

Phylogenomics reveals extensive misidentification of fungal strains from the genus *Aspergillus*

Jacob L. Steenwyk^{1,2,3}, Charu Balamurugan^{2,3}, Huzefa A. Raja⁴, Carla Gonçalves^{2,3}, Ningxiao Li^{5,6}, Frank Martin⁶, Judith Berman⁷, Nicholas H. Oberlies⁴, John G. Gibbons^{8,9,10}, Gustavo H. Goldman¹¹, David M. Geiser¹², David S. Hibbett¹³ & Antonis Rokas^{2,3,14,*}

¹ Howards Hughes Medical Institute and the Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA

² Vanderbilt University, Department of Biological Sciences, Nashville, TN, United States of America

³ Evolutionary Studies Initiative, Vanderbilt University, Nashville, TN, USA

⁴ Department of Chemistry & Biochemistry, University of North Carolina at Greensboro, Greensboro, NC, USA

⁵ Department of Plant Pathology, University of California, Davis, Davis, CA

⁶ USDA-ARS, Salinas, CA

⁷ Shmunis School of Biomedicine and Cancer Research, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv, Israel

⁸ Department of Food Science, University of Massachusetts, Amherst, MA, USA

⁹ Molecular and Cellular Biology Graduate Program, University of Massachusetts, Amherst, MA, USA

¹⁰ Organismic & Evolutionary Biology Graduate Program, University of Massachusetts, Amherst, MA, USA

¹¹ Faculdade de Ciencias Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, São Paulo, Brazil

¹² Department of Plant Pathology and Environmental Microbiology, Penn State University, University Park, PA USA 16802

¹³ Biology Department, Clark University, Worcester, MA, USA

¹⁴ Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnengasse, Heidelberg, Germany

*Correspondence should be addressed to: antonis.rokas@vanderbilt.edu

ORCiDs

J.L. Steenwyk: 0000-0002-8436-595X

A. Rokas: 0000-0002-7248-6551

H.A. Raja: 0000-0002-0824-9463

N. Li: 0000-0002-7847-3567

N.H. Oberlies: 0000-0002-0354-8464

J. Berman 0000-0002-8577-0084000

D.S. Hibbett: 0000-0002-9145-3165

Running title: Phylogenomics of species and populations of *Aspergillus*

Keywords: Taxogenomics, pathogenicity, pathogen, virulence, plant pathogen, genomics, taxonomy, Aspergillaceae, *Penicillium*

Abstract

Modern taxonomic classification is often based on phylogenetic analyses of a few molecular markers, although single-gene studies are still common. However, the use of one or few molecular markers can lead to inaccurate inferences of species history and errors in classification. Here, we leverage genome-scale molecular phylogenetics (phylogenomics) of species and populations to reconstruct evolutionary relationships in a dense dataset of 711 fungal genomes from the biomedically and technologically important genus *Aspergillus*. To do so, we generated a novel set of 1,362 high-quality molecular markers specific for *Aspergillus* and provide profile Hidden Markov Models for each, facilitating others to use these molecular markers. Examination of the resulting genome-scale phylogeny: (1) helped resolve ongoing taxonomic controversies and identified new ones; (2) revealed extensive strain misidentification, underscoring the importance of population-level sampling in species classification; and (3) identified novel lineages that may shed light on the early evolution of an important genus. These findings suggest that phylogenomics of species and populations can facilitate accurate taxonomic classifications and reconstructions of the tree of life.

Introduction

Accurate identification of fungal species underpins mycological research. Modern fungal taxonomy is based on diverse phenotypic traits and single or few molecular markers for species determination (Schoch et al., 2012; Houbraken et al., 2021). In addition, an increasing number of novel taxa, known from sequence data alone, are being discovered using environmental barcoding techniques (Lücking et al., 2021). For fungi, six molecular markers are used as universal barcodes: the ribosomal large and small subunit; ribosomal DNA array internal transcribed spacer region; the RNA polymerase II large subunit and core subunits; and minichromosome maintenance complex component 7 (Schoch et al., 2012; Raja et al., 2017). Two additional commonly used loci are transcription enhancer factor 1 α and β -tubulin (Bruns et al., 1991; Hibbett et al., 2016; Raja et al., 2017; Lücking et al., 2021). Secondary barcodes are also often used to characterize species complexes (Bradshaw et al., 2022)—such as calmodulin for *Aspergillus* and other species (Susca et al., 2020)—owing to the power of more loci in phylogenetic reconstructions.

Typically, the sequences of one or more barcode loci are obtained from a strain of interest. Next, orthologous sequences are inferred based on sequence similarity from databases such as the National Center for Biotechnology Information; then, software is used to infer phylogenetic trees for each sequence, representing a hypothesis of the evolutionary history of these species (Raja et al., 2017). Despite the potential for genome-scale data to facilitate fungal taxonomy, current practices typically do not rely on whole-genome data, in part because of sparse genome sequences available across the fungal tree of life (Xu, 2020; Aime et al., 2021; Arias et al., 2021; Fountain et al., 2021; Kanamasa, 2021; Kück and Dahlmann, 2021; Li and Yu, 2021; Ruiz and Radwan, 2021; Stevenson and Stamps, 2021; Venkateswaran et al., 2021; Yin et al., 2021). Here, we tested the hypothesis that phylogenomic approaches that leverage whole genome sequences will improve the accuracy of phylogenetic reconstructions that have been based on sequences from a small number of loci (Rokas et al., 2003; Jarvis et al., 2014).

Fungi from the genus *Aspergillus* are of medical, agricultural, and biotechnological significance. *Aspergillus fumigatus* and *Aspergillus flavus* are pathogens, allergens, and produce mycotoxins (Nierman et al., 2005; Hedayati et al., 2007). *Aspergillus niger* is an industrial workhorse and *Aspergillus oryzae* is used in the production of fermented foods like soy sauce and sake (Gibbons et al., 2012; Cairns et al., 2018). Accurate identification of *Aspergillus* fungi with barcode loci is often challenging (Houbraken et al., 2021; Li and Yu, 2021; Stevenson and

Stamps, 2021; Venkateswaran et al., 2021; Yin et al., 2021), but important because even closely related species can differ in drug resistance profiles and ability to cause disease (dos Santos et al., 2020; Steenwyk et al., 2020c). For example, clinical strains of *Aspergillus nomius* and *Aspergillus tamarii* have been misidentified as *Aspergillus flavus* (Tam et al., 2014). In the clinic, inaccurate species determination could lead to misguided disease management strategies due to differences in intrinsic drug resistance levels between species. For example, *A. latus*, commonly misidentified as *A. nidulans*, is more resistant to the antifungal drug caspofungin than *A. nidulans* (Steenwyk et al., 2020b).

Species are a taxonomic rank and serve as the foundation for comparative studies (De Queiroz, 2007; Houbraken et al., 2014; Xu, 2020; Aime et al., 2021). Several different species concepts have been proposed over the years (De Queiroz, 2007). Due to the microscopic nature of many fungi and lack of known sexual cycles for some species (Dyer and O’Gorman, 2012), species delimitation in the Kingdom Fungi has relied, in addition to cultural growth and micromorphological data, on molecular phylogenetics and the adoption of universal molecular barcodes (Schoch et al., 2012). Among *Aspergillus* fungi, multi-locus molecular phylogenetic approaches have become the predominant method (Samson et al., 2007, 2014; Houbraken and Samson, 2011; Houbraken et al., 2014). However, there is no current consensus for barcode similarity to designate a new species of fungi; for example, calmodulin gene sequences of *Aspergillus labruscus* and *Aspergillus oerlinghausenensis*—two recently described species of *Aspergillus*—share 85% and 97.3% sequence similarity to their closest relatives, *Aspergillus homomorphus* and *Aspergillus fumigatus*, respectively (Houbraken et al., 2016; Fungaro et al., 2017).

