

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

Jacob L. Steenwyk,<sup>1,2,3</sup> Charu Balamurugan,<sup>2,3</sup> Huzefa A. Raja,<sup>4</sup> Carla Gonçalves,<sup>2,3</sup> Ningxiao Li,<sup>5,6</sup> Frank Martin,<sup>6</sup> Judith Berman,<sup>7</sup> Nicholas H. Oberlies,<sup>4</sup> John G. Gibbons,<sup>8,9,10</sup> Gustavo H. Goldman,<sup>11</sup> David M. Geiser,<sup>12</sup> Jos Houbraken,<sup>13</sup> David S. Hibbett,<sup>14</sup> Antonis Rokas<sup>2,3,15</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 14.

**ABSTRACT** Modern taxonomic classification is often based on phylogenetic analyses of a few molecular markers, although single-gene studies are still common. Here, we leverage genome-scale molecular phylogenetics (phylogenomics) of species and populations to reconstruct evolutionary relationships in a dense data set of 710 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 provided profile Hidden Markov Models for each, facilitating their use by others. Examining the resulting phylogeny helped resolve ongoing taxonomic controversies, identified new ones, and revealed extensive strain misidentification (7.59% of strains were previously misidentified), underscoring the importance of population-level sampling in species classification. These findings were corroborated using the current standard, taxonomically informative loci. These findings suggest that phylogenomics of species and populations can facilitate accurate taxonomic classifications and reconstructions of the Tree of Life.

**IMPORTANCE** Identification of fungal species relies on the use of molecular markers. Advances in genomic technologies have made it possible to sequence the genome of any fungal strain, making it possible to use genomic data for the accurate assignment of strains to fungal species (and for the discovery of new ones). We examined the usefulness and current limitations of genomic data using a large data set of 710 publicly available genomes from multiple strains and species of the biomedically, agriculturally, and industrially important genus *Aspergillus*. Our evolutionary genomic analyses revealed that nearly 8% of publicly available *Aspergillus* genomes are misidentified. Our work highlights the usefulness of genomic data for fungal systematic biology and suggests that systematic genome sequencing of multiple strains, including reference strains (e.g., type strains), of fungal species will be required to reduce misidentification errors in public databases.

**KEYWORDS** taxogenomics, pathogenicity, pathogen, virulence, plant pathogen, genomics, taxonomy, *Aspergillaceae*, *Penicillium*

Species determination is the foundation for evolutionary studies (1–4). Over the years, several different species concepts have been proposed (1). Due to the microscopic nature of many fungi and the lack of known sexual cycles for some species (5), 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 (6) such as the ribosomal large and small subunit; the nuclear ribosomal internal transcribed spacer regions including the 5.8 S rDNA (ITS); the RNA

**Editor** Alexandre Alanio, Institut Pasteur, Paris, France

Address correspondence to Antonis Rokas, antonis.rokas@vanderbilt.edu.

J.L.S. is an advisor for ForensisGroup Incorporated. A.R. is a scientific consultant for LifeMine Therapeutics, Inc.

See the funding table on p. 14.

**Received** 20 November 2023

**Accepted** 18 February 2024

**Published** 6 March 2024

Copyright © 2024 Steenwyk et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

polymerase II large subunit and core subunits; and minichromosome maintenance complex component 7 (7, 8).

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 (NCBI). Software is subsequently used to infer phylogenetic trees for each sequence, representing a hypothesis of the evolutionary history of these species (8). 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 the sparsity of available genome sequences across the fungal tree of life (3, 4, 9–17).

Species in the genus *Aspergillus* are of medical, agricultural, and biotechnological significance. *Aspergillus fumigatus* and *Aspergillus flavus* are pathogens and allergens, and produce mycotoxins (18, 19). *Aspergillus niger* is an industrial workhorse, and *Aspergillus oryzae* is used to produce fermented foods like soy sauce and sake (20, 21). Accurate identification of *Aspergillus* fungi with barcode loci is often challenging (13, 15–17, 22, 23) but, importantly, because even closely related species can differ in drug resistance profiles and ability to cause disease (24–26). For example, clinical strains of *Aspergillus nomius* and *Aspergillus tamarii* have been misidentified as *A. flavus* (27). In the clinic, inaccurate species determination could lead to misguided disease management strategies due to differences in intrinsic drug resistance levels between species (23). For example, *Aspergillus latus*, commonly misidentified as *Aspergillus nidulans*, is more resistant to the antifungal drug caspofungin than *A. nidulans* (28). However, there is no current consensus for the levels of barcode sequence divergence required to consider two distinct fungal clades as different species; 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 *A. fumigatus*, respectively (29, 30).

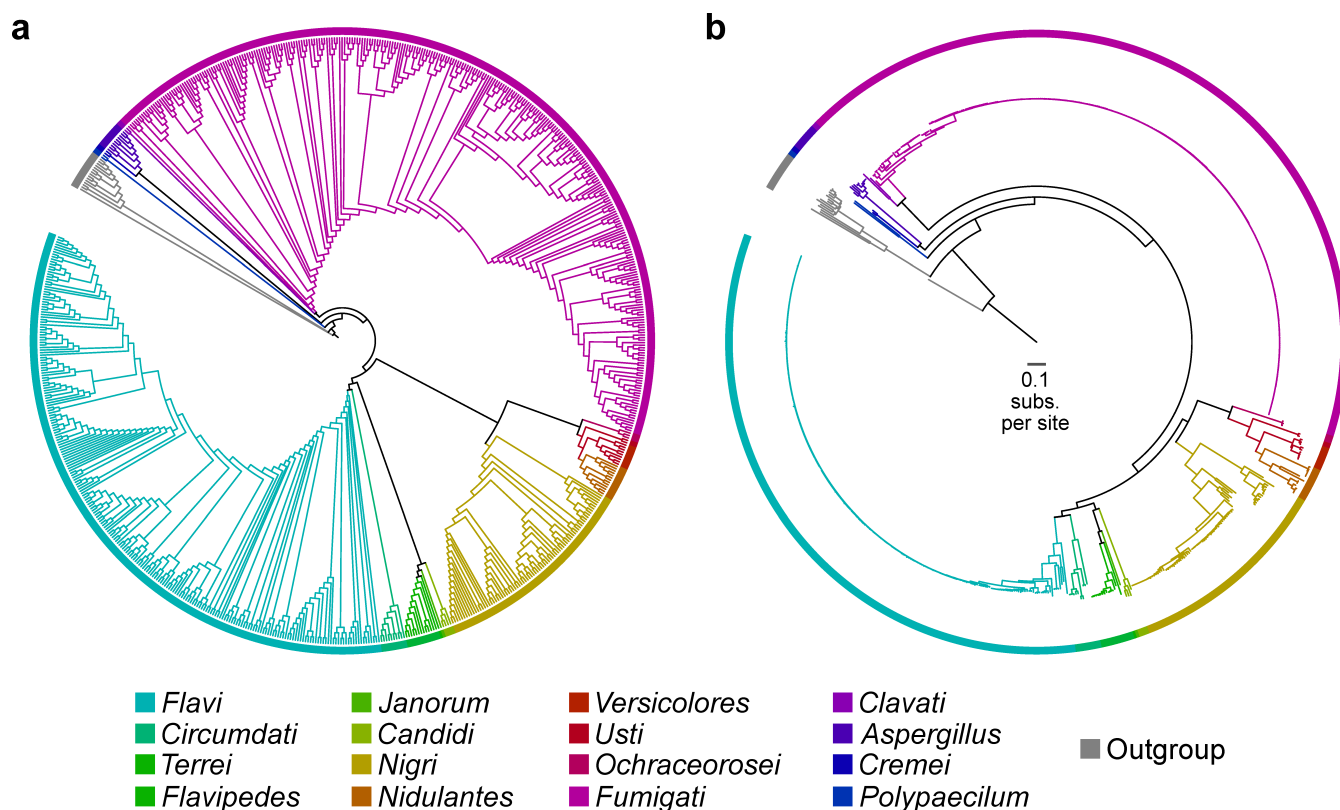
Reconstructing deeper evolutionary relationships from a few molecular markers can also be challenging. Divergences among sections—a secondary taxonomic rank above the species and below the genus ranks—have been debated. For example, the sections *Nigri*, *Ochraceorosei*, *Flavi*, *Circumdati*, *Candidi*, and *Terrei* were inferred to be monophyletic based on analyses of four loci from 81 taxa (31, 32) but topology tests using a 1,668-gene matrix from a different 81-taxon data set rejected the monophyly of these lineages (33). Accurate reconstructions of the *Aspergillus* phylogeny 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 710 *Aspergillus* genome sequences spanning 98 species and population data for 36 species, more than doubling the number of species analyzed in previous genome-scale studies (33–35) and capturing roughly one-quarter of all known species in the genus (36, 37). The new phylogeny reveals that phylogenomics using species and populations can facilitate strain classification and resolve taxonomic controversies while identifying new ones. Moreover, phylogenomic analyses revealed 7.59% of strains (55/725) were previously misidentified. These findings were further corroborated using taxonomically informative loci, the current gold standard in the field of systematics.

