive structures. Some lineages, such as the Agaricomycotina, produce complex above-ground multicellular reproductive structures (mushrooms) that disperse spores aerially, while other lineages have evolved diverse microscopic structures or underground fruiting bodies (truffles) that are sought after and dispersed by animals. Complex multicellularity has also emerged in unusual places like in *Neolecta*[134] (Taphrinomycotina), which has helped inform our understanding of multicellularity across Ascomycota. To elucidate the role gene regulatory changes play in driving evolution of fungal reproductive structures, Trail et al., 2017[135] applied an evolutionary transcriptomics approach, allowing them to explore shifts in expression levels over time. This approach revealed that genes with increased expression levels compared to the ancestor frequently displayed alterations in fruiting body development when knocked out[135,136]. A related observation was made which found that transcriptional patterns strongly depended on gene age and younger genes tended to display more allele-specific expression variation[137,138].

Comparative genomics has also helped illuminate fungal multicellularity and complex fruiting body development[139,140]. Also, by conducting comparative analyses of 72 fungal genomes across the kingdom, Kiss et al., 2019[140] determined that multicellularity evolved largely through co-option and exaptation of ancient eukaryotic genes, more consistent with evolutionary tinkering than a major innovation through gene duplication. Using transcriptomic approaches, Krizsán et al., 2019[139] detected ~400 gene families where expansion/contraction correlates with the origin of multicellularity, enabling further studies on the function of these genes. With respect to fruiting body development, they also found several gene families, for example, F-box proteins and expansin-like proteins, were also expanded in mushroom-forming fungi. This observation was consistent with solutions for complex multicellularity that evolved in other eukaryotes[139]. Interestingly, while many genes involved in multicellularity are ancient, others (especially small secreted proteins) are more recent and their functions are unknown.

Characterizing genes of unknown function

One gap in our understanding of fungi that has been made clear through genome sequencing is our general lack of knowledge on gene function. As of now, outside of model organisms such as *Saccharomyces cerevisiae*, the function is unknown for approximately half of genes predicted in fungal genomes. Some of these unknowns are conserved across the entire kingdom or beyond, strongly suggesting that they perform crucial roles, otherwise we may not expect them to persist over such long periods of time. The JGI has recently released a list of such highly conserved genes of unknown function (https://mycocosm.jgi.doe.gov/conserved-clusters), equipped with bioinformatics analyses such as remote homology detection and protein structure prediction, and seeks to build a community around characterizing these. Due to their conservation, characterizing individual genes and proteins from this list can help propagate annotations across much or all of the fungal tree of life.

In addition to engaging the community for characterization of individual genes (e.g. using gene disruption libraries like that in *Neurospora crassa*[141]), new technologies such as CRISPR and RB-TDNASeq offer us revolutionary opportunities to connect genes with traits at a huge scale. For example, in *Rhodosporidium toruloides* (Microbotryomycetes), barcoded TDNAs were randomly inserted across the genome to generate single-gene disruptions for all genes in the genome[142]. The resulting mutant population was then screened for performance in various conditions, allowing for rapid profiling of genes important within a given context[143,144]. Furthermore, the new high-throughput ability to synthesize long stretches of DNA makes it possible to scalably screen for gene function using forward genetics approaches. Even though such approaches may find limited success for non-compatible metabolic networks, they would be excellent for exploring genes like transporters and secreted proteins to understand widespread fungal traits like osmotrophy.

Fungal evolution through the lens of multi-omics

Epigenomic shifts across fungal tree of life

While genome-encoded gene sets define an organism's biological potential, its implementation can be controlled by many factors. For example, epigenomic modifications can perturb histone-DNA interactions, altering chromatin structure and changing access of DNA binding proteins (such as transcription factors) to their target sites. These include tools such as DNA modifications, histone modifications, histone variants and nucleosome remodeling complexes. While this review focuses on recent insights into fungal epigenetics using -omics based approaches, deep mechanistic knowledge and functional insights exists through exploration of model eukaryotes which are reviewed elsewhere[145–147].

Through exploring the distribution and variability in epigenomic machinery across taxa, we gain important insights into the evolution of gene regulation. For example, kingdom-wide comparative genomics revealed patterns of methyltransferase evolution and, combined with bisulfite-sequencing across 40 diverse fungi, correlated 5-methylcytosine (5mC) levels with methyltransferase types and transposon abundances[148]. 5mC is primarily associated with suppression of gene or transposon activity, but its role can vary by clade[148]. For example, *Cryptococcus neoformans* (Agaricomycotina) has retained 5mC DNA modifications to stabilize centromeres and subtelomeric regions for millions of years despite loss of de-novo methylation machinery, where it is evolving following a Darwinian process[149].

In addition to 5mC, we discovered that many EDF utilize a non-canonical DNA modification, N6-methyldeoxyadenine (6mA), at very high levels compared to other eukaryotes (with 6mA present at up to 2.8% of all adenines in the genome), and associated with promoters of actively expressed genes[91]. 6mA was also subsequently detected in the arbuscular mycorrhizal fungi (Glomeromycotina[150]). However, 6mA modifications in the Dikarya could not be validated using multiple methods, indicating low levels of this modification (if any). This suggests that some major epigenomic reorganization occurred when the Dikarya evolved. Recent work has now characterized the genes involved in EDF 6mA regulation, including a MTA-70 complex regulating symmetric 6mA methylation and a potentially endosymbiont-derived, EDF-specific methyltransferase regulating de-novo methylation[122].

Other types of epigenomic modifications can also have a direct influence on fungal adaptation and evolution[146,151,152]. For example, entire accessory chromosomes in some ascomycetes have been marked with elevated levels of H3K9me3 and H3K27me3 which have a destabilizing effect[153,154]. Also, nucleosome remodeling complexes and non-canonical histones can be involved in regulation of fungal mitosis, pathogenicity, and carbon assimilation[155,156].

Gene and pathway regulation

While epigenetics can provide us clues about DNA accessibility, new techniques such as DNA Affinity Purification followed by sequencing (DAPseq[157]) bring us one step closer to a comprehensive understanding of the processes underlying gene and pathway regulation. Challenging exposed DNA with synthesized transcription factors, DAPseq maps transcription factor binding sites in a high-throughput, *in-vitro* environment. Through assessing their proximity to genes, potential targets regulated by these transcription factors can be identified. Combined with RNASeq transcriptomics, this approach has proven effective to address several challenges, for example, mapping regulatory networks controlling plant biomass breakdown pathways in *Neurospora crassa*[158], elucidating the roles of specific transcription factors[159] and understanding specific biological processes (e.g. nitrogen regulation[160]).

While transcriptomics and DAPseq can provide us with the pieces, completing the puzzle requires accurate reconstruction of associated metabolic pathways. Unfortunately, availability of metabolic models is limited primarily to biotech-relevant taxa, pathogens and models[161–163]. Unfortunately, we cannot expect models to translate well to distantly related taxa. Without this crucial information, we cannot fully understand how fungi evolved to process metabolites of interest, how pathways differ across clades, and identify efficiencies which can be utilized to improve industrial production of metabolites of interest. Some work has been done to close this gap in some taxa[161] but much remains to be done. To overcome such challenges, community efforts are needed for building accurate metabolic models across the kingdom. Having higher-resolution pathway information on a per-clade basis will then enable thorough exploration of differences (and innovations) between closely-related species and within populations.

Population genomics, adaptive evolution and pan-genomics

With the increase in high quality fungal genomes and annotations, pan-genomic and population genetics-based approaches help analyze related genomes in a holistic way. Both methods provide complementary sets of information to effectively drill down to a set of genes for further analysis, helping explain population dynamics and geographic migrations over time and explore genes associated with traits[164]. For example, analysis of ~1000 isolates of *S. cerevisiae*

revealed an origin out of China and phenotypic differences largely driven by copy number variation, as opposed to accessory genes[164].

Population genomics can also be very powerful when applied in reverse through exploring signatures of adaptive evolution. Analyzing conserved single-copy genes within a population to identify genes experiencing positive, purifying or neutral selection proved valuable for detecting genes associated with adaptation of *Zymoseptoria* to infect wheat[165,166] and other symbiotic systems (see HGT section above).

While positive selection analyses can be powerful, they do not work for lineage-specific or recently duplicated genes. For these genes, pan-genomic methods can be used to further parse genes into sets: core (conserved across all), accessory (present in more than one), and lineage-specific, to enable a variety of different analyses. Although a relatively new technique, pan-genomic methods have already been applied in fungi to study model organisms[167], plant pathogens[168], and fungal mitochondria[169].

By exploring 4 model systems, McCarthy & Fitzpatrick et al., 2019[167] reveal that fungal pan-genomes are typically composed primarily of core genes (>80%) and accessory genes were often duplicated core genes localized to subterminal regions of chromosomes. Others found accessory genes frequently packaged on entirely dispensable chromosomes[168,170]. However, accessory genes sometimes constituted large portions of the pan-genome, for example, ~40% in *Zymoseptoria tritici*[170]. The accessory pan-genome is often enriched in important lifestyle-related genes such as effectors and secondary metabolites[168,170] but can also include genes associated with primary metabolism[171].

While typically applied to multiple genomes within a single species, pan-genome analyses can conceptually be expanded to analyze any taxonomic level of organization (genus, family, etc). Several large-scale whole-genus sequencing projects have been launched recently where such methods could be applied[43,99,107,108,172]. For example, sequencing ~300 *Aspergillus* species[108], combined with metabolic data revealed hundreds of new BGCs important for niche adaptation in competition with other microbes.

Uncovering fungi in microbial communities

While their genomes are often simpler than other eukaryotes, accurate assembly and binning of fungi within metagenomic datasets still lags behind that of bacteria and archaea. Nonetheless, given that fungi play a key role in shaping microbial communities[173], their documentation is essential to understand these systems. To increase eukaryotic representation, contig classification or extraction of fungal bins using tools such as EukCC[174] could represent a path forward. While likely incomplete, such metagenome-derived fungal genomes would help fill in gaps in the tree of life (especially for uncultivable fungi) and improve future taxonomy assignment efforts[173,175].