Reconstructing deeper evolutionary relationships from a few molecular markers can also be challenging. For example, divergences among sections—a secondary taxonomic rank above the species and below the genus ranks—have been debated. For example, the sections *Nigri*, *Ochraceorosei*, *Flavi*, *Circumdati*, *Candidi*, and *Terrei* were inferred to be monophyletic based on analyses of four loci from 81 taxa (Pitt and Taylor, 2014; Taylor et al., 2016) but topology tests using a 1,668-gene matrix from a different 81-taxon dataset rejected monophyly of these lineages (Steenwyk et al., 2019). Accurate reconstructions of the *Aspergillus* phylogeny will enable more accurate strain identification and will facilitate our understanding of how biomedically and technologically relevant traits—such as antimicrobial resistance—evolved.

Here, we present a dense phylogenomic tree inferred from a 1,362-gene matrix from 711 *Aspergillus* genome sequences spanning 106 species and population data for 36 species, more than doubling the number of species analyzed in previous genome-scale studies (de Vries et al., 2017; Steenwyk et al., 2019; Kjærboelling et al., 2020) and capturing roughly one quarter of all known species in the genus (Balajee, 2009; Sayers et al., 2009). The new phylogeny yields three key insights by: (1) demonstrating that population-level sequencing can facilitate strain classification; (2) resolving taxonomic controversies and identifying new ones; and (3) identifying a previously uncharacterized *Aspergillus* lineage that holds promise for elucidating the early evolution of this biomedically and technologically important genus.

Results and Discussion

A phylogenomic tree of *Aspergillus*

The evolutionary history of 725 genomes (711 *Aspergillus* genomes; 14 outgroup *Penicillium* genomes) was reconstructed using maximum likelihood analysis of a 1,362-gene data matrix with 6,378,237 nucleotide sites (Fig 1; Fig S1). The 725 genomes represent publicly available whole genome assemblies available through the National Center for Biotechnology Information (NCBI, <https://www.ncbi.nlm.nih.gov/>) that also passed quality-control measures (see *Methods*; Table S1 & S2). Based on the metadata provided to NCBI, the dataset includes 115 *Aspergillus* and 14 *Penicillium* species. The genomes of two or more strains were available for 36 *Aspergillus* species, but the depth of strain sampling varied (Table S3). Sampling was densest for *Aspergillus fumigatus* (N = 275), *Aspergillus flavus* (N = 105), *Aspergillus oryzae* (N = 97), and *Aspergillus niger* (N = 24). Sixteen species had genome sequences from two representative strains available. Twelve strains were of unknown species but were reported to be from the genus *Aspergillus*.

The dataset spanned 17 *Aspergillus* sections. Higher-order relationships among sections were examined using both concatenation- and coalescence-based tree inference approaches (Rokas et al., 2003; Mirarab et al., 2014). The phylogenies inferred with both approaches were highly congruent, differing by only one bipartition (Fig 2). Both approaches revealed the presence of a previously undescribed lineage, herein termed *Clade A* and represented by *Aspergillus* sp. MCCF 102 (Lekshmi et al., 2020). This lineage, by virtue of its placement as sister to all other sampled *Aspergillus* genomes, holds promise for understanding the early evolution of the genus *Aspergillus*.

Phylogenomics sheds new light on *Aspergillus* strain identification

Phylogenomic analyses of the 711 strains reveal that half of the 36 species with multiple strains were not monophyletic (Table S4). Although monophyly is not required for species delimitation under certain species concepts (Rieseberg and Brouillet, 1994; Rieppel, 2010), this observation was surprising and suggests that cryptic speciation and / or species misidentification may be rampant. Here, we highlight six cases that illustrate this issue.

The phylogenetic placement of *Aspergillus neoellipticus* has been debated and bears on disease management strategies. Some studies suggest that *A. neoellipticus* is distinct from the major human pathogen *Aspergillus fumigatus*, based on the analysis of five loci (Li et al., 2014). Other studies suggested that *A. neoellipticus* is conspecific to *A. fumigatus* based on partial single-gene sequences and random amplified polymorphic DNA (Hong et al., 2005) (Peterson, 2008; Li et al., 2014, 2018). Phylogenomic analyses of species in section *Fumigati* (including four strains of *A. fumigatus* and one strain of *A. neoellipticus*) failed to resolve this ongoing debate, due to inferring *A. neoellipticus* as sister to *A. fumigatus* (Mead et al., 2021). The combined use of genome-scale data and extensive population-level sampling of 275 *A. fumigatus* strains in the analyses found that *A. neoellipticus* NRRL5109 is nested within a clade of 16 isolates of *A. fumigatus* (Fig 3a); the remaining 261 *A. fumigatus* isolates form the sister lineage.

These results can be interpreted in two ways. Either *A. neoellipticus* NRRL5109 is an isolate of *A. fumigatus* or the clade formed by the 16 *A. fumigatus* strains and *A. neoellipticus* NRRL5109 is, in fact, *A. neoellipticus*, a species sister to *A. fumigatus*. If the latter, this would suggest that the 16 *A. fumigatus* isolates that form a clade with *A. neoellipticus* should be renamed to *A. neoellipticus*. Either way, these results underscore the importance of dense population-level representation for accurate strain identification.

A similar situation, but with much lower levels of sampling, exists for two other pairs of species: *Aspergillus terreus* and *Aspergillus pseudoterreus* (Fig 3b) and *A. nomiae* and *A. pseudonomiae* (Fig 3c). For *A. terreus* strains TN-484 and IFO 6365, our data corroborate recent findings that these are misidentified and should be *A. pseudoterreus*, suggesting misidentified information has persisted in databases (Houbraken et al., 2021).

Species from section *Nigri* are similar both in their morphology and in their molecular barcode sequences (Susca et al., 2016). *Aspergillus niger*, a species from the section *Nigri* (Coutinho et al., 2006), has been considered a prominent pathogen of sisal (*Agave sisalana*), an industrial crop used in the textile industry (Duarte et al., 2018). However, recent molecular phylogenetic analysis suggests that the main etiological agent of sisal bole rot disease is *Aspergillus welwitschiae*, rather than *A. niger* (Duarte et al., 2018). Examination of the evolutionary history of *A. niger* and *A. welwitschiae*, using phylogenomic analyses that sampled multiple strains from both species, identified four putative cases of strain misidentification (Fig 3d). Three strains (F3 4F1 F, L14, and F3 4F2 F), previously identified as *A. niger*, are conspecific with the majority of *A. welwitschiae* strains, suggesting that they are, indeed, *A. welwitschiae* and not *A. niger*. The fourth strain, CCMB 674, which was previously identified as *A. welwitschiae*, is sister to a lineage of 19 *A. niger* strains; therefore, isolate CCMB 674 is likely *A. niger* and not *A. welwitschiae*. These findings suggest that strain misidentification confounds studies of sisal bole rot, which could have major economic ramifications.