## RESULTS AND DISCUSSION

### A phylogenomic tree of *Aspergillus*

The evolutionary history of 725 genomes (710 *Aspergillus* genomes; 15 outgroup genomes) was reconstructed using maximum likelihood analysis of a 1,362-gene data matrix with 6,378,237 nucleotide sites (Fig. 1; Fig. S1 at <https://doi.org/10.6084/m9.figshare.21382131>). The 725 genomes represent public whole genome assemblies available through the NCBI (<https://www.ncbi.nlm.nih.gov/>) that also passed quality-control measures (see Materials and Methods; Tables S1 and S2 at <https://doi.org/10.6084/m9.figshare.21382131>). Based on the NCBI-provided taxonomic information, the data set

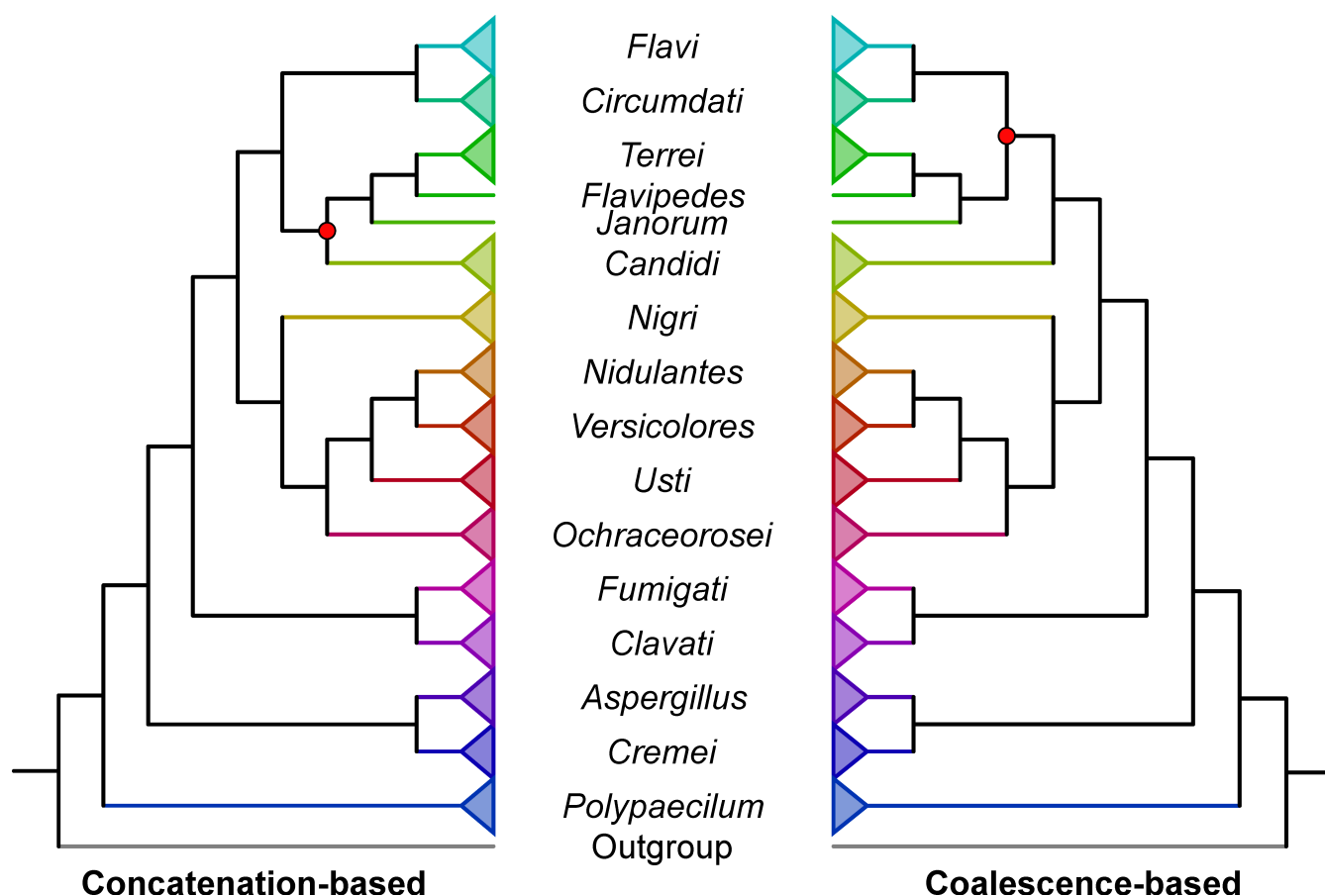


**FIG 1** Phylogenomic tree of 725 genomes based on analysis of 1,362 genes (6,378,237 nucleotide sites). The evolutionary history of 710 *Aspergillus* species and 15 outgroup genomes was reconstructed from a 1,362-gene matrix. The phylogeny is depicted without branch lengths (a) and with branch lengths, representing substitutions per site (b). Colors represent different sections—taxonomic ranks above species and below genus. Note, *Versicolores* is a series within the section *Nidulantes*.

includes 115 *Aspergillus* and 14 *Penicillium* species. Note that analysis presented later in our manuscript identified one *Aspergillus* genome (strain MCF 102) as *Paecilomyces formosus*; thus, the total number of outgroup genomes is 15. The genomes of two or more strains were available for 36 *Aspergillus* species, but the depth of strain sampling varied (Table S4 at <https://doi.org/10.6084/m9.figshare.21382131>). Sampling was densest for *A. fumigatus* ( $N = 275$ ), *A. flavus* ( $N = 105$ ), *A. oryzae* ( $N = 97$ ), and *A. niger* ( $N = 24$ ). Sixteen species had genome sequences from two representative strains available. Twelve strains were of unknown species but were reportedly from the genus *Aspergillus*. The data set spanned 17 *Aspergillus* sections. Higher-order relationships among sections were examined using concatenation- and coalescence-based tree inference approaches (38, 39). The phylogenies inferred with both approaches were highly congruent, differing by only one bipartition (Fig. 2).

### Phylogenomics sheds new light on *Aspergillus* strain identification

Phylogenomic analyses reveal that half of the 36 species with multiple strains were not monophyletic (Table S5 at <https://doi.org/10.6084/m9.figshare.21382131>). Although monophyly is not required for species delimitation under certain species concepts (40, 41), this observation was surprising for *Aspergillus* because new species descriptions typically include molecular phylogenetic evidence, and suggests that cryptic speciation and/or species misidentification may be rampant. Cross-referencing these findings with strain determination via taxonomically informative loci indicated that 55 strains (55/725 or 7.59%) had been misidentified (Table 1). The high percentage of misidentified strains underscores both the magnitude of the problem as well as the importance of



**FIG 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. Note, *Versicolores* is a series within the section *Nidulantes*.

population sampling for accurate strain identification. Here, we highlight eight cases that illustrate this issue.

### *Aspergillus fumigatus* and *Aspergillus neoellipticus*

The phylogenetic placement of *A. neoellipticus* has been debated and bears on disease management strategies. Some studies suggest that *A. neoellipticus* is distinct from the major human pathogen *A. fumigatus*, based on the analysis of five loci (42). Other studies suggested that *A. neoellipticus* is conspecific to *A. fumigatus* based on partial single-gene sequences and random amplified polymorphic DNA (42–45). Resolving if *A. neoellipticus* is a separate species bears on our understanding of the evolution of pathogenicity in *Aspergillus*. Specifically, if *A. neoellipticus* can cause disease, as implicated in the ex-type strain NRRL 5109 (45), which was isolated from a case of chronic emphysema, the ability to cause disease would have evolved in the ancestor between *A. fumigatus* and *A. neoellipticus*; if not, pathogenicity would have only evolved in *A. fumigatus*.

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 because *A. neoellipticus* was inferred as a sister to *A. fumigatus* (46). The combined use of genome-scale data and extensive population-level sampling of 275 *A. fumigatus* strains in the analyses found that *A. neoellipticus* NRRL 5109 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* NRRL 5109

