OXFORD GENETICS

GENETICS, 2021, 218(2), iyab066
DOI: 10.1093/genetics/iyab066
Advance Access Publication Date: 4 May 2021
Fungal Genetics and Genomics

An evolutionary genomic approach reveals both conserved and species-specific genetic elements related to human disease in closely related *Aspergillus* fungi

Matthew E. Mead [1,†] Jacob L. Steenwyk [1,†] Lilian P. Silva,[2] Patrícia A. de Castro,[2] Nauman Saeed [3] Falk Hillmann [3] Gustavo H. Goldman [2] and Antonis Rokas [1,*]

[1]Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA,
[2]Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brazil
[3]Junior Research Group Evolution of Microbial Interactions, Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute (HKI), Jena, Germany
[†]These authors contributed equally to this work.

*Corresponding author: Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA. Email: antonis.rokas@vanderbilt.edu

Abstract

Aspergillosis is an important opportunistic human disease caused by filamentous fungi in the genus *Aspergillus*. Roughly 70% of infections are caused by *Aspergillus fumigatus*, with the rest stemming from approximately a dozen other *Aspergillus* species. Several of these pathogens are closely related to *A. fumigatus* and belong in the same taxonomic section, section *Fumigati*. Pathogenic species are frequently most closely related to nonpathogenic ones, suggesting *Aspergillus* pathogenicity evolved multiple times independently. To understand the repeated evolution of *Aspergillus* pathogenicity, we performed comparative genomic analyses on 18 strains from 13 species, including 8 species in section *Fumigati*, which aimed to identify genes, both ones previously connected to virulence as well as ones never before implicated, whose evolution differs between pathogens and nonpathogens. We found that most genes were present in all species, including approximately half of those previously connected to virulence, but a few genes were section- or species-specific. Evolutionary rate analyses identified over 1700 genes whose evolutionary rate differed between pathogens and nonpathogens and dozens of genes whose rates differed between specific pathogens and the rest of the taxa. Functional testing of deletion mutants of 17 transcription factor-encoding genes whose evolution differed between pathogens and nonpathogens identified eight genes that affect either fungal survival in a model of phagocytic killing, host survival in an animal model of fungal disease, or both. These results suggest that the evolution of pathogenicity in *Aspergillus* involved both conserved and species-specific genetic elements, illustrating how an evolutionary genomic approach informs the study of fungal disease.

Keywords: evolution; *Aspergillus fumigatus*; evolutionary rate; virulence; pathogenicity; fungal disease; aspergillosis; convergent evolution

Introduction

The ability of a microbe to cause disease is a multifactorial trait that is dependent upon diverse genomic loci, including genes and noncoding regulatory elements. For opportunistic fungal pathogens, whose "accidental" infections of humans are not a part of their normal life cycle (Casadevall and Pirofski 2007), the evolution of genomic loci contributing to virulence is thought to have been shaped by diverse evolutionary and ecological pressures, such as avoiding predation from soil-dwelling amoebae and surviving in warm and stressful environmental niches similar to those found inside human hosts (Tekaia and Latgé 2005; Nielsen *et al.* 2007; Hillmann *et al.* 2015). However, the genetic differences between fungal pathogens and their nonpathogenic relatives have only recently begun to be understood (Fedorova *et al.* 2008; Butler *et al.* 2009; Sharpton *et al.* 2009; Moran *et al.* 2011; Taylor 2015; Gabaldón *et al.* 2016; Gupta *et al.* 2020; Rokas *et al.* 2020). This is especially true for filamentous fungi in the genus *Aspergillus*, which infect hundreds of thousands of humans each year (Brown *et al.* 2012; Bongomin *et al.* 2017).

Aspergillosis, the spectrum of diseases caused by fungi in the genus *Aspergillus*, afflicts a broad range of animals, including humans (Seyedmousavi *et al.* 2015). In humans, aspergillosis is a major global health issue and primarily affects individuals with compromised immune systems or who have other lung diseases or conditions (Barrs *et al.* 2013; Gregg and Kauffman 2015; Frisvad and Larsen 2016). Approximately 70% of aspergillosis patients are infected with *Aspergillus fumigatus*, but other members of the genus cause the rest of the infections, with no individual species responsible for a disproportionately large amount of cases (Alastruey-Izquierdo *et al.* 2014; Perlin *et al.* 2017; Latgé and

Received: March 11, 2021. **Accepted:** April 20, 2021

Chamilos 2019). Some of these pathogenic species are very closely related to *A. fumigatus* and belong to the same taxonomic section, section *Fumigati* (Balajee *et al.* 2009; Alastruey-Izquierdo *et al.* 2013; Houbraken *et al.* 2020). In contrast, most of the ~60 species in section *Fumigati* do not cause disease or have rarely been found in the clinic, suggesting that the ability to cause disease in humans evolved multiple times independently (or convergently) within *Aspergillus* (Rokas *et al.* 2020). For example, *Aspergillus oerlinghausenensis* and *Aspergillus fischeri*, the two closest relatives of *A. fumigatus* are both considered nonpathogenic (Houbraken *et al.* 2016; Mead *et al.* 2019; Steenwyk *et al.* 2020b).

Why some *Aspergillus* species routinely infect humans whereas their very close relatives never or rarely do remains an open question (Rokas *et al.* 2020). To date, studies addressing this question have focused on comparing *A. fumigatus* to one or a few closely related (usually pathogenic) species (Fedorova *et al.* 2008; Wiedner *et al.* 2013; Sugui *et al.* 2014a, 2014b; Mead *et al.* 2019; Dos Santos *et al.* 2020; Knowles *et al.* 2020; Steenwyk *et al.* 2020b) or to larger numbers of very distantly related species (Kjærbølling *et al.* 2018). Many individual genes and pathways are known to contribute to *A. fumigatus* virulence (Abad *et al.* 2010; Bignell *et al.* 2016; Brown and Goldman 2016; Steenwyk *et al.* 2020c), but if they are present or function in the same manner in other section *Fumigati* species, including in nonpathogens, has rarely been studied (Knowles *et al.* 2020; Steenwyk *et al.* 2020b). Genomic loci (*i.e.*, genes and noncoding regulatory elements, and their variants) associated with pathogenicity could be shared or absent amongst all pathogens, including *A. fumigatus*, following a "conserved pathogenicity" model, or be uniquely present (or absent) in each pathogen ("species-specific pathogenicity" model; Rokas *et al.* 2020). The two models are not mutually exclusive, and the limited evidence available suggests that some genomic loci likely follow the conserved pathogenicity model (Fedorova *et al.* 2008; Kjærbølling *et al.* 2018; Knowles *et al.* 2020; Steenwyk *et al.* 2020b), whereas others follow the species-specific pathogenicity model (Fedorova *et al.* 2008; Kowalski *et al.* 2019; Mead *et al.* 2019; Steenwyk *et al.* 2020b).

To study the signatures of the repeated evolution of human pathogenicity we conducted diverse evolutionary analyses on the genomes of 18 *Aspergillus* strains representing 13 species, including both pathogenic and nonpathogenic species from section *Fumigati*. Our results show that previously identified virulence-related genes are largely conserved throughout section *Fumigati* and outgroups. Consistent with the species-specific pathogenicity model, we found dozens of gene families that were present only in a given pathogen as well as dozens of genes whose evolutionary rates differed between a given pathogen and the rest of the taxa. For example, we identified 72 *A. fumigatus*-specific gene families and 34 genes whose evolutionary rate was uniquely different in *A. fumigatus*. Consistent with the conserved pathogenicity model, we identified over 1700 genes that showed pathogen-specific evolutionary rates; however, we did not identify any gene families that were shared only by pathogenic taxa. To test whether our approach could identify loci that contribute to *Aspergillus* disease-related traits, we carried out functional assays of deletion mutants of 17 transcription factor (TF)-encoding genes identified in our bioinformatic analyses as consistent with either pathogenicity model. We found that eight genes (four consistent with the conserved model and four consistent with the species-specific model) significantly affected pathogenicity-related traits, suggesting that the evolution of *Aspergillus* pathogenicity involved both conserved and species-specific genetic contributors. More broadly, our study shows that an evolutionary genomic approach is a useful framework for gaining insights into the molecular mechanisms by which *Aspergillus* species impact human health.

Materials and methods

Genome procurement, assembly, and annotation

Genomes and annotations for *A. fumigatus* strains Af293 and A1163, along with all non-*A. fumigatus*, publicly available (as of July 2019) annotated genomes from section *Fumigati* were downloaded for analyses (see Supplementary Table S1 for NCBI accession numbers). We also obtained genomes and annotations for four outgroup species to facilitate phylogenetic analyses and comparisons. To expand the number of genomes analyzed, we assembled and/or annotated five additional *Aspergillus* genomes. More specifically, raw genomic reads for *A. fumigatus* strains F16311 and 12-7505446 were downloaded from NCBI for genome assembly and annotation. These strains were chosen because they, together with *A. fumigatus* strains Af293 and A1163, span the known diversity of *A. fumigatus* (Lind *et al.* 2017); additionally, available genomes for *Aspergillus cejpii* FS110, *Aspergillus neoellipticus* NRRL 5109, and *Aspergillus viridinutans* FRR 0576 were downloaded from NCBI and annotated (Abdolrasouli *et al.* 2015; Li *et al.* 2018; Urquhart *et al.* 2019; Supplementary Table S1). To quality trim sequence reads, we used Trimmomatic, version 0.36 (Bolger *et al.* 2014) using parameters described elsewhere (Steenwyk and Rokas 2017). The resulting high-quality reads were used as input to the genome assembly software SPAdes, version 3.8.1 (Bankevich *et al.* 2012), with the "careful" parameter to reduce mismatches and short indels and the "cov-cutoff" parameter set to "auto." Partial and complete gene models were predicted using Augustus, version 2.5.5 (Stanke and Waack 2003), with the "minexonintronprob" and "minmeanexonintronprob" parameters set to 0.1 and 0.4, respectively. Genome annotation quality was assessed using BUSCO, version 2.0.1 (Waterhouse *et al.* 2018), with the Pezizomycotina database of orthologs from OrthoDB, version 9 (Waterhouse *et al.* 2013). Genome annotation quality was similar between publicly available genomes and genomes assembled and/or annotated in the present project. For example, the publicly available assembly and annotation for *A. fumigatus* strain A1163 had 94.0% of BUSCO genes present in single copy while the assembled and annotated genome for *A. fumigatus* strain F16311 had 93.9% of BUSCO genes present in a single copy.

Inference of gene families

We first identified orthologous genes by clustering genes with high sequence similarity into orthologous groups using Markov clustering (van Dongen 2000) as implemented in OrthoMCL, version 1.4 (Li *et al.* 2003), with an inflation parameter of 2.8. Gene sequence similarity was determined using a blastp "all-vs-all" using NCBI's Blast+, version 2.3.0 (Camacho *et al.* 2009) with an *e*-value cutoff of 1*e*-10, a 30% identity cutoff, and a 70% match cutoff. In subsequent analyses, these 14,294 orthogroups were used as proxies for gene families. In total, 3,601 of the 14,294 orthogroups had all 18 taxa represented by a single sequence and are hereafter referred to as single-copy orthologous genes. Finally, 16 orthogroups had the same number of family members in each taxon but in more than one copy, and 10,677 orthogroups had a different number of family members in at least one taxon.

Phylogenomic data matrix construction and analyses

To construct a phylogenomic data matrix, we retrieved the protein sequences of the 3,601 single-copy orthologous genes and individually aligned them with Mafft, version 7.402 (Katoh and Standley 2013), using the same parameters as described elsewhere (Steenwyk *et al.* 2019b). Nucleotide sequences were threaded onto the protein alignments using the thread_dna function in PhyKIT, version 0.1 (Steenwyk *et al.* 2021). The codon-based sequences were subsequently trimmed using trimAl, version 1.2rev59 (Capella-Gutierrez *et al.* 2009), using the "automated1" parameter. The resulting single-gene alignments were concatenated into a single data matrix using the create_concat function in PhyKIT, version 0.1 (Steenwyk *et al.* 2021).