Another instance of putative misidentification concerns *Aspergillus versicolor* ATCC 9577, which is conspecific with *Aspergillus sydowii* strains, suggesting that strain ATCC 9577 is in fact *A. sydowii*. In contrast, the type strain of *A. sydowii*, CBS 583.65^T, is closely related to *A. amoenus*, an observation that is corroborated by previous taxonomic studies (Jurjevic et al., 2012). These findings suggest that multiple strains from multiple species may be misidentified. Furthermore, ATCC 9577 and CBS 583.65 are reportedly synonymous strain names (Sklenář et al., 2022); the fact that there are two different genome sequences available is suggestive of contamination or erroneous metadata for at least one of these strains. These findings suggest that the entire clade may benefit from taxonomic revision and closer scrutiny of strain identity. To this point, a recent analysis of a five-gene, 213-taxon dataset proposed species in the series *Versicolores* be reduced from 17 species to four, citing intraspecific variation as a driver of over splitting of species boundaries by taxonomists (Sklenář et al., 2022). Evolutionary relationships under this new analysis differ slightly from our genome-scale phylogeny (Fig S2). Additional genome sequences of species and strains may shed light on the evolutionary history of section *Versicolores*.

Accurate classification of *A. flavus*, *A. oryzae*, and *Aspergillus parasiticus* is important because of biomedical and food safety concerns (Fig 3f). *Aspergillus flavus* is a human pathogen, post-harvest food pathogen, and mycotoxin producer (Gourama and Bullerman, 1995; Geiser et al.,

1998; Hedayati et al., 2007). *Aspergillus oryzae* is the domesticate of *A. flavus* used for sake production (Gibbons et al., 2012). Successive short branch lengths and non-monophyly of *A. flavus* and *A. oryzae* strains may suggest multiple domestication events or the introduction of domesticated isolates into the wild. Another interpretation is that the two species are distinct ecotypes (or populations), rather than distinct species. Notably, *A. oryzae* strains are known to produce fewer mycotoxins than *A. flavus* (Gibbons et al., 2012), which may be a diagnostic signature (i.e., phenotype) of an ecotype appropriate for food production. Interestingly, the aflatoxigenic postharvest pathogen *A. parasiticus* is conspecific with *A. flavus* strains, and the two *A. parasiticus* strains do not form a monophyletic group; strain NRRL 2999 is misidentified as *A. parasiticus* but is, in fact, *A. flavus* (Houbraken et al., 2021). Five isolates of *A. flavus* are sister to *A. minisclerotigenes* and more distantly related to other *A. flavus* isolates; our genome-scale analyses are consistent with inferences based on a recent examination of individual barcode loci (Houbraken et al., 2021) in suggesting that the five *A. flavus* isolates are misidentified. Lastly, five isolates with unknown species designations were determined, whereas six remain unknown (Table S5).

A roadmap for studies of *Aspergillus*

We present a comprehensive genome-scale phylogeny of *Aspergillus* (Fig 1 and 4). Our results underscore the need for further research into *Aspergillus* species delimitation and that strain misidentification may be a more common problem among publicly available data than previously appreciated (Houbraken et al., 2021). Strain misidentification can be reduced by two factors: (1) genome-scale data, which are less prone to errors in phylogenetic inference compared to single or a few loci (Rokas et al., 2003; Jarvis et al., 2014; Steenwyk et al., 2019) and (2) population-level sampling from diverse environmental niches and geographic locations. Combined, these two factors facilitated identifying cases where strains may represent new species or should be conspecific with another (such as *A. neoellipticus*) or uncovered instances of more rampant strain misidentification (such as the case of *A. niger* and *A. welwitschiae* and species in section *Flavi*). Lastly, our analyses identified a new lineage (*Aspergillus* sp. MCCC 102) that is sister to all other *Aspergillus* strains whose genomes have been sequenced and may hold clues to the early evolution of *Aspergillus* species. To facilitate others using our findings, we summarize higher-order and species relationships among *Aspergillus* species (Fig 4).

Resolving these issues and facilitating future identification of strains and species requires novel strategies and prioritizes analysis of type strains. Similar to databases of internal transcribed

spacer regions of fungi that help facilitate isolate and strain classification (Nilsson et al., 2019), genome-scale resources can be utilized for strain classification, an approach that is of growing interest among budding yeast biologists (Libkind et al., 2020), but yet to be well adopted for other fungal lineages. To facilitate strain identification among *Aspergillus* species, we have produced profile Hidden Markov Models of the 1,362 molecular markers used in the present study (see *Data Availability*). Adding newly sequenced strains to our proposed phylogenetic tree of *Aspergillus* may be a helpful step toward accurate strain identification. Furthermore, additional research is needed to identify and characterize cryptic species, that is organisms that are morphologically highly similar to known species but genetically and physiologically distinct (Geiser et al., 1998; Alastruey-Izquierdo et al., 2014). Accurate strain identification and elucidation of species boundaries will greatly benefit from increased genome sequencing of type strains.

Methods

Genome data acquisition and quality control

All publicly available *Aspergillus* genomes (N = 717) were downloaded from NCBI (National Center for Biotechnology Information; <https://www.ncbi.nlm.nih.gov/>; date accessed: January 9th, 2022). Publicly available genome annotations were also downloaded. For genomes without available annotations, gene boundaries were predicted using AUGUSTUS, v3.3.2 (Stanke and Waack, 2003), with the “species” parameter set to “*aspergillus_nidulans*.” To determine if the genomes were suitable to phylogenomic analyses, gene prediction completeness was examined using BUSCO, v4.0.4 (Waterhouse et al., 2018), and the Eurotiales database of 4,191 near-universally single-copy orthologs (or BUSCO genes) from OrthoDB, v10 (Kriventseva et al., 2019). Six genomes with less than 75% single-copy complete BUSCO genes were removed, resulting in 711 *Aspergillus* genomes. The resulting sets of gene predictions were highly complete (mean $\pm$ standard deviation: $95.74 \pm 2.25\%$). For outgroup taxa, 14 *Penicillium* genomes and annotations were downloaded from NCBI. The completeness of *Penicillium* gene predictions were assessed using the same protocol and were highly complete (mean $\pm$ standard deviation: $94.73 \pm 4.04\%$). The final dataset had 725 genomes.

Single-copy orthologous gene identification

Phylogenomics often relies on single-copy orthologous genes. OrthoFinder, v.2.3.8 (Emms and Kelly, 2019), was used to identify single-copy orthologous genes by clustering protein sequences into groups of orthologs. Clustering sequences was based on protein sequence

similarity and calculated using DIAMOND, v2.0.13.151 (Buchfink et al., 2021). To reduce computation time and memory, orthology predictions were conducted among 40 representative species that span the diversity of *Aspergillus* species (Table S6) (Steenwyk et al., 2019). The impact of 41 different inflation parameter settings (one through five with a step of 0.1) on the number of single-copy orthologs identified was examined. The inflation parameter that resulted in the highest number of single-copy orthologs was 3.6. The resulting 7,882 single-copy orthologs with at least 50% occupancy (N = 20) were used for downstream analysis.