TABLE 1 Summary of 55 misidentified isolates<sup>a</sup>

Updated genus and species	Strain identifier	Updated section	NCBI accession	Genus and species on NCBI	Section based on NCBI data
<i>Aspergillus pseudoglaucus</i>	CBS 112.26	<i>Aspergillus</i>	GCA_020284055.1	<i>Aspergillus brunneus</i>	<i>Aspergillus</i>
<i>Aspergillus neotritici</i>	UdeA_Ac1	<i>Candidi</i>	GCA_009812425.1	<i>Aspergillus tritici</i>	<i>Candidi</i>
<i>Aspergillus pallidofulvus</i>	AO.MF010	<i>Circumdati</i>	GCA_005784425.1	<i>Aspergillus ochraceus</i>	<i>Circumdati</i>
<i>Aspergillus westerdijkiae</i>	fc-1	<i>Circumdati</i>	GCA_004849945.1	<i>Aspergillus ochraceus</i>	<i>Circumdati</i>
<i>Aspergillus sclerotiorum</i>	NIBRFGC000004109	<i>Circumdati</i>	GCA_002215965.1	<i>Aspergillus persii</i>	<i>Circumdati</i>
<i>Aspergillus subramanianii</i>	HBR18	<i>Circumdati</i>	GCA_000530345.1	<i>Aspergillus sclerotiorum</i>	<i>Circumdati</i>
<i>Aspergillus europaeus</i>	IBT 35662	<i>Cremeri</i>	GCA_019976455.1	<i>Aspergillus</i> sp. subgen. <i>Cremeri</i>	<i>Cremeri</i>
<i>Paecilomyces formosus</i>	MCCF 102	Not applicable	GCA_018167545.1	<i>Aspergillus</i> sp.	Unknown
<i>Aspergillus salisburgensis</i>	HF37	<i>Polypaecilum</i>	GCA_003698115.1	<i>Aspergillus</i> sp.	Unknown
<i>Aspergillus luteovirescens</i>	NRRL 26010	<i>Flavi</i>	GCF_001792695.1	<i>Aspergillus bombycis</i>	<i>Flavi</i>
<i>Aspergillus alliaceus</i>	FRR 5400	<i>Flavi</i>	GCA_013421405.2	<i>Aspergillus burnettii</i>	<i>Flavi</i>
<i>Aspergillus minisclerotigenes</i>	E1293	<i>Flavi</i>	GCA_013145855.1	<i>Aspergillus flavus</i>	<i>Flavi</i>
<i>Aspergillus minisclerotigenes</i>	E1406	<i>Flavi</i>	GCA_013146155.1	<i>Aspergillus flavus</i>	<i>Flavi</i>
<i>Aspergillus minisclerotigenes</i>	E1316	<i>Flavi</i>	GCA_013145865.1	<i>Aspergillus flavus</i>	<i>Flavi</i>
<i>Aspergillus minisclerotigenes</i>	E1288	<i>Flavi</i>	GCA_011420435.1	<i>Aspergillus flavus</i>	<i>Flavi</i>
<i>Aspergillus minisclerotigenes</i>	E1376	<i>Flavi</i>	GCA_013146015.1	<i>Aspergillus flavus</i>	<i>Flavi</i>
<i>Aspergillus pseudonomiae</i>	HBR9	<i>Flavi</i>	GCA_000531055.1	<i>Aspergillus nomiae</i>	<i>Flavi</i>
<i>Aspergillus flavus</i> or <i>Aspergillus oryzae</i>	NRRL 2999	<i>Flavi</i>	GCA_012897115.1	<i>Aspergillus parasiticus</i>	<i>Flavi</i>
<i>Aspergillus flavus</i> or <i>Aspergillus oryzae</i>	E1365	<i>Flavi</i>	GCA_013145995.1	<i>Aspergillus parasiticus</i>	<i>Flavi</i>
<i>Aspergillus flavus</i> or <i>Aspergillus oryzae</i>	GbtcF2	<i>Flavi</i>	GCA_019176375.1	<i>Aspergillus</i> sp.	Unknown
<i>Aspergillus flavus</i> or <i>Aspergillus oryzae</i>	ATCC 12892	<i>Flavi</i>	GCA_002894705.1	<i>Aspergillus</i> sp.	Unknown
<i>Aspergillus parasiticus</i> or <i>Aspergillus sojae</i> (should be <i>A. transmontanensis</i> ; strain contaminated)	CBS 130015	<i>Flavi</i>	GCA_009193505.1	<i>Aspergillus transmontanensis</i>	<i>Flavi</i>
<i>Aspergillus fumigatus</i>	NRRL 5109	<i>Fumigati</i>	GCA_003116565.1	<i>Aspergillus neoellipticus</i>	<i>Fumigati</i>
<i>Aspergillus fumigatus</i>	GbtcF1	<i>Fumigati</i>	GCA_019176365.1	<i>Aspergillus</i> sp.	Unknown
<i>Aspergillus unguis</i>	JS3-R5	<i>Nidulantes</i>	GCA_019775315.1	<i>Aspergillus</i> sp.	Unknown
<i>Aspergillus luchuensis</i>	JCM 22320	<i>Nigri</i>	GCA_001599415.1	<i>Aspergillus awamori</i>	<i>Nigri</i>
<i>Aspergillus niger</i>	IFM 58123	<i>Nigri</i>	GCA_003850985.1	<i>Aspergillus awamori</i>	<i>Nigri</i>
<i>Aspergillus tubingensis</i>	CBS 115574	<i>Nigri</i>	GCF_003184835.1	<i>Aspergillus costaricensis</i>	<i>Nigri</i>
<i>Aspergillus brunneoviolaceus</i>	CBS 313.89	<i>Nigri</i>	GCF_003184825.1	<i>Aspergillus fijiensis</i>	<i>Nigri</i>
<i>Aspergillus tubingensis</i>	CBS 115656	<i>Nigri</i>	GCF_003184625.1	<i>Aspergillus neoniger</i>	<i>Nigri</i>
<i>Aspergillus luchuensis</i>	RAF106	<i>Nigri</i>	GCA_011316255.1	<i>Aspergillus niger</i>	<i>Nigri</i>
<i>Aspergillus tubingensis</i>	3.316	<i>Nigri</i>	GCA_013618955.1	<i>Aspergillus niger</i>	<i>Nigri</i>
<i>Aspergillus tubingensis</i>	An76	<i>Nigri</i>	GCA_001515345.1	<i>Aspergillus niger</i>	<i>Nigri</i>
<i>Aspergillus niger</i>	ATCC 13157	<i>Nigri</i>	GCA_003344505.1	<i>Aspergillus phoenicis</i>	<i>Usti</i>
<i>Aspergillus luchuensis</i>	CBS 112811	<i>Nigri</i>	GCF_003184755.1	<i>Aspergillus piperis</i>	<i>Nigri</i>
<i>Aspergillus japonicus</i>	CBS 115571	<i>Nigri</i>	GCA_003184705.1	<i>Aspergillus violaceofuscus</i>	<i>Nigri</i>
<i>Aspergillus niger</i>	CCMB 674	<i>Nigri</i>	GCA_009761105.1	<i>Aspergillus welwitschiae</i>	<i>Nigri</i>
<i>Aspergillus niger</i>	IHEM 2864	<i>Nigri</i>	GCA_012275225.1	<i>Aspergillus welwitschiae</i>	<i>Nigri</i>
<i>Aspergillus niger</i>	CBS 139.54b	<i>Nigri</i>	GCF_003344945.1	<i>Aspergillus welwitschiae</i>	<i>Nigri</i>
<i>Aspergillus niger</i>	ITEM 11945	<i>Nigri</i>	GCA_012365075.1	<i>Aspergillus welwitschiae</i>	<i>Nigri</i>
<i>Aspergillus ochraceoroseus</i>	SRRC1468	<i>Ochraceorosei</i>	GCA_000986645.1	<i>Aspergillus rambellii</i>	<i>Ochraceorosei</i>
<i>Aspergillus floccosus</i>	IMV 01167	<i>Terrei</i>	GCA_001931935.1	<i>Aspergillus</i> aff. <i>floccosus</i>	Unknown
<i>Aspergillus</i> sp. nov.	MEXU 27854	<i>Terrei</i>	GCA_019721355.1	<i>Aspergillus</i> sp.	Unknown
<i>Aspergillus terreus</i>	A31	<i>Terrei</i>	GCA_016162245.1	<i>Aspergillus</i> sp.	Unknown
<i>Aspergillus pseudoterreus</i>	IFO 6365	<i>Terrei</i>	GCA_009932835.1	<i>Aspergillus terreus</i>	<i>Terrei</i>
<i>Aspergillus pseudoterreus</i>	TN-484	<i>Terrei</i>	GCA_009014675.2	<i>Aspergillus terreus</i>	<i>Terrei</i>
<i>Aspergillus pseudodefectus</i>	ADI1	<i>Usti</i>	GCA_020615375.1	<i>Aspergillus</i> sp.	Unknown
<i>Aspergillus versicolor</i>	UdeA_Aid1	<i>Nidulantes</i>	GCA_009812435.1	<i>Aspergillus amoenus</i>	<i>Versicolores</i>
<i>Aspergillus creber</i>	MK2	<i>Nidulantes</i>	GCF_016861865.1	<i>Aspergillus puulaauensis</i>	<i>Versicolores</i>

(Continued on next page)



TABLE 1 Summary of 55 misidentified isolates<sup>a</sup> (Continued)

Updated genus and species	Strain identifier	Updated section	NCBI accession	Genus and species on NCBI	Section based on NCBI data
<i>Aspergillus creber</i>	MA 6037	<i>Nidulantes</i>	GCA_003138035.1	<i>Aspergillus</i> sp.	<i>Versicolores</i>
<i>Aspergillus creber</i>	2663	<i>Nidulantes</i>	GCA_016880755.1	<i>Aspergillus</i> sp.	<i>Versicolores</i>
<i>Aspergillus sydowii</i>	ATCC 9577	<i>Nidulantes</i>	GCA_020284045.1	<i>Aspergillus versicolor</i>	<i>Versicolores</i>
<i>Penicillium rubens</i>	P2niaD18	<i>Chrysogena</i>	GCA_000710275.1	<i>Penicillium chrysogenum</i>	<i>Chrysogena</i>
<i>Penicillium brevistipitatum</i>	IBT 31321	<i>Robsamsonia</i>	GCA_002072405.1	<i>Penicillium coprophilum</i>	<i>Robsamsonia</i>

<sup>a</sup>The first column refers to the updated genus and species name. The second column is the strain identifier. The third column is the updated section. The fourth column is the NCBI accession for the genome assembly. The fifth and sixth columns are the original genus, species, and corresponding section designation. Note, *Versicolores* is a series within the section *Nidulantes*.

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*.

Resolving this taxonomic controversy may require additional taxonomically informative features. These features may include secondary metabolite production, minimum inhibit concentration profiles, and morphological features (e.g., spore size and shape). Moreover, examinations of genome sequencing among additional strains coupled with concordance analysis of gene histories among representative *A. fumigatus*/*A. neoellipticus* strains and other closely related species, such as *A. oerlinghausenensis* and *Aspergillus fischeri* (25) may help elucidate species boundaries. Moreover, biological species concepts, examined through mating compatibility tests (47), may also prove helpful.

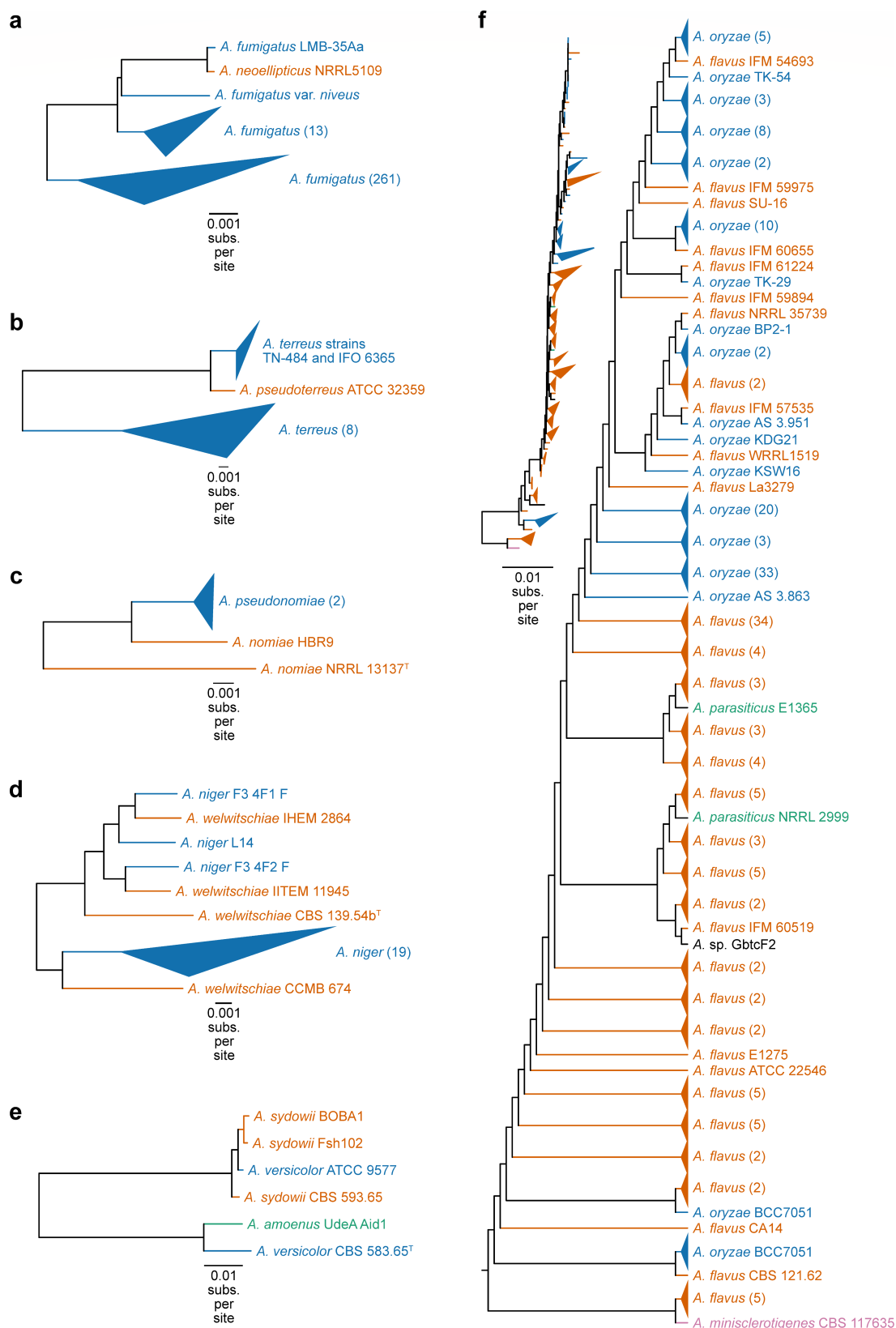
### *Aspergillus pseudoterreus* and *Aspergillus pseudonomiae*

