



# Fungal taxonomy and sequence-based nomenclature

Robert Lücking<sup>1,2</sup>, M. Catherine Aime<sup>1</sup> , Barbara Robbertse<sup>4</sup>, Andrew N. Miller<sup>1</sup> , Takayuki Aoki<sup>2,6</sup>, Hiran A. Ariyawansa<sup>2,7</sup>, Gianluigi Cardinali<sup>8</sup>, Pedro W. Crous<sup>2,9,10</sup>, Irina S. Druzhinina<sup>1</sup> , David M. Geiser<sup>2,13</sup>, David L. Hawksworth<sup>1</sup> , Kevin D. Hyde<sup>2,18,19,20,21</sup>, Laszlo Irinyi<sup>1</sup> , Rajesh Jeewon<sup>26</sup>, Peter R. Johnston<sup>1</sup> , Paul M. Kirk<sup>28</sup>, Elaine Malosso<sup>2,29</sup>, Tom W. May<sup>1</sup> , Wieland Meyer<sup>22,23,24,25</sup>, Henrik R. Nilsson<sup>1</sup> , Maarja Öpik<sup>2,32</sup>, Vincent Robert<sup>8</sup>, Marc Stadler<sup>1</sup> , Marco Thines<sup>1</sup> , Duong Vu<sup>9</sup>, Andrey M. Yurkov<sup>1</sup> , Ning Zhang<sup>2,37</sup> and Conrad L. Schoch<sup>1</sup>  

**The identification and proper naming of microfungi, in particular plant, animal and human pathogens, remains challenging. Molecular identification is becoming the default approach for many fungal groups, and environmental metabarcoding is contributing an increasing amount of sequence data documenting fungal diversity on a global scale. This includes lineages represented only by sequence data. At present, these taxa cannot be formally described under the current nomenclature rules. By considering approaches used in bacterial taxonomy, we propose solutions for the nomenclature of taxa known only from sequences to facilitate consistent reporting and communication in the literature and public sequence repositories.**

The kingdom *Fungi* forms a highly diverse lineage of eukaryotes that shares a common ancestor with animals. Both comprise heterotrophic organisms, but *Fungi* form (chitinous) cell walls and are exclusively osmotrophic; that is, nutrient uptake is extracellular<sup>1</sup>. Although nearly 150,000 species of *Fungi* have been described, between 2.2 and 3.8 million are estimated to exist<sup>2</sup>. *Fungi* are ubiquitous and perform essential ecosystem processes and are of economic importance as agents of diseases or sources of biocontrol agents, biofuel, food and food additives, industrial enzymes and pharmaceuticals.

*Fungi* follow a simple body plan that underwent convergent evolution<sup>3</sup>, which makes it challenging to identify them with accuracy

and precision<sup>4</sup>. Methods in fungal taxonomy depend on whether a species is in culture, available as a dried fungarium sample or assessed in situ, and diagnostic tools encompass phenotype-based identification, physiological profiling and sequence-based DNA barcoding or phylogenetic reconstruction, including phylogenomics. In addition, the recent past has seen a shift towards laboratory-based approaches for many fungal groups, which has resulted in an increasing methodological overlap with prokaryote taxonomy<sup>4</sup>.

Some unrelated microbial groups have evolved life forms similar to *Fungi* and are studied by mycologists<sup>5</sup>. The most important are the *Oomycota* ('egg fungi'), which include the causal agents of late blight on potatoes (*Phytophthora infestans*) and white rust on mustards

<sup>1</sup>Botanischer Garten und Botanisches Museum, Freie Universität Berlin, Berlin, Germany. <sup>2</sup>International Commission on the Taxonomy of Fungi (ICTF).

<sup>3</sup>Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN, USA. <sup>4</sup>National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA. <sup>5</sup>Illinois Natural History Survey, University of Illinois, Champaign, IL, USA. <sup>6</sup>National Agriculture and Food Research Organization, Genetic Resources Center, Ibaraki, Japan. <sup>7</sup>Department of Plant Pathology and Microbiology, College of Bio-Resources and Agriculture, National Taiwan University, Taipei, Taiwan. <sup>8</sup>Department of Pharmaceutical Sciences, University of Perugia, Perugia, Italy. <sup>9</sup>Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands. <sup>10</sup>Laboratory of Phytopathology, Wageningen University and Research Centre (WUR), Wageningen, The Netherlands. <sup>11</sup>Fungal Genomics Laboratory (FungiG), Nanjing Agricultural University, Nanjing, China. <sup>12</sup>Institute of Chemical, Environmental and Bioscience Engineering (ICEBE), TU Wien, Vienna, Austria. <sup>13</sup>Department of Plant Pathology and Environmental Microbiology, The Pennsylvania State University, University Park, PA, USA. <sup>14</sup>Department of Life Sciences, The Natural History Museum, London, UK. <sup>15</sup>Comparative Plant and Fungal Biology, Royal Botanic Gardens, Kew, Richmond, UK. <sup>16</sup>Geography and Environment, University of Southampton, Southampton, UK. <sup>17</sup>Jilin Agricultural University, Changchun, China. <sup>18</sup>CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Science, Kunming, China. <sup>19</sup>Center of Excellence in Fungal Research, Mae Fah Luang University, Chiang Rai, Thailand. <sup>20</sup>World Agroforestry Centre, East and Central Asia, Kunming, China. <sup>21</sup>Mushroom Research Foundation, Chiang Rai, Thailand. <sup>22</sup>Molecular Mycology Research Laboratory, Centre for Infectious Diseases and Microbiology, Westmead Institute for Medical Research, Westmead, New South Wales, Australia. <sup>23</sup>Faculty of Medicine and Health, Sydney Medical School, Westmead Clinical School, The University of Sydney, Sydney, New South Wales, Australia. <sup>24</sup>Marie Bashir Institute for Infectious Diseases and Biosecurity, The University of Sydney, Sydney, New South Wales, Australia. <sup>25</sup>Westmead Hospital (Research and Education Network), Westmead, New South Wales, Australia. <sup>26</sup>Department of Health Sciences, Faculty of Medicine and Health Sciences, University of Mauritius, Reduit, Mauritius. <sup>27</sup>Manaaki Whenua—Landcare Research, Auckland, New Zealand. <sup>28</sup>Royal Botanic Gardens, Kew, Richmond, UK. <sup>29</sup>Departamento de Micología, Laboratório de Hifomicetos de Folheda, Centro de Biotecnologia, Universidade Federal de Pernambuco, Recife, Brazil. <sup>30</sup>Royal Botanic Gardens Victoria, Melbourne, Victoria, Australia. <sup>31</sup>Department of Biological and Environmental Sciences, Gothenburg Global Biodiversity Centre, University of Gothenburg, Göteborg, Sweden. <sup>32</sup>Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia. <sup>33</sup>Department Microbial Drugs, Helmholtz Center for Infection Research, Braunschweig, Germany. <sup>34</sup>German Centre for Infection Research (DZIF), Partner site Hannover-Braunschweig, Braunschweig, Germany. <sup>35</sup>Institute of Ecology, Evolution and Diversity, Goethe University, Frankfurt (Main), Germany. <sup>36</sup>Leibniz Institute DSMZ—German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany. <sup>37</sup>Department of Plant Biology, Rutgers University, New Brunswick, NJ, USA. <sup>✉</sup>e-mail: [schoch2@ncbi.nlm.nih.gov](mailto:schoch2@ncbi.nlm.nih.gov)

(*Albugo candida*) (Fig. 1a,b). Oomycota belong to the Straminipila, forming a heterotrophic lineage closely related to brown algae and diatoms<sup>5,6</sup>. The non-taxonomic term ‘fungi’, without italics and capitalization, is used to encompass *Fungi*, Oomycota and several other unrelated but fungus-like organisms. This is similar to the usage of ‘algae’ for unrelated organisms with an algal-like habit, for example, in the Archaeplastida, Straminipila and Rhizaria.