To identify orthologs in the full 725-genome dataset, sequence similarity searches were conducted in each proteome. To do so, the 7,882 single-copy orthologs were aligned using MAFFT, v7.402 (Katoh and Standley, 2013), with the auto parameter. Profile Hidden Markov Models (HMMs) were then built for each alignment using the hmmbuild function in HMMER, v3.1b2 (Eddy, 2011). The resulting HMMs were used to identify single-copy orthologs in the 725 proteomes using orthofisher, v.1.0.3 (Steenwyk and Rokas, 2021), and a bitscore fraction threshold of 0.95.

To generate single-gene phylogenies, the protein sequences of the 7,882 single-copy orthologs, identified using orthofisher (Steenwyk and Rokas, 2021), were aligned using MAFFT as described above (Katoh and Standley, 2013). The corresponding nucleotide sequences were threaded onto the protein alignment using the thread_dna function in PhyKIT, v1.11.12 (Steenwyk et al., 2021) and trimmed using ClipKIT, v1.3.0 (Steenwyk et al., 2020a). Excessive trimming of multiple sequence alignments worsens single-gene phylogenetic inference (Tan et al., 2015; Steenwyk et al., 2020a); thus, multiple sequence alignments wherein 40% or more of the original alignment length was maintained after trimming were retained resulting in 4,300 single-copy orthologs. The evolutionary histories of the 4,300 single-copy orthologs were inferred using IQ-TREE 2 (Minh et al., 2020, 2). The best-fitting substitution model was selected using ModelFinder (Kalyaanamoorthy et al., 2017). To remove potential instances of hidden paralogy, the monophyly of the five well-established lineages was examined using PhyKIT (Steenwyk et al., 2021). Specifically, the single-gene phylogenies were examined for the monophyly of five well established lineages: sections *Flavi* (N = 246), *Fumigati* (N = 316), *Nidulantes* (N = 12), and *Versicolores* (N = 7) as well as the outgroup lineage of 14 *Penicillium* species. Genes wherein one or more of the five lineages were not monophyletic were removed, resulting in a final set of 1,362 single-copy orthologous genes. The average occupancy for each single-copy ortholog was 0.98 ± 0.07 .

Phylogenomic tree inference

The final set of single-copy orthologs was used to infer the evolutionary history of the 725 species and strains. Specifically, the 1,362 single-copy orthologs were concatenated into a single matrix using the `create_concat` function in PhyKIT (Steenwyk et al., 2021). The resulting supermatrix had 6,378,237 sites (2,846,432 parsimony informative sites). Alignment length and number of parsimony informative sites were calculated using BioKIT, v0.1.2 (Steenwyk et al., 2022). IQ-TREE 2 was used for tree inference (Minh et al., 2020, 2). The best fitting substitution model was selected using ModelFinder (Kalyaanamoorthy et al., 2017). Bipartition support was assessed using 1,000 ultrafast bootstrap approximations (Hoang et al., 2018). The total central processing unit (or CPU) hours required for this analysis was 2,763 (approximately 4 months). An additional coalescence-based consensus tree approach was used to infer the evolutionary history among species and strains. Specifically, the 1,362 single-copy orthologous gene phylogenies were used as input to ASTRAL, v5.6.3 (Zhang et al., 2018). Phylogenies were visualized using iTOL, v6 (Letunic and Bork, 2019).

Data availability

Results and data presented in this study will be available from figshare upon manuscript acceptance (doi: 10.6084/m9.figshare.21382131). The following link is provided for review purposes only <https://figshare.com/s/6b91c4f3f575c0ef9883>.

Acknowledgements

We thank the Rokas lab for helpful discussion and comments.

Conflicts of interest

JLS is a scientific consultant for Latch AI Inc. AR is a scientific consultant for LifeMine Therapeutics, Inc.

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. Research in AR's lab is supported by grants from the National Science Foundation (DEB-1442113 and DEB-2110404), the National Institutes of Health/National Institute of Allergy and Infectious Diseases (R56 AI146096 and R01 AI153356), and the Burroughs Wellcome Fund. Research in DSH's lab is supported by National

Science Foundation (DEB-1456588). Research in JGG's lab is supported by the National Science Foundation (1942681). AR acknowledges support from a Klaus Tschira Guest Professorship from the Heidelberg Institute for Theoretical Studies and from a Visiting Research Fellowship from Merton College of the University of Oxford. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Aime, M. C., Miller, A. N., Aoki, T., Bensc, K., Cai, L., Crous, P. W., et al. (2021). How to publish a new fungal species, or name, version 3.0. *IMA Fungus* 12, 11. doi: 10.1186/s43008-021-00063-1.
- Alastruey-Izquierdo, A., Alcazar-Fuoli, L., and Cuenca-Estrella, M. (2014). Antifungal Susceptibility Profile of Cryptic Species of *Aspergillus*. *Mycopathologia* 178, 427–433. doi: 10.1007/s11046-014-9775-z.
- Arias, R. S., Ormer, V. A., Martinez-Castillo, J., and Sobolev, V. S. (2021). *Aspergillus* Section Flavi, Need for a Robust Taxonomy. *Microbiol Resour Announc* 10, e00784-21. doi: 10.1128/MRA.00784-21.
- Balajee, S. A. (2009). *Aspergillus terreus* complex. *Med Mycol* 47, S42–S46. doi: 10.1080/13693780802562092.
- Bradshaw, M. J., Guan, G.-X., Nokes, L., Braun, U., Liu, S.-Y., and Pfister, D. H. (2022). Secondary DNA Barcodes (CAM, GAPDH, GS, and RpB2) to Characterize Species Complexes and Strengthen the Powdery Mildew Phylogeny. *Front. Ecol. Evol.* 10, 918908. doi: 10.3389/fevo.2022.918908.
- Bruns, T. D., White, T. J., and Taylor, J. W. (1991). Fungal Molecular Systematics. *Annu. Rev. Ecol. Syst.* 22, 525–564. doi: 10.1146/annurev.es.22.110191.002521.
- Buchfink, B., Reuter, K., and Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods* 18, 366–368. doi: 10.1038/s41592-021-01101-x.
- Cairns, T. C., Nai, C., and Meyer, V. (2018). How a fungus shapes biotechnology: 100 years of *Aspergillus niger* research. *Fungal Biol Biotechnol* 5, 13. doi: 10.1186/s40694-018-0054-5.
- Coutinho, W. M., Suassuna, N. D., Luz, C. M., Suinaga, F. A., and Silva, O. R. R. F. (2006). Bole rot of sisal caused by *Aspergillus niger* in Brazil. *Fitopatol. bras.* 31, 605–605. doi: 10.1590/S0100-41582006000600014.
- De Queiroz, K. (2007). Species Concepts and Species Delimitation. *Systematic Biology* 56, 879–886. doi: 10.1080/10635150701701083.
- de Vries, R. P., Riley, R., Wiebenga, A., Aguilar-Osorio, G., Amillis, S., Uchima, C. A., et al. (2017). Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*. *Genome Biology* 18, 28. doi: 10.1186/s13059-017-1151-0.
- dos Santos, R. A. C., Steenwyk, J. L., Rivero-Menendez, O., Mead, M. E., Silva, L. P., Bastos, R. W., 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. doi: 10.3389/fgene.2020.00459.