A similar situation, but with much lower levels of sampling, exists for two other pairs of species: *Aspergillus terreus* and *A. 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 (22). For *A. pseudonomiae*, phylogenomics revealed strain HBR9—which was initially labeled as *A. nomiae*—is monophyletic with *A. pseudonomiae* isolates suggesting the strain is mislabeled (Fig. 3c). Sequence similarity searches of the taxonomically informative locus,  $\beta$ -tubulin, also suggest HBR9 is *A. pseudonomiae* (Table S3 at <https://doi.org/10.6084/m9.figshare.21382131>).

### *Aspergillus niger*

Species from section *Nigri* are similar both in their morphology and in their molecular barcode sequences (48). Although additional barcode sequences, such as beta-tubulin and calmodulin gene sequences help differentiate species within the section, the high similarity of certain taxonomically informative loci presents an opportunity to demonstrate the power and utility of genome-scale data. As an example, we highlight strain misidentification between *A. niger* and *Aspergillus welwitschiae*.

*A. niger*, a species from the section *Nigri* (49), has been considered a prominent pathogen of sisal (*Agave sisalana*), an industrial crop used in the textile industry (50). However, recent molecular phylogenetic analysis suggests that the main etiological agent of sisal bole rot disease is *A. welwitschiae*, not *A. niger* (50). Examination of the evolutionary history of *A. niger* and *A. welwitschiae*, using phylogenomic analyses that sampled multiple strains from both species, identified four cases of strain misidentification, concerning strains CCMB 674, IHEM 2864, CBS 139.54b, and ITEM 11945 (Fig. 3d). Sequencing similarity searches of taxonomically informative loci further corroborated this finding (Table S3 at <https://doi.org/10.6084/m9.figshare.21382131>). Although some reports have recently suggested to unite *A. niger* and *A. welwitschiae* into a single species (51), this proposal has yet to be widely adopted. Our phylogenomic analyses support the presence of two distinct clades; however, a broader strain sampling covering a



**FIG 3** Phylogenomics underscores known taxonomic uncertainties and reveals new ones. (a) *A. neoellipticus* NRRL 5109 is an *A. fumigatus* strain or the clade formed by the 16 *A. fumigatus* strains and *A. neoellipticus* NRRL 5109 is, in fact, *A. neoellipticus*, a species sister to *A. fumigatus*. (b) Strain misidentification occurs between *Aspergillus pseudoterreus* and *Aspergillus terreus* and (c) *Aspergillus pseudonomiae* and *Aspergillus nomiae*. (d) Strains identified as *Aspergillus welwitschiae* (Continued on next page)

**FIG 3 (Continued)**

are *A. niger*. (e) *Aspergillus* strain ATCC 9577 is misidentified as *Aspergillus versicolor* but is *Aspergillus sydowii*. (f) Strains of *A. oryzae*, *A. 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 Houbraken et al. (22), detailing misidentification of five strains of *A. minisclerotigenes* as *A. flavus* (22)]. Topologies presented were inferred using the concatenation approach. See Fig. S3 at <https://doi.org/10.6084/m9.figshare.21382131> 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 is shown next to the species name in parentheses. See Table 1 for the revised taxonomy.

higher diversity should be conducted to gain better insight. An alternative interpretation is that *A. niger* may have two distinct populations (Fig. 3d). Either way, results from phylogenomic analyses highlight how genome-wide analyses inform our understanding of intra-species genetic diversity and species boundaries. Until the taxonomy is resolved, uncertainty confounds studies of sisal bole rot, which could have economic ramifications.

***Aspergillus sydowii***

Another instance of putative misidentification concerns *Aspergillus versicolor* ATCC 9577, which is conspecific with *A. sydowii* strains, suggesting that strain ATCC 9577 is, in fact, *A. sydowii* (Fig. 3e). This observation was further supported using sequence similarity searches of taxonomically informative loci (Table S3 at <https://doi.org/10.6084/m9.figshare.21382131>). ATCC 9577 and CBS 583.65 are derived from the same sample material and serve as ex-neotype strains of *A. versicolor* (52); the availability of two different genome sequences suggests contamination or erroneous metadata for the former strain. The close relationship of CBS 583.65<sup>T</sup> with *Aspergillus amoenus* is supported by previous taxonomic studies (52, 53).

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 data set proposed species in the series *Versicolores* (section *Nidulantes*) be reduced from 17 species to four, citing intraspecific variation as a driver for over-splitting species boundaries by taxonomists (52). Evolutionary relationships under this new analysis differ slightly from our genome-scale phylogeny (Fig. S2 at <https://doi.org/10.6084/m9.figshare.21382131>). Additional genome sequences of species and strains may shed light on the evolutionary history of these species (54).

***Aspergillus flavus* and *Aspergillus oryzae***

Accurate classification of *A. flavus*, *A. oryzae*, and *Aspergillus parasiticus* is important because of biomedical and food safety concerns (Fig. 3f). *A. flavus* is a human pathogen, post-harvest food pathogen, and mycotoxin producer (19, 55, 56). *A. oryzae* is the domesticated *A. flavus* used for food fermentation (e.g., sake and miso production) (20). 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, rather than distinct species or populations. Notably, *A. oryzae* strains are known to produce fewer mycotoxins than *A. flavus* (20), which may be a diagnostic signature (i.e., phenotype) of an ecotype appropriate for food production.

Interestingly, the genomes of NRRL 2999 and E136, which represent strains of the aflatoxigenic post-harvest pathogen *A. parasiticus*, reside in a clade with *A. flavus* and *A. oryzae* strains. Notably, the two *A. parasiticus* strains do not form a monophyletic group. Similar to a previous report, strain NRRL 2999 is misidentified as *A. parasiticus* but is, in fact, *A. flavus* (22). Five strains of *A. flavus* are sister to *A. minisclerotigenes* and more distantly related to other *A. flavus* strains; our genome-scale analyses are also consistent with inferences based on a recent examination of individual barcode loci (22) in suggesting that the five *A. flavus* isolates are misidentified. Reexamination of the taxonomic loci by conducting sequence similarity searches against a gold standard



database among these five isolates supported this finding (Table S3 at <https://doi.org/10.6084/m9.figshare.21382131>).

### Misidentified/mislabeled reference genomes

Surprisingly, we identified instances where the type strain of a species (i.e., the representative isolate for a species) was misidentified/mislabeled. For example, phylogenomics indicated *Aspergillus bombycis* NRRL 26010 is mislabeled and is, in fact, *Aspergillus luteovirescens* (57). This finding raises awareness about how formerly used taxonomic names persist in databases. A similar observation was made for five species from the section *Nigri* and one species from the series *Versicolores* (Table 1). Thus, the genomes of seven strains are mislabeled as type strains in NCBI, which may misguide various experiments.

### Species determination for unlabeled genomes

We also identified 11 isolates with unknown species designations (Table 1 and S6 at <https://doi.org/10.6084/m9.figshare.21382131>). Phylogenomics confidently provided species labels for most isolates. For example, strain GbtcF2 was confidently determined to be *A. flavus* because phylogenomics revealed GbtcF2 is monophyletic with other *A. flavus* isolates (Fig. 3f). However, we identified instances where population-level genome sequences were unavailable, which limited insights from phylogenomic analyses. For example, although sequence similarity searches of taxonomically informative loci against a gold standard database indicated that strain ADI1 is *Aspergillus pseudodeflectus*, the lack of genome sequences from this species made it impossible to make this inference from our phylogenomic analyses.

Similarly, there is no phylogenomic-based guidance for identifying novel species. For example, phylogenomics was unable to determine the species designation of strain MEXU 27854; however, sequence similarity searches of taxonomically informative loci against other *Aspergillus* genomes indicated that strain MEXU 27854 is likely a novel *Aspergillus* species (Table S3 at <https://doi.org/10.6084/m9.figshare.21382131>).