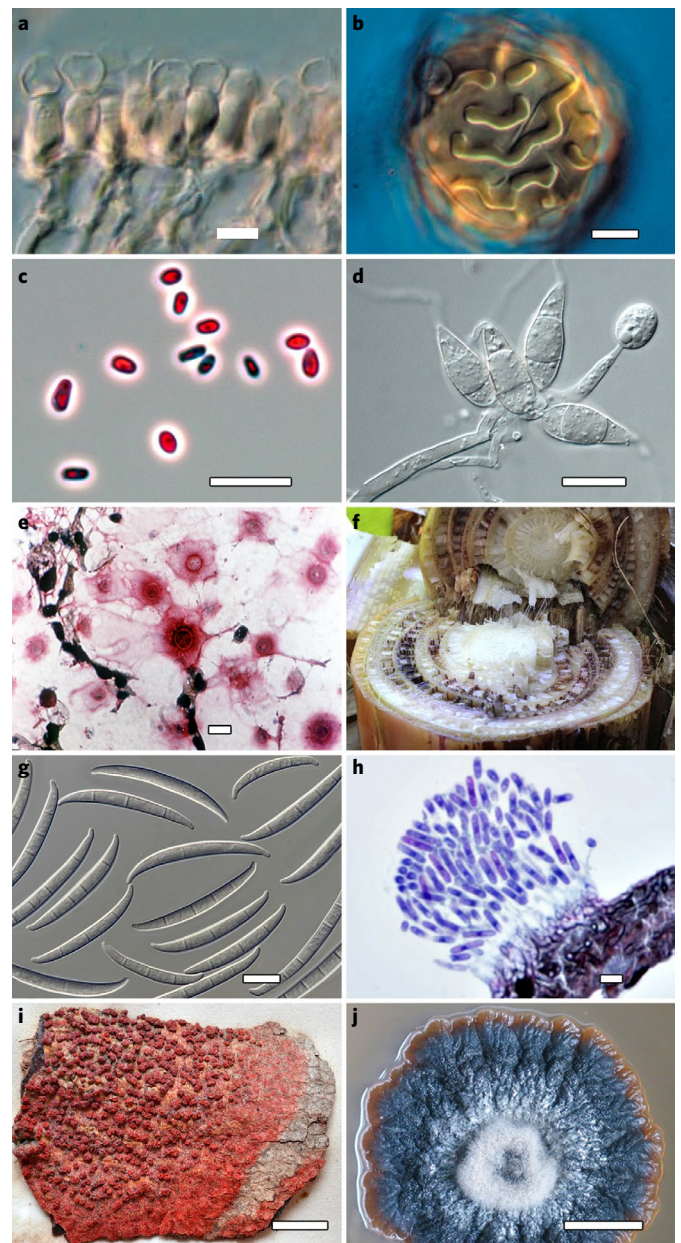
*Fungi* and fungus-like organisms are treated under the same nomenclatural rules set forth in the *International Code of Nomenclature for algae, fungi, and plants*<sup>7</sup> (ICNafp; Box 1). The ICNafp reflects advancements made along a historical timeline spanning three centuries (Fig. 2 and Supplementary Information). With fungal taxonomy beginning before the official start of scientific nomenclature in the middle of the eighteenth century, the nineteenth century brought substantial advancements in cataloguing fungal diversity and witnessed the discovery of lichen symbiosis. During the twentieth century, the mycological community became increasingly networked, and fungal systematics organized itself in platforms such as *Index of Fungi* and *Systema Ascomycetum*. The increasing importance and refinement of molecular approaches after the turn of the millennium led to substantial changes in fungal classifications and nomenclature (Fig. 2 and Supplementary Information).

Many fungi form phenotypically and biologically divergent life stages that reproduce sexually or asexually. Sexual and asexual morphs were traditionally named and classified separately under the concept of ‘dual nomenclature’<sup>8</sup>. Molecular systematics has documented numerous instances whereby differently named fungi represent different morphs of the same species. Therefore, mycologists adopted the principle of ‘one fungus=one name’<sup>8</sup>. Consequently, well-known fungal names, for example, in clinical mycology or plant pathology, may end up as synonyms of older, unfamiliar names. Such changes can, however, be held to a minimum due to new provisions in the ICNafp that allow greater flexibility in selecting the preferred name (Box 1).

### Providers and users of fungal taxonomy

Owing to the diversity of fungal biology, ranging from simple to complex life cycles and including multiple morphs within one species, a universal approach to unambiguously identify all fungi is currently not feasible. Identification is also often confused with the underlying processes of species delimitation and recognition<sup>4</sup>. Species delimited through molecular phylogenies may exhibit diagnostic phenotypic characteristics, which makes identifications feasible without molecular data. Likewise, resolving species complexes may require phylogenomic approaches, while the individual taxa may be identifiable with DNA barcoding markers. Consequently, species identification by a broad user community should require fewer, more readily available resources than species delimitation performed by a small number of taxonomic experts.

The level of accuracy and precision in fungal taxonomy depends on the group under study, the sample size, the available tools and the underlying objectives, and is particularly critical in plant pathology, food safety and clinical mycology<sup>4</sup>. While species name labels imply accuracy and precision, identifications may be incorrect or imprecise, which has consequences for understanding the biology, the distribution ranges and the conservation status of the underlying taxa. Therefore, verification of identifications is of crucial importance<sup>4</sup>. While physiological profiling is routinely used in food microbiology and clinical settings, molecular identification has become a major tool in fungal taxonomy. In addition to accuracy and precision, it offers universal coverage and broad applicability<sup>9</sup>. Regrettably, the increasing availability of molecular tools generates the misleading impression that DNA barcoding can replace taxonomic expertise, thereby overlooking the fact that such expertise remains indispensable to establish comprehensive



**Fig. 1 | Fungal diversity.** Although *Fungi* and fungus-like organisms exhibit striking phenotypic diversity, accurate and precise identification often requires molecular approaches or specific tools such as metabolic profiling owing to widespread cryptic diversification and a lack of diagnostic features in microscopic vegetative structures. **a**, *Albugo laibachii* (Oomycota) sporogenous hyphae. **b**, *Albugo candida* oospore. **c**, *Candida auris* cells. **d**, *Pyricularia oryzae* conidiophore with conidia. **e**, *Cryptococcus neoformans* cells in tissue. **f**, Banana plant infected with fusarium wilt. **g**, *Fusarium odoratissimum* macroconidia. **h**, *Colletotrichum siamense* section of acervulus. **i**, *Trypethelium purpurinum* (also known as *Marcelaria purpurina*) physical type specimen. **j**, *Chaetocapnodium tanzanicum* culture. Credit: photographs courtesy of Young-Joon Choi (**a,b**), Nani Maryani (**f,g**), Min Fu (**h**) and Jafar Abdollahzadeh (**j**). Scale bars, 10  $\mu$ m (**a-e**, **g** and **h**) or 10 mm (**i** and **j**).

and accurate reference databases for molecular identifications, a task that is still in its infancy.