- Duarte, E. A. A., Damasceno, C. L., de Oliveira, T. A. S., Barbosa, L. de O., Martins, F. M., de Queiroz Silva, J. R., et al. (2018). Putting the Mess in Order: *Aspergillus welwitschiae* (and Not *A. niger*) Is the Etiological Agent of Sisal Bole Rot Disease in Brazil. *Front. Microbiol.* 9, 1227. doi: 10.3389/fmicb.2018.01227.
- Dyer, P. S., and O’Gorman, C. M. (2012). Sexual development and cryptic sexuality in fungi: insights from *Aspergillus* species. *FEMS Microbiol Rev* 36, 165–192. doi: 10.1111/j.1574-6976.2011.00308.x.
- Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS Computational Biology* 7, e1002195. doi: 10.1371/journal.pcbi.1002195.
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* 20, 238. doi: 10.1186/s13059-019-1832-y.
- Fountain, J. C., Clevenger, J. P., Vaughn, J. N., and Guo, B. (2021). Lessons Learned: the Importance of Biological Curation. *Microbiol Resour Announc* 10, e00473-21. doi: 10.1128/MRA.00473-21.
- Fungaro, M. H. P., Ferranti, L. S., Massi, F. P., da Silva, J. J., Sartori, D., Taniwaki, M. H., et al. (2017). *Aspergillus labruscus* sp. nov., a new species of *Aspergillus* section Nigri discovered in Brazil. *Sci Rep* 7, 6203. doi: 10.1038/s41598-017-06589-y.
- Geiser, D. M., Pitt, J. I., and Taylor, J. W. (1998). Cryptic speciation and recombination in the aflatoxin-producing fungus *Aspergillus flavus*. *Proceedings of the National Academy of Sciences* 95, 388–393. doi: 10.1073/pnas.95.1.388.
- Gibbons, J. G., Salichos, L., Slot, J. C., Rinker, D. C., McGary, K. L., King, J. G., et al. (2012). The Evolutionary Imprint of Domestication on Genome Variation and Function of the Filamentous Fungus *Aspergillus oryzae*. *Current Biology* 22, 1403–1409. doi: 10.1016/j.cub.2012.05.033.
- Gourama, H., and Bullerman, L. B. (1995). *Aspergillus flavus* and *Aspergillus parasiticus*: Aflatoxigenic Fungi of Concern in Foods and Feeds†: A Review. *Journal of Food Protection* 58, 1395–1404. doi: 10.4315/0362-028X-58.12.1395.
- Hedayati, M. T., Pasqualotto, A. C., Warn, P. A., Bowyer, P., and Denning, D. W. (2007). *Aspergillus flavus*: human pathogen, allergen and mycotoxin producer. *Microbiology* 153, 1677–1692. doi: 10.1099/mic.0.2007/007641-0.
- Hibbett, D., Abarenkov, K., Kõljalg, U., Öpik, M., Chai, B., Cole, J., et al. (2016). Sequence-based classification and identification of Fungi. *Mycologia* 108, 1049–1068. doi: 10.3852/16-130.
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* 35, 518–522. doi: 10.1093/molbev/msx281.

- Hong, S.-B., Go, S.-J., Shin, H.-D., Frisvad, J. C., and Samson, R. A. (2005). Polyphasic taxonomy of *Aspergillus fumigatus* and related species. *Mycologia* 97, 1316–1329. doi: 10.1080/15572536.2006.11832738.
- Houbraken, J., de Vries, R. P., and Samson, R. A. (2014). “Modern Taxonomy of Biotechnologically Important *Aspergillus* and *Penicillium* Species,” in *Advances in Applied Microbiology*, 199–249. doi: 10.1016/B978-0-12-800262-9.00004-4.
- Houbraken, J., and Samson, R. A. (2011). Phylogeny of *Penicillium* and the segregation of Trichocomaceae into three families. *Studies in Mycology* 70, 1–51. doi: 10.3114/sim.2011.70.01.
- Houbraken, J., Visagie, C. M., and Frisvad, J. C. (2021). Recommendations To Prevent Taxonomic Misidentification of Genome-Sequenced Fungal Strains. *Microbiol Resour Announc* 10, e01074-20. doi: 10.1128/MRA.01074-20.
- Houbraken, J., Weig, M., Groß, U., Meijer, M., and Bader, O. (2016). *Aspergillus oerlinghausenensis*, a new mould species closely related to *A. fumigatus*. *FEMS Microbiology Letters* 363, fnv236. doi: 10.1093/femsle/fnv236.
- Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., et al. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346, 1320–1331. doi: 10.1126/science.1253451.
- Jurjevic, Z., Peterson, S. W., and Horn, B. W. (2012). *Aspergillus* section Versicolores: nine new species and multilocus DNA sequence based phylogeny. *IMA Fungus* 3, 59–79. doi: 10.5598/imafungus.2012.03.01.07.
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14, 587–589. doi: 10.1038/nmeth.4285.
- Kanamasa, S. (2021). Opinion on Updating the Taxonomic Names of *Aspergillus terreus* Strains IFO 6365 and TN-484. *Microbiol Resour Announc* 10, e00472-21. doi: 10.1128/MRA.00472-21.
- Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 30, 772–780. doi: 10.1093/molbev/mst010.
- Kjærboelling, I., Vesth, T., Frisvad, J. C., Nybo, J. L., Theobald, S., Kildgaard, S., et al. (2020). A comparative genomics study of 23 *Aspergillus* species from section Flavi. *Nature Communications* 11, 1106. doi: 10.1038/s41467-019-14051-y.
- Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., et al. (2019). OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research* 47, D807–D811. doi: 10.1093/nar/gky1053.

- Kück, U., and Dahlmann, T. A. (2021). The Confusion in Renaming Species: *Penicillium chrysogenum* and *Penicillium rubens*. *Microbiol Resour Announc* 10, e00464-21. doi: 10.1128/MRA.00464-21.
- Lekshmi, N., Umar, M. D., Dhaneesha, M., Rojin, J., Ravinesh, R., and Sajeevan, T. P. (2020). Endophytic Fungi Isolated from the Marine Sponges as a Source of Potential Bioactive Compounds. *Journal of Aquatic Biology & Fisheries* 8, 56–66.
- Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* 47, W256–W259. doi: 10.1093/nar/gkz239.
- Li, F., Wang, B., Wang, L., and Cao, B. (2014). Phylogenetic Analyses on the Diversity of *Aspergillus fumigatus* Ssensu 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., and Yu, Z. (2018). Draft Genome Sequence of NRRL 5109, an Ex-Type Isolate of *Aspergillus neoellipticus*. *Microbiology Resource Announcements* 7. doi: 10.1128/MRA.01262-18.
- Li, L., and Yu, Z. (2021). Changing the Taxonomic Classification of Strain NRRL5109 to *Aspergillus fumigatus*. *Microbiol Resour Announc* 10, e00475-21. doi: 10.1128/MRA.00475-21.
- Libkind, D., Čadež, N., Ofulente, D. A., Langdon, Q. K., Rosa, C. A., Sampaio, J. P., et al. (2020). Towards yeast taxogenomics: lessons from novel species descriptions based on complete genome sequences. *FEMS Yeast Research* 20. doi: 10.1093/femsyr/foaa042.
- Lücking, R., Aime, M. C., Robbertse, B., Miller, A. N., Aoki, T., Ariyawansa, H. A., et al. (2021). Fungal taxonomy and sequence-based nomenclature. *Nat Microbiol* 6, 540–548. doi: 10.1038/s41564-021-00888-x.
- Mead, M. E., Steenwyk, J. L., Silva, L. P., de Castro, P. A., Saeed, N., Hillmann, F., et al. (2021). An evolutionary genomic approach reveals both conserved and species-specific genetic elements related to human disease in closely related *Aspergillus* fungi. *Genetics* 218. doi: 10.1093/genetics/iyab066.
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* 37, 1530–1534. doi: 10.1093/molbev/msaa015.
- Mirarab, S., Reaz, R., Bayzid, Md. S., Zimmermann, T., Swenson, M. S., and Warnow, T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* 30, i541–i548. doi: 10.1093/bioinformatics/btu462.
- Nierman, W. C., Pain, A., Anderson, M. J., Wortman, J. R., Kim, H. S., Arroyo, J., et al. (2005). Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* 438, 1151–1156. doi: 10.1038/nature04332.