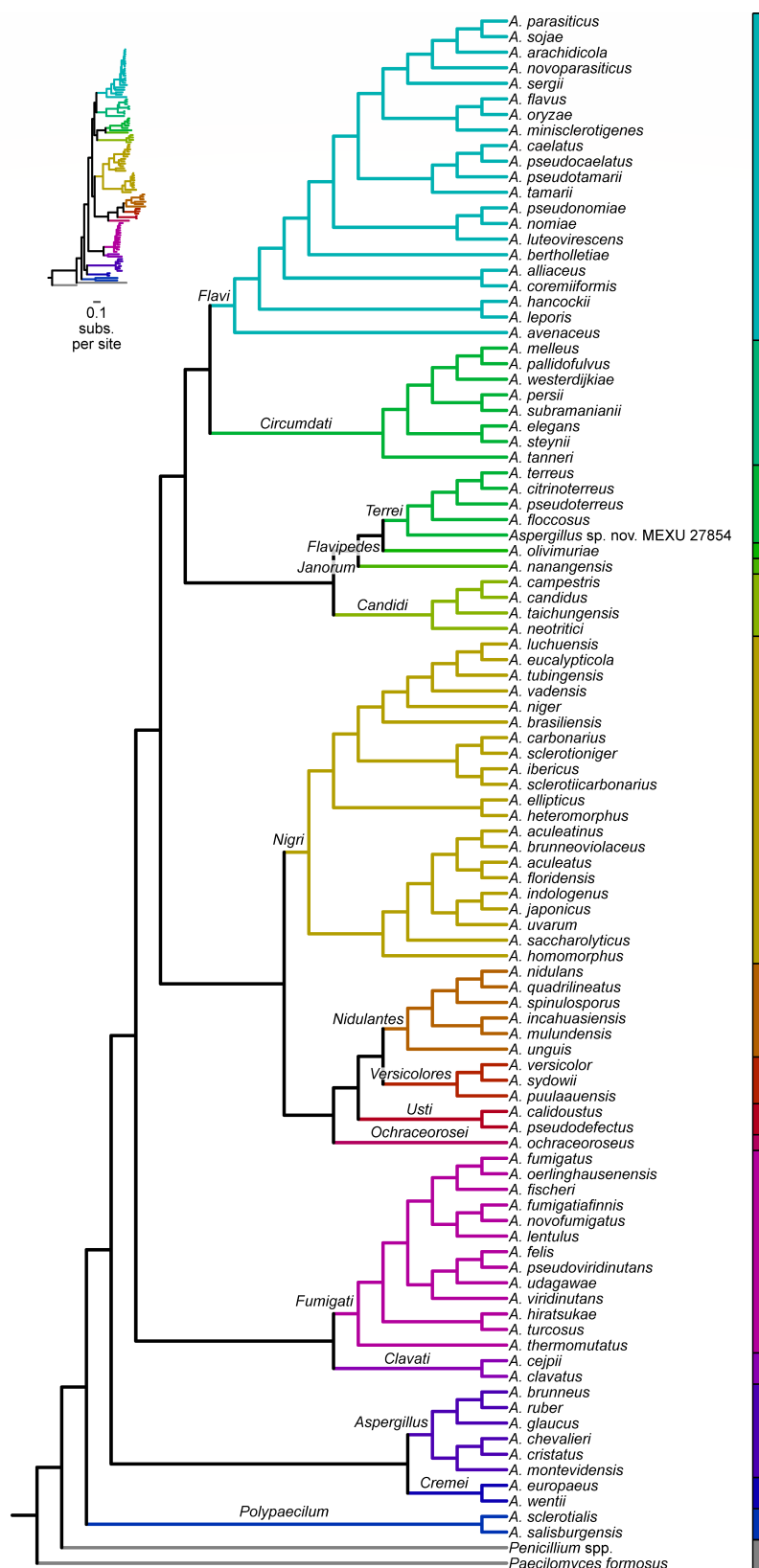
### Mislabeled among outgroup taxa

While confirming species designations using taxonomically informative loci, misidentified outgroup taxa were also identified. Specifically, *Penicillium chrysogenum* P2niaD18 and *Penicillium coprophilum* IBT 31321 are misidentified; instead, isolates P2niaD18 and IBT 31321 are *Penicillium rubens* and *Penicillium brevistipitatum*, respectively. Similarly, strain MCCF 102 is not an *Aspergillus* but a *P. formosus* (Table 1).

These findings suggest that issues of species, even genus, misidentification are prevalent both within and beyond *Aspergillus*, raising the possibility that misidentification is pervasive across microbes. Indeed, misidentification has been reported in bacterial genera such as *Burkholderia* and *Streptococcus* (58, 59) and genera of *Saccharomycotina* yeasts such as *Candida* and *Naumovozyma* (60, 61). Genome sequencing of multiple strains, including the type strains, of more *Aspergillus* species will further facilitate phylogenomic-based taxonomy.

### 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 (22). Strain misidentification can be reduced by two factors: genome-scale data, which are less prone to errors in phylogenetic inference compared to single or a few loci (23, 33, 38, 62–64), and population-level sampling from diverse environmental niches and geographic locations. Combined, these two factors facilitated identifying cases where strains may represent distinct species (such as *A. neoellipticus*) or were misidentified (such as the case of *A. niger* and *A. welwitschiae* and



**FIG 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. Note, *Versicolores* is a series within the section *Nidulantes*.

species in section *Flavi*). 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 prioritizing the genome sequencing of type strains. However, we identified that seven reference genomes were mislabeled in NCBI, suggesting that genome sequencing of multiple strains for each species will also be helpful. Genome-scale resources can be utilized for strain classification, similar to databases of internal transcribed spacer regions of fungi that help facilitate isolate and strain classification (6, 65–68). This approach is of growing interest among biologists working on *Saccharomycotina* yeasts and cryptic species of *Aspergillus* pathogens (23, 26, 28, 69, 70), but has yet to be well adapted for other fungal lineages. Currently, analysis of taxonomically informative loci remains the gold standard and should be cross-referenced with the latest, accepted taxonomic information available (54, 71). Working with NCBI curators to correct known mislabeling issues should also be a priority for members of the research community.

Adopting a genome-scale approach will require extensive genome sequencing efforts, such that publicly available genome sequence information is on par with publicly available single-locus information and well-curated genome-scale markers. To this end, we have produced profile Hidden Markov Models of the 1,362 molecular markers used in the present study that can facilitate genome-scale determination of *Aspergillus* species (see Data Availability). These markers, combined with our phylogenomic tree, may be helpful resources for accurate species determination among newly sequenced *Aspergillus* genomes. Coupling genome-scale data with phenotypic information may also help resolve certain controversies where DNA alone is insufficient. Incorporating phenotypic information will also help leverage copious amounts of data and taxonomic information generated prior to the genomics era (51, 72). Moreover, integrating phylogenomic and high-throughput phenotypic information may unite traditional taxonomic approaches with big data, allowing for multiple dimensions of information to elucidate species boundaries and determinations, an approach referred to, by us and others, as “taxogenomics” (73–76).

Additional research is needed to identify and characterize cryptic species, organisms morphologically highly similar to known species but genetically and physiologically distinct (23, 56, 77). Accurate strain identification and elucidation of species boundaries will greatly benefit from increased genome sequencing of under-represented species and populations thereof.

## Conclusion

Using a phylogenomic framework, dense sampling of genes, species, and populations is helpful for species identification. We demonstrate this using biomedically and technologically important fungi from the genus *Aspergillus* using 725 genomes and 1,362 high-quality molecular markers. In doing so, we resolved ongoing taxonomic controversies, identified new ones, and revealed extensive strain misidentification. Together, these analyses underscored the considerable amounts of inaccurate information concerning *Aspergillus* genomes and the importance of population-level and species-level sampling for identifying these inaccuracies. Our finding that some outgroup taxa were misidentified and that one isolate labeled as *Aspergillus* was *Paecilomyces* indicates that the issue of strain misidentification may be pervasive among microbes. Our ability to identify and resolve taxonomic controversies was aided by a multi-disciplinary approach, leaning on expertise from mycologists, taxonomists, bioinformaticians, and phylogenomicists. Although our approach was efficacious, it relies on the availability of genome sequences from type strains and populations. Thus, we advocate for additional genome sequencing across the genus and suggest that our approach is best supported by the traditional identification method. However, as the number of genomes increases across the Tree of Life rapidly, we anticipate that our framework and collaborative approach can be broadly applied to ensure accurate species identification.

## MATERIALS AND 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: 9 January 2022). Publicly available genome annotations were also downloaded. For genomes without available annotations, gene boundaries were predicted using AUGUSTUS, v3.3.2 (78), with the “species” parameter set to “*aspergillus\_nidulans*.” To determine if the genomes were suitable for phylogenomic analyses, gene prediction completeness was examined using BUSCO, v4.0.4 (79), and the Eurotiales database of 4,191 near-universally single-copy orthologs (or BUSCO genes) from OrthoDB, v10 (80). Six genomes with less than 75% single-copy complete BUSCO genes were removed, resulting in 711 genomes labeled as *Aspergillus* in NCBI. 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 was assessed using the same protocol and was highly complete (mean  $\pm$  standard deviation: 94.73%  $\pm$  4.04%). The final data set had 725 genomes.

### Single-copy orthologous gene identification

Phylogenomics often relies on single-copy orthologous genes. OrthoFinder, v2.3.8 (81), was used to identify single-copy orthologous genes by clustering protein sequences into groups of orthologs. Clustering sequences were based on protein sequence similarity and calculated using DIAMOND, v2.0.13.151 (82). To reduce computation time and memory, orthology predictions were conducted among 40 representative species that span the diversity of *Aspergillus* species (Table S7 at <https://doi.org/10.6084/m9.figshare.21382131>) (33). 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 data set, sequence similarity searches were conducted in each proteome. To do so, the 7,882 single-copy orthologs were aligned using MAFFT, v7.402 (83), with the auto parameter. Profile Hidden Markov Models (HMMs) were then built for each alignment using the hmmbuild function in HMMER, v3.1b2 (84). The resulting HMMs were used to identify single-copy orthologs in the 725 proteomes using orthofisher, v1.0.3 (85), 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 (85), were aligned using MAFFT as described above (83). The corresponding nucleotide sequences were threaded onto the protein alignment using the thread\_dna function in PhyKIT, v1.11.12 (86) and trimmed using ClipKIT, v1.3.0 (87). Excessive trimming of multiple sequence alignments worsens single-gene phylogenetic inference (87, 88); 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 (89). The best-fitting substitution model was selected using ModelFinder (90). To remove potential instances of hidden paralogy, the monophyly of the five well-established lineages was examined using PhyKIT (86). 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  $0.98 \pm 0.07$ .

## 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 (86). The resulting supermatrix had 6,378,237 sites (2,846,432 parsimony informative sites). The alignment length and number of parsimony informative sites were calculated using BioKIT, v0.1.2 (91). IQ-TREE 2 was used for tree inference (89). The best-fitting substitution model was selected using ModelFinder (90). Bipartition support was assessed using 1,000 ultrafast bootstrap approximations (92). 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 input to ASTRAL, v5.6.3 (93). Phylogenies were visualized using iTOL, v6 (94).