A particular challenge for DNA-based identifications of fungi is the choice of genetic markers. While mycologists routinely use the nuclear internal transcribed spacer (ITS) region of the

**Box 1 | Regulating the naming of fungi**

The nomenclature of *Fungi* and fungus-like organisms is regulated by the ICNafp (<https://www.iapt-taxon.org/nomen/main.php>), which historically goes back to the first *Lois de la Nomenclature Botanique* (Laws of Botanical Nomenclature) published in 1867. A new edition of the ICNafp is prepared every 6 years following the deliberations of the Nomenclature Section of an International Botanical Congress. Since 2018, provisions that specifically relate to *Fungi* and organisms treated as fungi have been included in a special Chapter F of the ICNafp, which can be modified only during an International Mycological Congress (IMC), which occurs every 4 years. The IMC now appoints the Nomenclature Committee for Fungi (NCF) with around 20 members, which operates between IMCs and considers and votes on formal nomenclatural matters submitted to it, including those related to the conservation or rejection of names (<http://www.ima-mycology.org/nomenclature/nomenclature-committee-fungi>).

The International Commission on the Taxonomy of Fungi (ICTF), established in 1982, complements the NCF and has the primary remit of promoting sound taxonomic practice through disseminating guidelines and recommendations as to best practice and developing proposals for consideration by the NCF or IMCs. Membership is also decided at IMCs, and the ICTF now has 23 members supplemented by five subcommissions and a series of start-and-finish task-related working groups (<https://www.fungaltaxonomy.org>).

The ICNafp contains provisions that allow retention of names of fungi in current use even if they do not have priority, either through proposals to conserve or reject names or through lists of names proposed for protection, all of which are reviewed by the NCF. Other fungal-specific provisions include the sanctioning of names; that is, giving priority to names of fungi (except slime moulds) adopted in Persoon's *Synopsis Methodica Fungorum* (1801) and Fries' *Systema Mycologicum* (1821–1832), and the obligate registration of new fungal names (since 2012) and new typifications (since 2019).

ribosomal RNA operon as a standard DNA barcode<sup>9</sup>, the *Fusarium* and *Trichoderma* communities employ translation elongation factor 1 alpha (*TEF-1α*) as secondary barcode. In arbuscular mycorrhizal (AM) and in rust fungi, the small subunit (nuSSU) and the large subunit (nuLSU) of the nuclear ribosomal RNA operon are preferred, whereas the *Oomycota* community focuses on the mitochondrial cytochrome oxidase *c* subunits (*COX1* and *COX2*)<sup>4</sup>. Users must be aware of such community-specific approaches to prevent inaccurate or imprecise identifications, which is a problem arising particularly in environmental metabarcoding approaches<sup>1</sup>.

**Species identification in applied mycology**

Problems of correct identification and nomenclature are notorious in economically and medically important species, whereby different methods may result in conflicting identifications. The use of physiological profilers, such as API 20C, API ID32C and VITEK 2, has led to misidentification of the emerging, multidrug-resistant yeast *Candida auris* (Fig. 1c). It can accurately be identified using matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF) mass spectrometry or the ITS barcoding marker<sup>10</sup>. This example also shows how outdated taxonomy may result in misinformation. Classification of yeasts in vast, polyphyletic genera, such as *Candida*, gives the misleading impression that these species are closely related<sup>11</sup>. Multimarker and genome data indicate that *C. auris* is a member of the *Candida haemulonis* species complex, which is

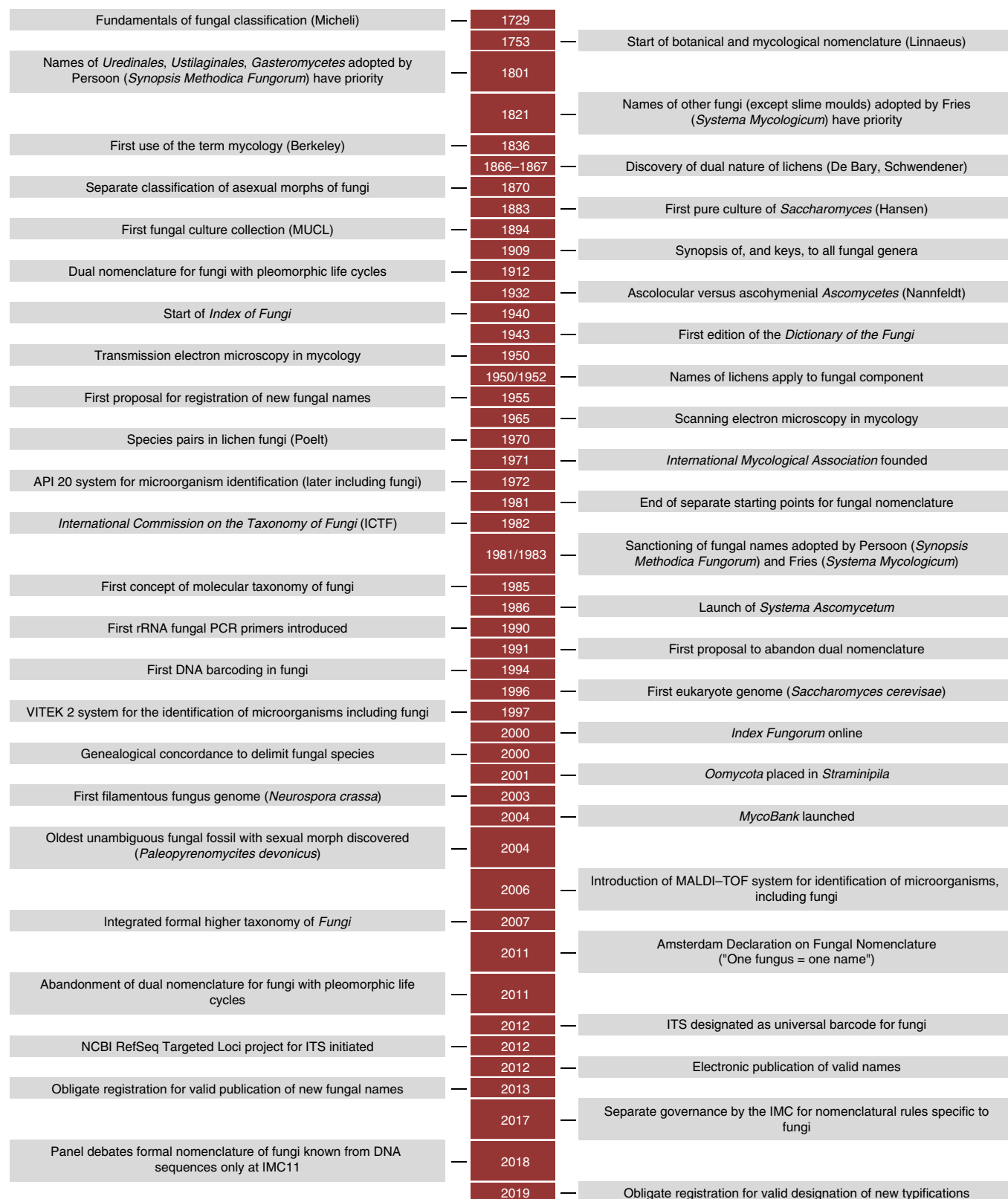
more closely related to *Clavispora* than to *Candida* in the strict sense, which is typified by *Candida albicans*<sup>12</sup>. Although the introduction of new genus names to accurately reflect such relationships may cause short-term confusion for users, it provides long-term, community-wide benefits. For instance, antifungal-resistance profiles of yeasts currently classified under the genus name *Candida* correlate with their phylogenetic position in different families, including *C. albicans* (*Debaryomycetaceae*), *C. auris* (*Metschnikowiaceae*) and *Candida glabrata* (*Saccharomycetaceae*)<sup>10</sup>. Separating distantly related species into different genera or families therefore provides a clear signal that these fungi can be expected to have different properties and treatment options.