To infer the evolutionary history of *Aspergillus* species in section *Fumigati* and the outgroup taxa, we used concatenation without gene-based partitioning, concatenation with gene-based partitioning, and gene-based coalescence in a maximum likelihood framework (Felsenstein 1981; Rokas *et al.* 2003; Edwards 2009; Zhang *et al.* 2018). For concatenation without gene-based partitioning, we used the 3,601-gene matrix as input to IQ-TREE (Nguyen *et al.* 2015) and inferred the best-fitting model of substitutions according to Bayesian information criterion values using the "-m TEST" parameter. The best-fitting model was determined to be a general time-reversal model with invariable sites, empirical nucleotide frequencies, and a discrete gamma model with four rate categories or "GTR + F+I + G4" (Tavaré 1986; Yang 1994; Gu *et al.* 1995). Last, we increased the number of candidate trees used during maximum likelihood search by setting the "-nbest" parameter to 10. Bipartition support was assessed using 5,000 ultrafast bootstrap approximations (Hoang *et al.* 2018). We refer to the tree inferred using this method as the reference tree topology depicted in Figure 1.

To infer the evolutionary history of *Aspergillus* species in section *Fumigati* and the outgroup strains using concatenation with gene-based partitioning and coalescence, we first determined the best-fitting model of substitution using the "-m TEST" parameter and reconstructed the phylogeny of the 3,601 single-copy orthologous genes individually using default IQ-TREE parameters (Nguyen *et al.* 2015). For concatenation with gene-based partitioning, we created a nexus-format partition file that describes gene boundaries in the 3,601-gene matrix and the best-fitting model of substitutions for each partition. We used the nexus-format partition file as input using the "-spp" parameter along with the concatenated 3,601-gene matrix to reconstruct the *Fumigati* phylogeny. Bipartition support was assessed using 5,000 ultrafast bootstrap approximations (Hoang *et al.* 2018). For coalescence, we first collapsed lowly supported bipartitions in all single-gene trees defined as less than 80% ultrafast bootstrap approximation support to reduce signal from poorly supported bipartitions. To do so, we assessed bipartition support using 5,000 ultrafast bootstrap approximations for individual single-gene trees (Hoang *et al.* 2018). To infer a coalescence-based phylogeny, we combined all single-gene trees with collapsed bipartitions into a single file and used it as input to ASTRAL-III, version 5.6.3 (Zhang *et al.* 2018), with default parameters. Bipartition support was assessed using posterior probabilities.

Gene family history

To determine the evolutionary history of the 14,294 gene families across section *Fumigati* species and outgroups, we implemented a maximum likelihood framework with a birth-death innovation model and gamma-distributed rates across families as implemented in DupliPHY-ML (Ames *et al.* 2012). DupliPHY-ML takes as input a matrix of gene family copy number and a phylogeny. To construct a matrix of gene family copy number, we used all orthologous groups of genes constructed as part of our

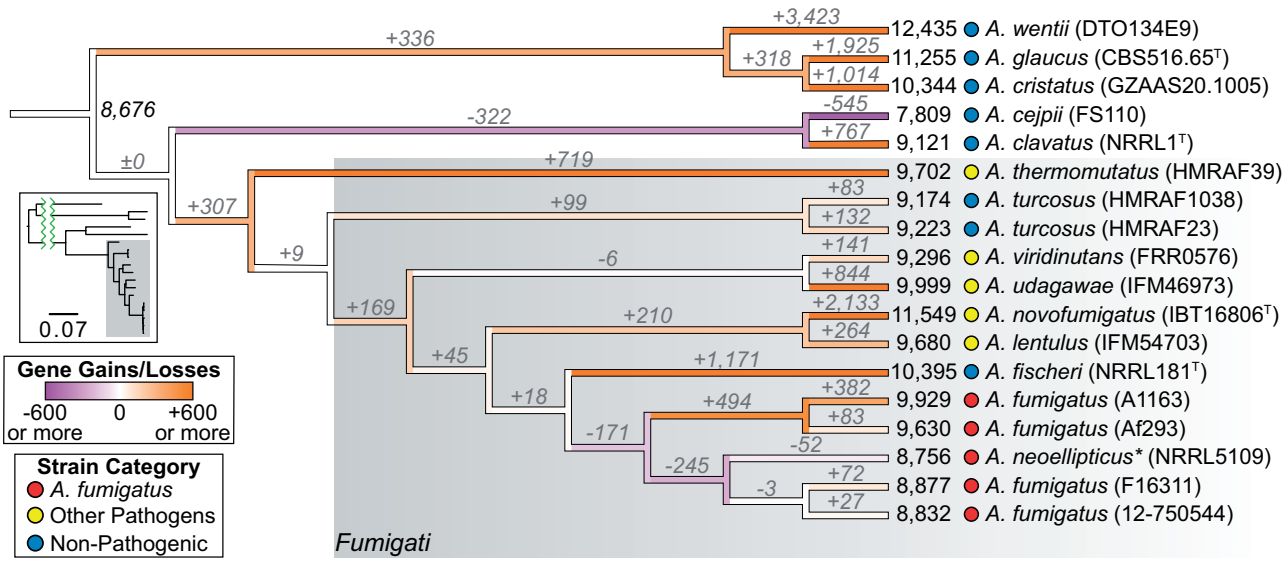

Figure 1 Genome-scale phylogeny and evolution of net gene gains or losses across *Aspergillus* section *Fumigati*. Relationships among taxa in section *Fumigati* inferred from a concatenation-based, maximum likelihood approach. The asterisk next to *A. neoellipticus* denotes that in our coalescence approach this taxon was placed sister to all *A. fumigatus* strains. Gene gains and losses were calculated based on a maximum likelihood framework implemented in DupliPHY-ML (Ames *et al.* 2012) that utilized the 14,294 orthogroups we constructed as part of our phylogenomic analyses as proxies for gene families. Branches are colored based on the number of net gene gains or losses, and 8,676 genes were inferred at the last common ancestor of all taxa studied. Numbers at branch tips represent the total number of genes in that genome. Strain designations are in parenthesis next to species names and type strains are denoted by a superscript "T" next to their strain designations. Insert shows the phylogeny with branch lengths reflective of the estimated number of nucleotide substitutions per site (scale bar is 0.07 substitutions/site); taxa are in the same order as the larger cladogram. The number of gene gains, losses, and the net gain or loss are shown in Supplementary Table S2.

phylogenetic analyses as proxies for gene families and used the number of gene sequences for a given species as the copy number information per gene family. For the phylogeny, we used the reference phylogeny described previously.

Gene ontology enrichment analyses

To determine if lists of genes of interest contained enriched Gene Ontology (GO) terms, we used GOATOOLs version 0.9.7 (Klopfenstein *et al.* 2018). Annotations for the *A. fumigatus* Af293 genome were downloaded from version 45 of FungiDB (Basenko *et al.* 2018), and the July 1, 2019 version of the basic Gene Ontology (Ashburner *et al.* 2000; The Gene Ontology Consortium 2021) was used for all analyses. A term was considered enriched if it had an adjusted *P*-value (using the Benjamini–Hochberg method) < 0.05.

Gene family expansions and contractions

To study if the number of gene family members is expanded or contracted in classes of strains (pathogens or nonpathogens) or in specific strains, we carried out a phylogenetically informed analysis of variance with the phylANOVA function located within version 0.7-70 of the phytools package (Revell 2012) with the 10,677 orthogroups that had a different number of family members in at least one taxon. Taxon relationships were provided from the phylogenetic tree that resulted from the concatenation without gene based partitioning approach. The simulation-based ANOVA was performed for each gene family and run with 10,000 simulations in order to derive a *P*-value reflecting if the average number of genes was different in the three groups of strains: *A. fumigatus* strains, other pathogens, and nonpathogens (see '*Results*' section for how these groups were defined). *P*-values were then corrected using the Benjamini–Hochberg method found within the "p.adjust" function in R (R Core Team 2018). Gene families were considered significantly different if their adjusted *P*-values were < 0.05. Tukey's range post-hoc test from the Python module "statsmodels" version 0.10.0 (Seabold and Perktold 2010) was then carried out on significantly different gene families in order to determine if the average number of gene family members differed in any of the pairwise comparison (*e.g.*, the number of genes in *A. fumigatus vs.* nonpathogenic species).

Estimating rates of molecular evolution

To determine the rate of sequence evolution across the evolutionary history of *Fumigati* species on a per gene basis, we used measures of the rate of nonsynonymous substitutions (dN) over the rate of synonymous substitutions (dS; (hereafter referred to as dN/dS or ω)) using an approach described elsewhere (Steenwyk *et al.* 2019a). To do so, we used untrimmed codon-based alignments generated during the construction of the 3,601-gene matrix used for phylogenomic analyses. For each of the 3,601 genes, we calculated ω using PAML, version 4.9 (Yang 2007), under two hypotheses: a null hypothesis (H_O) and an alternative hypothesis (H_A). For H_O, we allowed a single ω value to represent the rate of sequence evolution across the reference phylogeny. For the first H_A, we tested if different groups (*A. fumigatus*, other pathogens, or nonpathogens) were associated with different rates of sequence evolution. For the second H_A, we tested if each gene was evolving at a unique rate in each pathogen, relative to the other branches in the tree. For each comparison, to determine if H_A significantly differed from H_O, we used a likelihood ratio test ($\alpha = 0.01$).

Amoeba predation assays

To test whether our evolutionary genomic analyses could identify loci that contribute to resistance to phagocytosis, we conducted amoeba predation assays. Asexual spores (conidia) of *A. fumigatus* [either TF mutants previously described (Furukawa *et al.* 2020) and obtained as described in Zhao *et al.* (2019) or the background strain CEA17] were incubated 4 h at 37 °C in Czapek-Dox (CZD) medium (Sigma-Aldrich Chemie, Munich, Germany) to induce swelling, and confronted with *Protostelium aurantium* at a prey–predator ratio of 10:1 (10^5 conidia and 10^4 trophozoites of *P. aurantium*) for 18 h at 22°C. Mutants were chosen based on the traits their genes possessed in the lists of genes identified in the evolutionary genomic analyses (*e.g.*, only in *A. fumigatus* or fast-evolving in pathogens). After confrontation, the assay plate was incubated for 1 h at 37°C to inactivate the amoebae. Subsequently, 0.002% [w/v] resazurin (Sigma-Aldrich, Taufkirchen, Germany) was added and metabolic rates were calculated from the time dependent reduction of resazurin to the fluorescent resorufin over 3 h at 37°C using an Infinite M200 Pro fluorescence plate reader (Tecan, Männedorf, Switzerland). Survival was determined from the difference in the metabolic rates of the fungus after amoeba confrontation and amoeba-free controls. These controls were also used to determine the fitness of each strain in CZD-medium. Essentially, the same assay was carried out to determine the survival of germlings of *A. fumigatus*, except that conidia of *A. fumigatus* were pre-grown to germlings for 10 h at 37°C in CZD medium before the addition of trophozoites of *P. aurantium*.