## Species confirmation using taxonomic loci

The current standard for species identification is sequence similarity using taxonomically informative loci. To conduct species determination using taxonomically informative loci, the best nucleotide-to-nucleotide BLAST hit to beta-tubulin, calmodulin, and/or RNA polymerase II second largest subunit (RPB2) gene regions were extracted from the genome assembly of each isolate. The query sequence for each sequence is as follows: calmodulin, GenBank identifier: [EF669865.1](#), Description: *Neosartorya fischeri* isolate NRRL 181 calmodulin gene, partial cds; beta-tubulin, GenBank identifier: [EF669796.1](#), Description: *Neosartorya fischeri* isolate NRRL 181 beta-tubulin gene, partial cds; and RPB2, GenBank identifier: [XM\\_677297.2](#), Description: *Aspergillus nidulans* FGSC A4 DNA-directed RNA polymerase II core subunit RPB2 (ANIA\_09120), partial mRNA. A sequence similarity search was conducted for each extracted sequence against a local database of taxonomically informative fungal sequences. The database comprised reference sequences sourced from GenBank, encompassing the accepted *Aspergillus* species listed by Houbraken et al. (54), as well as the new *Aspergillus* species described afterward.

## ACKNOWLEDGMENTS

We thank the Rokas lab for helpful discussions and comments.

J.L.S. and A.R. were supported by the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study program. J.L.S. is a Howard Hughes Medical Institute Awardee of the Life Sciences Research Foundation. Research in A.R.'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 D.S.H.'s lab is supported by the National Science Foundation (DEB-1456588). Research in J.G.G.'s lab is supported by the National Science Foundation (1942681). A.R. 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.

Conceptualization—J.L.S. and A.R. Data curation—J.L.S. Formal analysis—J.L.S., C.B., and H.A.R. Funding acquisition—J.L.S. and A.R. Investigation—J.L.S., C.B., and H.A.R. Methodology—J.L.S. and A.R. Project administration—J.L.S. and A.R. Resources—J.L.S. and A.R. Software—J.L.S. Supervision—G.H.G., D.M.G., J.H., D.S.H., and A.R. Validation—J.L.S. and J.H. Visualization—J.L.S. and A.R. Writing—original draft—J.L.S. and A.R. Writing—review and editing—all authors.



# AUTHOR AFFILIATIONS

- <sup>1</sup>Howards Hughes Medical Institute and the Department of Molecular and Cell Biology, University of California, Berkeley, California, USA
- <sup>2</sup>Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, USA
- <sup>3</sup>Evolutionary Studies Initiative, Vanderbilt University, Nashville, Tennessee, USA
- <sup>4</sup>Department of Chemistry and Biochemistry, University of North Carolina at Greensboro, Greensboro, North Carolina, USA
- <sup>5</sup>Department of Plant Pathology, University of California, Davis, California, USA
- <sup>6</sup>USDA-ARS, Salinas, California, USA
- <sup>7</sup>Shmunis School of Biomedicine and Cancer Research, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv, Israel
- <sup>8</sup>Department of Food Science, University of Massachusetts, Amherst, Massachusetts, USA
- <sup>9</sup>Molecular and Cellular Biology Graduate Program, University of Massachusetts, Amherst, Massachusetts, USA
- <sup>10</sup>Organismic and Evolutionary Biology Graduate Program, University of Massachusetts, Amherst, Massachusetts, USA
- <sup>11</sup>Faculdade de Ciencias Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, São Paulo, Brazil
- <sup>12</sup>Department of Plant Pathology and Environmental Microbiology, Penn State University, University Park, Pennsylvania, USA
- <sup>13</sup>Food and Indoor Mycology, Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands
- <sup>14</sup>Biology Department, Clark University, Worcester, Massachusetts, USA
- <sup>15</sup>Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnengasse, Heidelberg, Germany

# AUTHOR ORCID*s*

- Jacob L. Steenwyk  <http://orcid.org/0000-0002-8436-595X>
- Huzefa A. Raja  <http://orcid.org/0000-0002-0824-9463>
- Ningxiao Li  <http://orcid.org/0000-0002-7847-3567>
- Judith Berman  <http://orcid.org/0000-0002-8577-0084>
- Nicholas H. Oberlies  <http://orcid.org/0000-0002-0354-8464>
- John G. Gibbons  <http://orcid.org/0000-0003-1282-9345>
- Gustavo H. Goldman  <http://orcid.org/0000-0002-2986-350X>
- David S. Hibbett  <http://orcid.org/0000-0002-9145-3165>
- Antonis Rokas  <http://orcid.org/0000-0002-7248-6551>

# FUNDING

Funder	Grant(s)	Author(s)
National Science Foundation (NSF)	DEB-1442113	Antonis Rokas
National Science Foundation (NSF)	DEB-2110404	Antonis Rokas
Life Sciences Research Foundation (LSRF)		Jacob L. Steenwyk
HHS   NIH   National Institute of Allergy and Infectious Diseases (NIAID)	R56 AI146096	Antonis Rokas
HHS   NIH   National Institute of Allergy and Infectious Diseases (NIAID)	R01 AI153356	Antonis Rokas
National Science Foundation (NSF)	1942681	John G. Gibbons

# AUTHOR CONTRIBUTIONS

Jacob L. Steenwyk, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Validation, Visualization, Writing – original

draft | Charu Balamurugan, Data curation, Formal analysis, Writing – review and editing | Huzefa A. Raja, Methodology, Resources, Writing – review and editing | Carla Gonçalves, Methodology, Resources, Writing – review and editing | Ningxiao Li, Methodology, Resources, Writing – review and editing | Frank Martin, Methodology, Resources, Writing – review and editing | Judith Berman, Methodology, Resources, Writing – review and editing | Nicholas H. Oberlies, Methodology, Resources, Writing – review and editing | John G. Gibbons, Methodology, Resources, Writing – review and editing | Gustavo H. Goldman, Methodology, Resources, Writing – review and editing | David M. Geiser, Methodology, Resources, Writing – review and editing | Jos Houbroken, Methodology, Resources, Writing – review and editing | David S. Hibbett, Methodology, Resources, Writing – review and editing.

## DATA AVAILABILITY

Results and supplementary data presented in this study are available from figshare at <https://doi.org/10.6084/m9.figshare.21382131>.

## REFERENCES

- De Queiroz K. 2007. Species concepts and species delimitation. *Syst Biol* 56:879–886. <https://doi.org/10.1080/10635150701701083>
- Houbroken J, de Vries RP, Samson RA. 2014. Modern taxonomy of biotechnologically important *Aspergillus* and *Penicillium* species, p 199–249. In *Advances in applied microbiology*
- Xu J. 2020. Fungal species concepts in the genomics era. *Genome* 63:459–468. <https://doi.org/10.1139/gen-2020-0022>
- Aime MC, Miller AN, Aoki T, Bensch K, Cai L, Crous PW, Hawksworth DL, Hyde KD, Kirk PM, Lücking R, May TW, Malosso E, Redhead SA, Rossman AY, Stadler M, Thines M, Yurkov AM, Zhang N, Schoch CL. 2021. How to publish a new fungal species, or name, version 3.0. *IMA Fungus* 12:11. <https://doi.org/10.1186/s43008-021-00063-1>
- Dyer PS, O’Gorman CM. 2012. Sexual development and cryptic sexuality in fungi: insights from *Aspergillus* species. *FEMS Microbiol Rev* 36:165–192. <https://doi.org/10.1111/j.1574-6976.2011.00308.x>
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Bolchacova E, Voigt K, Crous PW, 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. <https://doi.org/10.1073/pnas.1117018109>
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Fungal Barcoding Consortium, Fungal Barcoding Consortium Author List. 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. <https://doi.org/10.1073/pnas.1117018109>
- Raja HA, Miller AN, Pearce CJ, Oberlies NH. 2017. Fungal identification using molecular tools: a primer for the natural products research community. *J Nat Prod* 80:756–770. <https://doi.org/10.1021/acs.jnatprod.6b01085>
- Arias RS, Orner VA, Martinez-Castillo J, Sobolev VS. 2021. *Aspergillus* section *Flavi*, need for a robust taxonomy. *Microbiol Resour Announc* 10:e0078421. <https://doi.org/10.1128/MRA.00784-21>
- Fountain JC, Clevenger JP, Vaughn JN, Guo B. 2021. Lessons learned: the importance of biological curation. *Microbiol Resour Announc* 10:e0047321. <https://doi.org/10.1128/MRA.00473-21>
- Kanamasa S. 2021. Opinion on updating the taxonomic names of *Aspergillus terreus* strains IFO 6365 and TN-484. *Microbiol Resour Announc* 10:e0047221. <https://doi.org/10.1128/MRA.00472-21>
- Kück U, Dahlmann TA. 2021. The confusion in renaming species: *Penicillium chrysogenum* and *Penicillium rubens*. *Microbiol Resour Announc* 10:e0046421. <https://doi.org/10.1128/MRA.00464-21>
- Li L, Yu Z. 2021. Changing the taxonomic classification of strain NRRL5109 to *Aspergillus fumigatus*. *Microbiol Resour Announc* 10:e0047521. <https://doi.org/10.1128/MRA.00475-21>
- Ruiz ON, Radwan O. 2021. Difficulty in assigning fungal identity based on DNA sequences. *Microbiol Resour Announc* 10:e0046021. <https://doi.org/10.1128/MRA.00460-21>
- Stevenson B, 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:e0045821. <https://doi.org/10.1128/MRA.00458-21>
- Venkateswaran K, Wood JM, Simpson A, Singh N. 2021. Reliable identification of environmental fungi requires multigene-based phylogenetic analyses. *Microbiol Resour Announc* 10:e0046921. <https://doi.org/10.1128/MRA.00469-21>
- Yin G, Jurick WM, Zhao G, Bennett JW. 2021. New names for three *Penicillium* strains based on updated barcoding and phylogenetic analyses. *Microbiol Resour Announc* 10:e0046621. <https://doi.org/10.1128/MRA.00466-21>
- Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, Berriman M, Abe K, Archer DB, Bermejo C, et al. 2005. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* 438:1151–1156. <https://doi.org/10.1038/nature04332>
- Hedayati MT, Pasqualotto AC, Warn PA, Bowyer P, Denning DW. 2007. *Aspergillus flavus*: human pathogen, allergen and mycotoxin producer. *Microbiology (Reading)* 153:1677–1692. <https://doi.org/10.1099/mic.0.2007/007641-0>
- Gibbons JG, Salichos L, Slot JC, Rinker DC, McGary KL, King JG, Klich MA, Tabb DL, McDonald WH, Rokas A. 2012. The evolutionary imprint of domestication on genome variation and function of the filamentous fungus *Aspergillus oryzae*. *Curr Biol* 22:1403–1409. <https://doi.org/10.1016/j.cub.2012.05.033>
- Cairns TC, Nai C, Meyer V. 2018. How a fungus shapes biotechnology: 100 years of *Aspergillus niger* research. *Fungal Biol Biotechnol* 5:13. <https://doi.org/10.1186/s40694-018-0054-5>
- Houbroken J, Visagie CM, Frisvad JC. 2021. Recommendations to prevent taxonomic misidentification of genome-sequenced fungal strains. *Microbiol Resour Announc* 10:e01074-20. <https://doi.org/10.1128/MRA.01074-20>
- Steenwyk JL, Rokas A, Goldman GH. 2023. Know the enemy and know yourself: addressing cryptic fungal pathogens of humans and beyond. *PLoS Pathog* 19:e1011704. <https://doi.org/10.1371/journal.ppat.1011704>
- dos Santos RAC, Steenwyk JL, Rivero-Menendez O, Mead ME, Silva LP, Bastos RW, Alastruey-Izquierdo A, Goldman GH, Rokas A. 2020. Genomic and phenotypic heterogeneity of clinical isolates of the human pathogens *Aspergillus fumigatus*, *Aspergillus lentulus*, and *Aspergillus fumigatiformis*. *Front Genet* 11:459. <https://doi.org/10.3389/fgene.2020.00459>
- Steenwyk JL, Mead ME, Knowles SL, Raja HA, Roberts CD, Bader O, Houbroken J, Goldman GH, Oberlies NH, Rokas A. 2020. Variation among biosynthetic gene clusters, secondary metabolite profiles, and cards of virulence across *Aspergillus* species. *Genetics* 216:481–497. <https://doi.org/10.1534/genetics.120.303549>
- Steenwyk JL, Knowles S, Bastos RW, Balamurugan C, Rinker D, Mead ME, Roberts CD, Raja HA, Li Y, Colabardini AC, de Castro PA, Dos Reis TF,