The causal agent of the invasive Asian soybean rust, *Phakopsora pachyrhizi*, posed a similar problem in plant pathology, since species classified in *Phakopsora* represent a polyphyletic assemblage<sup>13</sup>. Revised tools provided by the ICNafp (Box 1) allow these issues to be resolved, but name changes are sometimes unavoidable or deemed acceptable by the user community. In the case of *P. pachyrhizi*, the genus and species name were proposed for conservation so that this fungus will retain its name. A contrasting case is the rice blast fungus, which has been placed in the genus *Magnaporthe* based on its sexual stage. Molecular studies showed that *Magnaporthe* is polyphyletic and that the fungus, previously known under the name *Magnaporthe grisea*, encompasses distinct, host-specific lineages. The rice blast fungus is now correctly referred to by its original name *Pyricularia oryzae*<sup>14</sup> (Fig. 1d).

Taxonomic interpretations may also be challenging in a broader context. The human pathogen *Cryptococcus neoformans* (Fig. 1e), best identified with MALDI–TOF mass spectrometry and molecular tools<sup>15,16</sup>, is classified in a genus that traditionally included diverse members of *Tremellomycetes*; however, that genus is now restricted to *C. neoformans* and its relatives<sup>17</sup>. As a result, fungi in plant and soil samples that are identified with the name *Cryptococcus* may not actually represent that genus, and may give the erroneous impression that these substrata bear human-pathogenic yeasts. Therefore, ecological tools, such as the FUNGuild pipeline<sup>18</sup>, require continuous updates to reflect revised phylogenetic classifications.

Resolving species complexes remains a major challenge when it comes to obtaining accurate identifications. Panama disease (fusarium wilt) of banana (Fig. 1f) is caused by a species of *Fusarium*, in particular tropical race 4 (TR4) of *Fusarium oxysporum* f. sp. *cubense* (Foc), which affects the popular Cavendish cultivar of banana. Using secondary barcoding markers, this *Fusarium* lineage was recognized as a separate species, *Fusarium odoratissimum* (Fig. 1g), which is phenotypically characterized by conidial septation and peculiar volatile secondary compounds<sup>19</sup>. However, it may be unclear where to draw the limit between species and infraspecific lineages. In *Colletotrichum siamense* (Fig. 1h), a plant pathogen that causes postharvest rot of tropical crops, an integrative taxonomic approach encompassing phenotyping, multiple molecular markers and mating tests found evidence of genetic recombination and cross fertility, which therefore did not support the separation of lineages as separate species<sup>20</sup>.

*Oomycota* exhibit similar challenges as *Fungi* when it comes to accurate identifications. Basil downy mildew has become a global pandemic that was at first attributed to a common species, *Peronospora lamii*, with an assumed broad host range. Through morphological and phylogenetic analyses, it was shown that the pandemic was caused by a previously unrecognized species, *Peronospora belbahrii*<sup>21</sup>, which has led to more effective phytosanitary precautions. These findings suggest that fine-scaled diversification in fungi is triggered by dynamic pathogen–host interactions that generate conditions for abundant diversification but also require sophisticated approaches to properly delimit, recognize and identify these species.



**Fig. 2 | Timeline of important events in fungal taxonomy and nomenclature.** See Supplementary Information for detailed references regarding each event.

### Classification of fungal 'dark taxa'

Environmental metabarcoding via high-throughput sequencing (HTS) has added a new dimension to assessing the diversity of fungi<sup>22</sup>. Many ecologically cryptic groups, such as anaerobic gut

fungi and chytrids, have seen improved sampling, and environmental sequencing has uncovered new lineages of evolutionary importance: while early diverging clades have expanded the range of organisms included in *Fungi*<sup>23,24</sup>, others have emerged within the

better studied *Ascomycota* and *Basidiomycota*<sup>25</sup>. For new lineages discovered through environmental metabarcoding, taxonomy is operating metaphorically in the dark because sequence-based identifications cannot be verified with voucher specimens, and culturing is necessary to bring these fungi 'to life'.

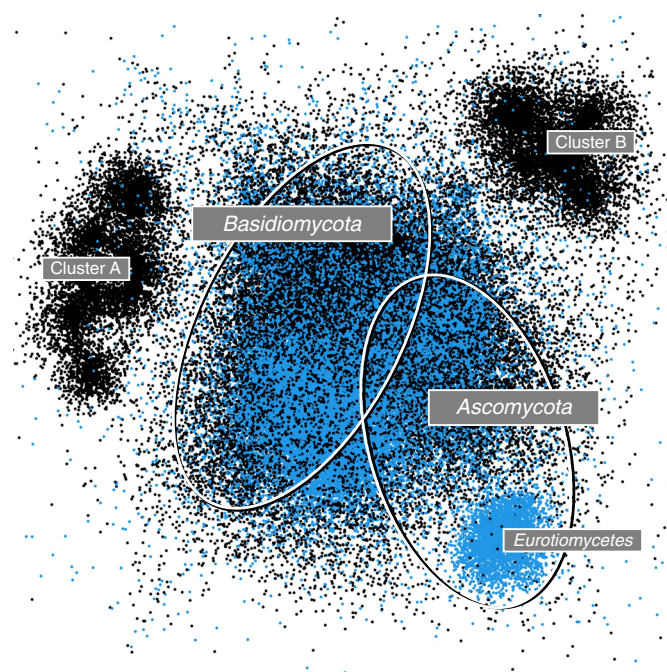
The term 'dark taxa' was originally coined for unnamed sequences<sup>26</sup>. However, some of these may represent known species but have not been linked to a published name. Therefore, dark taxa in a fungal context denote new lineages known from sequence data only, but for which no individual voucher specimens or cultures exist. Consequently, a lineage is no longer considered a dark taxon when a matching specimen or culture becomes available.