To complement metagenomics, metatranscriptomics can both describe composition of species in the microbial community and highlight active contributors. Complexity of fungal communities varies from few dominant species to huge diversity[176,177]. For anaerobic fungi, this approach (combined with metaproteomics) revealed that anaerobic fungi provide a specific set of CAZymes that complement those from bacteria for maximum digestion of switchgrass in the cow rumen[176]. In forest soils, metatranscriptomics studies demonstrated different ratios of saprotrophs, pathogens, and mycorrhizal symbionts in different forests[177]. Of course, the success of metatranscriptomic analyses highly depends upon the availability of reference

genomes to which RNAseq reads can be mapped. Auer et al., 2024[177] showed that a small fraction of metatranscriptomic reads had mapped to genomes currently available in MycoCosm, suggesting that observed environmental diversity requires many more sequenced genomes.

Conclusions

Exploration of the fungal tree of life with genomics has revolutionized our understanding of fungal evolution, diversity, and ecological roles. Many of the studies described here took an evolutionary approach when analyzing genomics data, which has led to discoveries of novel lineages, resolved phylogenetic relationships, and revealed the genetics underlying key innovations such as symbiosis, pathogenicity, and multicellularity. The growing repository of fungal genomes provides unprecedented insights into the molecular mechanisms driving fungal adaptation to diverse environments and interactions with other organisms including niche specialization in Saccharomycotina through genome reduction and expansion of host range and capabilities through both endosymbiosis and HGT across fungi. Genomics data have also allowed us to describe similarities and differences across divergent taxa. For example, smaller, intron-poor genomes in dimorphic and yeast lineages (Figure 4) and an evolutionary transition away from 6mA utilization in Dikarya. However, despite these advances, large portions of the fungal kingdom remain genetically uncharted (Figure 3) and over half of the sequenced fungal genes are of unknown function, some conserved across the entire fungal tree of life. With the JGI and projects like Y1000+ providing the core of annotated reference genomes (Figure 2), scaling up sequencing requires coordinated efforts across the research community. Recently, JGI indicated their intent to move towards 10,000 sequenced and annotated genomes[178]. Deeper exploration of ecological groups of fungi is also necessary, including (among others) mycorrhizae, plant pathogens, endophytes, extremophiles and aquatic fungi. As our genome representation of fungi deepens across taxonomic scales, ideally in combination with phenotypic data, our ability to annotate genes responsible for phenotype-defining traits will also improve.

While crucial, actually covering these gaps remains a challenge. One possible solution could be large-scale sequencing of fungal strains available through culture collections. While funding reductions have impacted some collections[179], many of these contain lineages only sampled once and never again, making their preservation essential to achieving this goal. In addition to cultivable strains, much of the fungal kingdom is uncultivable or challenging to isolate and grow. Thankfully, several new methods have been developed which can successfully recover genomic information from these lineages. These include single-cell approaches[94], various enrichment techniques[173], and extracting metagenome-assembled genomes[175]. Also, while eukaryotic genomes are challenging to extract from metagenomes, mitochondria are comparatively smaller and often higher copy number than their hosts and may be a suitable proxy to catalog environmental fungi. Such methods have already proven useful to some communities, for example, in studying arthropod and protist population genetics, species delimitation, and phylogenetics[180,181] and are currently being explored at the JGI for larger collections of fungal genomes.

Furthermore, even if they can be cultivated we still face major challenges when it comes to isolating environmental fungi, preparing materials for genome sequencing, and cultivating analysis expertise. To overcome these barriers, community-wide initiatives that engage and train larger audiences like Myco-Ed: The Mycological Curriculum for Education and Discovery (mycocosm.jgi.doe.gov/mycocosm/home/myco-ed), the Undergraduate Microbiome Scientist Program (https://walkerlabmtsu.weebly.com/undergraduate-microbiome-scientist-program.html),

or other citizen-science approaches (e.g. https://namyco.org/annual-foray) could potentially be transformative.

Sequencing this vast array of fungal genomes would provide catalogs of parts lists (genes, pathways, and traits) for biotechnology, reveal medically relevant natural products, and provide critical reference data for understanding microbial communities for improved environmental sustainability. Of course, we would ideally like to create complete parts lists for maximum usability, which would require more thorough characterization of gene function since right now it is unknown in many cases (including highly conserved genes). To accomplish this, high throughput options like RB-TDNAseq, CRISPRi, or other techniques are needed to connect genes with traits at massive scale. Performed across diverse fungal lineages and coupled with multi-omics techniques, a fingerprint of gene regulation and fitness across conditions can be constructed for elucidating gene functions. As comprehensive characterization of fungal traits is often more difficult than sequencing a genome nowadays, pairing such data with traits we could train Artificial Intelligence to predict these capabilities based on genome features, paving the way for a future where genome annotation comes with hypotheses about fungal lifestyle and role in ecosystems.

References

1. Betts, H. C. *et al.* Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat. Ecol. Evol.* **2**, 1556–1562 (2018).
2. Sebé-Pedrós, A., Degnan, B. M. & Ruiz-Trillo, I. The origin of Metazoa: a unicellular perspective. *Nat. Rev. Genet.* **18**, 498–512 (2017).
3. Lücking, R., Huhndorf, S., Pfister, D. H., Plata, E. R. & Lumbsch, H. T. Fungi evolved right on track. *Mycologia* **101**, 810–822 (2009).
4. Taylor, J. W. & Berbee, M. L. Dating divergences in the Fungal Tree of Life: review and new analyses. *Mycologia* **98**, 838–849 (2006).
5. Hawksworth, D. L. & Lücking, R. Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiol. Spectr.* **5**, (2017).
6. Shen, X.-X. *et al.* Tempo and mode of genome evolution in the budding yeast subphylum. *Cell* **175**, 1533-1545.e20 (2018).
7. Frisvad, J. C. Taxonomy, chemodiversity, and chemoconsistency of Aspergillus, Penicillium, and Talaromyces species. *Front. Microbiol.* **5**, 773 (2014).
8. Steenwyk, J. L., Shen, X.-X., Lind, A. L., Goldman, G. H. & Rokas, A. A robust

phylogenomic time tree for biotechnologically and medically important fungi in the genera Aspergillus and Penicillium. *mBio* **10**, e00925-19 (2019).

9. Kullnig-Gradinger, C. M., Szakacs, G. & Kubicek, C. P. Phylogeny and evolution of the genus Trichoderma: a multigene approach. *Mycol. Res.* **106**, 757–767 (2002).

10. Grine, F. E. & Fleagle, J. G. The first humans: A summary perspective on the origin and early evolution of the genus Homo. in *The First Humans – Origin and Early Evolution of the Genus Homo: Contributions from the Third Stony Brook Human Evolution Symposium and Workshop October 3 – October 7, 2006* (eds. Grine, F. E., Fleagle, J. G. & Leakey, R. E.) 197–207 (Springer Netherlands, Dordrecht, 2009). doi:10.1007/978-1-4020-9980-9_17.

11. Lee, S. C. *et al.* Microsporidia evolved from ancestral sexual fungi. *Curr. Biol. CB* **18**, 1675–1679 (2008).

12. Quandt, C. A. *et al.* Evaluating the diversity of the enigmatic fungal phylum Cryptomycota across habitats using 18S rRNA metabarcoding. *Fungal Ecol.* **64**, 101248 (2023).

13. James, T. Y. *et al.* Shared signatures of parasitism and phylogenomics unite Cryptomycota and microsporidia. *Curr. Biol. CB* **23**, 1548–1553 (2013).

14. James, T. Y. & Berbee, M. L. No jacket required--new fungal lineage defies dress code: recently described zoosporic fungi lack a cell wall during trophic phase. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **34**, 94–102 (2012).

15. Bass, D. *et al.* Clarifying the relationships between Microsporidia and Cryptomycota. *J. Eukaryot. Microbiol.* **65**, 773–782 (2018).

16. Letcher, P. M. & Powell, M. J. A taxonomic summary of Aphelidiaceae. *IMA Fungus* **10**, 4 (2019).

17. Mikhailov, K. V. *et al.* Genomic analysis reveals cryptic diversity in aphelids and sheds light on the emergence of Fungi. *Curr. Biol. CB* **32**, 4607-4619.e7 (2022).

18. Jones, M. D. M. *et al.* Discovery of novel intermediate forms redefines the fungal tree of life. *Nature* **474**, 200–203 (2011).

19. James, T. Y. *et al.* A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). *Mycologia* **98**, 860–871 (2006).
20. Solomon, K. V. *et al.* Early-branching gut fungi possess a large, comprehensive array of biomass-degrading enzymes. *Science* **351**, 1192–1195 (2016).
21. Reisert, P. S. & Fuller, M. S. Decomposition of chitin by Chytriomyces species. *Mycologia* **54**, 647–657 (1962).
22. Skerratt, L. F. *et al.* Spread of chytridiomycosis has caused the rapid global decline and extinction of frogs. *EcoHealth* **4**, 125–134 (2007).
23. Chang, Y. *et al.* Genome-scale phylogenetic analyses confirm Olpidium as the closest living zoosporic fungus to the non-flagellated, terrestrial fungi. *Sci. Rep.* **11**, 3217 (2021).
24. Gryganskyi, A. P. *et al.* Sequencing the Genomes of the First Terrestrial Fungal Lineages: What Have We Learned? *Microorganisms* **11**, 1830 (2023).
25. Shimizu, S., Ogawa, J., Kataoka, M. & Kobayashi, M. Screening of novel microbial enzymes for the production of biologically and chemically useful compounds. *Adv. Biochem. Eng. Biotechnol.* **58**, 45–87 (1997).
26. Jeffries, P., Gianinazzi, S., Perotto, S., Turnau, K. & Barea, J.-M. The contribution of arbuscular mycorrhizal fungi in sustainable maintenance of plant health and soil fertility. *Biol. Fertil. Soils* **37**, 1–16 (2003).
27. Ibrahim, A. S., Spellberg, B., Walsh, T. J. & Kontoyiannis, D. P. Pathogenesis of Mucormycosis. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **54**, S16–S22 (2012).
28. Gupta, Y. K. *et al.* Major proliferation of transposable elements shaped the genome of the soybean rust pathogen Phakopsora pachyrhizi. *Nat. Commun.* **14**, 1835 (2023).
29. Duplessis, S. *et al.* Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9166–9171 (2011).
30. Duplessis, S. *et al.* Host Adaptation and Virulence in Heteroecious Rust Fungi. *Annu. Rev. Phytopathol.* **59**, 403–422 (2021).