- Nilsson, R. H., Larsson, K.-H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., et al. (2019). The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research* 47, D259–D264. doi: 10.1093/nar/gky1022.
- Peterson, S. W. (2008). Phylogenetic analysis of *Aspergillus* species using DNA sequences from four loci. *Mycologia* 100, 205–226. doi: 10.3852/mycologia.100.2.205.
- Pitt, J. I., and Taylor, J. W. (2014). *Aspergillus*, its sexual states and the new International Code of Nomenclature. *Mycologia* 106, 1051–1062. doi: 10.3852/14-060.
- Raja, H. A., Miller, A. N., Pearce, C. J., and Oberlies, N. H. (2017). Fungal Identification Using Molecular Tools: A Primer for the Natural Products Research Community. *J. Nat. Prod.* 80, 756–770. doi: 10.1021/acs.jnatprod.6b01085.
- Rieppel, O. (2010). Species monophyly. *Journal of Zoological Systematics and Evolutionary Research* 48, 1–8. doi: 10.1111/j.1439-0469.2009.00545.x.
- Rieseberg, L. H., and Brouillet, L. (1994). Are many plant species paraphyletic? *TAXON* 43, 21–32. doi: 10.2307/1223457.
- Rokas, A., Williams, B. L., King, N., and Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425, 798–804. doi: 10.1038/nature02053.
- Ruiz, O. N., and Radwan, O. (2021). Difficulty in Assigning Fungal Identity Based on DNA Sequences. *Microbiol Resour Announc* 10, e00460-21. doi: 10.1128/MRA.00460-21.
- Samson, R. A., Hong, S., Peterson, S. W., Frisvad, J. C., and Varga, J. (2007). Polyphasic taxonomy of *Aspergillus* section *Fumigati* and its teleomorph *Neosartorya*. *Studies in Mycology* 59, 147–203. doi: 10.3114/sim.2007.59.14.
- Samson, R. A., Visagie, C. M., Houbraken, J., Hong, S.-B., Hubka, V., Klaassen, C. H. W., et al. (2014). Phylogeny, identification and nomenclature of the genus *Aspergillus*. *Studies in Mycology* 78, 141–173. doi: 10.1016/j.simyco.2014.07.004.
- Sayers, E. W., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., et al. (2009). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* 37, D5–D15. doi: 10.1093/nar/gkn741.
- Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., et al. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 6241–6246. doi: 10.1073/pnas.1117018109.
- Sklenář, F., Glässnerová, K., Jurjević, Ž., Houbraken, J., Samson, R. A., Visagie, C. M., et al. (2022). Taxonomy of *Aspergillus* series *Versicolores*: species reduction and lessons

- learned about intraspecific variability. *Studies in Mycology*. doi: 10.3114/sim.2022.102.02.
- Stanke, M., and 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. L., Buida, T. J., Gonçalves, C., Goltz, D. C., Morales, G., Mead, M. E., et al. (2022). BioKIT: a versatile toolkit for processing and analyzing diverse types of sequence data. *Genetics*. doi: 10.1093/genetics/iyac079.
- Steenwyk, J. L., Buida, T. J., Labella, A. L., Li, Y., Shen, X.-X., and Rokas, A. (2021). PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. *Bioinformatics (Oxford, England)*. doi: 10.1093/bioinformatics/btab096.
- Steenwyk, J. L., Buida, T. J., Li, Y., Shen, X.-X., and Rokas, A. (2020a). ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLOS Biology* 18, e3001007. doi: 10.1371/journal.pbio.3001007.
- Steenwyk, J. L., Lind, A. L., Ries, L. N. A., dos Reis, T. F., Silva, L. P., Almeida, F., et al. (2020b). Pathogenic Allodiploid Hybrids of *Aspergillus* Fungi. *Current Biology* 30, 2495–2507.e7. doi: 10.1016/j.cub.2020.04.071.
- Steenwyk, J. L., Mead, M. E., Knowles, S. L., Raja, H. A., Roberts, C. D., Bader, O., et al. (2020c). Variation Among Biosynthetic Gene Clusters, Secondary Metabolite Profiles, and Cards of Virulence Across *Aspergillus* Species. *Genetics* 216, 481–497. doi: 10.1534/genetics.120.303549.
- Steenwyk, J. L., and Rokas, A. (2021). orthofisher: a broadly applicable tool for automated gene identification and retrieval. *G3 Genes|Genomes|Genetics* 11. doi: 10.1093/g3journal/jkab250.
- Steenwyk, J. L., Shen, X.-X., Lind, A. L., Goldman, G. H., and Rokas, A. (2019). 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.
- Stevenson, B., and Stamps, B. (2021). It’s Complicated: Rapid Publication of Genomes in *Microbiology Resource Announcements* Can Be Both Part of the Problem and the Solution to Fungal Taxonomic Resolution. *Microbiol Resour Announc* 10, e00458-21. doi: 10.1128/MRA.00458-21.
- Susca, A., Proctor, R. H., Morelli, M., Haidukowski, M., Gallo, A., Logrieco, A. F., et al. (2016). Variation in Fumonisin and Ochratoxin Production Associated with Differences in Biosynthetic Gene Content in *Aspergillus niger* and *A. welwitschiae* Isolates from Multiple Crop and Geographic Origins. *Front. Microbiol.* 7. doi: 10.3389/fmicb.2016.01412.
- Susca, A., Villani, A., Moretti, A., Stea, G., and Logrieco, A. (2020). Identification of toxigenic fungal species associated with maize ear rot: Calmodulin as single informative gene.

- International Journal of Food Microbiology* 319, 108491. doi: 10.1016/j.ijfoodmicro.2019.108491.
- Tam, E. W. T., Chen, J. H. K., Lau, E. C. L., Ngan, A. H. Y., Fung, K. S. C., Lee, K.-C., et al. (2014). Misidentification of *Aspergillus nomius* and *Aspergillus tamarii* as *Aspergillus flavus*: characterization by internal transcribed spacer, β -Tubulin, and calmodulin gene sequencing, metabolic fingerprinting, and matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol* 52, 1153–1160. doi: 10.1128/JCM.03258-13.
- Tan, G., Muffato, M., Ledergerber, C., Herrero, J., Goldman, N., Gil, M., et al. (2015). Current Methods for Automated Filtering of Multiple Sequence Alignments Frequently Worsen Single-Genes Phylogenetic Inference. *Systematic Biology* 64, 778–791. doi: 10.1093/sysbio/syv033.
- Taylor, J. W., Göker, M., and Pitt, J. I. (2016). Choosing one name for pleomorphic fungi: The example of *Aspergillus* versus *Eurotium*, *Neosartorya* and *Emericella*. *Taxon* 65, 593–601. doi: 10.12705/653.10.
- Venkateswaran, K., Wood, J. M., Simpson, A., and Singh, N. (2021). Reliable Identification of Environmental Fungi Requires Multigene-Based Phylogenetic Analyses. *Microbiol Resour Announc* 10, e00469-21. doi: 10.1128/MRA.00469-21.
- Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., et al. (2018). BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Molecular Biology and Evolution* 35, 543–548. doi: 10.1093/molbev/msx319.
- Xu, J. (2020). Fungal species concepts in the genomics era. *Genome* 63, 459–468. doi: 10.1139/gen-2020-0022.
- Yin, G., Jurick, W. M., Zhao, G., and Bennett, J. W. (2021). New Names for Three *Penicillium* Strains Based on Updated Barcoding and Phylogenetic Analyses. *Microbiol Resour Announc* 10, e00466-21. doi: 10.1128/MRA.00466-21.
- Zhang, C., Rabiee, M., Sayyari, E., and 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.