Virulence assays in the great wax moth (*Galleria mellonella*) model of fungal disease

To test whether our evolutionary genomic analyses could identify loci that contribute to fungal disease, we conducted virulence assays using the greater wax moth (*G. mellonella*) model of fungal disease. *Galleria mellonella* larvae were obtained by breeding adult larvae (Fuchs *et al.* 2010) weighing 275–330 mg in starvation conditions in petri dishes at 37°C in the dark for 24 h prior to infection. All selected larvae were in the final stage of larval (sixth) stage development. Fresh asexual spores of each strain of *A. fumigatus* were obtained. For each strain, spores were counted using a hemocytometer and the initial concentration of the spore suspensions for the infections were 2×10^8 spores/ml. A total of 5 µl (1×10^6 spores) of each suspension was inoculated per larva. The control group was composed of larvae inoculated with 5 µl of PBS to observe death by physical trauma. The inoculation was performed using a Hamilton syringe (7000.5KH) through the last left proleg. After infection, the larvae were kept in petri dishes at 37°C in the dark and were scored daily. Larvae were considered dead when a lack of movement was observed in response to touch. The viability of the inoculum administered was determined by plating a serial dilution of the asexual spores in 37% YAG (yeast extract, agar, and glucose) medium. The statistical significance of the comparative survival values was calculated using the log rank analysis of Mantel–Cox and Gehan–Brestow–Wilcoxon found in the statistical analysis package Prism.

Data availability

All Supplementary material and their descriptions can be found on figshare at https://doi.org/10.6084/m9.figshare.14424386.

Results

A genome-scale phylogeny of *Aspergillus* section *Fumigati*

Phylogenetic relationships among taxa in section *Fumigati* (Supplementary Table S1) were examined using three maximum likelihood approaches—concatenation without gene-based partitioning, concatenation with gene-based partitioning, and coalescence—using 3,601 single-copy orthologous genes. These 3,601 genes were the subset of the 14,294 groups of orthologous genes inferred for these 18 taxa (see '*Materials and methods*'). Both concatenation approaches yielded the same topology, recovering *A. neoellipticus* nested within *A. fumigatus* (Figure 1). All bipartitions received full support except the split between *A. neoellipticus* and *A. fumigatus* strains F16311 and 12-750544, which received 98% ultrafast bootstrap approximation support. The coalescence approach inferred a fully supported alternative topology that placed *A. neoellipticus* sister to the four *A. fumigatus* strains. Whether *A. neoellipticus* is conspecific with *A. fumigatus* or a distinct species has been previously discussed in the literature (Li *et al.* 2014) and our genome-scale analyses reflect this debate. Given the close evolutionary relationship of the two species, we choose to refer to *A. neoellipticus* as a strain of *A. fumigatus* rather than a distinct species.

Broad conservation of genes and gene families, including those related to virulence, across section *Fumigati*

To understand variation in the distribution of genes, including ones known to be involved in *A. fumigatus* virulence (Abad *et al.* 2010; Bignell *et al.* 2016; Kjærbølling *et al.* 2018; Mead *et al.* 2019; Steenwyk *et al.* 2020c; Urban *et al.* 2020), we inferred gene and gene family gains and losses for every branch on the phylogeny (Figure 1 and Supplementary Figure S1). The number of gene family members at each node of the tree was estimated based on a maximum likelihood framework implemented in DupliPHY-ML (Ames *et al.* 2012) that utilized the 14,294 orthogroups we constructed as part of our phylogenomic analyses as proxies for gene families. The same dataset was used in our gene family analyses; in these analyses, we did not use the numbers of family members in each taxon but whether a specific gene family was present/absent in a given taxon. We inferred a net gain of 307 genes in the last common ancestor of section *Fumigati*. In addition, we found a net loss of 171 genes in the last common ancestor of *A. fumigatus* strains (Figure 1 and Supplementary Table S2). An estimated net gain of 494 genes occurred in the last common ancestor of the two *A. fumigatus* reference strains, A1163 and Af293. The same general patterns of genome expansion and contraction were observed when gene family gain and loss were estimated (Supplementary Figure S1).

To identify genes and gene families whose evolution was consistent with the conserved pathogenicity and species-specific pathogenicity models, we searched the 18 *Aspergillus* genomes for genes that were conserved across pathogens or specific to individual pathogens using both candidate and unbiased approaches. The candidate approach consisted of inferring the presence or absence pattern of 206 virulence-related genes (Steenwyk *et al.* 2020c) in each of the 18 genomes. Our gene family analysis placed the 206 virulence-related genes into 189 gene families. The largest virulence-related gene family (containing the transporter *abcC*—Afu1g14330; Paul *et al.* 2013) had 259 family members spread across all 18 genomes, and the smallest gene family (containing the terpene cyclase *fma-TC*—Afu8g00520—from the fumagillin biosynthetic gene cluster; Guruceaga *et al.* 2018) had six family members spread across only six of the genomes we analyzed. The same number of family members was found in every genome for 84/189 (~44%) of the virulence-related gene families, including for 81 families with one gene family member in each strain. Of those virulence-related gene families that differed in family member size across the 18 genomes, there were no virulence-related gene families with only members in pathogens, section *Fumigati* species, or *A. fumigatus* (Figure 2A).

We inferred that 164/189 virulence-related gene families were already present in the last common ancestor of all section *Fumigati* species. Similarly, we estimated that on average, 12 genes have been lost from virulence-related gene families during the evolution of *A. fumigatus* strains 12-750544, F16311, and *A. neoellipticus* and 2 genes have been gained from virulence-related gene families during the evolution of *A. fumigatus* strains Af293 and A1163 compared with the *A. fumigatus* last common ancestor. The finding that many virulence-related genes are conserved across both pathogens and nonpathogens in section *Fumigati* suggests that most known genetic determinants of virulence likely evolved for functions other than causing disease in humans and have been instead recruited into performing roles important for pathogenicity in certain species.

Our unbiased approach consisted of analyzing all 14,294 gene families that resulted from constructing orthologous groups of genes from all 18 *Aspergillus* genomes. Similar to what we observed with virulence-related genes, we found that 4,361/14,294 gene families (~31%) had family members in each of the 18 strains analyzed, and no gene families were present only in pathogens (Figure 2B). However, we found 98 gene families that were specific to section *Fumigati* (Figure 2B and Supplementary Table S4). Although the 98 gene families were not enriched for any Gene Ontology biological processes, molecular functions, or cellular compartments, the group contained genes associated with previously identified virulence-related traits, such as: a gene encoding a dimethylallyl tryptophan synthase (*cdpNPT*—Afu8g00620) located near the fumitremorgin-fumagillin-pseurotin supercluster (Yin *et al.* 2007; Wiemann *et al.* 2013), a major facilitator type transporter (*mdr3*—Afu3g03500) whose gene is highly expressed in *A. fumigatus* strains resistant to drugs (Nascimento *et al.* 2003; da Silva Ferreira *et al.* 2004), and a homolog of *mgtC* (Afu7g05060), a bacterial virulence factor required for survival in macrophages (Blanc-Potard 1997; Gastebois *et al.* 2011).

We found 72 gene families that were uniquely present in *A. fumigatus* (Figure 2B and Supplementary Table S4). These *A. fumigatus*-specific genes were not enriched for any GO terms and have not previously been tested for roles in virulence-related traits. The number of uniquely present gene families in other pathogens ranged from 1,280 in *Aspergillus novofumigatus* to 303 in *Aspergillus lentulus*. We also found two gene families (predicted glucose-methanol-choline oxidoreductase family members—NFIA_036190/NFIA_036210, and a membrane dipeptidase—NFIA_057190) that had members in all other taxa except *A. fumigatus* and one gene family found only in nonpathogenic taxa (a hypothetical protein with no identifiable domains—NFIA_057720). In summary, gene families are largely conserved across pathogens and nonpathogens in section *Fumigati*, but 72 gene families were found only in *A. fumigatus*; while these gene families have yet to be investigated for their potential roles in virulence, they represent candidates for the species-specific pathogenicity model.

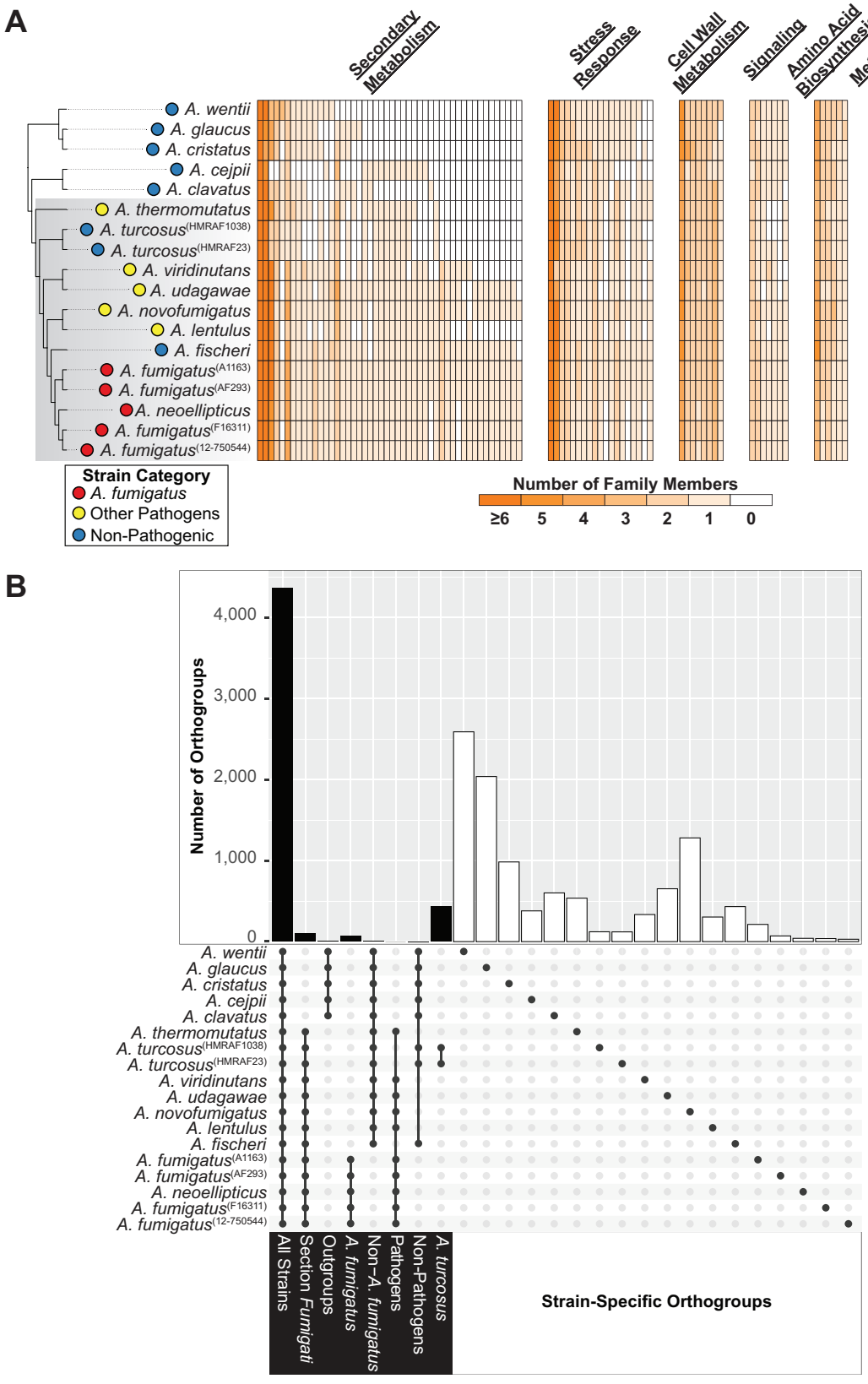

Figure 2 Gene families are largely conserved across section *Fumigati*, regardless of pathogenicity level. (A) Some virulence-related genes have different presence/absence patterns across strains in section *Fumigati*. Left, cladogram from Figure 1 showing the relationships amongst the 18 genomes studied. Gray box indicates taxa belonging to section *Fumigati*. Right, heatmaps of the 105/189 gene families related to virulence that exhibited at least one gene presence/absence change in at least one taxon, split into groups based on their general biological functions. Gene family labels can be found in Supplementary Table S3 in the same order presented here from left to right. (B) Gene families with representatives from all strains are the most prevalent. Upset plot (Conway *et al.* 2017) showing the number of all gene families present or absent in specific sets of strains. Black bars, gene family sets with members in more than one strain. White bars, strain-specific gene family sets.