31. Lass-Flörl, C. *et al.* Invasive candidiasis. *Nat. Rev. Dis. Primer* **10**, 20 (2024).
32. Cadena, J., Thompson, G. R. & Patterson, T. F. Aspergillosis: Epidemiology, Diagnosis, and Treatment. *Infect. Dis. Clin.* **35**, 415–434 (2021).
33. Nagy, L. G. *et al.* Six key traits of fungi: their evolutionary origins and genetic bases. *Microbiol. Spectr.* **5**, (2017).
34. James, T. Y. & Rokas, A. Use their names: there are no basal, lower, or early diverging fungi. *Mycologia* **117**, 246–254 (2025).
35. Lankiewicz, T. S. *et al.* Lignin deconstruction by anaerobic fungi. *Nat. Microbiol.* **8**, 596–610 (2023).
36. Cantrell, S. A., Dianese, J. C., Fell, J., Gunde-Cimerman, N. & Zalar, P. Unusual fungal niches. *Mycologia* **103**, 1161–1174 (2011).
37. Coleine, C., Stajich, J. E. & Selbmann, L. Fungi are key players in extreme ecosystems. *Trends Ecol. Evol.* **37**, 517–528 (2022).
38. Richards, T. A. & Talbot, N. J. Horizontal gene transfer in osmotrophs: playing with public goods. *Nat. Rev. Microbiol.* **11**, 720–727 (2013).
39. Nagy, L. G. *et al.* Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. *Nat. Commun.* **5**, 4471 (2014).
40. Miyauchi, S. *et al.* Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. *Nat. Commun.* **11**, 5125 (2020).
41. Lutzoni, F. *et al.* Contemporaneous radiations of fungi and plants linked to symbiosis. *Nat. Commun.* **9**, 5451 (2018).
42. Floudas, D. *et al.* The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* **336**, 1715–1719 (2012).
43. Franco, M. E. E. *et al.* Ecological generalism drives hyperdiversity of secondary metabolite gene clusters in xylarialean endophytes. *New Phytol.* **233**, 1317–1330 (2022).
44. Case, N. T. *et al.* The future of fungi: threats and opportunities. *G3*

GenesGenomesGenetics **12**, jkac224 (2022).

45. Kohler, A. *et al.* Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat. Genet.* **47**, (2015).

46. van de Peppel, L. J. J. *et al.* Ancestral predisposition toward a domesticated lifestyle in the termite-cultivated fungus Termitomyces. *Curr. Biol. CB* **31**, 4413-4421.e5 (2021).

47. Haridas, S. *et al.* 101 Dothideomycetes genomes: A test case for predicting lifestyles and emergence of pathogens. *Stud. Mycol.* **96**, 141–153 (2020).

48. Opulente, D. A. *et al.* Genomic factors shape carbon and nitrogen metabolic niche breadth across Saccharomycotina yeasts. *Science* **384**, eadj4503 (2024).

49. Steindorff, A. S. *et al.* Comparative genomic analysis of thermophilic fungi reveals convergent evolutionary adaptations and gene losses. *Commun. Biol.* **7**, 1124 (2024).

50. Dort, E. N. *et al.* Large-scale genomic analyses with machine learning uncover predictive patterns associated with fungal phytopathogenic lifestyles and traits. *Sci. Rep.* **13**, 17203 (2023).

51. Nguyen, N. H. *et al.* FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* **20**, 241–248 (2016).

52. Corradi, N., Pombert, J.-F., Farinelli, L., Didier, E. S. & Keeling, P. J. The complete sequence of the smallest known nuclear genome from the microsporidian Encephalitozoon intestinalis. *Nat. Commun.* **1**, 77 (2010).

53. Ma, L. *et al.* Genome analysis of three Pneumocystis species reveals adaptation mechanisms to life exclusively in mammalian hosts. *Nat. Commun.* **7**, 10740 (2016).

54. Venice, F. *et al.* At the nexus of three kingdoms: the genome of the mycorrhizal fungus Gigaspora margarita provides insights into plant, endobacterial and fungal interactions. *Environ. Microbiol.* **22**, 122–141 (2020).

55. Stajich, J. E. *et al.* Signatures of transposon-mediated genome inflation, host specialization, and photoentrainment in Entomophthora muscae and allied

entomophthoralean fungi. *eLife* **12**, RP92863 (2024).

56. Amses, K. R. *et al.* Diploid-dominant life cycles characterize the early evolution of Fungi. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2116841119 (2022).

57. Lamb, R. J. & Richards, B. N. Survival potential of sexual and asexual spores of ectomycorrhizal fungi. *Trans. Br. Mycol. Soc.* **62**, 181–191 (1974).

58. Aanen, D. K. & Hoekstra, R. F. Why Sex Is Good: On Fungi and Beyond. in *Sex in Fungi* 527–534 (John Wiley & Sons, Ltd, 2007). doi:10.1128/9781555815837.ch32.

59. Auxier, B., Czárán, T. L. & Aanen, D. K. Modeling the consequences of the dikaryotic life cycle of mushroom-forming fungi on genomic conflict. *eLife* **11**, e75917.

60. Baranova, M. A. *et al.* Extraordinary genetic diversity in a wood decay mushroom. *Mol. Biol. Evol.* **32**, 2775–2783 (2015).

61. Marian, I. M. *et al.* High phenotypic and genotypic plasticity among strains of the mushroom-forming fungus Schizophyllum commune. *Fungal Genet. Biol. FG B* **173**, 103913 (2024).

62. Gabaldón, T. Hybridization and the origin of new yeast lineages. *FEMS Yeast Res.* **20**, foaa040 (2020).

63. Mondo, S. J. *et al.* Genome expansion by allopolyploidization in the fungal strain Coniochaeta 2T2.1 and its exceptional lignocellulolytic machinery. *Biotechnol. Biofuels* **12**, 229 (2019).

64. Thierry, A. *et al.* Macrotene chromosomes provide insights to a new mechanism of high-order gene amplification in eukaryotes. *Nat. Commun.* **6**, 6154 (2015).

65. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546, 563–567 (1996).

66. Wood, V. *et al.* The genome sequence of Schizosaccharomyces pombe. *Nature* **415**, 871–880 (2002).

67. Galagan, J. E. *et al.* The genome sequence of the filamentous fungus Neurospora crassa. *Nature* **422**, 859–868 (2003).

68. Martinez, D. *et al.* Genome sequence of the lignocellulose degrading fungus Phanerochaete chrysosporium strain RP78. *Nat. Biotechnol.* **22**, 695–700 (2004).
69. Basenko, E. Y. *et al.* What is new in FungiDB: a web-based bioinformatics platform for omics-scale data analysis for fungal and oomycete species. *Genetics* **227**, iyae035 (2024).
70. Yates, A. D. *et al.* Ensembl Genomes 2022: an expanding genome resource for non-vertebrates. *Nucleic Acids Res.* **50**, D996–D1003 (2022).
71. Sayers, E. W. *et al.* GenBank. *Nucleic Acids Res.* **49**, D92–D96 (2021).
72. Grigoriev, I. V. *et al.* MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* **42**, D699-704 (2014).
73. Lewin, H. A. *et al.* The Earth BioGenome Project 2020: starting the clock. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2115635118 (2022).
74. Stajich, J. E. & Palmer, J. M. Automatic Assembly For The Fungi. (2023).
75. Palmer, J. M. & Stajich, J. E. Funannotate. (2023).
76. Linné, C. von, Linné, C. von & Salvius, L. *Caroli Linnaei ... Species Plantarum :Exhibentes Plantas Rite Cognitas, Ad Genera Relatas, Cum Differentiis Specificis, Nominibus Trivialibus, Synonymis Selectis, Locis Natalibus, Secundum Systema Sexuale Digestas...* vol. 2 1–673 (Impensis Laurentii Salvii, Holmiae, 1753).
77. Taylor, J. W. One Fungus = One Name: DNA and fungal nomenclature twenty years after PCR. *IMA Fungus Glob. Mycol. J.* **2**, 113–120 (2011).
78. James, T. Y. *et al.* Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* **443**, 818–822 (2006).
79. Spatafora, J. W. *et al.* A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* **108**, 1028–1046 (2016).
80. Smith, M. E. *et al.* Phylogenetic analysis of the genus Modicella reveals an independent evolutionary origin of sporocarp-forming fungi in the Mortierellales. *Fungal Genet. Biol. FG B* **61**, 61–68 (2013).

81. Desirò, A. *et al.* Multigene phylogeny of Endogonales, an early diverging lineage of fungi associated with plants. *IMA Fungus* **8**, 245–257 (2017).
82. Voigt, K. *et al.* Early-diverging fungal phyla: taxonomy, species concept, ecology, distribution, anthropogenic impact, and novel phylogenetic proposals. *Fungal Divers.* **109**, 59–98 (2021).
83. Galindo, L. J., López-García, P., Torruella, G., Karpov, S. & Moreira, D. Phylogenomics of a new fungal phylum reveals multiple waves of reductive evolution across Holomycota. *Nat. Commun.* **12**, 4973 (2021).
84. James, T. Y., Stajich, J. E., Hittinger, C. T. & Rokas, A. Toward a fully resolved fungal tree of life. *Annu. Rev. Microbiol.* **74**, 291–313 (2020).
85. Li, Y. *et al.* A genome-scale phylogeny of the kingdom Fungi. *Curr. Biol. CB* **31**, 1653-1665.e5 (2021).
86. Oliveira, J., Yildirir, G. & Corradi, N. From chaos comes order: genetics and genome biology of arbuscular mycorrhizal fungi. *Annu. Rev. Microbiol.* (2024) doi:10.1146/annurev-micro-041522-105143.
87. Thomé, P. C. *et al.* Phylogenomics including new sequence data of phytoplankton-infecting chytrids reveals multiple independent lifestyle transitions across the phylum. *Mol. Phylogenet. Evol.* **197**, 108103 (2024).
88. van de Vossenberg, B. T. L. H. *et al.* Comparative genomics of chytrid fungi reveal insights into the obligate biotrophic and pathogenic lifestyle of Synchytrium endobioticum. *Sci. Rep.* **9**, 8672 (2019).
89. Mondo, S. J., Toomer, K. H., Morton, J. B., Lekberg, Y. & Pawlowska, T. E. Evolutionary stability in a 400-million-year-old heritable facultative mutualism. *Evol. Int. J. Org. Evol.* **66**, 2564–2576 (2012).
90. Mathieu, D. *et al.* Multilevel analysis between Physcomitrium patens and Mortierellaceae endophytes explores potential long-standing interaction among land plants and fungi. *Plant*

J. Cell Mol. Biol. **118**, 304–323 (2024).

91. Mondo, S. J. *et al.* Widespread adenine N6-methylation of active genes in fungi. *Nat. Genet.* **49**, 964–968 (2017).