Figures and figure legends

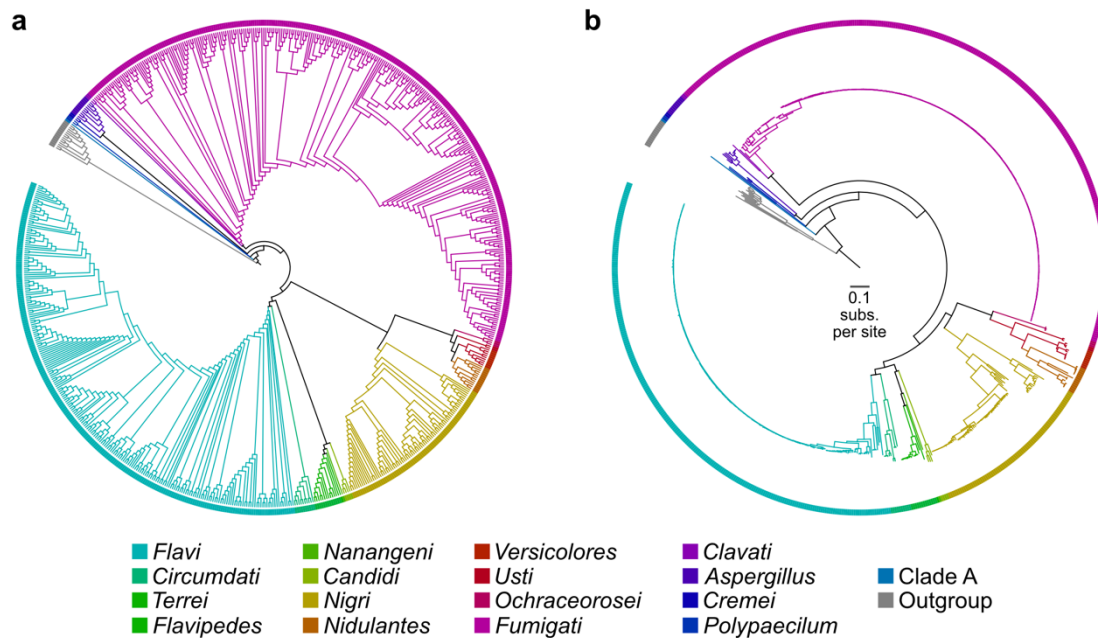


Figure 1. Phylogenomic tree of 725 genomes based on analysis of 1,362 genes (6,378,237 nucleotide sites). The evolutionary history of 711 *Aspergillus* species and 14 outgroup genomes from the genus *Penicillium* was reconstructed from a 1,362-gene matrix. The phylogeny is depicted without branch lengths (a) and with branch lengths, representing substitutions per site (b). Colors represent different sections—taxonomic ranks above species and below genus.

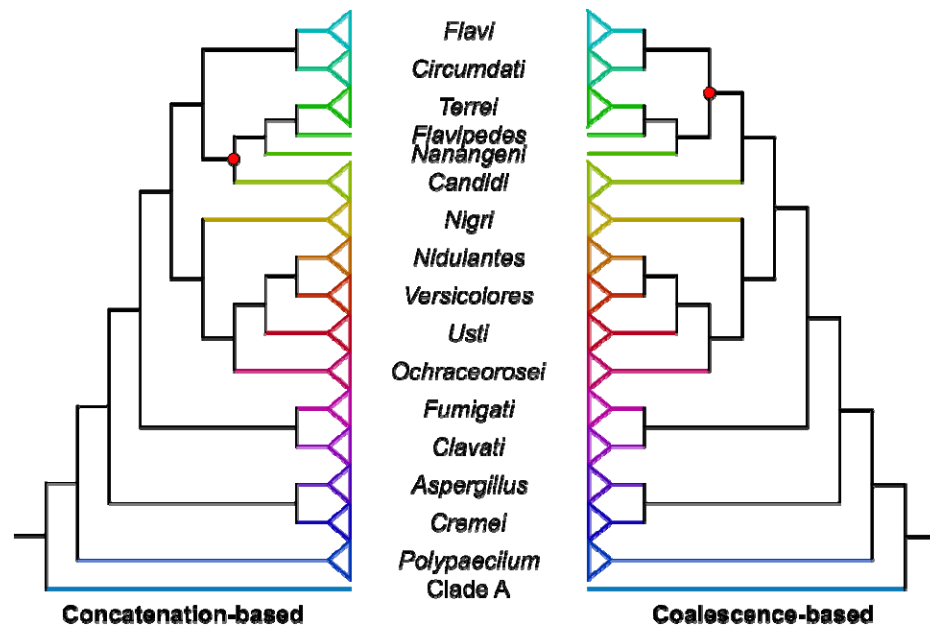


Figure 2. Concatenation- (left) and coalescence-based (right) phylogenies of taxonomic sections in the genus *Aspergillus* are highly congruent. The evolutionary relationships among sections are depicted. Species-level concatenation- and coalescence-based phylogenies differed at two bipartitions (represented by a red dot). Branch lengths and triangle sizes have no meaning.