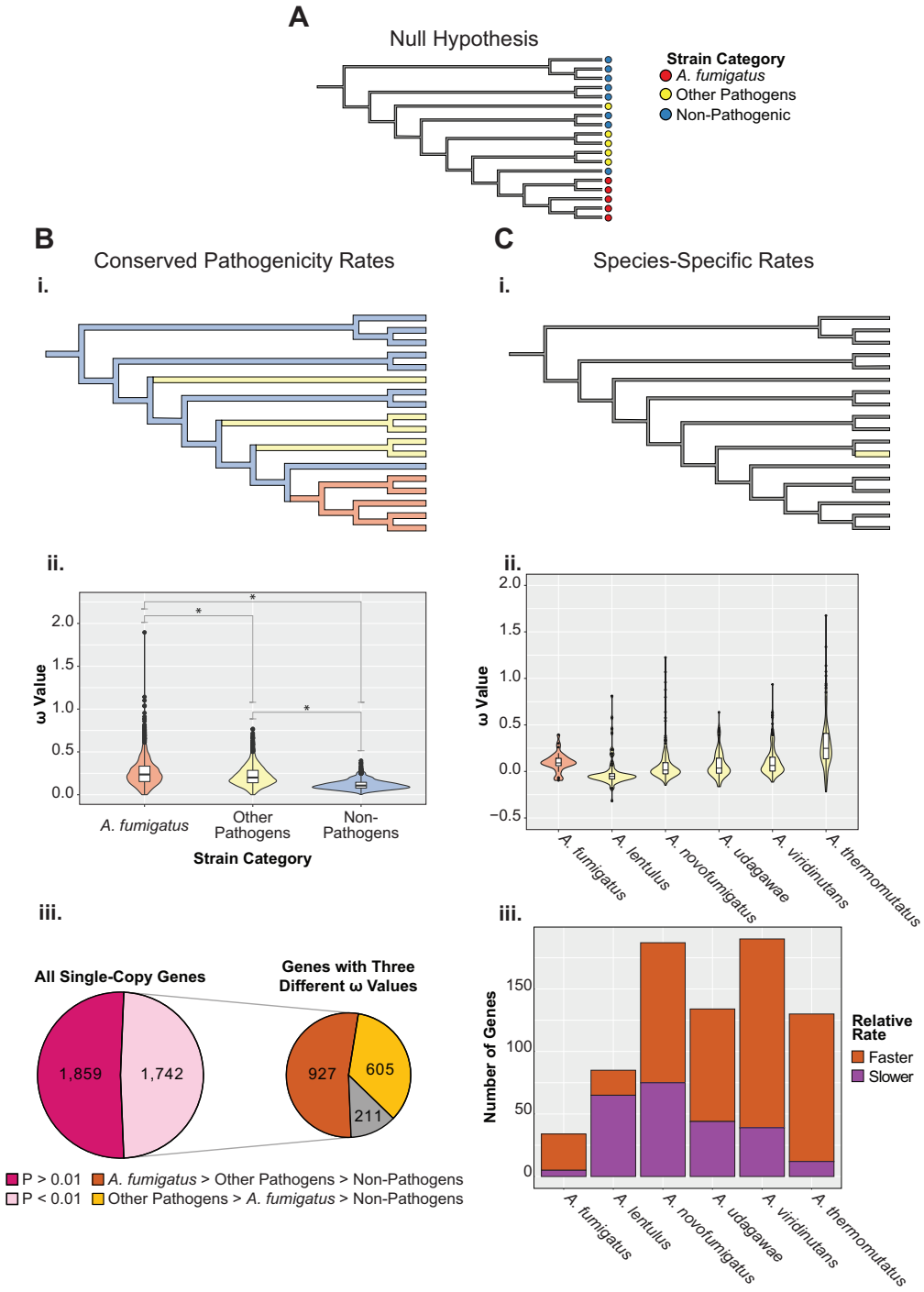

Figure 3 Genes in section *Fumigati* exhibit both pathogen- and species-specific rates of evolution. (A) The null hypothesis that all branches in the phylogeny have the same ω value. (Bi) Alternative hypothesis examining the conserved pathogenicity model of gene evolution where genes are evolving at rates that correspond to their pathogenicity level. Branches are colored based on their strain category (*A. fumigatus*—most pathogenic, other pathogens, and nonpathogens—least pathogenic) and evolutionary rate. (Bii) Violin and box plots showing the ω values for each of 1,742 genes (49% of all single-copy genes) that exhibited different ω values (*P*-value < 0.01) in the three groups of strains (*A. fumigatus*, other pathogens, and nonpathogens). Ten genes had ω values > 0.8 in *A. fumigatus* strains and are not shown here (see Supplementary Table S4). *adjusted *P*-value < 0.0001 in a Paired Wilcoxon Signed-Rank test. (Biii) Left, pie chart showing the number of genes that either did (light pink) or did not (dark pink) have different ω values in strains with different pathogenicity levels. Right, pie chart further detailing the relative magnitude of ω values of genes in strains with different pathogenicity levels. (Ci) A second alternative hypothesis examining the species-specific pathogenicity model of gene evolution where genes are evolving at one rate in one taxon and a different rate in all other taxa. Shown here is the model that was tested for *A. lentulus* where genes in *A. lentulus* experienced one rate of evolution (yellow), whereas their counterparts in all other taxa exhibited a different rate (gray). (Cii) Violin and box plots showing the difference between ω values of genes that had one ω in the pathogen, a different ω value in other taxa analyzed ($P < 0.01$), and were not found also to have pathogen-specific rates (*i.e.*, were not included in the set of 1,742 genes shown in B). The numbers of genes used to construct these plots were 34 for *A. fumigatus*, 85 for *A. lentulus*, 187 for *A. novofumigatus*, 130 for *A. thermomutatus*, 134 for *A. udagawae*, and 190 for *A. viridinutans*. ω value differences between the pathogen of interest and all other strains that were greater than two are not shown and constituted only 13/643 comparisons that exhibited $P < 0.01$ and were not found in the conserved pathogenicity analyses (B). Strains are colored the same as in (B). (Ciii) Bar chart showing the number of genes where the gene in one pathogen is evolving faster or slower than its counterparts in all other strains. "Faster," genes evolving faster in the pathogen compared with all other taxa. "Slower," genes evolving slower in the pathogen compared with all other taxa. Taxa in all phylogenies are in the same order as in Figure 1.

The distributions of few gene families are associated with pathogenicity

Our analyses did not identify gene families whose presence/absence patterns were conserved in all pathogens found in section *Fumigati*. An alternative hypothesis is that the number of gene family members in a given taxon could reflect the organisms' ability to cause disease. To test this hypothesis we carried out a phylogenetically informed ANOVA (Revell 2012) on all 10,677 gene families that displayed a different number of gene family members in at least one taxon. For this analysis we split the 18 *Aspergillus* taxa into three groups based on their pathogenicity levels (*i.e.*, how frequently they are generally found in the clinic): *A. fumigatus* (most pathogenic), other pathogens (that are not *A. fumigatus*), and nonpathogens (least pathogenic). Although we focus here on identifying pathogenicity-related genes, this approach will likely also identify genes important for *A. fumigatus*-specific traits unrelated to pathogenicity as *A. fumigatus* is the only species in its category.

We found 83 gene families that had statistically significant differences in the number of members between groups (Supplementary Figure S2). After conducting Tukey's post-hoc test on all 83 gene families, we observed that 72/83 gene families had more copies in *A. fumigatus* and were in fact those previously identified as "*A. fumigatus*-specific" in our strict gene presence/absence analysis (Supplementary Table S4). One of the remaining 11 gene families was the membrane dipeptidase (NFIA_057190) found during our gene presence/absence analysis in all genomes other than *A. fumigatus*. The hypothetical gene family (NFIA_057720) found only in nonpathogenic species with the same gene family presence/absence analysis (Figure 2B) was also identified via the phylogenetically informed ANOVA. The remaining nine gene families had the same number of genes in *A. fumigatus* and nonpathogens but a different number of family members in the other pathogens. Three (P174DRAFT_459701, P174DRAFT_448681, and P174DRAFT_440824) possessed no conserved domains, and the other six had a carbohydrate binding domain (P174DRAFT_502341—PF09362), three ankyrin repeat domains (P174DRAFT_501662—PF12796), a major facilitator superfamily domain (P174DRAFT_432793—PF07690), a domain of unknown function (P174DRAFT_509497—PF11905), a sulfur-carrier domain (P174DRAFT_347437—PF03473), and an aldehyde dehydrogenase family domain (P174DRAFT_378794—PF00171), respectively. Together, these data show that few gene families exhibit significant variation in their numbers across section *Fumigati* with respect to pathogenicity.

Many genes experienced faster rates of evolution in pathogenic species

Another way in which genomes evolve that can affect pathogenicity is through changes in the evolutionary rates of their constituent genes (Yang and Bielawski 2000). We carried out two evolutionary rate analyses to test whether our set of 3,601 single-copy orthologous genes exhibited different rates of evolution in pathogens compared with nonpathogens. For both analyses, our null hypothesis was that for a given single-copy gene, a single rate (ω) represented the rate of sequence evolution for each gene in every strain, regardless of the pathogenicity level of the organisms examined (Figure 3A). In the first analysis, our alternative hypothesis was that each gene evolved at a unique rate in each of our three groups (*A. fumigatus* strains, other pathogens, and nonpathogens; Figure 3Bi). We observed that 49% of genes tested (1,742/3,601) rejected the null hypothesis, suggesting that the evolutionary rate of these genes differs among the three groups. Of the 1,742 genes with three different ω values, 88% (1,532/1,742) had faster rates in pathogenic organisms (Figure 3Biii) and 10 had relatively high ω values (>0.8) in *A. fumigatus* (Supplementary Table S4). None of these 10 fastest-evolving genes have previously been studied and contain a variety of domains likely involved in diverse functions ranging from RNA binding to catalyzing oxidation/reduction reactions. Each group also exhibited its own statistically different distribution of ω values (Figure 3Bii). Of the 81/189 virulence-related gene families that were present in a single copy in each *Aspergillus* genome, 56% (45/81) exhibited different rates of evolution.

In the second analysis, we tested if each gene was evolving at a unique rate in each pathogen, relative to the other species we analyzed (Figure 3Ci). We found that, on average, 127 single-copy genes exhibited a different rate in the pathogen of interest than in the rest of species and were not found also to have pathogen-specific rates, with most evolving faster in the pathogens (Figure 3, Cii and Ciii). *A. fumigatus* had the smallest number of genes whose evolutionary rates differed from the rest of the species (34), whereas *A. viridinutans* had the most (190; Supplementary Table S5). Overall, our data show that genes in pathogens are evolving faster than in nonpathogens, both in a conserved and species-specific manner.