92. Vandepol, N. *et al.* Resolving the Mortierellaceae phylogeny through synthesis of multi-gene phylogenetics and phylogenomics. *Fungal Divers.* **104**, 267–289 (2020).

93. Reynolds, N. K. *et al.* Mycoparasites, gut dwellers, and saprotrophs: phylogenomic reconstructions and comparative analyses of Kickxellomycotina fungi. *Genome Biol. Evol.* **15**, evac185 (2023).

94. Ahrendt, S. R. *et al.* Leveraging single-cell genomics to expand the fungal tree of life. *Nat. Microbiol.* **3**, 1417–1428 (2018).

95. Baldrian, P., Větrovský, T., Lepinay, C. & Kohout, P. High-throughput sequencing view on the magnitude of global fungal diversity. *Fungal Divers.* **114**, 539–547 (2022).

96. Malar C, M. *et al.* The genome of Geosiphon pyriformis reveals ancestral traits linked to the emergence of the arbuscular mycorrhizal symbiosis. *Curr. Biol. CB* **31**, 1578–1580 (2021).

97. Strope, P. K. *et al.* The 100-genomes strains, an S. cerevisiae resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res.* **25**, 762–774 (2015).

98. Haitjema, C. H. *et al.* A parts list for fungal cellulosomes revealed by comparative genomics. *Nat. Microbiol.* **2**, 17087 (2017).

99. Petersen, C. *et al.* Comparative genomic study of the Penicillium genus elucidates a diverse pangenome and 15 lateral gene transfer events. *IMA Fungus* **14**, 3 (2023).

100. Kjærbølling, I. *et al.* A comparative genomics study of 23 Aspergillus species from section Flavi. *Nat. Commun.* **11**, 1106 (2020).

101. Kjærbølling, I. *et al.* Linking secondary metabolites to gene clusters through genome sequencing of six diverse Aspergillus species. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E753–

E761 (2018).

102. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, (2017).

103. Toronto International Data Release Workshop Authors *et al.* Prepublication data sharing. *Nature* **461**, 168–170 (2009).

104. Nanda, S. & Kowalczuk, M. K. Unpublished genomic data-how to share? *BMC Genomics* **15**, 5 (2014).

105. Wadi, L. & Reinke, A. W. Evolution of microsporidia: An extremely successful group of eukaryotic intracellular parasites. *PLoS Pathog.* **16**, e1008276 (2020).

106. Kijpornyongpan, T. *et al.* Broad genomic sampling reveals a smut pathogenic ancestry of the fungal clade Ustilaginomycotina. *Mol. Biol. Evol.* **35**, 1840–1854 (2018).

107. Hensen, N. *et al.* Genome-scale phylogeny and comparative genomics of the fungal order Sordariales. *Mol. Phylogenet. Evol.* **189**, 107938 (2023).

108. Vesth, T. C. *et al.* Investigation of inter- and intraspecies variation through genome sequencing of Aspergillus section Nigri. *Nat. Genet.* **50**, 1688–1695 (2018).

109. Ocaña-Pallarès, E. *et al.* Divergent genomic trajectories predate the origin of animals and fungi. *Nature* **609**, 747–753 (2022).

110. Stajich, J. E. Fungal genomes and insights into the evolution of the kingdom. *Microbiol. Spectr.* **5**, (2017).

111. Swift, C. L. *et al.* Anaerobic gut fungi are an untapped reservoir of natural products. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2019855118 (2021).

112. Cuomo, C. A. *et al.* Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. *Genome Res.* **22**, 2478–2488 (2012).

113. Druzhinina, I. S. *et al.* Massive lateral transfer of genes encoding plant cell wall-degrading enzymes to the mycoparasitic fungus Trichoderma from its plant-associated hosts. *PLoS Genet.* **14**, e1007322 (2018).

114. Tabima, J. F. *et al.* Phylogenomic analyses of non-Dikarya fungi supports horizontal gene transfer driving diversification of secondary metabolism in the amphibian gastrointestinal symbiont, Basidiobolus. *G3 Bethesda Md* **10**, 3417–3433 (2020).
115. Medina, E. M., Turner, J. J., Gordân, R., Skotheim, J. M. & Buchler, N. E. Punctuated evolution and transitional hybrid network in an ancestral cell cycle of fungi. *eLife* **5**, e09492 (2016).
116. Milner, D. S. *et al.* Environment-dependent fitness gains can be driven by horizontal gene transfer of transporter-encoding genes. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 5613–5622 (2019).
117. Sahu, N. *et al.* Vertical and horizontal gene transfer shaped plant colonization and biomass degradation in the fungal genus Armillaria. *Nat. Microbiol.* **8**, 1668–1681 (2023).
118. Kominek, J. *et al.* Eukaryotic acquisition of a bacterial operon. *Cell* **176**, 1356-1366.e10 (2019).
119. Gluck-Thaler, E. *et al.* Giant starship elements mobilize accessory genes in fungal genomes. *Mol. Biol. Evol.* **39**, msac109 (2022).
120. Bucknell, A. H. & McDonald, M. C. That's no moon, it's a Starship: Giant transposons driving fungal horizontal gene transfer. *Mol. Microbiol.* **120**, 555–563 (2023).
121. Liu, F., Wang, S.-H., Cheewangkoon, R. & Zhao, R.-L. Uneven distribution of prokaryote-derived horizontal gene transfer in fungi: a lifestyle-dependent phenomenon. *mBio* **16**, e0285524 (2025).
122. Lax, C. *et al.* Symmetric and asymmetric DNA N6-adenine methylation regulates different biological responses in Mucorales. *Nat. Commun.* **15**, 6066 (2024).
123. Pawlowska, T. E. *et al.* Biology of fungi and their bacterial endosymbionts. *Annu. Rev. Phytopathol.* **56**, 289–309 (2018).
124. Longley, R. *et al.* Comparative genomics of Mollicutes-related endobacteria supports a late invasion into Mucoromycota fungi. *Commun. Biol.* **6**, 948 (2023).

125. Snyder, A. B. & Worobo, R. W. Risk Mitigation for Immunocompromised Consumers of Mucormycete Spoiled and Fermented Foods: Germain Guidance and Remaining Needs. *Microorganisms* **6**, 45 (2018).
126. Partida-Martinez, L. P. & Hertweck, C. Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature* **437**, 884–888 (2005).
127. Lumini, E. *et al.* Presymbiotic growth and sporal morphology are affected in the arbuscular mycorrhizal fungus Gigaspora margarita cured of its endobacteria. *Cell. Microbiol.* **9**, 1716–1729 (2007).
128. Venice, F. *et al.* Symbiotic responses of Lotus japonicus to two isogenic lines of a mycorrhizal fungus differing in the presence/absence of an endobacterium. *Plant J. Cell Mol. Biol.* **108**, 1547–1564 (2021).
129. Mondo, S. J., Salvioli, A., Bonfante, P., Morton, J. B. & Pawlowska, T. E. Nondegenerative evolution in ancient heritable bacterial endosymbionts of fungi. *Mol. Biol. Evol.* **33**, 2216–2231 (2016).
130. Lastovetsky, O. A. *et al.* Molecular dialogues between early divergent fungi and bacteria in an antagonism versus a mutualism. *mBio* **11**, e02088-20 (2020).
131. Mondo, S. J. *et al.* Bacterial endosymbionts influence host sexuality and reveal reproductive genes of early divergent fungi. *Nat. Commun.* **8**, 1843 (2017).
132. Uehling, J. *et al.* Comparative genomics of Mortierella elongata and its bacterial endosymbiont Mycoavidus cysteinexigens. *Environ. Microbiol.* **19**, 2964–2983 (2017).
133. Yamamura, N. Vertical Transmission and Evolution of Mutualism from Parasitism. *Theor. Popul. Biol.* **44**, 95–109 (1993).
134. Nguyen, T. A. *et al.* Innovation and constraint leading to complex multicellularity in the Ascomycota. *Nat. Commun.* **8**, 14444 (2017).
135. Trail, F., Wang, Z., Stefanko, K., Cubba, C. & Townsend, J. P. The ancestral levels of transcription and the evolution of sexual phenotypes in filamentous fungi. *PLoS Genet.* **13**,

e1006867 (2017).

136. Kim, W. *et al.* Transcriptional divergence underpinning sexual development in the fungal class Sordariomycetes. *mBio* **13**, e0110022 (2022).

137. Merényi, Z. *et al.* Gene age shapes the transcriptional landscape of sexual morphogenesis in mushroom-forming fungi (Agaricomycetes). *eLife* **11**, e71348 (2022).

138. Merényi, Z. *et al.* Genomes of fungi and relatives reveal delayed loss of ancestral gene families and evolution of key fungal traits. *Nat. Ecol. Evol.* **7**, 1221–1231 (2023).

139. Krizsán, K. *et al.* Transcriptomic atlas of mushroom development reveals conserved genes behind complex multicellularity in fungi. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 7409–7418 (2019).

140. Kiss, E. *et al.* Comparative genomics reveals the origin of fungal hyphae and multicellularity. *Nat. Commun.* **10**, 4080 (2019).

141. Dunlap, J. C. *et al.* Enabling a community to dissect an organism: overview of the Neurospora functional genomics project. *Adv. Genet.* **57**, 49–96 (2007).

142. Coradetti, S. T. *et al.* Functional genomics of lipid metabolism in the oleaginous yeast Rhodosporidium toruloides. *eLife* **7**, e32110 (2018).