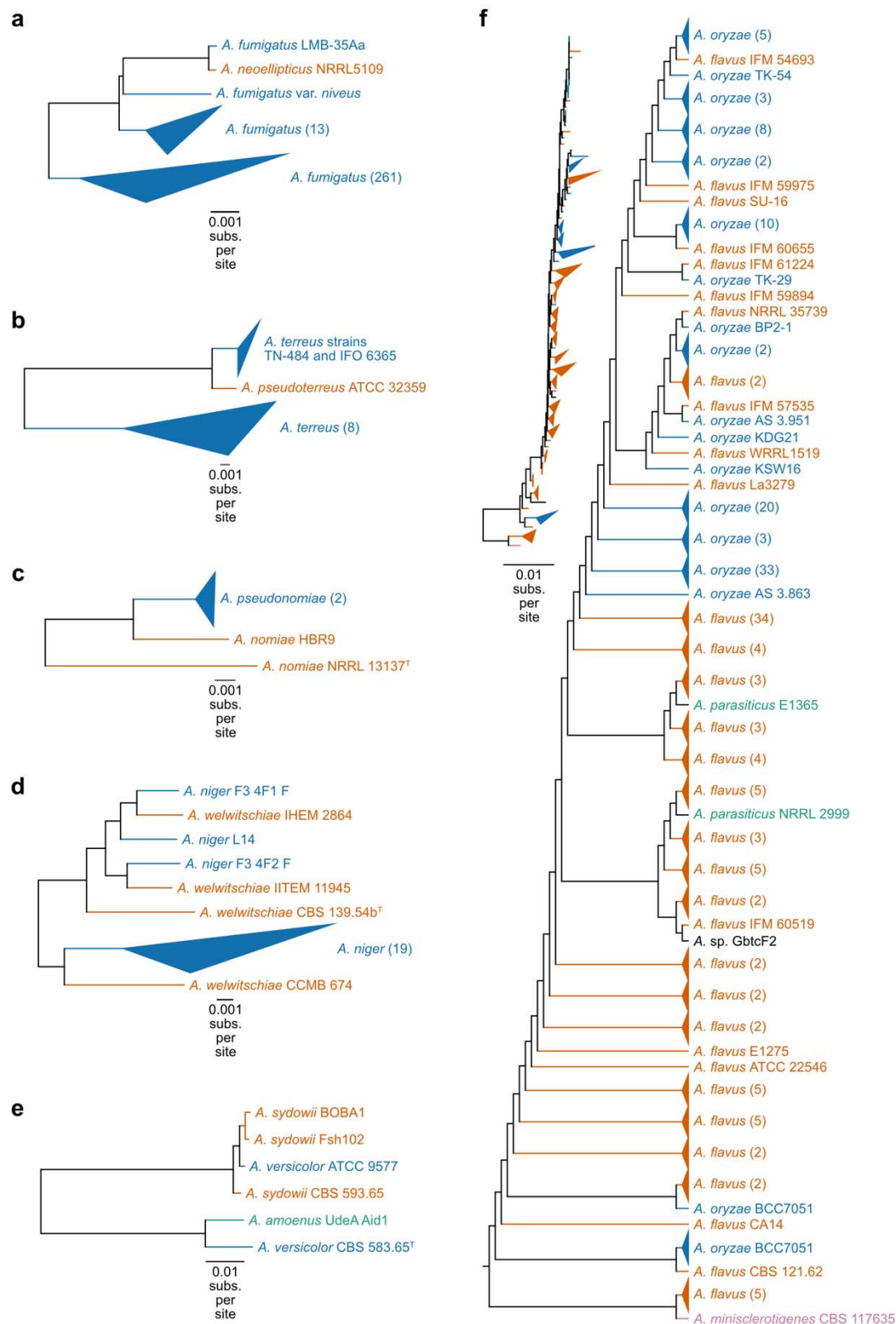


Figure 3. Phylogenomics underscores known taxonomic uncertainties and reveals new

ones. (a) *Aspergillus neoellipticus* is a misidentified strain of *Aspergillus fumigatus*; alternatively, several *A. neoellipticus* strains are misidentified as *A. fumigatus*. (b) Strain misidentification occurs between *Aspergillus pseudoterreus* and *Aspergillus terreus* and (c) *Aspergillus pseudonomiae* and *Aspergillus nomiae*. (d) Isolates identified as *Aspergillus niger* and *Aspergillus welwitschiae* are likely misidentified. (e) *Aspergillus* strain ATCC 9577 is misidentified as *Aspergillus versicolor* but is likely *Aspergillus sydowii*. (f) Strains of *Aspergillus oryzae*, *Aspergillus flavus*, *Aspergillus parasiticus*, and *Aspergillus minisclerotigenes* appear to have polyphyletic origins, a result that is likely due to extensive strain misidentification (e.g., see also Houbroken et al 2021, detailing misidentification of five strains of *A. minisclerotigenes* as *A. flavus* (Houbroken et al., 2021)). Topologies presented were inferred using the concatenation approach. See Figure S3 for topologies inferred using coalescence. Different colors represent different species. Isolates with no known species are depicted in black. Triangles represent collapsed lineages with multiple isolates. The number of isolates in each collapsed lineage are shown next to the species name in parentheses.

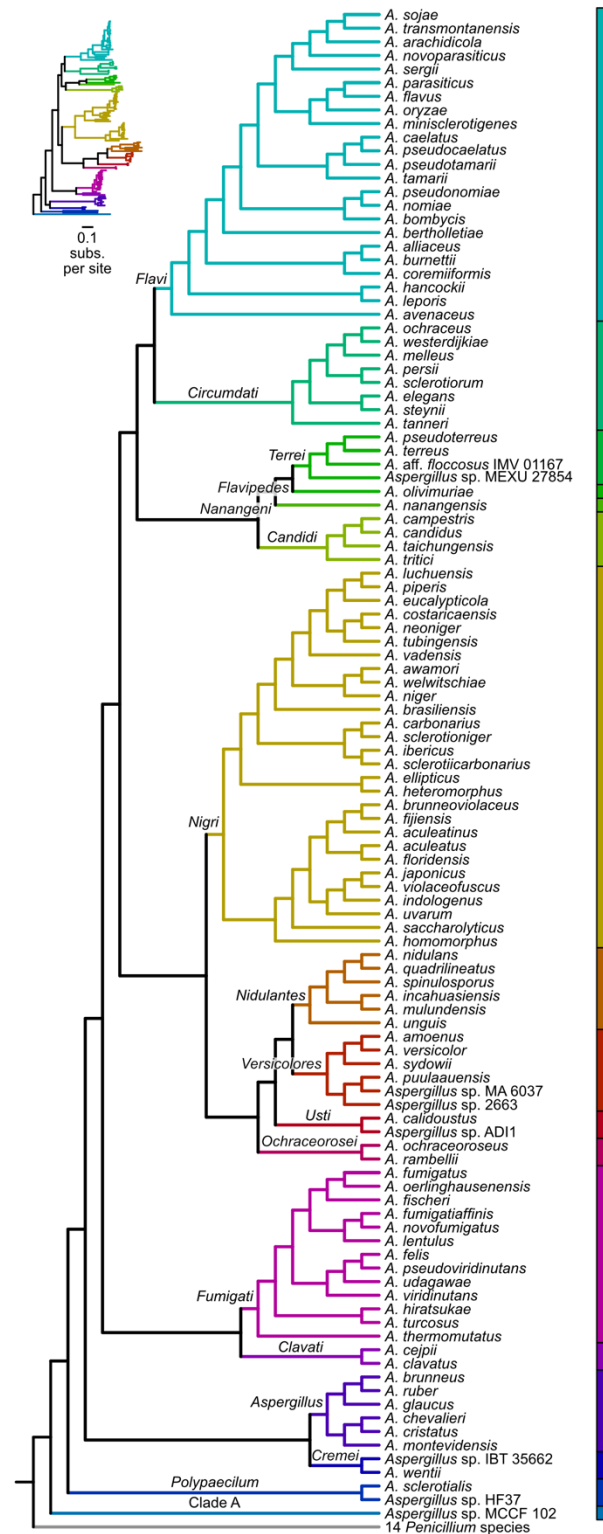


Figure 4. A species-level phylogeny of the genus *Aspergillus*. A genome-scale view of the *Aspergillus* phylogeny and the identification of a new sister lineage, clade A, to a clade of the rest of *Aspergillus* species whose genomes have been sequenced may help understand the

early evolution of this biomedically and technologically relevant lineage. Inset represents the same phylogeny with branch lengths representing substitutions per site.