Transcription Factors with pathogenicity-related patterns of evolution have diverse effects on virulence

To test if any of the genes whose evolutionary signatures differed between pathogens and nonpathogens directly affected either fungal or host survival, we tested 17 knockout strains of TF-encoding genes (Furukawa *et al.* 2020) in two virulence-related assays. One TF (Afu7g00210) was found only in *A. fumigatus* (Figure 2B), one (Afu6g08540) was identified as being fast-evolving in *A. fumigatus*, four (Afu2g17895, Afu3g02160, Afu7g04890, and *gliZ*—Afu6g09630) were members of gene families with a statistically significant higher number of family members in pathogens in a preliminary one-way ANOVA (but not in our phylogenetically informed ANOVA), five (Afu1g11000, Afu2g00470, Afu6g11750, Afu3g00210, and Afu8g05750) were members of gene families with a statistically significant higher number of family members in *A. fumigatus* in the same preliminary one-way ANOVA (but also not in our phylogenetically-informed ANOVA), and six (Afu2g17860, Afu5g01065, Afu5g14530, Afu1g01340, Afu3g01640, and Afu2g16310) were found only section *Fumigati*. None of the TF mutants exhibited a growth defect compared with their parent strain (CEA17) when grown in conventional lab conditions (Supplementary Figure S3).

In the first assay, asexual spores (conidia) or germlings from either a background strain of *A. fumigatus* (CEA17) or one of the *A. fumigatus* knockout mutants of TFs were incubated with *P. aurantium*, a fungivorous amoeba used to study how fungi may have evolved the ability to evade or survive phagocytosis by human immune cells (Radosa *et al.* 2019). We found that mutant asexual spores of four genes (Afu2g17860, Afu2g17895, Afu7g04890, and Afu2g00470) exhibited an increase in survival relative to CEA17 (Figure 4A). Given that overall spore killing of all strains tested was >90%, to independently confirm the increase in viability of these four mutants, we conducted predation assays with germlings. Although germlings of many mutants showed a qualitative difference in survival relative to CEA17 (including three out of the four with statistically significant increases in viability during the spore predation assays), those differences were not statistically significant (Dunn's test-adjusted *P*-value > 0.1).

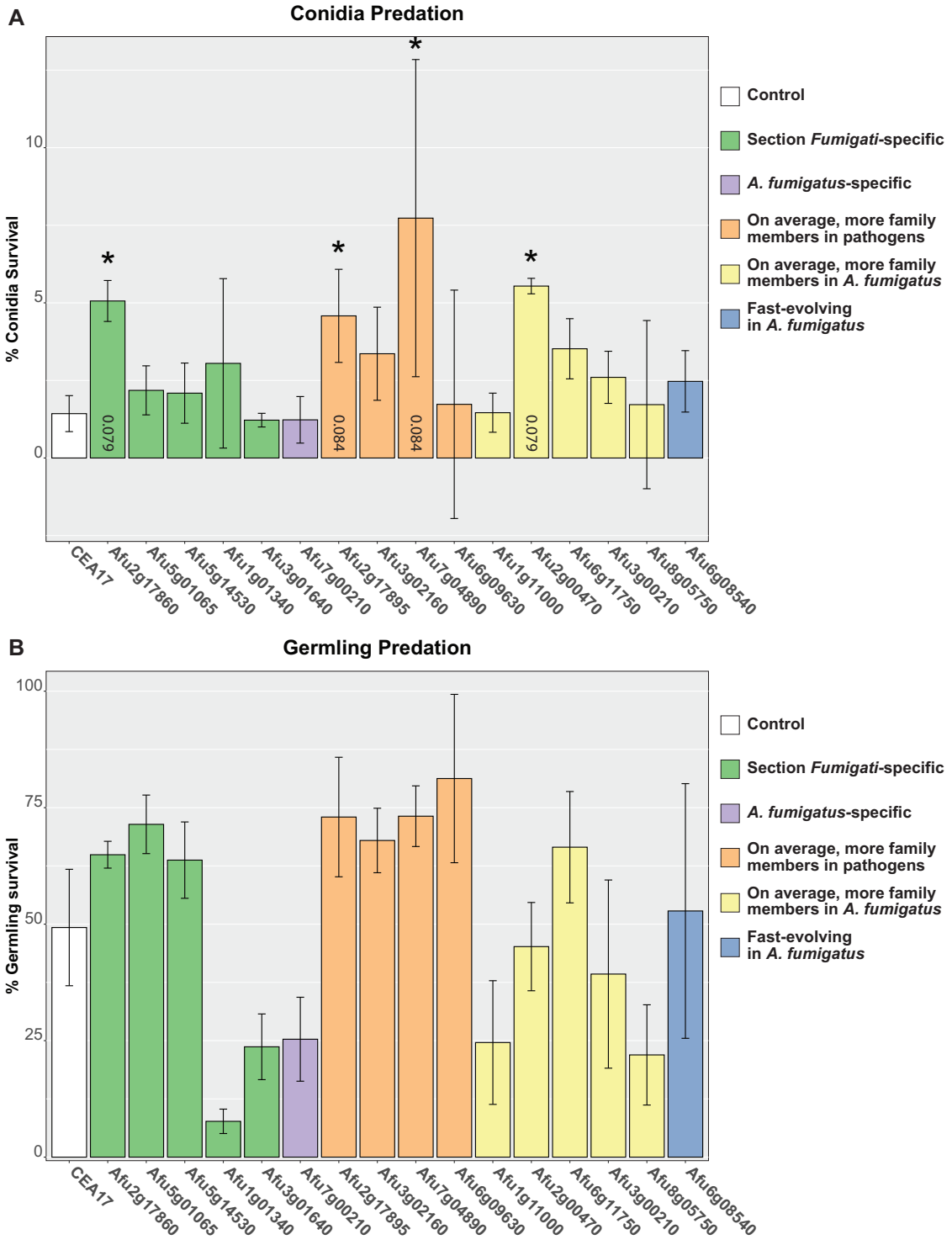

Figure 4 Multiple transcription factors whose evolution varies with respect to *Aspergillus* pathogenicity affect the survival of *A. fumigatus* during amoeba predation. (A) Survival of swollen *A. fumigatus* asexual spores (conidia) after interaction with *P. aurantium*. Spores of *A. fumigatus* were incubated 4 h at 37°C in CZD medium and confronted with *P. aurantium* at a prey–predator ratio of 10:1 (10^5 spores and 10^4 trophozoites of *P. aurantium*). Survival is expressed as the relative reduction in the metabolic rate of the fungus in comparison to amoeba-free controls over 3 h. Data represent the mean and SD of three biological replicates. *$P < 0.1$ in an adjusted Dunn Test comparing the survival of the mutant strain to the parental strain CEA17. (B) Survival of *A. fumigatus* germlings after interaction with *P. aurantium*. Asexual spores of *A. fumigatus* were pre-grown to germlings for 10 h at 37°C in CZD medium and confronted with *P. aurantium*. All other assay parameters are the same as in (A). No mutant strain exhibited statistically significant difference in survival relative to CEA17 in an adjusted Dunn test. Both asexual spores and germling confrontation assays were confirmed to have significant P-values (<0.05) in Kruskal–Wallis tests before carrying out the post-hoc test and mutant strains did not exhibit large growth phenotypes in the absence of amoeba (Supplementary Figure S4). Mutants are color-coded based on genomic traits related to pathogenicity that they possess. For the "On average, more family members in pathogens" and "On average, more family members in *A. fumigatus*" groups, the family members did not exhibit a statistically significant different number of family members in our phylogenetically informed ANOVA (Supplementary Figure S2). Knockout mutants were constructed in the CEA17 background (Furukawa *et al.* 2020), but *A. fumigatus* strain Af293 gene ids for the corresponding orthologous genes are shown here and in Figure 5. Gene absence patterns were confirmed with tblastn. A potential, low-confidence ortholog of Afu5g01065 was found in *A. wentii*; however, all other *A. fumigatus* genes were found missing in the species listed. The mutant of Afu2g16310 could not be assayed due to technical reasons.

In the second assay, we measured virulence in the greater wax moth (*G. mellonella*) model of *A. fumigatus* disease. We found that almost one third of all knockout mutants tested (5/17) exhibited a statistically significant decrease in virulence (Figure 5 and Supplementary Figure S5). One of the TF mutants that resulted in a significant decrease in larval killing was that of *gliZ* (Afu6g09630), a regulator of gliotoxin production (Bok *et al.* 2006) that we observed was found in all pathogenic taxa but missing in all nonpathogenic ones except *A. cejpii* and *A. fischeri*. To our knowledge, this is the first time *gliZ* has been tested in the greater wax moth model of fungal disease or shown to contribute to fungal pathogenesis. None of the other TFs whose mutants exhibited a decrease in virulence have previously been studied, and for three (Afu3g00210, Afu8g05750, and Afu7g00210) the only PFAM domain they contained was a "Fungal Zn(2)-Cys(6) binuclear cluster domain" (PF00172), whereas the other predicted TF (Afu7g04890) contained both the binuclear cluster domain and a "Fungal-specific TF domain" (PF04082). In summary, the data from both functional assays suggest that genes whose evolution differs between pathogens and nonpathogens are likely to contribute to disease-related traits.

Discussion

Our examination of *Aspergillus* genes whose evolution is associated with the observed differences in pathogenicity among section *Fumigati* taxa identified candidate genes that support the conserved pathogenicity model as well as candidates that support the species-specific pathogenicity model. Our results also show that previously described virulence-related genes are largely present in both pathogens and nonpathogens (Figure 2A), suggesting that most known genetic determinants of virulence are not likely to explain the observed pathogenicity differences between *Aspergillus* species.

Multiple TFs we identified as having virulence-related genomic traits also displayed roles in different virulence-related assays (Figures 4 and 5). All four spore mutants that were significant in the amoeba predation assays showed increased viability compared with the control strain. Survival differences in the presence of a phagocytic predator can occur at several levels, such as recognition, uptake, or intracellular fate. It is possible that a mutant that has acquired an advantage against a phagocyte has a tradeoff in the complex environment of the host. For example, modified cell surface components may allow escape from recognition by a phagocyte, but also result in better adhesion to surface structures in the host. Mutants of *Cryptococcus*, another fungal pathogen of humans, can undergo filamentation and then survive better against amoebae, but in the host, the increase in filamentation results in a decrease in virulence due to a reduction of fungal dissemination (Magditch *et al.* 2012). It is possible that some of the *A. fumigatus* mutants have a slightly extended resting stage and thus escape the predator over the limited time of the assay (dormant wildtype spores are inert to the amoebae; Ferling *et al.* 2020). We observed no major differences in growth between the mutants and the wildtype strain (Supplementary Figure S3), but these resting stage effects may be very small, and much is still unknown regarding how they may impact pathogenicity. In addition, our amoeba predation assays measure one or a few events during disease progression, namely phagocytic cell interactions, but the greater wax moth model tested the entirety of disease progression in a susceptible host; those results showed the expected outcome of decreased virulence in the knockout mutants.

Of the eight TFs whose null mutants exhibited at least one phenotype in our two assays, half of them (including one of the genes, Afu2g17860, whose mutant increased viability of spores in the amoeba predation assay) were downregulated and none were upregulated during the switch to human body temperature (Lind *et al.* 2016). Additionally, *gliZ*, a regulator of gliotoxin production whose gene family we found to be largely pathogen-specific and whose mutant was less virulent than the WT strain, was heavily upregulated in *A. fumigatus* germlings that were extracted 12–14 h after mouse infection (McDonagh *et al.* 2008) while the seven other TFs we studied were not differentially regulated during the early events of mouse infection. Taken together, our functional assays show that our evolutionary genomic approach is useful for uncovering the molecular mechanisms underpinning the evolution of pathogenicity and also has the power to identify genes both previously connected to *A. fumigatus* virulence in addition to novel ones.