143. Robertson, N. R. *et al.* Optimized genome-wide CRISPR screening enables rapid engineering of growth-based phenotypes in Yarrowia lipolytica. *Metab. Eng.* **86**, 55–65 (2024).

144. Tafrishi, A. *et al.* Functional genomic screening in Komagataella phaffii enabled by high-activity CRISPR-Cas9 library. *Metab. Eng.* **85**, 73–83 (2024).

145. Selker, E. U. Epigenetic phenomena in filamentous fungi: useful paradigms or repeat-induced confusion? *Trends Genet. TIG* **13**, 296–301 (1997).

146. Madhani, H. D. Unbelievable but true: epigenetics and chromatin in fungi. *Trends Genet. TIG* **37**, 12–20 (2021).

147. Smith, K. M. *et al.* Epigenetics of Filamentous Fungi. in *Reviews in Cell Biology and*

Molecular Medicine (John Wiley & Sons, Ltd, 2012). doi:10.1002/3527600906.mcb.201100035.

148. Bewick, A. J. *et al.* Diversity of cytosine methylation across the fungal tree of life. *Nat. Ecol. Evol.* **3**, 479–490 (2019).

149. Catania, S. *et al.* Evolutionary persistence of DNA methylation for millions of years after ancient loss of a de novo methyltransferase. *Cell* **180**, 263-277.e20 (2020).

150. Chaturvedi, A. *et al.* The methylome of the model arbuscular mycorrhizal fungus, Rhizophagus irregularis, shares characteristics with early diverging fungi and Dikarya. *Commun. Biol.* **4**, 901 (2021).

151. Brimacombe, C. A. *et al.* A natural histone H2A variant lacking the Bub1 phosphorylation site and regulated depletion of centromeric histone CENP-A foster evolvability in Candida albicans. *PLoS Biol.* **17**, e3000331 (2019).

152. Chang, Z., Yadav, V., Lee, S. C. & Heitman, J. Epigenetic mechanisms of drug resistance in fungi. *Fungal Genet. Biol. FG B* **132**, 103253 (2019).

153. Möller, M. *et al.* Destabilization of chromosome structure by histone H3 lysine 27 methylation. *PLoS Genet.* **15**, e1008093 (2019).

154. Galazka, J. M. & Freitag, M. Variability of chromosome structure in pathogenic fungi--of 'ends and odds'. *Curr. Opin. Microbiol.* **20**, 19–26 (2014).

155. Balachandra, V. K. & Ghosh, S. K. Emerging roles of SWI/SNF remodelers in fungal pathogens. *Curr. Genet.* **68**, 195–206 (2022).

156. Portnoy, T. *et al.* The CRE1 carbon catabolite repressor of the fungus Trichoderma reesei: a master regulator of carbon assimilation. *BMC Genomics* **12**, 269 (2011).

157. Bartlett, A. *et al.* Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat. Protoc.* **12**, 1659–1672 (2017).

158. Wu, V. W. *et al.* The regulatory and transcriptional landscape associated with carbon utilization in a filamentous fungus. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 6003–6013 (2020).

159. Bai, N. *et al.* AoSte12 is required for mycelial development, conidiation, trap morphogenesis, and secondary metabolism by regulating hyphal fusion in nematode-trapping fungus Arthrobotrys oligospora. *Microbiol. Spectr.* **11**, e0395722 (2023).
160. Huberman, L. B. *et al.* DNA affinity purification sequencing and transcriptional profiling reveal new aspects of nitrogen regulation in a filamentous fungus. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2009501118 (2021).
161. Wilken, S. E. *et al.* Experimentally validated reconstruction and analysis of a genome-scale metabolic model of an anaerobic Neocallimastigomycota fungus. *mSystems* **6**, e00002-21 (2021).
162. Fonseca-Fernández, A. L., González Barrios, A. F. & Celis Ramírez, A. M. Genome-scale metabolic models in fungal pathogens: past, present, and future. *Int. J. Mol. Sci.* **25**, 10852 (2024).
163. Han, Y., Tafur Rangel, A., Pomraning, K. R., Kerkhoven, E. J. & Kim, J. Advances in genome-scale metabolic models of industrially important fungi. *Curr. Opin. Biotechnol.* **84**, 103005 (2023).
164. Peter, J. *et al.* Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature* **556**, 339–344 (2018).
165. Feurtey, A. *et al.* A thousand-genome panel retraces the global spread and adaptation of a major fungal crop pathogen. *Nat. Commun.* **14**, 1059 (2023).
166. Poppe, S., Dorsheimer, L., Happel, P. & Stukenbrock, E. H. Rapidly evolving genes are key players in host specialization and virulence of the fungal wheat pathogen Zymoseptoria tritici (Mycosphaerella graminicola). *PLoS Pathog.* **11**, e1005055 (2015).
167. McCarthy, C. G. P. & Fitzpatrick, D. A. Pan-genome analyses of model fungal species. *Microb. Genomics* **5**, e000243 (2019).
168. Badet, T., Oggenfuss, U., Abraham, L., McDonald, B. A. & Croll, D. A 19-isolate reference-quality global pangenome for the fungal wheat pathogen Zymoseptoria tritici. *BMC Biol.* **18**,

12 (2020).

169. Brankovics, B. *et al.* First steps towards mitochondrial pan-genomics: detailed analysis of Fusarium graminearum mitogenomes. *PeerJ* **6**, e5963 (2018).

170. Plissonneau, C., Hartmann, F. E. & Croll, D. Pangenome analyses of the wheat pathogen Zymoseptoria tritici reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biol.* **16**, 5 (2018).

171. Lofgren, L., Ross, B., Cramer, R. & Stajich, J. The pan-genome of Aspergillus fumigatus provides a high-resolution view of its population structure revealing high levels of lineage-specific diversity driven by recombination. *PLoS Biol.* **20**, (2022).

172. Lofgren, L. A. *et al.* Comparative genomics reveals dynamic genome evolution in host specialist ectomycorrhizal fungi. *New Phytol.* **230**, 774–792 (2021).

173. Peng, X. *et al.* Genomic and functional analyses of fungal and bacterial consortia that enable lignocellulose breakdown in goat gut microbiomes. *Nat. Microbiol.* **6**, 499–511 (2021).

174. Saary, P., Mitchell, A. L. & Finn, R. D. Estimating the quality of eukaryotic genomes recovered from metagenomic analysis with EukCC. *Genome Biol.* **21**, 244 (2020).

175. Nelson, A. R. *et al.* Wildfire-dependent changes in soil microbiome diversity and function. *Nat. Microbiol.* **7**, 1419–1430 (2022).

176. Lh, H. *et al.* Proteome specialization of anaerobic fungi during ruminal degradation of recalcitrant plant fiber. *ISME J.* **15**, (2021).

177. Auer, L. *et al.* Metatranscriptomics sheds light on the links between the functional traits of fungal guilds and ecological processes in forest soil ecosystems. *New Phytol.* **242**, 1676–1690 (2024).

178. Joint Genome Institute. Innovating Genomics to Serve the Changing Planet: 5-Year Strategic Plan, U.S. Department of Energy Joint Genome Institute. (2024) doi:10.71468/P14W25.

179. Boundy-Mills, K. *et al.* Preserving US microbe collections sparks future discoveries. *J. Appl. Microbiol.* **129**, 162–174 (2020).
180. Crampton-Platt, A., Yu, D. W., Zhou, X. & Vogler, A. P. Mitochondrial metagenomics: letting the genes out of the bottle. *GigaScience* **5**, 15 (2016).
181. Wideman, J. G. *et al.* Unexpected mitochondrial genome diversity revealed by targeted single-cell genomics of heterotrophic flagellated protists. *Nat. Microbiol.* **5**, 154–165 (2020).
182. Steenwyk, J. L., Li, Y., Zhou, X., Shen, X.-X. & Rokas, A. Incongruence in the phylogenomics era. *Nat. Rev. Genet.* **24**, 834–850 (2023).
183. Ovchinnikov, S., Kamisetty, H. & Baker, D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife* **3**, e02030 (2014).
184. Redecker, D., Kodner, R. & Graham, L. E. Glomalean fungi from the Ordovician. *Science* **289**, 1920–1921 (2000).
185. Vankuren, N. W., den Bakker, H. C., Morton, J. B. & Pawlowska, T. E. Ribosomal RNA gene diversity, effective population size, and evolutionary longevity in asexual glomeromycota. *Evol. Int. J. Org. Evol.* **67**, 207–224 (2013).
186. Todd, R. T., Forche, A. & Selmecki, A. Ploidy Variation in Fungi: Polyploidy, Aneuploidy, and Genome Evolution. *Microbiol. Spectr.* **5**, (2017).
187. Selmecki, A. M. *et al.* Polyploidy can drive rapid adaptation in yeast. *Nature* **519**, 349–352 (2015).
188. Ohm, R. A. *et al.* Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. *PLoS Pathog.* **8**, e1003037 (2012).
189. Khaldi, N. *et al.* SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genet. Biol. FG B* **47**, 736–741 (2010).
190. Idnurm, A., Walton, F. J., Floyd, A. & Heitman, J. Identification of the sex genes in an early diverged fungus. *Nature* **451**, 193–196 (2008).

191. Marcet-Houben, M. & Gabaldón, T. Evolutionary and functional patterns of shared gene neighbourhood in fungi. *Nat. Microbiol.* **4**, 2383–2392 (2019).
192. Dhillon, B., Kema, G. H. J., Hamelin, R. C., Bluhm, B. H. & Goodwin, S. B. Variable genome evolution in fungi after transposon-mediated amplification of a housekeeping gene. *Mob. DNA* **10**, 37 (2019).
193. Gluck-Thaler, E. & Vogan, A. A. Systematic identification of cargo-mobilizing genetic elements reveals new dimensions of eukaryotic diversity. *Nucleic Acids Res.* **52**, 5496–5513 (2024).
194. Mat Razali, N., Cheah, B. H. & Nadarajah, K. Transposable elements adaptive role in genome plasticity, pathogenicity and evolution in fungal phytopathogens. *Int. J. Mol. Sci.* **20**, 3597 (2019).
195. Freitag, M., Williams, R. L., Kothe, G. O. & Selker, E. U. A cytosine methyltransferase homologue is essential for repeat-induced point mutation in Neurospora crassa. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 8802–8807 (2002).
196. Gladyshev, E. Repeat-Induced Point Mutation (RIP) and other genome defense mechanisms in fungi. *Microbiol. Spectr.* **5**, 10.1128/microbiolspec.FUNK-0042–2017 (2017).
197. Galagan, J. E. & Selker, E. U. RIP: the evolutionary cost of genome defense. *Trends Genet. TIG* **20**, 417–423 (2004).