Environmental sequencing has generated more than 99.9% of the available molecular sequence data for fungi. The Sequence Read Archive (SRA), which collects data from metabarcoding studies separately from the nucleotide archive in GenBank<sup>27</sup>, has amassed more than 16,000,000,000 HTS-derived fungal ITS reads (<https://ftp.ncbi.nih.gov/sra/fungi>). Compared with about 1,400,000 mostly Sanger-derived fungal ITS sequences in GenBank, this results in a ratio of over 11,000:1. These numbers are not comparable, as HTS reads represent multiple instances from single individuals, but even so, only a fraction of fungal sequence data are derived from voucher specimens or cultures.

While the staggering amount of HTS data encompass substantial fungal diversity not captured by specimen-based or culture-based sequencing (Fig. 3), environmental metabarcoding faces important challenges<sup>4</sup>. Because this approach targets a broad range of higher taxa across fungal phyla, accuracy and precision cannot be achieved for all fungal groups with a single DNA barcode, and different markers sequenced from the same environmental samples cannot be linked back to the same individual sources. Analytical HTS pipelines are optimized to rapidly process hundreds of thousands of sequences<sup>28</sup>; however, they are not able to satisfactorily recognize species-level lineages or derive formal classifications, thereby resulting in a trade-off between speed and accuracy<sup>4</sup>. A substantial proportion of sequences in public repositories is incorrectly or incompletely named and should be treated with caution<sup>4</sup>. Matching but misidentified or unidentified sequences will lead to inaccurate names or imprecise identifications, and new lineages may not be properly recognized.

Curated sequence databases, such as UNITE, MaarjAM and NCBI RefSeq Targeted Loci, attempt to overcome this problem. UNITE has emerged as the most widely used resource for curated fungal ITS sequences, including data from vouchered specimens (indexed with a UDB prefix) and reference sequences for so-called 'species hypotheses', that is, phylogroups of sequences that are interpreted as species. These data are curated by taxonomy experts during taxonomic workshops<sup>29</sup>. Its General FASTA Release (<https://unite.ut.ee/repository.php>) provides vetted reference sequences. Using the nuSSU marker, MaarjAM provides curated reference sequences for phylogroups of arbuscular mycorrhizal fungi, so-called 'virtual taxa'<sup>30</sup>. The NCBI RefSeq Targeted Loci project (<https://www.ncbi.nlm.nih.gov/bioproject/41209>) provides curated reference sequences from type material annotated as part of the NCBI Taxonomy<sup>31</sup> and keeps track of biocollections and collection identifiers in the sequence records to facilitate verification.

A major issue regarding the documentation of fungal diversity from environmental metabarcoding is their formal naming. The ICNafp requires fungal names to have a type: either a physical voucher specimen (Fig. 1i) or alternatively a culture (Fig. 1j) cryopreserved in a metabolically inactive state or, in certain cases, an illustration. The requirement for a type enforces the permanent linking of a name to an individual, thereby providing context for the stability of nomenclature and for taxonomic comparisons. Similarly, the *International Code of Nomenclature of Prokaryotes* (ICNP)<sup>32</sup> requires a preserved pure culture to represent a new species. Since



**Fig. 3 | Visualization of new lineages.** Environmental samples of ITS amplicons generated from the fungal microbiome of switchgrass (*Panicum virgatum*) at a site in North Carolina, USA (SRA accession number SRS7144269; see ref. <sup>50</sup> for the corresponding paper). The approximately 60,000 original fungal ITS reads from the switchgrass sample (black dots) were analysed together with 84,000 ITS sequences (blue dots) from the UNITE General FASTA Release 8.2. (<https://unite.ut.ee/repository.php>). Visualization was done in the rgl package in R (<https://r-forge.r-project.org/projects/rgl>) after employing a sparse similarity matrix approach, with multidimensional ordination using LargeVis<sup>51</sup> based on fast multilevel clustering (fMLC; thresholds of 0.95 and 0.98)<sup>52</sup>. The two ellipses indicate the approximate location of *Ascomycota* (separate smaller cluster represents *Eurotiomycetes*) and *Basidiomycota* sequences; the large centre cluster also contains representatives of the other phyla present in *Fungi*. The reads from the *Panicum* sample form two additional clusters (A and B) in the periphery of the UNITE reference sequences, which indicate phyla or classes only sparsely captured in the UNITE General FASTA Release.

dark taxa do not immediately provide such types, there is no possibility to formally name species known from sequence data alone under the current ICNafp rules<sup>33,34</sup>. New lineages first detected through metabarcoding can only be validly described using laboratory approaches. One option is subsequent culturing from the environment (Table 1, Option 1), as was done for *Archaeorhizomyces*<sup>25</sup> and *Bifiguratus*<sup>35</sup>. Another is visualizing fungal structures corresponding to sequence data, thereby resulting in an illustration as the type (Table 1, Option 2), for example, with fluorescent in situ hybridization (FISH), which was employed in the case of the new phylum *Cryptomycota*<sup>24</sup>. Additionally, fluorescence-activated cell sorting (FACS) can be utilized to obtain an actual sample from the environment.

Fungal taxonomists may argue that techniques such as FISH or FACS are difficult to access, but since they are widely used in microbiology, the formal naming of dark taxa may offer new opportunities for interdisciplinary collaboration. While culturing has a long history of use in microbiology and mycology, it faces various challenges. Not all fungi grow readily in culture, especially in early diverging fungal clades such as *Aphelidiomycota* and *Cryptomycota*, or in arbuscular mycorrhizal *Glomeromycotina* or parasitic rust fungi (*Pucciniomycotina*). Since the physiological requirements

**Table 1 | Proposed alternatives to name fungi discovered through sequence data (dark taxa) and their assessment based on selected parameters**

Parameter	Option 1	Option 2	Option 3	Option 4	Option 5
Approach to obtain or visualize fungus	Culturing	FISH	None (options 3–5)		
Nature of type	Metabolically inactive culture	FISH image	DNA sequence (options 3–5)		
Nature of name	Formal binomial (with priority)	Formal binomial (with priority)	Formal binomial (with priority)	Provisional binomial (nom. seq. without priority)	Informal (non-binomial with identifier)
Compliant with ICNaftp	Yes	Yes	No (pending further action)	No (pending further action)	Not required
Enforcing rules	ICNaftp Divisions I–III, Chapter F	ICNaftp Divisions I–III, Chapter F	ICNaftp Divisions I–III, Chapter F	ICNaftp Appendix to Chapter F	Community agreement (for example, through UNITE)
Integration of names into fungal classifications	Immediate	Immediate	Immediate	Adoption as formal binomial through matching specimen or culture or vetted lists	Provisional, does not replace formal binomial
Competition with specimen-based names	Yes (desired)	Yes (not necessarily desired)	Yes (not necessarily desired)	No (desired)	No (desired)
Resource requirements	Very high (laboratory facilities, space, labour)	High (laboratory facilities)	Low (computational)		
Speed of cataloguing dark taxa	Very low	Low	High		
Taxonomic bias	Low to medium	Low	Absent		
Risk of establishing synonyms	Low	Low to medium	Low to medium if restricted to specific marker (or whole-genome approach) (medium to high if more than one marker is used in parallel)		
Risk of establishing artefactual species	Absent	Very low	Low to medium (depending on quality control)		
Quality control	Intrinsic in approach	Intrinsic in approach	Methodological guidelines (can be enforced through registration)		
Phenotype assessment	Limited (culture characteristics)	Strongly limited (cell or hyphal outlines)	Not possible (metadata only)		
Multigene or genome-wide assessment	Unlimited	Strongly limited (matching FISH patterns)	Currently not feasible (will probably become possible with technological advancements)		