We analyzed all sequenced species in section *Fumigati* (as of July 2019) and a representative sampling of strains from *A. fumigatus*, carried out a diverse set of evolutionary genomic analyses, and functionally tested our identified genes in multiple assays, thus building on previous studies that used smaller numbers of section *Fumigati* species and close relatives in *Aspergillus* and focused on strict gene presence/absence (Fedorova *et al.* 2008; Mead *et al.* 2019). Previous work also compared *A. novofumigatus*, one of the section *Fumigati* species we considered here, to its relative *A. fumigatus* (Kjærbølling *et al.* 2018), and while that study used a broader and less stringent list of virulence-related genes that also included allergens, they also saw high levels of gene conservation between the two species. This previous work and our own support the hypothesis that *A. novofumigatus* could be nearly as pathogenic as *A. fumigatus* due to this conservation of almost all virulence-related genes. In section *Flavi*, another taxonomic section in genus *Aspergillus* that contains the human and plant pathogen *Aspergillus flavus*, it has been hypothesized that TFs may be linked to pathogenicity (Kjærbølling *et al.* 2020), and similarly, we saw that one of the 72 *A. fumigatus*-specific genes we identified is a TF whose deletion reduces virulence in our invertebrate model of fungal disease (Figure 5).

As more genomes from strains and species in section *Fumigati* become available, our power to detect quantitative differences will increase, and allow us to more robustly test the conserved pathogenicity model and expand our species-specific pathogenicity model to include "strain-specific" elements. This will be especially important considering the continued and growing appreciation for strain-specific traits and differences in *Aspergillus* genomes and pathogenicity (Keller 2017; Ries *et al.* 2019; Bastos *et al.* 2020; Dos Santos *et al.* 2020; Kjærbølling *et al.* 2020; Steenwyk *et al.* 2020a, 2020b; Kowalski *et al.* 2021). Similarly, we recognize that it is unlikely that all of the *A. fumigatus*-specific genes (Figure 2B) or genomic attributes (Figure 3) we discovered may be directly connected to pathogenicity but may instead be connected to other *A. fumigatus*-specific traits. This caveat notwithstanding, these *A. fumigatus*-specific genes constitute a useful list of targets for beginning to understand why *A. fumigatus* evolved to be pathogenic whereas its closest relatives did not. Future studies will also place *A. oerlinghausenensis*, another species closely related to *A. fumigatus* (Houbraken *et al.* 2016), within this evolutionary framework of pathogenicity, but based on our recent genome-wide phylogenomic analyses of *A. oerlinghausenensis*, *A. fischeri*, and *A. fumigatus* (Steenwyk *et al.* 2020b), we do not anticipate that inclusion of *A. oerlinghausenensis* will drastically change our findings.

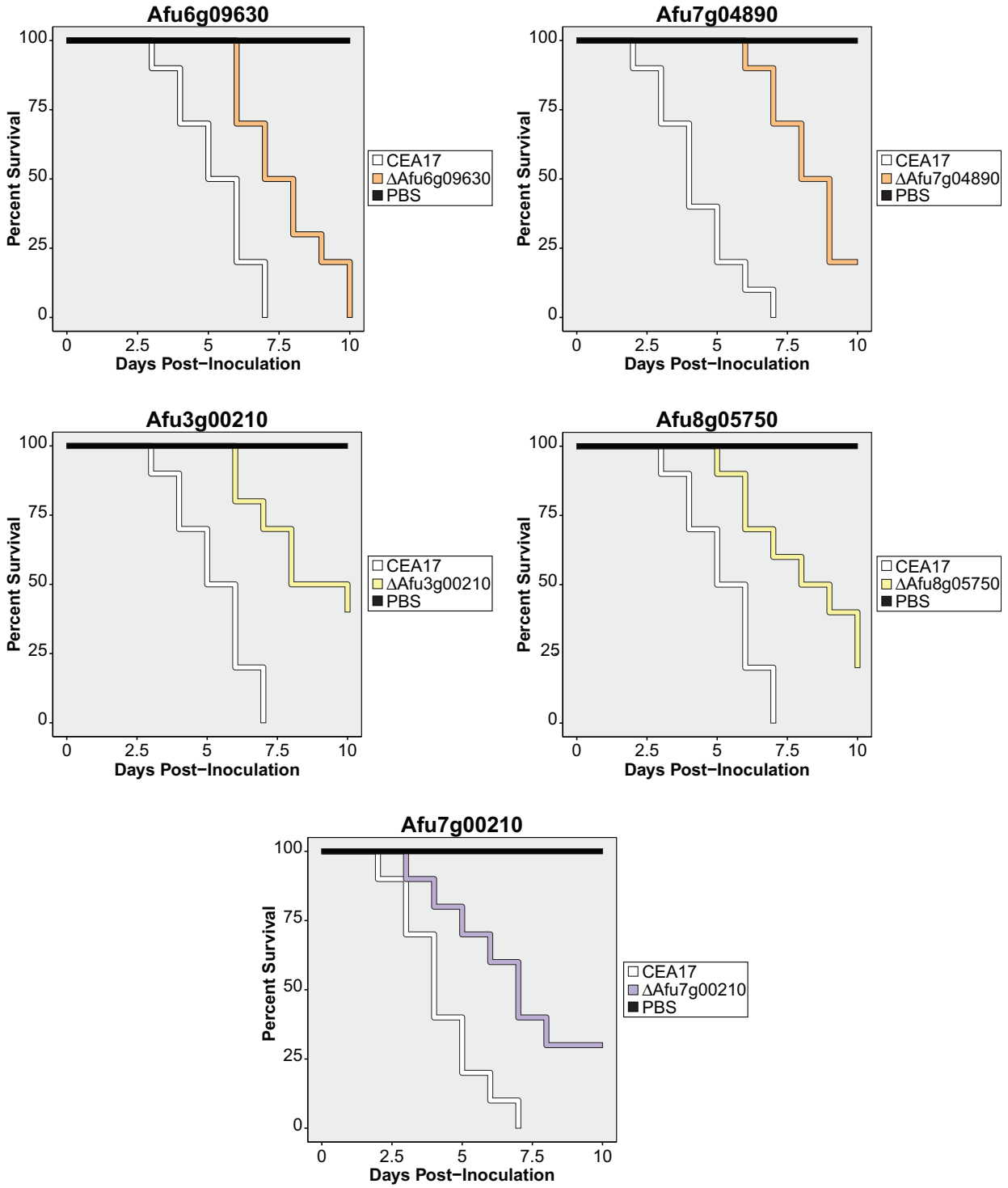

Figure 5 Multiple transcription factors in *A. fumigatus* whose evolution differs with respect to pathogenicity affect virulence in the greater wax moth model of disease. Cumulative survival of *G. mellonella* larvae inoculated with phosphate buffered saline (control; black), asexual spores of the parental strain CEA17 (white), and asexual spores from null mutants of TFs whose evolution is associated with the observed differences in *Aspergillus* pathogenicity (various colors). Ten larvae were used per inoculation in all assays. Color scheme is the same as in Figure 4. All mutant survival curves shown here were statistically different ($P < 0.008$ in a Log-rank test) from the CEA17 survival curve. Mutants whose survival curves are shown in orange on average have more gene family members in pathogens, those in yellow have on average more family members in *A. fumigatus*, and the mutant in purple is *A. fumigatus*-specific. Note that the results of the mutants in orange support the conserved pathogenicity model, whereas those of the mutants in yellow and purple support the species-specific pathogenicity model and that the mutants with orange and yellow survival curves did not exhibit a statistically significant different number of family members in our phylogenetically informed ANOVA (Supplementary Figure S2). The Afu6g09630 gene is *gliZ*, a regulator of the biosynthesis of the secondary metabolite gliotoxin, a known modulator of host biology.

The strains whose genomes we analyzed were isolated from both environmental and clinical locations (Supplementary Table S1), and based on published literature, we do not anticipate isolate setting to play a large role or confound our results analyzing pathogens and nonpathogens. For example, previous work (Ashu *et al.* 2017) reported that the ecological niche of *A. fumigatus* strains (including whether or not they were isolated from the clinic or environment) contributed a very small but statistically significant amount to the overall amounts of observed diversity between 2026 isolates. This suggests that the environmental and clinical strains used in our study are likely representative of *A. fumigatus* strain diversity. Furthermore, consistent with our results (Figure 2A), Puértolas-Balint *et al.* (2019) reported, using their own set of virulence-related genes, that both clinical and environmental strains of *A. fumigatus* have similar "virulence genetic content."

In general, it appears that clinical isolates of *A. fumigatus* are slightly more pathogenic than environmental isolates (Mondon *et al.* 1996; Alshareef and Robson 2014), perhaps due to within-host microevolution of clinical isolates, but this issue is still under active investigation in the field. For example, Kowalski *et al.* (2016) showed that on average, clinical strains are indeed slightly more pathogenic in their sample of *A. fumigatus* clinical and environmental strains, but the difference is relatively small compared with the heterogeneity of pathogenicity observed between *A. fumigatus* strains. We are unaware if the genomic traits and pathogenicity of environmental and clinical strains have been compared in species other than *A. fumigatus*. We are currently designing these experiments with these and other strains.

Evolutionary studies have also been carried out in fungal pathogens outside of the genus *Aspergillus*, and when our results are placed in the context of this literature, a diverse set of mechanisms have driven the evolution of fungal pathogenicity (Taylor 2015). The ability to infect humans has also evolved multiple times in *Candida* species found within the fungal subphylum Saccharomycotina (Gabaldón *et al.* 2016); however, gene family expansion and interspecies hybridization were much more important for the evolution of pathogenicity in that clade (Butler *et al.* 2009) compared with the results we present here in section *Fumigati* where there was little evidence of dramatic changes in gene family member number between pathogens and nonpathogens (Supplementary Figure S2). Similarly, gene family size was hypothesized to be an important factor in the evolution of *Coccidioides* pathogens, and just as we only saw 84 genes with *A. fumigatus*-specific evolutionary rates, this group of pathogens had a relatively small number of genes with species-specific evolutionary rates (Sharpton *et al.* 2009). In pathogenic *Cryptococcus* species, mating-type loci and the switch from a tetrapolar to bipolar mating system have been suggested as being key in producing the genomic environment necessary for pathogenicity to evolve (Sun *et al.* 2019), but in *A. fumigatus*, mating-type loci do not appear to contribute to virulence (Losada *et al.* 2015) and the contribution of mating across section *Fumigati* has only rarely been studied (Rydholm *et al.* 2007).

Worldwide mortality rates for aspergillosis infections are estimated to range from as high as 95% to as low as 30%, and drug resistance is a frequent worry for clinicians (Brown *et al.* 2012). To combat this global health issue, more must be understood about *Aspergillus* biology and evolution. Here, we showed how an evolutionary approach can guide the identification of pathogenicity-associated genetic elements in *Aspergillus* fungi, presented many promising, novel candidates for future study, and placed them within an evolutionary context that will also guide their study with relation to non-*A. fumigatus* pathogenic species found within section *Fumigati*. Our data also provide clues on how *Aspergillus* pathogenicity evolved through the contribution of genetic elements that fit both the conserved pathogenicity and species-specific pathogenicity models. Furthermore, genes that fit the conserved pathogenicity model may be useful as targets for the treatment of disease caused by all section *Fumigati* species, whereas genes that fit the species-specific pathogenicity model may be useful for species-specific treatment strategies. More generally, this work provides the basis for an evolutionary framework that can inform multiple aspects of the study of both *Aspergillus* species and the diseases they cause.

Acknowledgments

We thank members of the Rokas lab for enlightening discussions and comments on the project, especially Dr. Xing-Xing Shen and Dr. Abigail LaBella. We also thank Dr. Xing-Xing for conducting preliminary evolutionary analyses that were not included in this article.