- Canóvas D, Sanchez RL, Lagrou K, Torrado E, Rodrigues F, Oberlies NH, Zhou X, Goldman GH, Rokas A. 2023. Evolutionary origin, population diversity, and diagnostics for a cryptic hybrid pathogen. *bioRxiv*:2023.07.03.547508. <https://doi.org/10.1101/2023.07.03.547508>
27. Tam EWT, Chen JHK, Lau ECL, Ngan AHY, Fung KSC, Lee K-C, Lam C-W, Yuen K-Y, Lau SKP, Woo PCY. 2014. Misidentification of *Aspergillus nomius* and *Aspergillus tamarii* as *Aspergillus flavus*: characterization by internal transcribed spacer,  $\beta$ -tubulin, and calmodulin gene sequencing, metabolic fingerprinting, and matrix-assisted laser desorption/ionization-time of flight mass spectrometry. *J Clin Microbiol* 52:1153–1160. <https://doi.org/10.1128/JCM.03258-13>
  28. Steenwyk JL, Lind AL, Ries LNA, Dos Reis TF, Silva LP, Almeida F, Bastos RW, Fraga da Silva T de C, Bonato VLD, Pessoni AM, Rodrigues F, Raja HA, Knowles SL, Oberlies NH, Lagrou K, Goldman GH, Rokas A. 2020. Pathogenic allopolyploid hybrids of *Aspergillus* fungi. *Curr Biol* 30:2495–2507. <https://doi.org/10.1016/j.cub.2020.04.071>
  29. Houbbraken 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. <https://doi.org/10.1093/femsle/fnv236>
  30. Fungaro MHP, Ferranti LS, Massi FP, da Silva JJ, Sartori D, Taniwaki MH, Frisvad JC, Iamanaka BT. 2017. *Aspergillus labruscus* sp. nov., a new species of *Aspergillus* section *Nigri* discovered in Brazil. *Sci Rep* 7:6203. <https://doi.org/10.1038/s41598-017-06589-y>
  31. Pitt JI, Taylor JW. 2014. *Aspergillus*, its sexual states and the new international code of nomenclature. *Mycologia* 106:1051–1062. <https://doi.org/10.3852/14-060>
  32. Taylor JW, Göker M, Pitt JI. 2016. Choosing one name for pleomorphic fungi: the example of *Aspergillus* versus *Eurotium*, *Neosartorya* and *Emmericella*. *Taxon* 65:593–601. <https://doi.org/10.12705/653.10>
  33. Steenwyk JL, Shen X-X, Lind AL, Goldman GH, Rokas A. 2019. A robust phylogenomic time tree for biotechnologically and medically important fungi in the genera *Aspergillus* and *Penicillium*. *mBio* 10:e00925-19. <https://doi.org/10.1128/mBio.00925-19>
  34. de Vries RP, Riley R, Wiebenga A, Aguilar-Osorio G, Amillis S, Uchima CA, Anderluh G, Asadollahi M, Askin M, Barry K, et al. 2017. Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*. *Genome Biol* 18:28. <https://doi.org/10.1186/s13059-017-1151-0>
  35. Kjærboelling I, Vesth T, Frisvad JC, Nybo JL, Theobald S, Kildgaard S, Petersen TI, Kuo A, Sato A, Lyhne EK, et al. 2020. A comparative genomics study of 23 *Aspergillus* species from section *Flavi*. *Nat Commun* 11:1106. <https://doi.org/10.1038/s41467-019-14051-y>
  36. Balajee SA. 2009. *Aspergillus terreus* complex. *Med Mycol* 47 Suppl 1:S42–S46. <https://doi.org/10.1080/13693780802562092>
  37. Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, et al. 2009. Database resources of the national center for biotechnology information. *Nucleic Acids Res* 37:D5–D15. <https://doi.org/10.1093/nar/gkn741>
  38. Rokas A, Williams BL, King N, Carroll SB. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425:798–804. <https://doi.org/10.1038/nature02053>
  39. Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* 30:i541–i548. <https://doi.org/10.1093/bioinformatics/btu462>
  40. Rieseberg LH, Brouillet L. 1994. Are many plant species paraphyletic? *Taxon* 43:21–32. <https://doi.org/10.2307/1223457>
  41. Rieppel O. 2010. Species monophyly. *J Zool Syst Evol Res* 48:1–8. <https://doi.org/10.1111/j.1439-0469.2009.00545.x>
  42. 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. <https://doi.org/10.1007/s11046-014-9790-0>
  43. Hong S-B, Go S-J, Shin H-D, Frisvad JC, Samson RA. 2005. Polyphasic taxonomy of *Aspergillus fumigatus* and related species. *Mycologia* 97:1316–1329. <https://doi.org/10.3852/mycologia.97.6.1316>
  44. Peterson SW. 2008. Phylogenetic analysis of *Aspergillus* species using DNA sequences from four loci. *Mycologia* 100:205–226. <https://doi.org/10.3852/mycologia.100.2.205>
  45. 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:e01262-18. <https://doi.org/10.1128/MRA.01262-18>
  46. Mead ME, Steenwyk JL, Silva LP, de Castro PA, Saeed N, Hillmann F, Goldman GH, Rokas A. 2021. An evolutionary genomic approach reveals both conserved and species-specific genetic elements related to human disease in closely related *Aspergillus* fungi. *Genetics* 218:iyab066. <https://doi.org/10.1093/genetics/iyab066>
  47. Sugui JA, Peterson SW, Figat A, Hansen B, Samson RA, Mellado E, Cuenca-Estrella M, Kwon-Chung KJ. 2014. Genetic relatedness versus biological compatibility between *Aspergillus fumigatus* and related species. *J Clin Microbiol* 52:3707–3721. <https://doi.org/10.1128/JCM.01704-14>
  48. Susca A, Proctor RH, Morelli M, Haidukowski M, Gallo A, Logrieco AF, Moretti A. 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* 7:1412. <https://doi.org/10.3389/fmicb.2016.01412>
  49. Coutinho WM, Suassuna ND, Luz CM, Suinaga FA, Silva ORRF. 2006. Bole rot of sisal caused by *Aspergillus niger* in Brazil. *Fitopatol Bras* 31:605–605. <https://doi.org/10.1590/S0100-41582006000600014>
  50. Duarte EAA, Damasceno CL, de Oliveira TAS, Barbosa L de O, Martins FM, de Queiroz Silva JR, de Lima TEF, da Silva RM, Kato RB, Bortolini DE, Azevedo V, Góes-Neto A, Soares ACF. 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. <https://doi.org/10.3389/fmicb.2018.01227>
  51. Bian C, Kusuya Y, Sklenář F, D'hooge E, Yaguchi T, Ban S, Visagie CM, Houbbraken J, Takahashi H, Hubka V. 2022. Reducing the number of accepted species in *Aspergillus* series *Nigri*. *Stud Mycol* 102:95–132. <https://doi.org/10.3114/sim.2022.102.03>
  52. Sklenář F, Glässnerová K, Jurjević Ž, Houbbraken J, Samson RA, Visagie CM, Yilmaz N, Gené J, Cano J, Chen AJ, Nováková A, Yaguchi T, Kolářik M, Hubka V. 2022. Taxonomy of *Aspergillus* series *Versicolores*: species reduction and lessons learned about intraspecific variability. *Stud Mycol* 102:53–93. <https://doi.org/10.3114/sim.2022.102.02>
  53. Jurjević Z, Peterson SW, Horn BW. 2012. *Aspergillus* section *Versicolores*: nine new species and multilocus DNA sequence based phylogeny. *IMA Fungus* 3:59–79. <https://doi.org/10.5598/ima fungus.2012.03.01.07>
  54. Houbbraken J, Kocsu S, Visagie CM, Yilmaz N, Wang X-C, Meijer M, Kraak B, Hubka V, Bensch K, Samson RA, Frisvad JC. 2020. Classification of *Aspergillus*, *Penicillium*, *Talaromyces* and related genera (*Eurotiales*): an overview of families, genera, subgenera, sections, series and species. *Stud Mycol* 95:5–169. <https://doi.org/10.1016/j.simyco.2020.05.002>
  55. Gourama H, Bullerman LB. 1995. *Aspergillus flavus* and *Aspergillus parasiticus*: aflatoxinogenic fungi of concern in foods and feeds: a review. *J Food Prot* 58:1395–1404. <https://doi.org/10.4315/0362-028X-58.12.1395>
  56. Geiser DM, Pitt JI, Taylor JW. 1998. Cryptic speciation and recombination in the aflatoxin-producing fungus *Aspergillus flavus*. *Proc Natl Acad Sci U S A* 95:388–393. <https://doi.org/10.1073/pnas.95.1.388>
  57. Frisvad JC, Hubka V, Ezekiel CN, Hong S-B, Nováková A, Chen AJ, Arzanlou M, Larsen TO, Sklenář F, Mahakarnchanakul W, Samson RA, Houbbraken J. 2019. Taxonomy of *Aspergillus* section *Flavi* and their production of aflatoxins, ochratoxins and other mycotoxins. *Stud Mycol* 93:1–63. <https://doi.org/10.1016/j.simyco.2018.06.001>
  58. Zong Z, Wang X, Deng Y, Zhou T. 2012. Misidentification of *Burkholderia pseudomallei* as *Burkholderia cepacia* by the VITEK 2 system. *J Med Microbiol* 61:1483–1484. <https://doi.org/10.1099/jmm.0.041525-0>
  59. Sadowy E, Bojarska A, Kuch A, Skoczyska A, Jolley KA, Maiden MCJ, van Tonder AJ, Hammerschmidt S, Hryniewicz W. 2020. Relationships among streptococci from the mitis group, misidentified as *Streptococcus pneumoniae*. *Eur J Clin Microbiol Infect Dis* 39:1865–1878. <https://doi.org/10.1007/s10096-020-03916-6>
  60. Huang Y-S, Wang F-D, Chen Y-C, Huang Y-T, Hsieh M-H, Hii I-M, Lee Y-L, Ho M-W, Liu C-E, Chen Y-H, Liu W-L. 2021. High rates of misidentification of uncommon *Candida* species causing bloodstream infections using conventional phenotypic methods. *J Formos Med Assoc* 120:1179–1187. <https://doi.org/10.1016/j.jfma.2020.11.002>
  61. Stavrou AA, Mixão V, Boekhout T, Gabaldón T. 2018. Misidentification of genome assemblies in public databases: the case of *Naumovozyma*