Options 1 and 2 are implemented in the ICNaftp and available for formal descriptions of dark taxa, whereas options 3–5, which share similar methodological approaches and challenges, require amendments to the ICNaftp (options 3 and 4) or a community-wide agreement outside the ICNaftp (option 5).

of dark taxa are not known, culturing requires resource-intensive approaches, a situation that is comparable to prokaryote microbiology. For example, the cultivation of an Asgard archaeon took more than 5 years<sup>36</sup>, and cultivation of a bacterial strain may cost 10,000 euros<sup>37</sup>. Even for fungi that can be readily cultured, the quantity of predicted undiscovered species poses a logistical challenge. According to the *World Directory of Culture Collections* (<http://www.wfcc.info/ccinfo/statistics>), culture collections currently hold about 870,000 fungal strains, representing around 30,000 known species. This number represents just 3% compared with the 1 million unrecognized fungi estimated to be hidden in environmental samples<sup>2</sup>. During the past two decades, about 10,000 new fungi have been described based on cultures (*Index Fungorum*; <http://www.indexfungorum.org>). While this is an impressive number, a comparable effort would need another 2,000 years to formally catalogue dark fungal taxa through culturing alone, or else the underlying resources would have to be increased by two orders of magnitude. Prokaryote taxonomy faces the same problem: some 16,000 Bacteria and Archaea have been formally described<sup>38</sup>, whereas around

1 million are estimated to exist<sup>39</sup>, thereby challenging their formal classification based on pure cultures alone<sup>37,40</sup>.

While these issues limit the options for formally naming dark taxa under the ICNaftp, unambiguous naming of all fungal species is desirable for precise communication in science, conservation and applications<sup>41</sup>. It also complies with the obligation to make associated data findable, accessible, interoperable and reusable (FAIR data principles<sup>42</sup>). Therefore, alternative options for sequence-based nomenclature are being explored, for which proof-of-concept studies are needed to gain broad acceptance among the mycological community. Some earlier proposed options either circumvented the ICNaftp, creating invalid names, or attempted a flexible interpretation of existing provisions. de Beer et al.<sup>43</sup> deliberately described a new species with a non-ICNaftp-compliant, invalid name, *Hawksworthiomyces sequentia* sp. nov. ENAS, adding the acronym ENAS (environmental nucleic acid sequences). Khan et al.<sup>44</sup> established two invalid names in the genus *Archaorhizomyces* using the denomination 'nom. seq.'. Arguing that a sequence illustration would constitute a valid type under the ICNaftp, Lücking and

Moncada<sup>45</sup> introduced seven new species in the newly proposed genus *Lawreymyces*. This approach has been rejected because a DNA sequence is an abstract letter code, and so its depiction does not represent an actual fungal feature<sup>7</sup>. A further strategy consisted of the deposition of the environmental sample, a so-called 'bag type'. The species thus published, *Piromyces cryptodigmaticus*<sup>46</sup>, was ruled invalid by the Nomenclature Committee for Fungi (Box 1) because of an insufficient diagnosis, but it was based on a technically permissible type as the ICNafp allows admixtures with other organisms. However, the 'bag type' is not a desired strategy as the individual fungus corresponding to the sequence data is not discernible within the sample or might have even been used up for DNA extraction.

The straightforward solution would be to amend the ICNafp to allow formal names based on sequence types (Table 1, option 3). Two such proposals have been rejected and referred to internationally appointed committees, with reports expected at the International Mycological Congress (IMC) in Amsterdam 2022 or the next International Botanical Congress in Rio de Janeiro 2023 (ref. <sup>47</sup>). This is comparable to the situation in prokaryotes, for which proposals for the formal naming of uncultivated taxa based on sequence data have likewise been rejected, thereby prompting an alternative proposal for a separate nomenclatural code for uncultivated species<sup>40</sup>. The main arguments against formal sequence-based nomenclature in fungi are the potential establishment of artefactual taxa based on compromised sequence data, the inability to link sequences from different markers back to the same lineage and the notion that sequence-based nomenclature cannot be subjected to verification based on mycological expertise<sup>34</sup>.

The *Candidatus* status in prokaryote nomenclature is one model that aims to overcome this conflict. It allows provisional names under the guidance of an appendix to a nomenclatural code, and such names do not have priority until a pure culture becomes available<sup>32</sup>. Analogously, the ICNafp could, either in a commentary in the introductory matter to Chapter F or in an appendix, refer to provisional fungal names that have no priority over names based on currently permissible types (Table 1, option 4). Recommendations could then be put forward through a resolution at an IMC plenary session, under governance by the International Commission on the Taxonomy of Fungi (ICTF; Box 1). As outlined above, such names could be indicated by a specific denomination such as nom. seq. A corresponding amendment to the ICNafp was proposed, but has been rejected and referred for further discussion to a special purpose committee<sup>47</sup>.

A fifth option would be to establish provisional names independently of the ICNafp (Table 1, option 5). UNITE offers the tools to implement such an informal classification<sup>29</sup>, and the strings applied to its species hypotheses could be used as a naming convention. For instance, the unnamed species hypothesis SH1566369.08FU ([https://unite.ut.ee/bl\\_forw\\_sh.php?sh\\_name=SH1566369.08FU](https://unite.ut.ee/bl_forw_sh.php?sh_name=SH1566369.08FU)), represented by sequences in the genus *Archaeorhizomyces*, could be given the informal dark taxon name *Archaeorhizomyces* SH1566369.08FU, with a corresponding digital object identifier (doi:10.15156/BIO/SH1566369.08FU). Such alphanumeric identifiers are not meant to do away with formal Latinized names, but they speed up the recognition of taxa, thereby allowing precise communication while awaiting formal solutions and avoiding invalid Latinized binomials.

Databases such as GlobalFungi<sup>48</sup> have adopted the UNITE species hypotheses as searchable name strings, so there is potential for broad acceptance of this option. This approach is analogous to species clusters in the prokaryotic Genome Taxonomy Database (<https://gtdb.ecogenomic.org>), which provides species hypotheses based on genome sequences in the form of alphanumeric designations<sup>49</sup>.