Funding

J.L.S. and A.R. were supported by the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study Program. A.R.'s laboratory received additional support from a Discovery grant from Vanderbilt University, the Burroughs Wellcome Fund, the National Science Foundation (DEB-1442113), and the National Institutes of Health/National Institute of Allergy and Infectious Diseases (R56AI146096). LPS, PAdC, and GHG were supported by FAPESP (Fundação de Apoio à Pesquisa do Estado de São Paulo; grant number 2016/07870-9), and CNPq (Conselho Nacional de Desenvolvimento Científico e Technológico), both from Brazil. FH was supported by the Deutsche Forschungsgemeinschaft (DFG, research grant HI1574/4-1).

Conflicts of interest

A.R. is a Scientific Consultant for LifeMine Therapeutics, Inc.

Literature cited

Abad A, Fernández-Molina JV, Bikandi J, Ramírez A, Margareto J, *et al.* 2010. What makes *Aspergillus fumigatus* a successful pathogen? Genes and molecules involved in invasive aspergillosis. Rev Iberoam Micol. 27:155–182.

Abdolrasouli A, Rhodes J, Beale MA, Hagen F, Rogers TR, *et al.* 2015. Genomic context of azole resistance mutations in *Aspergillus fumigatus* determined using whole-genome sequencing. mBio. 6: e00536. doi:10.1128/mBio.00536-15.

Alastruey-Izquierdo A, Mellado E, Peláez T, Pemán J, Zapico S, *et al.* 2013. Population-based survey of filamentous fungi and antifungal resistance in Spain (FILPOP Study). Antimicrob Agents Chemother. 57:3380–3387.

Alastruey-Izquierdo A, Alcazar-Fuoli L, Cuenca-Estrella M. 2014. Antifungal susceptibility profile of cryptic species of *Aspergillus*. Mycopathologia. 178:427–433.

Alshareef F, Robson GD. 2014. Genetic and virulence variation in an environmental population of the opportunistic pathogen *Aspergillus fumigatus*. Microbiology. 160:742–751. doi: 10.1099/mic.0.072520-0.

Ames RM, Money D, Ghatge VP, Whelan S, Lovell SC. 2012. Determining the evolutionary history of gene families. Bioinformatics. 28:48–55. doi:10.1093/bioinformatics/btr592.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, *et al.* 2000. Gene Ontology: tool for the unification of biology. Nat Genet. 25:25–29. doi:10.1038/75556.

Ashu EE, Hagen F, Chowdhary A, Meis JF, Xu J. 2017. Global population genetic analysis of *Aspergillus fumigatus*. mSphere. 2: e00019-17. doi:10.1128/mSphere.00019-17.

Balajee SA, Kano R, Baddley JW, Moser SA, Marr KA, *et al.* 2009. Molecular identification of *Aspergillus* species collected for the Transplant-Associated Infection Surveillance Network. J Clin Microbiol. 47:3138–3141.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, *et al.* 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19:455–477. doi: 10.1089/cmb.2012.0021.

Barrs VR, van Doorn TM, Houbraken J, Kidd SE, Martin P, *et al.* 2013. *Aspergillus felis* sp. nov., an emerging agent of invasive aspergillosis in humans, cats, and dogs. PLoS One. 8:e64871. doi: 10.1371/journal.pone.0064871.

Basenko EY, Pulman JA, Shanmugasundram A, Harb OS, Crouch K, *et al.* 2018. FungiDB: an integrated bioinformatic resource for fungi and oomycetes. J Fungi (Basel). 4:39.

Bastos RW, Valero C, Silva LP, Schoen T, Drott M, *et al.* 2020. Functional characterization of clinical isolates of the opportunistic fungal pathogen *Aspergillus nidulans*. mSphere. 5:e00153-20. doi:10.1128/mSphere.00153-20

Bignell E, Cairns TC, Throckmorton K, Nierman WC, Keller NP. 2016. Secondary metabolite arsenal of an opportunistic pathogenic fungus. Philos Trans R Soc B. 371:20160023–20160029.

Blanc-Potard A-B. 1997. The Salmonella selC locus contains a pathogenicity island mediating intramacrophage survival. EMBO J. 16: 5376–5385. doi:10.1093/emboj/16.17.5376

Bok JW, Chung D, Balajee SA, Marr KA, Andes D, *et al.* 2006. GliZ, a transcriptional regulator of gliotoxin biosynthesis, contributes to *Aspergillus fumigatus* virulence. Infection Immunity. 74:6761–6768.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120. doi:10.1093/bioinformatics/btu170

Bongomin F, Gago S, Oladele RO, Denning DW. 2017. Global and multi-national prevalence of fungal diseases-estimate precision. J Fungi (Basel). 3:57.doi:10.3390/jof3040057

Brown GD, Denning DW, Gow NAR, Levitz SM, Netea MG, *et al.* 2012. Hidden killers: human fungal infections. Sci Transl Med. 4: 165rv13.

Brown NA, Goldman GH. 2016. The contribution of *Aspergillus fumigatus* stress responses to virulence and antifungal resistance. J Microbiol. 54:243–253.

Butler G, Rasmussen MD, Lin MF, Santos MAS, Sakthikumar S, *et al.* 2009. Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature. 459:657–662. doi:10.1038/nature08064

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, *et al.* 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10:421. doi:10.1186/1471-2105-10-421

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25:1972–1973. doi:10.1093/bioinformatics/btp348

Casadevall A, Pirofski L. 2007. Accidental virulence, cryptic pathogenesis, martians, lost hosts, and the pathogenicity of environmental microbes. Eukaryot Cell. 6:2169–2174.

Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 33:2938–2940. doi:10.1093/bioinformatics/btx364

da Silva Ferreira ME, Capellaro JL, dos Reis Marques E, Malavazi I, Perlin D, *et al.* 2004. In vitro evolution of itraconazole resistance in *Aspergillus fumigatus* involves multiple mechanisms of resistance. Antimicrob Agents Chemother. 48:4405–4413.

Dos Santos RAC, Steenwyk JL, Rivero-Menendez O, Mead ME, Silva LP, *et al.* 2020. Genomic and phenotypic heterogeneity of clinical isolates of the human pathogens *Aspergillus fumigatus*, *Aspergillus lentulus*, and *Aspergillus fumigatiaffinis*. Front Genet. 11:459.

Edwards SV. 2009. Is a new and general theory of molecular systematics emerging? Evolution. 63:1–19. doi:10.1111/j.1558-5646.2008.00549.x

Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, *et al.* 2008. Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. PLoS Genet. 4:e1000046.

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol. 17:368–376. doi: 10.1007/BF01734359

Ferling I, Dunn JD, Ferling A, Soldati T, Hillmann F. 2020. Conidial melanin of the human-pathogenic fungus *Aspergillus fumigatus* disrupts cell autonomous defenses in amoebae. mBio. 11: e00862-20. doi:10.1128/mBio.00862-20

Frisvad JC, Larsen TO. 2016. Extrolites of *Aspergillus fumigatus* and other pathogenic species in *Aspergillus* section *Fumigati*. Front Microbiol. 6:1485.

Fuchs BB, O'Brien E, Khoury JBE, Mylonakis E. 2010. Methods for using *Galleria mellonella* as a model host to study fungal pathogenesis. Virulence. 1:475–482.

Furukawa T, van Rhijn N, Fraczek M, Gsaller F, Davies E, *et al.* 2020. The negative cofactor 2 complex is a key regulator of drug resistance in *Aspergillus fumigatus*. Nat Commun. 11:427. doi: 10.1038/s41467-019-14191-1

Gabaldón T, Naranjo-Ortíz MA, Marcet-Houben M. 2016. Evolutionary genomics of yeast pathogens in the Saccharomycotina. FEMS Yeast Res. 16:fow064. doi:10.1093/femsyr/fow064

Gastebois A, Blanc Potard A-B, Gribaldo S, Beau R, Latgé JP, *et al.* 2011. Phylogenetic and functional analysis of *Aspergillus fumigatus* MGTC, a fungal protein homologous to a bacterial virulence factor. Appl Environ Microbiol. 77:4700–4703. doi:10.1128/AEM.00243-11

Gregg K, Kauffman C. 2015. Invasive aspergillosis: epidemiology, clinical aspects, and treatment. Semin Respir Crit Care Med. 36: 662–672. doi:10.1055/s-0035-1562893

Gu X, Fu YX, Li WH. 1995. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. Mol Biol Evol. 12:546–557. doi:10.1093/oxfordjournals.molbev.a040235

Gupta S, Paul K, Kaur S. 2020. Diverse species in the genus *Cryptococcus*: pathogens and their non-pathogenic ancestors. IUBMB Life. 72:2303–2312. doi:10.1002/iub.2377

Guruceaga X, Ezpeleta G, Mayayo E, Sueiro-Olivares M, Abad-Diaz-De-Cerio A, *et al.* 2018. A possible role for fumagillin in cellular damage during host infection by *Aspergillus fumigatus*. Virulence. 9:1548–1561. doi:10.1080/21505594.2018.1526528

Hillmann F, Novohradská S, Mattern DJ, Forberger T, Heinekamp T, *et al.* 2015. Virulence determinants of the human pathogenic fungus *Aspergillus fumigatus* protect against soil amoeba predation. Environ Microbiol. 17:2858–2869.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 35:518–522. doi:10.1093/molbev/msx281

Houbraken J, Weig M, Groß U, Meijer M, Bader O. 2016. *Aspergillus oerlinghausenensis*, a new mould species closely related to *A. fumigatus*. FEMS Microbiol Lett. 363:fnv236.

Houbraken J, Kocsubé S, Visagie CM, Yilmaz N, Wang XC, *et al.* 2020. Classification of Aspergillus, Penicillium, Talaromyces and related genera (Eurotiales): an overview of families, genera, subgenera, sections, series and species. Stud Mycol. 95:5–165.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30:772–780.

Keller NP. 2017. Heterogeneity confounds establishment of "a" model microbial strain. mBio. 8:141.

Kjærbølling I, Vesth TC, Frisvad JC, Nybo JL, Theobald S, *et al.* 2018. Linking secondary metabolites to gene clusters through genome sequencing of six diverse *Aspergillus* species. Proc Natl Acad Sci USA. E753–E761. 115:

Kjærbølling I, Vesth T, Frisvad JC, Nybo JL, Theobald S, *et al.* 2020. A comparative genomics study of 23 *Aspergillus* species from section Flavi. Nat Commun. 11:1106. doi:10.1038/s41467-019-14051-y

Klopfenstein DV, Zhang L, Pedersen BS, Ramírez F, Warwick Vesztrocy A, *et al.* 2018. GOATOOLS: a python library for gene ontology analyses. Sci Rep. 8:10872. doi:10.1038/s41598-018-28948-z

Knowles SL, Mead ME, Silva LP, Raja HA, Steenwyk JL, *et al.* 2020. Gliotoxin, a known virulence factor in the major human pathogen *Aspergillus fumigatus*, is also biosynthesized by its nonpathogenic relative *Aspergillus fischeri*. mBio. 11:e03361-19. doi:10.1128/mBio.03361-19

Kowalski CH, Beattie SR, Fuller KK, McGurk EA, Tang Y-W, *et al.* 2016. Heterogeneity among isolates reveals that fitness in low oxygen correlates with *Aspergillus fumigatus* virulence. mBio. 7:e01515-16.