Acknowledgements

The work conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE- AC02-05CH11231. We thank Tatyana Smirnova for assistance with figure design and Igor Shabalov for MycoCosm portal development associated with this review. We additionally thank Tim James, David Hibbett, and Pedro Crous for their feedback on lifestyles and topology for Figure 1. We would

like to thank László Nagy and two anonymous reviewers for their insightful feedback on this review.

Author contributions

Both authors researched data for the article and contributed substantially to discussion of the content. S.J.M wrote the article and generated the figures, with input from I.V.G. Both authors edited the article before submission.

Competing interests

The authors declare no competing interests.

Supplementary information

Supplementary information is available for this paper at https://doi.org/10.1038/s415XX-XXX-XXXX-X (see below)

Box 1: Phylogenomics methods

Thanks to the large volume of genomic data available today, we are able to move beyond alignment of marker genes and instead can reconstruct phylogenomic relationships using thousands of genes and tens of thousands of informative sites. This allows us to generate high quality species trees to structure our research. By concatenating aligned sequences together and building a single tree, we can generate a ‘species’ tree, which provides us with a consensus of relationships amongst organisms. While species trees provide us with a framework for understanding evolution of organisms and traits, it is important to keep in mind that they are summaries across many genes, and will wash away the less frequent cases where individual gene evolution disagrees with the overall signal. Biologically, these incongruent genes can be very important and worth exploring, and are crucial for detecting events such as horizontal gene transfers. Incongruities can also be caused by a host of analytical factors, including taxon sampling, incorrect ortholog inferences, differences between substitution models, and more. For an in-depth review on phylogenetic incongruence, see Steenwyk et al., 2023[182].

In addition to reconciling incongruities, phylogenetic trees can be used for a variety of different purposes, such as testing coevolutionary hypotheses through comparing patristic distances, be it between different organisms (e.g. Mondo et al., 2012[89]) or different protein families within the same genome[183]. Phylogenies can also be used in conjunction with the fossil record for molecular dating[184], which allows us to estimate mutation rate and effective population size[185]. Phylogenies can also serve as a backbone for studying adaptive change using population genetics approaches, exploring gene gains/losses, protein family evolution, and more.

Box 2. Mechanisms and implications of genome rearrangement

Fungal genomes display remarkable variability in architecture, driven by several mechanisms that cause the reshuffling of DNA. Along with the usual culprit, transposable elements (see Box 3), polyploidy, aneuploidy and whole genome duplications are common, particularly in yeasts and EDF[56], which can be drivers of accelerated genome rearrangement and evolution. Specifically, elevated ploidy has been linked to increased genome instability and rates of chromosomal rearrangements, crossover events, etc[186]. It is also a mechanism some fungi use to adapt under stress[187]. Elevated ploidy has also been tied to enhanced capabilities, for example, more aggressive plant biomass degradation following a hybridization event in *Coniochaeta* sp. 2T2.1[63]. In the Dothideomycetes, while ploidy changes are rare, rampant intra-chromosomal inversions drive a pattern of mesosynteny[188] where chromosome content is generally preserved but gene order is highly shuffled.

However, some of these rearrangements are likely deleterious, as we know that gene neighborhood is important for certain types of genes such as biosynthetic gene clusters (BGCs; which produce secondary metabolites) and mating type loci, which have been clustered together across long periods of time. In fact, preservation of gene neighborhoods is so consistent for these genes that it is used as a feature used to predict these genes and their boundaries[189,190]. Recently, Marcet-Houben & Gabaldón, 2019[191] used ~350 fungal genomes to assess how often gene neighborhoods were shared across taxa. They concluded that approximately one third of fungal gene space is clustered into shared gene neighborhoods. Additionally, they predicted the origin and inheritance of those neighborhoods and determined that most are transmitted vertically, but there was variability based on the type of gene cluster. For example, BGCs showed higher rates of horizontal gene transfer[191].

Box 3. Transposons as drivers of evolution

Recent -omics research has revealed quite a bit of variability in repeat content across the Fungi (Figure 4). Originally considered ‘junk’, we now know that transposable elements (TEs) and the mechanisms hosts have evolved to control them play critical and lasting roles in host evolution. In addition to themselves, during jumps TE can carry neighboring gene content with them, causing duplications, inversions and more[192]. Other, massive 400kb ‘Starship’ can carry ecologically relevant ‘cargo’ with important ecological consequences. For example, copy number of a Voyager element carrying virulence-related cargo is correlated with growth on soybean[119]. Found primarily in Pezizomycotina[193], Starships are also hotspots for genome rearrangement and have been coevolving with fungi for at least 400 million years[119,193].

Depending on where TEs integrate, they can disrupt gene functions. Within this context, it is curious that we often observe increased TE content in biotrophic, plant-associated fungi (e.g. rusts[28,29] and AMF[54]; Figures 1 and 4). Given the substantial divergence between these groups (>500My), this trait likely arose through convergent evolution, suggesting some benefit to the accelerated evolutionary rates TEs provide their hosts [194]. However, for many fungi, TEs are something the host wants to purge. To combat them, fungi have evolved a special process called Repeat-Induced Point mutation (RIP). Primarily in ascomycetes, this process activates during sexual reproduction where duplicated content is identified and specific nucleotides are mutated[195], rendering the underlying sequence non-functional (see Gladyshev 2017[196] for detailed review). One might consider this an extreme response to deal with transposons, since RIP does not discriminate between transposons and protein-coding genes[197]. Essentially forbidding any form of duplication, RIP gives rise to an important trade-off in fungi - the choice between recombination and gene duplication, with potentially long-term impacts on evolutionary trajectory, effective population sizes and genetic drift**.**

Figure 1:

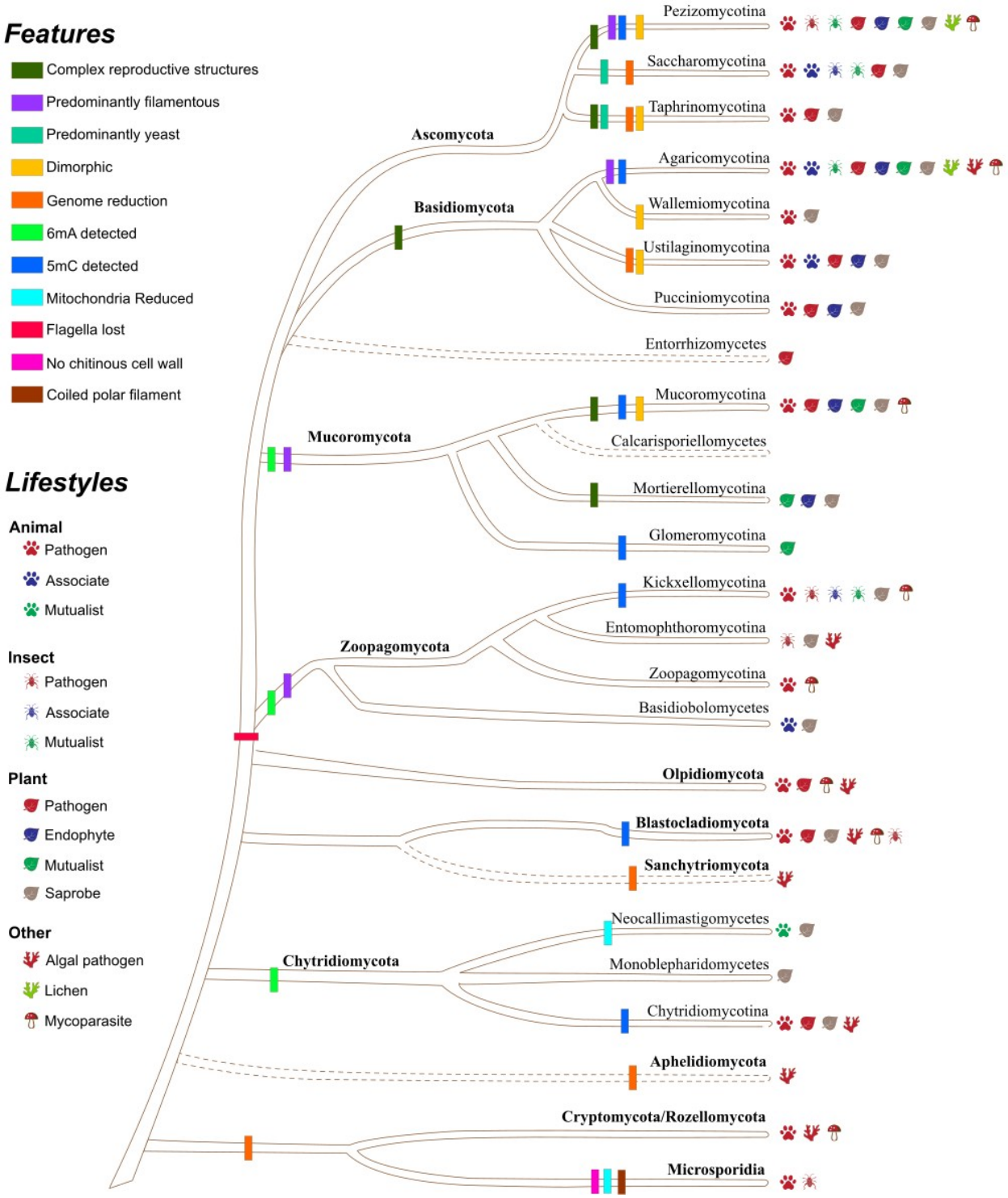


Figure 1. Cartoon representation of the fungal tree of life, decorated with notable lifestyles and traits within clades. Here, to keep consistent with genome-based taxonomic nomenclature and information in NCBI, we used the taxonomic classification as described in Spatafora et al.,

2016[79]. Branching order follows that of James & Rokas, 2025[34]. Taxa without available annotated reference genomes are denoted with dashed branches and were placed based on positions in Voigt et al., 2021[82] (Aphelidiomycota, Calcarisporiellomycetes, Entorrhizomycetes), Galindo et al., 2021[83] (Sanchytriomycota) and consultation with experts. Icons denote various lifestyles, colored by the type of interaction. Pathogenic: red; associate: blue (including endophytes); mutualist (including mycorrhizal): green; saprobe (including wood decay fungi): grey. Tick marks on branches show the most recent common ancestor harboring specific features (e.g. 6mA, complex multicellular structures, etc).