Formal requirement for registration of fungal names of dark taxa under options 4 and 5 would allow quality control measures at the

registration step that are difficult to enforce through peer review. The mycological community would have to agree on the requirements, with rules put forward by the ICTF (Box 1). These could include a minimum length for underlying sequence data, a minimum number of independent recoveries and an underlying phylogenetic analysis based on multiple sequence alignments, together with metadata to assess distribution and ecology. A subcommittee of the ICTF could then check published names and add them to a vetted list. There are therefore several instruments that the mycological community can use to advance the naming of dark taxa while maintaining high quality standards and opening new pathways to cataloguing fungal biodiversity.

## Conclusions

Fungal taxonomy has entered a new phase with methodological and conceptual advances. A single, universal approach to the identification of fungi is not feasible because of different needs by communities focusing on different groups, species limits remaining in flux and the majority not having been catalogued yet. Although the increased availability of whole-genome data will shift approaches towards genome-based identification, phenotype-based identifications, physiological profiling and DNA barcoding will remain important, which emphasizes the importance of broadly employed standards for these tools. The ICNafp has become increasingly flexible in dealing with nomenclatural challenges, mediating between revised phylogenetic classifications of fungi and the need for stable nomenclature.

Environmental metabarcoding allows for the rapid documentation of unknown fungal diversity. The mycological community is divided as to how this diversity should be named. Obtaining physical types, cultures or visualization of these fungi to meet formal requirements for valid publication of names may be impracticable, and discovery rates of dark taxa greatly exceed the current culturing capacities. Proposals to allow DNA sequence data as types have met with limited support, but sequence-based nomenclature appears to be the only alternative to name the vast diversity of dark taxa within a reasonable time frame. Therefore, alternative solutions that allow provisional names under a standardized approach need to be sought, either as an appendix to or outside the ICNafp, mirroring challenges in prokaryotic taxonomy. Before the adoption of a unified approach for fungi, case studies should assess potential problems. The mycological community would then put forward a strategy leading up to a decision at the 12th or the 13th IMCs planned for 2022 and 2026, respectively. Simultaneously, further technical advancements are expected that will allow substantially longer sequence reads of higher quality in metabarcoding, therefore improving the basis for sequence-based nomenclature.

Received: 2 June 2020; Accepted: 5 March 2021;

## References

- Richards, T. A., Leonard, G. & Wideman, J. G. What defines the "kingdom" Fungi? *Microbiol. Spectr.* **5**, 57–77 (2017).
- Hawksworth, D. L. & Lücking, R. in *The Fungal Kingdom* (eds Heitman, J. et al.) 79–95 (ASM Press, 2017).
- Berbee, M. L., James, T. Y. & Strullu-Derrien, C. Early diverging fungi: diversity and impact at the dawn of terrestrial life. *Annu. Rev. Microbiol.* **71**, 41–60 (2017).
- Lücking, R. et al. Unambiguous identification of fungi: where do we stand and how accurate and precise is fungal barcoding? *IMA Fungus* **11**, 14 (2020).
- Wijayawardene, N. N. et al. Outline of Fungi and fungus-like taxa. *Mycosphere* **11**, 1060–1456 (2020).
- Beakes, G. W. & Thines, M. in *Handbook of the Protists* 2nd edn (eds Archibald, J. M. et al.) 435–505 (Springer, 2017).
- Turland, N. J. et al. *International Code of Nomenclature for Algae, Fungi, and Plants (Shenzhen Code) 2018* (Koeltz Botanical Books, 2018).
- Hawksworth, D. L. et al. The Amsterdam declaration on fungal nomenclature. *IMA Fungus* **2**, 105–112 (2011).