Kowalski CH, Kerkaert JD, Liu K-W, Bond MC, Hartmann R, *et al.* 2019. Fungal biofilm morphology impacts hypoxia fitness and disease progression. Nat Microbiol. 4:2430–2441. doi:10.1038/s41564-019-0558-7

Kowalski CH, Morelli KA, Stajich JE, Nadell CD, Cramer RA. 2021. A heterogeneously expressed gene family modulates the biofilm architecture and hypoxic growth of *Aspergillus fumigatus*. mBio. 12:e03579-20. doi:10.1128/mBio.03579-20

Latgé J-P, Chamilos G. 2019. *Aspergillus fumigatus* and Aspergillosis in 2019. Clin Microbiol Rev. 33:e00140-18. doi:10.1128/CMR.00140-18

Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. 13:2178–2189. doi:10.1101/gr.1224503

Li F, Wang B, Wang L, Cao B. 2014. Phylogenetic analyses on the diversity of *Aspergillus fumigatus* sensu lato based on five orthologous loci. Mycopathologia. 178:163–176. doi:10.1007/s11046-014-9790-0

Li L, Hsiang T, Li Q, Wang L, Yu Z. 2018. Draft genome sequence of NRRL 5109, an ex-type isolate of *Aspergillus neoellipticus*. Microbiol Resour Announc. 7: doi:10.1128/MRA.01262-18

Lind AL, Smith TD, Saterlee T, Calvo AM, Rokas A. 2016. Regulation of secondary metabolism by the velvet complex is temperature-responsive in aspergillus. G3 (Bethesda). 6:4023–4033.

Lind AL, Wisecaver JH, Lameiras C, Wiemann P, Palmer JM, *et al.* 2017. Drivers of genetic diversity in secondary metabolic gene clusters within a fungal species. PLoS Biol. e2003583.15: doi:10.1371/journal.pbio.2003583

Losada L, Sugui JA, Eckhaus MA, Chang YC, Mounaud S, *et al.* 2015. Genetic analysis using an isogenic mating pair of *Aspergillus fumigatus* identifies azole resistance genes and lack of MAT locus's role in virulence. PLoS Pathog. 11:e1004834. doi:10.1371/journal.ppat.1004834

Magditch DA, Liu T-B, Xue C, Idnurm A. 2012. DNA mutations mediate microevolution between host-adapted forms of the pathogenic fungus *Cryptococcus neoformans*. PLoS Pathog. 8:e1002936. doi:10.1371/journal.ppat.1002936

McDonagh A, Fedorova ND, Crabtree J, Yu Y, Kim S, *et al.* 2008. Sub-telomere directed gene expression during initiation of invasive aspergillosis. PLoS Pathogens. 4:e1000154. doi:10.1371/journal.ppat.1000154

Mead ME, Knowles SL, Raja HA, Beattie SR, Kowalski CH, *et al.* 2019. Characterizing the pathogenic, genomic, and chemical traits of *Aspergillus fischeri*, a close relative of the major human fungal pathogen *Aspergillus fumigatus*. mSphere. 4:e00018-19.

Mondon P, De Champs C, Donadille A, Ambroise-Thomas P, Grillot R. 1996. Variation in virulence of *Aspergillus fumigatus* strains in a murine model of invasive pulmonary aspergillosis. J Med Microbiol. 45:186–191. doi:10.1099/00222615-45-3-186

Moran GP, Coleman DC, Sullivan DJ. 2011. Comparative genomics and the evolution of pathogenicity in human pathogenic fungi. Eukaryot Cell. 10:34–42. doi:10.1128/EC.00242-10

Nascimento AM, Goldman GH, Park S, Marras SAE, Delmas G, *et al.* 2003. Multiple resistance mechanisms among *Aspergillus fumigatus* mutants with high-level resistance to itraconazole. Antimicrob Agents Chemother. 47:1719–1726. doi:10.1128/AAC.47.5.1719-1726.2003

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32:268–274. doi:10.1093/molbev/msu300

Nielsen K, De Obaldia AL, Heitman J. 2007. *Cryptococcus neoformans* mates on pigeon guano: implications for the realized ecological niche and globalization. Eukaryotic Cell. 6:949–959. doi:10.1128/EC.00097-07

Paul S, Diekema D, Moye-Rowley WS. 2013. Contributions of *Aspergillus fumigatus* ATP-binding cassette transporter proteins to drug resistance and virulence. Eukaryotic Cell. 12:1619–1628. doi:10.1128/EC.00171-13

Perlin DS, Rautemaa-Richardson R, Alastruey-Izquierdo A. 2017. The global problem of antifungal resistance: prevalence, mechanisms, and management. Lancet Infect Dis. 17:e383–e392.

Puértolas-Balint F, Rossen JWA, Oliveira dos Santos C, Chlebowicz MA, Raangs E, *et al.* 2019. Revealing virulence potential of clinical and environmental *Aspergillus fumigatus* isolates using whole-genome sequencing. bioRxiv. 1:e00227-38.

R Core Team. 2018. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Radosa S, Ferling I, Sprague JL, Westermann M, Hillmann F. 2019. The different morphologies of yeast and filamentous fungi trigger distinct killing and feeding mechanisms in a fungivorous amoeba. Environ Microbiol. 21:1809–1820. doi:10.1111/1462-2920.14588

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things): *phytools: R package*. Methods Ecol Evol. 3:217–223. doi:10.1111/j.2041-210X.2011.00169.x

Ries LN, Steenwyk JL, de Castro PA, de Lima PBA, Almeida F, *et al.* 2019. Nutritional heterogeneity among *Aspergillus fumigatus* strains has consequences for virulence in a strain- and host-dependent manner. Front Microbiol. 10:1–40.

Rokas A, Williams BL, King N, Carroll SB. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature. 425:798–804. doi:10.1038/nature02053

Rokas A, Mead ME, Steenwyk JL, Oberlies NH, Goldman GH. 2020. Evolving moldy murderers: Aspergillus section Fumigati as a model for studying the repeated evolution of fungal pathogenicity. PLoS Pathogens. 16:e1008315.

Rydholm C, Dyer PS, Lutzoni F. 2007. DNA Sequence characterization and molecular evolution of MAT1 and MAT2 mating-type loci of the self-compatible ascomycete mold *Neosartorya fischeri*. Eukaryotic Cell. 6:868–874. doi:10.1128/EC.00319-06

Seabold S, Perktold J. 2010. statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*.

Seyedmousavi S, Guillot J, Arné P, de Hoog GS, Mouton JW, *et al.* 2015. *Aspergillus* and aspergilloses in wild and domestic animals: a global health concern with parallels to human disease. Med. Myco. 53:765–797. doi:10.1093/mmy/myv067

Sharpton TJ, Stajich JE, Rounsley SD, Gardner MJ, Wortman JR, *et al.* 2009. Comparative genomic analyses of the human fungal pathogens Coccidioides and their relatives. Genome Research. 19: 1722–1731. doi:10.1101/gr.087551.108

Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics. 19:ii215–ii225. doi:10.1093/bioinformatics/btg1080

Steenwyk J, Rokas A. 2017. Extensive copy number variation in fermentation-related genes among *Saccharomyces cerevisiae* wine strains. G3 (Bethesda). 7:1475–1486.

Steenwyk JL, Opulente DA, Kominek J, Shen X-X, Zhou X, *et al.* 2019a. Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts. PLoS Biol. 17:e3000255. doi: 10.1371/journal.pbio.3000255

Steenwyk JL, Shen X-X, Lind AL, Goldman GH, Rokas A. 2019b. A robust phylogenomic time tree for biotechnologically and medically important fungi in the genera Aspergillus and Penicillium. mBio. 10: doi:10.1128/mBio.00925-19

Steenwyk JL, Lind AL, Ries LNA, dos Reis TF, Silva LP, *et al.* 2020a. Pathogenic allodiploid hybrids of aspergillus fungi. Curr Biol. 30: 2495–2507.e7.

Steenwyk JL, Mead ME, Knowles SL, Raja HA, Roberts CD, *et al.* 2020b. Variation among biosynthetic gene clusters, secondary metabolite profiles, and cards of virulence across *Aspergillus* species. Genetics. 216:481–497.

Steenwyk JL, Mead ME, de Castro PA, Valero C, Damasio A, *et al.* 2020c. Genomic and phenotypic analysis of COVID-19-associated pulmonary aspergillosis isolates of *Aspergillus fumigatus*. [Preprint]. 2020 Nov 6: 2020.11.06.371971. doi: 10.1101/2020.11.06.371971. PMID: 33173866; PMCID: PMC7654854.

Steenwyk JL, Buida TJ, Labella AL, Li Y, Shen X-X, *et al.* 2021. PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. Bioinformatics. btab096. doi:10.1093/bioinformatics/btab096

Sugui JA, Peterson SW, Figat A, Hansen B, Samson RA, *et al.* 2014a. Genetic relatedness versus biological compatibility between *Aspergillus fumigatus* and related species. J Clin Microbiol. 52:3707–3721.

Sugui JA, Kwon-Chung KJ, Juvvadi PR, Latgé J-P, Steinbach WJ. 2014b. *Aspergillus fumigatus* and related species. Cold Spring Harb Perspect Med. 5:a019786.

Sun S, Coelho MA, David-Palma M, Priest SJ, Heitman J. 2019. The evolution of sexual reproduction and the mating-type locus: links to pathogenesis of *Cryptococcus* human pathogenic fungi. Annu Rev Genet. 53:417–444. doi:10.1146/annurev-genet-120116-024755

Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lect Math Life Sci. 17:57–86.

Taylor JW. 2015. Evolutionary perspectives on human fungal pathogens. Cold Spring Harb Perspect Med. 5:a019588.doi: 10.1101/cshperspect.a019588

Tekaia F, Latgé J-P. 2005. *Aspergillus fumigatus*: saprophyte or pathogen? Curr Opin Microbiol. 8:385–392.

The Gene Ontology Consortium. 2021. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 49:D325–D334. doi: 10.1093/nar/gkaa1113

Urban M, Cuzick A, Seager J, Wood V, Rutherford K, *et al.* 2020. PHI-base: the pathogen–host interactions database. Nucleic Acids Res. 48:D613–D620. doi:10.1093/nar/gkz904

Urquhart AS, Hu J, Chooi Y-H, Idnurm A. 2019. The fungal gene cluster for biosynthesis of the antibacterial agent viriditoxin. Fungal Biol Biotechnol. 6:9.doi:10.1186/s40694-019-0072-y

van Dongen S. 2000. Graph Clustering by Flow Simulation, PhD thesis, University of Utrecht, May 2000.

Waterhouse RM, Tegenfeldt F, Li J, Zdobnov EM, Kriventseva EV. 2013. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. Nucleic Acids Res. 41:D358–D365. doi:10.1093/nar/gks1116

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, *et al.* 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 35:543–548. doi: 10.1093/molbev/msx319

Wiedner SD, Ansong C, Webb-Robertson B-J, Pederson LM, Fortuin S, *et al.* 2013. Disparate proteome responses of pathogenic and non-pathogenic aspergilli to human serum measured by activity--based protein profiling (ABPP). Mol Cell Proteomics.12:1791–1805.

Wiemann P, Guo C-J, Palmer JM, Sekonyela R, Wang CCC, *et al.* 2013. Prototype of an intertwined secondary-metabolite supercluster. Proc Natl Acad Sci USA. 110:17065–17070.

Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol. 39:306–314. doi:10.1007/BF00160154

Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol Evol. 15:496–503. doi: 10.1016/s0169-5347(00)01994-7

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591. doi:10.1093/molbev/msm088

Yin W-B, Ruan H-L, Westrich L, Grundmann A, Li S-M. 2007. CdpNPT, an N-prenyltransferase from *Aspergillus fumigatus*: overproduction, purification and biochemical characterisation. ChemBioChem. 8:1154–1161. doi:10.1002/cbic.200700079

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics. 19:153. doi: 10.1186/s12859-018-2129-y

Zhao C, Fraczek MG, Dineen L, Lebedinec R, Macheleidt J, *et al.* 2019. High-throughput gene replacement in *Aspergillus fumigatus*. Curr Protoc Microbiol. 54:e88.doi:10.1002/cpmc.88

Communicating editor: A. Mitchell