- dairenensis* and proposal of a protocol to correct misidentifications. *Yeast* 35:425–429. <https://doi.org/10.1002/yea.3303>
62. Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320–1331. <https://doi.org/10.1126/science.1253451>
  63. Steenwyk JL, Li Y, Zhou X, Shen X-X, Rokas A. 2023. Incongruence in the phylogenomics era. *Nat Rev Genet* 24:834–850. <https://doi.org/10.1038/s41576-023-00620-x>
  64. Björnson S, Upham N, Verbruggen H, Steenwyk J. 2023. Phylogenomic inference, divergence-time calibration, and methods for characterizing reticulate evolution. Preprints. <https://doi.org/10.20944/preprints202309.0905.v1>
  65. Hibbett D, Abarenkov K, Kõljalg U, Öpik M, Chai B, Cole J, Wang Q, Crous P, Robert V, Helgason T, Herr JR, Kirk P, Lueschow S, O'Donnell K, Nilsson RH, Oono R, Schoch C, Smyth C, Walker DM, Porras-Alfaro A, Taylor JW, Geiser DM. 2016. Sequence-based classification and identification of fungi. *Mycologia* 108:1049–1068. <https://doi.org/10.3852/16-130>
  66. Lücking R, Aime MC, Robbertse B, Miller AN, Aoki T, Ariyawansa HA, Cardinali G, Crous PW, Druzhinina IS, Geiser DM, et al. 2021. Fungal taxonomy and sequence-based nomenclature. *Nat Microbiol* 6:540–548. <https://doi.org/10.1038/s41564-021-00888-x>
  67. Bradshaw MJ, Guan G-X, Nokes L, Braun U, Liu S-Y, Pfister DH. 2022. Secondary DNA barcodes (CAM, GAPDH, GS, and Rpb2) to characterize species complexes and strengthen the powdery mildew phylogeny. *Front Ecol Evol* 10:918908. <https://doi.org/10.3389/fevo.2022.918908>
  68. Bradshaw MJ, Aime MC, Rokas A, Maust A, Moparthi S, Jellings K, Pane AM, Hendricks D, Pandey B, Li Y, Pfister DH. 2023. Extensive intragenomic variation in the internal transcribed spacer region of fungi. *iScience* 26:107317. <https://doi.org/10.1016/j.isci.2023.107317>
  69. Libkind D, Čadež N, Opulente DA, Langdon QK, Rosa CA, Sampaio JP, Gonçalves P, Hittinger CT, Lachance MA. 2020. Towards yeast taxogenomics: lessons from novel species descriptions based on complete genome sequences. *FEMS Yeast Res* 20:foaa042. <https://doi.org/10.1093/femsyr/foaa042>
  70. Opulente DA, LaBella AL, Harrison M-C, Wolters JF, Liu C, Li Y, Kominek J, Steenwyk JL, Stoneman HR, VanDenAvond J, et al. 2023. Genomic and ecological factors shaping specialism and generalism across an entire subphylum. *bioRxiv*. <https://doi.org/10.1101/2023.06.19.545611>
  71. Visagie CM, Yilmaz N, Kocsubé S, Frisvad JC, Hubka V, Samson RA, Houbraken J. 2024. A review of recently introduced *Aspergillus*, *Penicillium*, *Talaromyces* and other *Eurotiales* species. *Stud Mycol*. <https://doi.org/10.3114/sim.2024.107.01>
  72. Glässerová K, Sklenář F, Jurjević Ž, Houbraken J, Yaguchi T, Visagie CM, Gené J, Siqueira JPZ, Kubátová A, Kolařík M, Hubka V. 2022. A monograph of *Aspergillus* section *Candidi*. *Stud Mycol* 102:1–51. <https://doi.org/10.3114/sim.2022.102.01>
  73. Suresh G, Lodha TD, Indu B, Sasikala C, Ramana CV. 2019. Taxogenomics resolves conflict in the genus *Rhodobacter*: a two and half decades pending thought to reclassify the genus *Rhodobacter*. *Front Microbiol* 10:2480. <https://doi.org/10.3389/fmicb.2019.02480>
  74. Ramírez-Durán N, de la Haba RR, Vera-Gargallo B, Sánchez-Porro C, Alonso-Carmona S, Sandoval-Trujillo H, Ventosa A. 2021. Taxogenomic and comparative genomic analysis of the genus *Saccharomonospora* focused on the identification of biosynthetic clusters PKS and NRPS. *Front Microbiol* 12:603791. <https://doi.org/10.3389/fmicb.2021.603791>
  75. Groenewald M, Hittinger CT, Bensch K, Opulente DA, Shen X-X, Li Y, Liu C, LaBella AL, Zhou X, Limtong S, et al. 2023. A genome-informed higher rank classification of the biotechnologically important fungal subphylum *Saccharomycotina*. *Stud Mycol* 105:1–22. <https://doi.org/10.3114/sim.2023.105.01>
  76. Libkind D, Čadež N, Opulente DA, Langdon QK, Rosa CA, Sampaio JP, Gonçalves P, Hittinger CT, Lachance MA. 2020. Towards yeast taxogenomics: lessons from novel species descriptions based on complete genome sequences. *FEMS Yeast Res* 20:foaa042. <https://doi.org/10.1093/femsyr/foaa042>
  77. Balajee SA, Gribskov JL, Hanley E, Nickle D, Marr KA. 2005. *Aspergillus lentulus* sp. nov., a new sibling species of *A. fumigatus*. *Eukaryot Cell* 4:625–632. <https://doi.org/10.1128/EC.4.3.625-632.2005>
  78. Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19 Suppl 2:ii215–ii225. <https://doi.org/10.1093/bioinformatics/btg1080>
  79. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* 35:543–548. <https://doi.org/10.1093/molbev/msx319>
  80. Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, Zdobnov EM. 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res* 47:D807–D811. <https://doi.org/10.1093/nar/gky1053>
  81. Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20:238. <https://doi.org/10.1186/s13059-019-1832-y>
  82. Buchfink B, Reuter K, Drost H-G. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* 18:366–368. <https://doi.org/10.1038/s41592-021-01101-x>
  83. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>
  84. Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>
  85. Steenwyk JL, Rokas A. 2021. orthofisher: a broadly applicable tool for automated gene identification and retrieval. *G3 (Bethesda)* 11:jkab250. <https://doi.org/10.1093/g3journal/jkab250>
  86. Steenwyk JL, Buida TJ, Labella AL, Li Y, Shen X-X, Rokas A. 2021. PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. *Bioinformatics* 37:2325–2331. <https://doi.org/10.1093/bioinformatics/btab096>
  87. Steenwyk JL, Buida TJ, Li Y, Shen X-X, Rokas A. 2020. ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. *PLOS Biol*. 18:e3001007. <https://doi.org/10.1371/journal.pbio.3001007>
  88. Tan G, Muffato M, Ledergerber C, Herrero J, Goldman N, Gil M, Dessimoz C. 2015. Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Syst Biol* 64:778–791. <https://doi.org/10.1093/sysbio/syv033>
  89. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37:1530–1534. <https://doi.org/10.1093/molbev/msaa015>
  90. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>
  91. Steenwyk JL, Buida TJ, Gonçalves C, Goltz DC, Morales G, Mead ME, LaBella AL, Chavez CM, Schmitz JE, Hadjifrangiskou M, Li Y, Rokas A. 2022. BioKIT: a versatile toolkit for processing and analyzing diverse types of sequence data. *Genetics* 221:iyac079. <https://doi.org/10.1093/genetics/iyac079>
  92. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 35:518–522. <https://doi.org/10.1093/molbev/msx281>
  93. 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. <https://doi.org/10.1186/s12859-018-2129-y>
  94. Letunic I, Bork P. 2019. Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 47:W256–W259. <https://doi.org/10.1093/nar/gkz239>