9. Schoch, C. L. et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl Acad. Sci. USA* **109**, 6241–6246 (2012).
10. Forsberg, K. et al. *Candida auris*: the recent emergence of a multidrug-resistant fungal pathogen. *Med. Mycol.* **57**, 1–12 (2019).
11. Daniel, H.-M., Lachance, M.-A. & Kurtzman, C. P. On the reclassification of species assigned to *Candida* and other anamorphic ascomycetous yeast genera based on phylogenetic circumscription. *Antonie van Leeuwenhoek* **106**, 67–84 (2014).
12. Shen, X. X. et al. Tempo and mode of genome evolution in the budding yeast subphylum. *Cell* **175**, 1533–1545 (2018).
13. Rush, T. A. et al. Variation in the internal transcribed spacer region of *Phakopsora pachyrhizi* and implications for molecular diagnostic assays. *Plant Dis.* **103**, 2237–2245 (2019).
14. Zhang, N. et al. Generic names in Magnaporthales. *IMA Fungus* **7**, 155–159 (2016).
15. Cogliati, M. et al. Genotypes and population genetics of *Cryptococcus neoformans* and *Cryptococcus gattii* species complexes in Europe and the Mediterranean area. *Fungal Genet. Biol.* **129**, 16–29 (2019).
16. Firacative, C., Trilles, L. & Meyer, W. MALDI-TOF MS enables the rapid identification of the major molecular types within the *Cryptococcus neoformans/C. gattii* species complex. *PLoS ONE* **7**, e37566 (2012).
17. Liu, X. Z. et al. Towards an integrated phylogenetic classification of the *Tremellomycetes*. *Stud. Mycol.* **81**, 85–147 (2015).
18. Nguyen, N. H. et al. FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* **20**, 241–248 (2016).
19. Maryani, N. et al. Phylogeny and genetic diversity of the banana Fusarium wilt pathogen *Fusarium oxysporum* f. sp. *ubense* in the Indonesian centre of origin. *Stud. Mycol.* **92**, 155–194 (2019).
20. Liu, F., Wang, M., Damm, U., Crous, P. W. & Cai, L. Species boundaries in plant pathogenic fungi: a *Colletotrichum* case study. *BMC Evol. Biol.* **16**, 81–81 (2016).
21. Thines, M., Telle, S., Ploch, S. & Runge, F. Identity of the downy mildew pathogens of basil, coleus, and sage with implications for quarantine measures. *Mycol. Res.* **113**, 532–540 (2009).
22. Ruppert, K. M., Kline, R. J. & Rahman, M. S. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: a systematic review in methods, monitoring, and applications of global eDNA. *Glob. Ecol. Conserv.* **17**, e00547 (2019).
23. James, T. Y., Stajich, J. E., Hittinger, C. T. & Rokas, A. Toward a fully resolved fungal tree of life. *Annu. Rev. Microbiol.* **74**, 291–313 (2020).
24. Jones, M. D. M. et al. Discovery of novel intermediate forms redefines the fungal tree of life. *Nature* **474**, 200–203 (2011).
25. Rosling, A. et al. Archaeorhizomycetes: unearthing an ancient class of ubiquitous soil Fungi. *Science* **333**, 876–879 (2011).
26. Page, R. D. DNA barcoding and taxonomy: dark taxa and dark texts. *Philos. Trans. R. Soc. B* **371**, 20150334 (2016).
27. Sayers, E. W. et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **D1**, D9–D16 (2020).
28. Anslan, S. et al. Great differences in performance and outcome of high-throughput sequencing data analysis platforms for fungal metabarcoding. *MycKeys* **39**, 29–40 (2018).
29. Nilsson, R. H. et al. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* **47**, D259–D264 (2019).
30. Öpik, M. et al. The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). *New Phytol.* **188**, 223–241 (2010).
31. Schoch, C. L. et al. NCBI taxonomy: a comprehensive update on curation, resources and tools. *Database* **2020**, baaa062 (2020).
32. Parker, C. T., Tindall, B. J. & Garrity, G. M. International Code of Nomenclature of Prokaryotes. Prokaryotic Code (2008 revision). *Int. J. Syst. Evol. Microbiol.* **69**, S7–S111 (2019).
33. Lücking, R. & Hawksworth, D. L. Formal description of sequence-based voucherless Fungi: promises and pitfalls, and how to resolve them. *IMA Fungus* **9**, 143–166 (2018).
34. Thines, M. et al. Ten reasons why a sequence-based nomenclature is not useful for fungi anytime soon. *IMA Fungus* **9**, 177–183 (2018).
35. Torres-Cruz, T. J. et al. *Bifiguratus adelaidae*, gen. et sp. nov., a new member of Mucoromycotina in endophytic and soil-dwelling habitats. *Mycologia* **109**, 363–378 (2017).
36. Imachi, H. et al. Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature* **577**, 519–525 (2020).
37. Overmann, J. Significance and future role of microbial resource centers. *Syst. Appl. Microbiol.* **38**, 258–265 (2015).
38. Parte, A. C., Sarda Carbasse, J., Meier-Kolthoff, J. P., Reimer, L. C. & Goker, M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* **70**, 5607–5612 (2020).
39. Louca, S., Mazel, F., Doebeli, M. & Parfrey, L. W. A census-based estimate of Earth's bacterial and archaeal diversity. *PLoS Biol.* **17**, e3000106 (2019).
40. Murray, A. E. et al. Roadmap for naming uncultivated Archaea and Bacteria. *Nat. Microbiol.* **5**, 987–994 (2020).
41. Ryberg, M. & Nilsson, R. H. New light on names and naming of dark taxa. *MycKeys* **30**, 31–39 (2018).
42. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
43. de Beer, Z. W. et al. *Hawksworthiomyces* gen. nov. (Ophiostomatales), illustrates the urgency for a decision on how to name novel taxa known only from environmental nucleic acid sequences (ENAS). *Fungal Biol.* **120**, 1323–1340 (2016).
44. Khan, F. et al. Naming the untouchable—environmental sequences and niche partitioning as taxonomical evidence in fungi. *IMA Fungus* **11**, 23 (2020).
45. Lücking, R. & Moncada, B. Dismantling *Marchandiomphalina* into *Agonimia* (Verrucariaceae) and *Lawreyomyces* gen. nov. (Corticaceae): setting a precedent to the formal recognition of thousands of voucherless fungi based on type sequences. *Fungal Divers.* **84**, 119–138 (2017).
46. Kirk, P. M. Nomenclatural novelties. *Index Fungi* **1**, 1 (2012).
47. May, T. W., Redhead, S. A., Lombard, L. & Rossman, A. Y. XI International Mycological Congress: report of Congress action on nomenclature proposals relating to fungi. *IMA Fungus* **9**, xxii–xxvii (2018).
48. Větrovský, T. et al. GlobalFungi: global database of fungal records from high-throughput-sequencing metabarcoding studies. *Sci. Data* **7**, 228 (2020).
49. Parks, D. H. et al. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat. Biotechnol.* **38**, 1098–1098 (2020).
50. Lee, M. R., & Hawkes, C. V. Plant and soil drivers of whole-plant microbiomes: variation in switchgrass fungi from coastal to mountain sites. *Phytobiomes J.* <https://doi.org/10.1094/PBIOMES-07-20-0056-FI> (2020).
51. Tang, J., Liu, J., Zhang, M. Y. & Mei, Q. Visualizing large-scale and high-dimensional data. In *WWW '16: Proc. 25th International Conference on World Wide Web* 287–297 (International World Wide Web Conferences Steering Committee, 2016); <https://doi.org/10.1145/2872427.2883041>
52. Vu, D., Groenewald, M. & Verkley, G. Convolutional neural networks improve fungal classification. *Sci. Rep.* **10**, 12628 (2020).

## Acknowledgements

Work by C.L.S. and B.R. was supported by the Intramural Research Program of the National Library of Medicine at the National Institutes of Health in Bethesda, Maryland, USA. D.M.G. received support through the National Science Foundation (NSF) grant DEB-1655980 and Project 4655 of the Pennsylvania State Agricultural Experiment Station. E.M. acknowledges CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil) and FACEPE (Fundação de Amparo à Ciência e Tecnologia de Pernambuco, Brazil). K.D.H. thanks the Thailand Research Fund, grant RDG6130001, entitled “Impact of Climate Change on Fungal Diversity and Biogeography in the Greater Mekong Subregion”. The USDA Hatch project 1010662 is acknowledged for support to M.C.A. M.Ö. was supported by the European Regional Development Fund (Centre of Excellence EcoChange). M.T. acknowledges LOEWE for funding in the framework of the Centre for Translational Biodiversity Genomics (TBG) and the German Science Foundation. N.Z. acknowledges the NSF of the United States (DEB-1452971). P.R.J. was supported through the Manaaki Whenua Biota Portfolio with funding from the Science and Innovation Group of the New Zealand Ministry of Business, Innovation and Employment. R.J. thanks the University of Mauritius for research support. We thank S. Redhead for nomenclatural advice. R. Sanders provided the update for the fungal ITS data in the SRA.

## Author contributions

The present manuscript was first discussed among members of the ICTF. Based on contributions from this initial discussion, a first version of the manuscript was drafted by R.L., C.L.S., M.C.A., B.R. and A.N.M. This version was distributed among the ICTF and to selected colleagues outside the ICTF, and all comments were recorded and incorporated. Based on these initial comments, lead and co-authorship was determined, additional co-authors then including T.A., H.A.A., G.C., P.W.C., I.S.D., D.M.G., D.L.H., K.D.H., L.I., R.J., P.R.J., P.M.K., E.M., T.W.M., W.M., M.Ö., V.R., M.S., M.T., D.V., A.M.Y. and N.Z. The revised draft was circulated two more times among all authors for additional comments before submission. After the first review, H.R.N. was invited as an additional co-author to provide specific input regarding dark taxa and the role of the UNITE database in the proposed alternatives for dark taxa nomenclature. Apart from contributing generally to the manuscript, D.L.H. and T.W.M. revised the nomenclatural details included in Box 1 and Fig. 1. B.R. and D.V. also assisted in technical aspects regarding the SRA and Fig. 3. P.W.C., R.L., W.M., M.T. and A.M.Y. organized the photographs used in Fig. 2 from their working groups. The final draft was approved by all authors.



**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41564-021-00888-x>.

**Correspondence** should be addressed to C.L.S.

**Peer review information** *Nature Microbiology* thanks Michaela Lackner, Emma Steenkamp and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature Limited 2021