Figure 2:

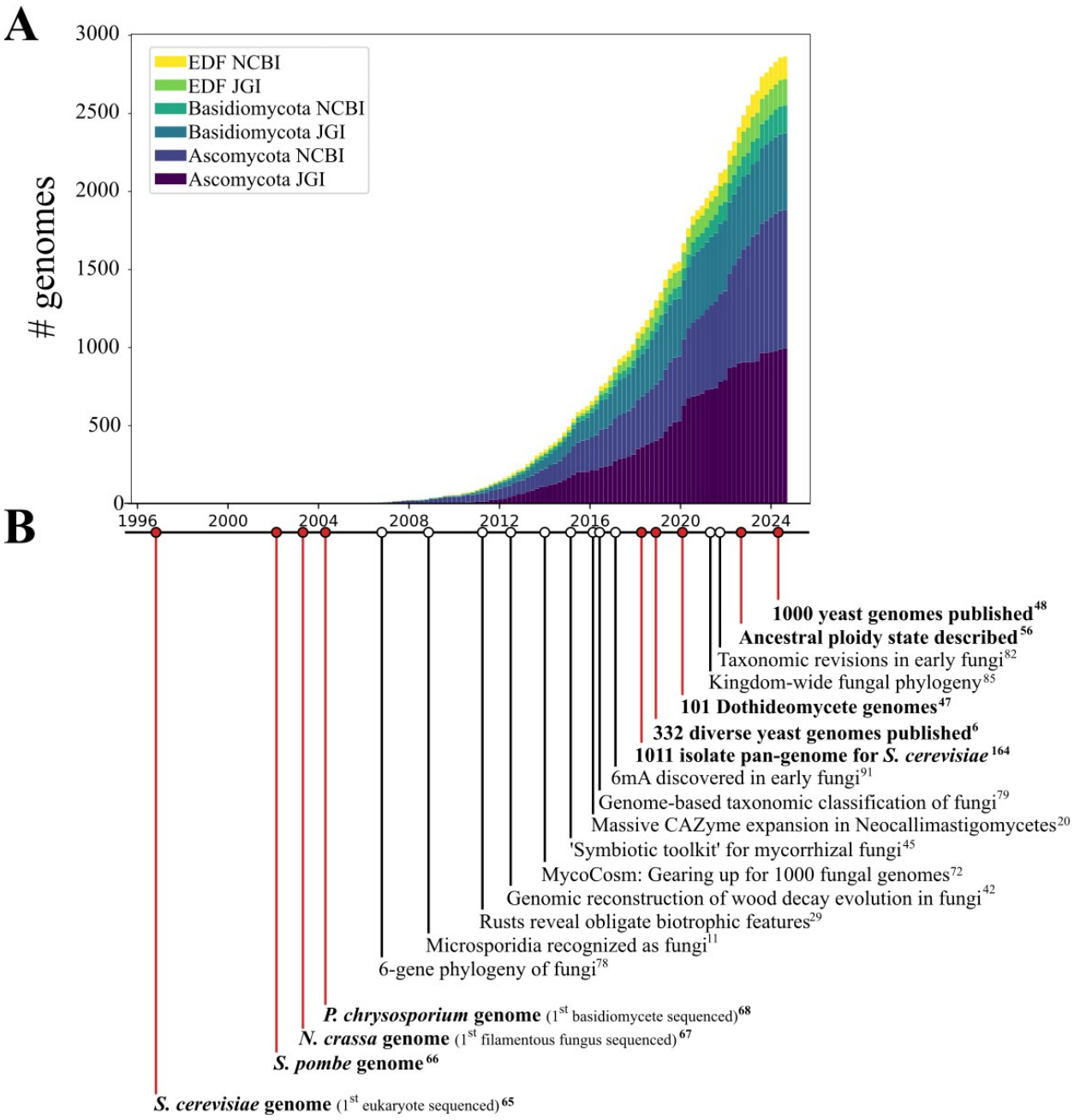


Figure 2. Timeline of discoveries and availability of annotated fungal genomes. A) Number of species with annotated reference genomes available over time across two resources, JGI's MycoCosm database (darker shades of color) and non-JGI genomes deposited in NCBI (lighter shades). If multiple genomes existed for the same species within resource (as determined using the NCBI taxonomy database), only the earliest annotated genome is represented. B) Timeline of some major discoveries in the field of fungal genomics. Publications which released important datasets are shown in red with bold text. Superscript numbers next to each discovery refer to the associated publication in the references section.

Figure 3:

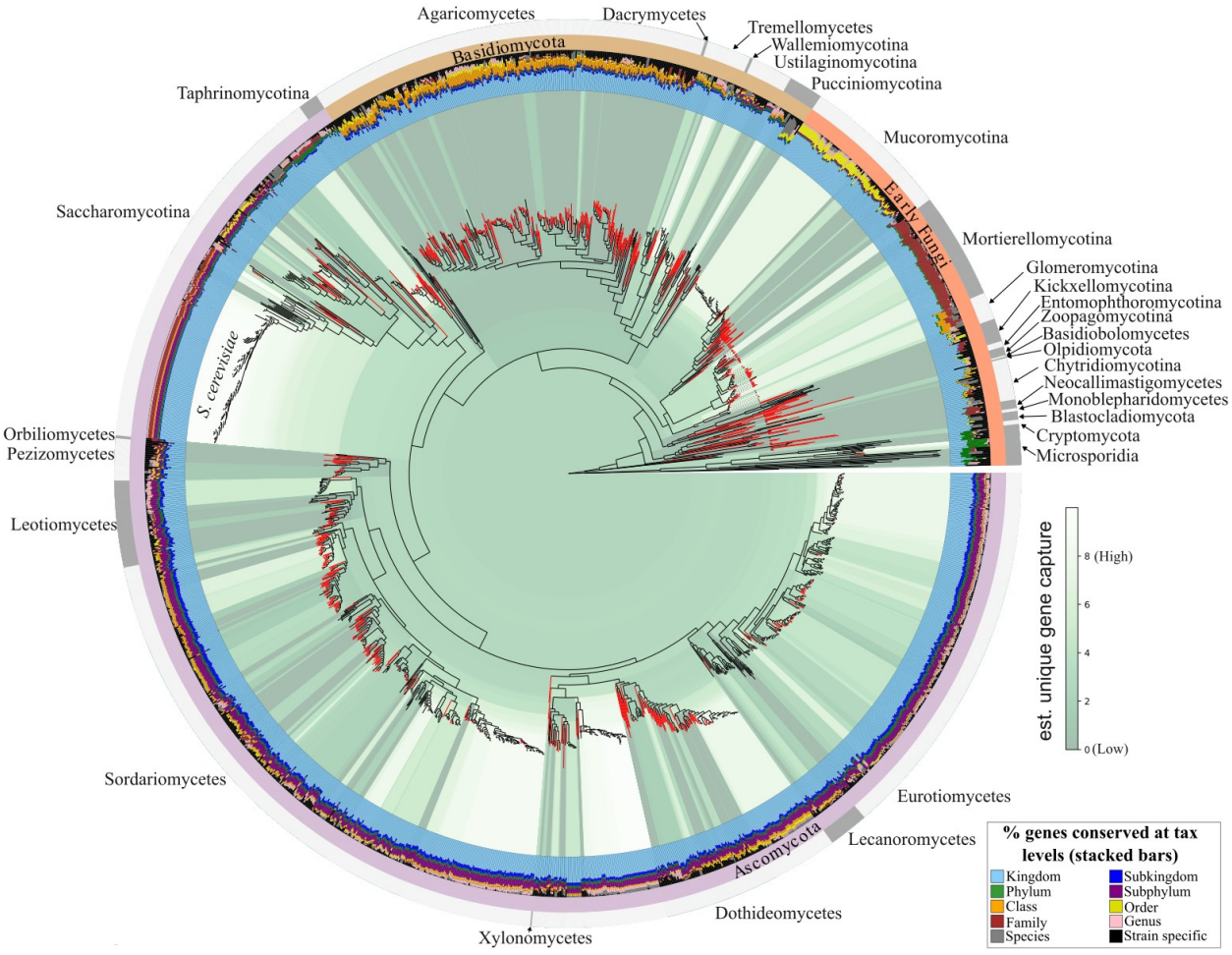


Figure 3. Overview of gene conservation across the fungal kingdom using ~2000 published references (see Supplementary Information, Supplementary Table 1 and http://mycocosm.jgi.doe.gov/ncbi-combined-tree for additional details, including lineage and publication information). The outermost two rings display taxonomic information at various

scales, followed by a stacked barchart showing % genes conserved at specific taxonomic levels (kingdom, subkingdom, phylum, subphylum, class, order, family, genus, species and strain specific). To highlight clades where unique content is high (and therefore may be worth sequencing deeper), we colored the background based on the proportion of taxa with elevated levels of lineage-specific genes (darker green shades = more lineage specific genes). This analysis suggests deeper sampling is needed, especially of EDF and Basidiomycota, but also highlights several undersampled clades of Ascomycota (e.g. Pezizomycetes).

Figure 4:

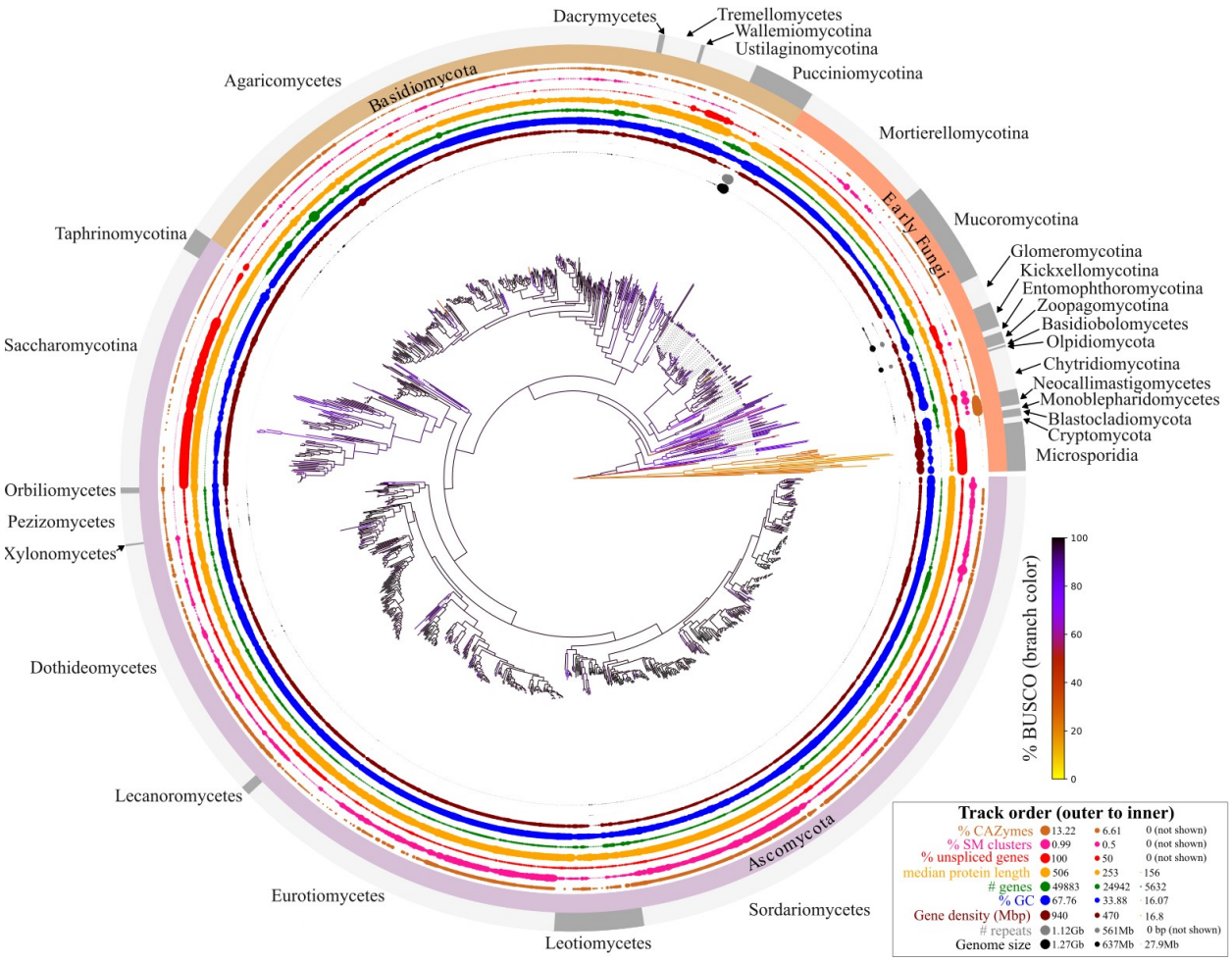


Figure 4. Genome features across the fungal kingdom. Using ~1440 published genomes available via MycoCosm (see Supplementary Table 2), we overlay assembly and annotation information, including (from outer-most track to inner): percent CAZymes, percent SM clusters, percent unspliced genes, median protein length, number of genes, percent GC, gene density, genome assembly repeat length, and genome size. While some features remain relatively consistent (e.g. median protein length), many features display lineage-specific enrichment or depletion (CAZymes and GC content in Neocallimastigomycetes, SM clusters in filamentous ascomycetes, etc). We also notice a general increase in unspliced genes and smaller genome sizes associated with lineages that exist as yeasts either entirely or during major life stages (e.g. Saccharomycotina, Taphrinomycotina and Ustilaginomycotina). Branches are colored according to BUSCO completeness (using the fungi_odb10 dataset; see Supplementary Information). While most annotations are fairly complete, reduced BUSCO completeness is observed in a clade-specific fashion, often those with smaller genomes (e.g. Ustilaginomycotina) or EDF.

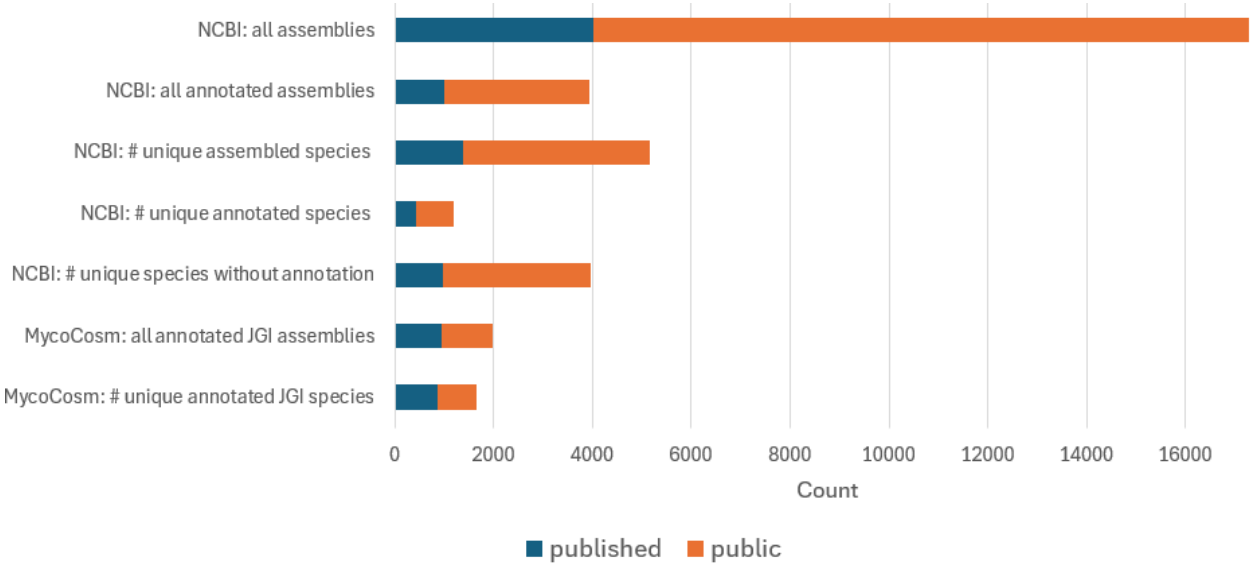
Supplementary information

To construct a phylogeny of genomic representatives across fungi, we used published references (see Supplementary Table 1 and Supplementary Table 2 for lineage and publication details) following a modified JGI tree-building workflow[1], then used the ete3[2] python package to visualize various types of information (Figures 3 and 4). Briefly, to build the tree, conserved genes were identified using mmseqs2[3] (version 0188988235c6f1a8e90f327827c73f981db8a19a; default parameters), then 2677 orthogroups (Figure 3) and 2839 orthogroups (Figure 4) were aligned using Mafft[4] version 7.475 and trimmed to informative sites using Gblocks[5]. FastTree[6] version 2.1.10 SSE3 was then used for phylogeny reconstruction. To identify conserved genes, genes within mmseqs2 clusters were assigned a conservation level based on the taxonomic breadth within those clusters (Figure 3). To display various genome features, data were extracted from MycoCosm. Percent BUSCO[7] completeness (version 5.0) was used to estimate annotation completeness using the fungi_odb10 dataset. To get a sense for how much unique content we might be missing within clades, the background for Figure 3 is colored based on the proportion of taxa with fewer than 3% lineage-specific genes under a given node, normalized by node length (distance from node to furthest tip). A minimum branch length from node to farthest tip of 0.05 was imposed for applying background color, as this represented the phylogenetic distance within *Saccharomyces cerevisiae*. As a well represented species, our aim through imposing this cutoff was to provide an assessment of where future sequencing efforts should be focused to maximize discovery of novel genes (ignoring possible extinction events).

For analysis of genomic data available through NCBI's Genbank repository, we downloaded the assembly summary report available via ftp (https://ftp.ncbi.nlm.nih.gov/genomes/ASSEMBLY_REPORTS/assembly_summary_genbank.txt) on August 23rd, 2024, filtered out any non-fungal assemblies and used the provided species taxonomy ID to determine species with available data and the earliest release date per species (Figure 2a). Assemblies were considered to be published only if they had associated PubMed IDs. Annotated assemblies were determined by filtering the protein_coding_gene_count column for counts > 500 proteins. Counts of assemblies that fit into certain categories (published, annotated, etc) were then reported in Figure S1. To calculate similar metrics for data available via the JGI MycoCosm database, on August 23rd, 2024 we similarly extracted release date, NCBI taxonomy ID and associated PubMed IDs for all portals, then filtered them in the same way as done for NCBI-sourced data.

Figure S1. Availability and annotation state of genome assemblies in NCBI's GenBank repository and JGI's MycoCosm database as of August, 2024. Genomes are additionally marked with publication status based on the presence of associated PubMed IDs. Blue: published; orange: publicly available but not published. While NCBI hosts an impressive amount of fungal assemblies, corresponding annotations were not deposited for the vast majority of these. Further, the number of unique species with reference genomes is substantially smaller than the total number of assemblies, indicating a bias in sampling to specific species of interest.

In contrast, MycoCosm hosts substantially fewer datasets overall, but they cover more unique species and are all annotated.



Supplementary Table 1. Published, annotated genomes included in Figure 3, including links to original data and associated PubMed IDs. Similar information can be viewed in an interactive phylogenetic tree format at http://mycocosm.jgi.doe.gov/ncbi-combined-tree.

Supplementary Table 2. Published, annotated genomes included in Figure 4, including links to MycoCosm portals and associated PubMed IDs. The data source - JGI versus External (which includes imported data from NCBI and other sources) - is also reported here.

Supplementary References:

1. Ahrendt, S. R., Mondo, S. J., Haridas, S. & Grigoriev, I. V. MycoCosm, the JGI's Fungal Genome Portal for Comparative Genomic and Multiomics Data Analyses. *Methods Mol. Biol. Clifton NJ* **2605**, 271–291 (2023).

2. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).

3. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, (2017).

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

(2002).

5. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).

6. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).

7. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).