REVIEW

Species Identification in Plant-Associated Prokaryotes and Fungi Using DNA

Patrik Inderbitzin,^{1,†} Barbara Robbertse,² and Conrad L. Schoch²¹ Indigo Ag, Inc., 500 Rutherford Avenue, Boston, MA 02129² National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 45 Center Drive, Bethesda, MD 20892

Accepted for publication 11 February 2020.

ABSTRACT

Species names are fundamental to managing biological information. The surge of interest in microbial diversity has resulted in an increase in the number of microbes that need to be identified and assigned a species name. This article provides an introduction to the principles of DNA-based identification of *Archaea* and *Bacteria* traditionally known as prokaryotes, and *Fungi*, the *Oomycetes* and other protists, collectively referred to as fungi. The prokaryotes and fungi are the most commonly

studied microbes from plants, and we introduce the most relevant concepts of prokaryote and fungal taxonomy and nomenclature. We first explain how prokaryote and fungal species are defined, delimited, and named, and then summarize the criteria and methods used to identify prokaryote and fungal organisms to species.

Keywords: bacteriology, microbiome, mycology

Recent advances in high-throughput DNA sequencing and microbial isolation techniques have renewed interest in the diversity of microbes associated with plants in natural and agricultural environments (Bai et al. 2015; Toju et al. 2018). Prokaryote genomes are routinely reconstructed from environmental DNA (Parks et al. 2017; Stewart et al. 2018) and never before cultured microbes are isolated and grown in the lab (Cross et al. 2019; Epstein et al. 2013; Imachi et al. 2020). The influx of microbial data are unprecedented in scope and shapes our understanding of various aspects of science related to plants and agriculture, including plant growth promotion, plant-microbe interactions, disease control, remediation, and carbon sequestration (Nelkner et al. 2019; Orellana et al. 2018; Wu et al. 2009).

In order to access and manage biological information, organisms need to be identified. Identification is critical in numerous contexts besides agriculture, including health care, quarantine regulations, and research (Crous et al. 2016; Pendleton et al. 2017; Pritchard et al. 2016). There are various ways to identify microbes such as phenotypic comparison and protein profiling (Cornut et al. 2019;

Monteiro et al. 2016), which are typically used within small, well-defined groups of microbes from culture. This review focuses mainly on DNA-based identification that is suitable for the diverse communities of plant-associated microbes both cultured and uncultured.

Prokaryotes and fungi are the most commonly studied microbes from plants and plant-associated habitats (Hassani et al. 2018). The terms “prokaryotes” and “fungi” each designates superficially similar but evolutionary divergent groups, which have traditionally been studied using similar methodologies. Prokaryotes include the *Archaea* that share traits with both *Bacteria* and eukaryotes (Eme et al. 2017), and *Fungi* consist of the true fungi and protists such as the *Oomycetes* and slime molds (Cavalier-Smith 2001).

In order to accurately identify prokaryotes and fungi, some knowledge of how species are discovered and named is required. Species discovery involves investigation of species boundaries following best taxonomic practices, whereas naming of species is governed by rules of nomenclature (Turland 2019) (Fig. 1). Prokaryote and fungal species are discovered and named according to different taxonomic traditions and nomenclatural rules. These must be considered when identifying prokaryotes and fungi, as we explain in the following sections.

HOW PROKARYOTE AND FUNGAL SPECIES ARE DEFINED, DELIMITED, AND NAMED

Despite the importance of species to biology, there is no universal agreement on what constitutes a species and how species should be defined and delimited. Charles Darwin wrote “... I look

[†]Corresponding author: P. Inderbitzin; pinderbitzin@indigoag.com

Funding: This work was supported in part by the Intramural Research Program of the National Library of Medicine, National Institutes of Health.

The author(s) declare no conflict of interest.

This article is in the public domain and not copyrightable. It may be freely reprinted with customary crediting of the source. The American Phytopathological Society, 2020.

at the term species, as one arbitrarily given for the sake of convenience to a set of individuals closely resembling each other, and that it does not essentially differ from the term variety, which is given to less distinct and more fluctuating forms.” (Darwin 1859). At least 32 species concepts have been proposed (Zachos 2016), and the debate about how to best group organisms into species is still in progress (Reydon and Kunz 2019). It is therefore important to keep in mind that species cannot be precisely defined, and that boundaries between species are dependent on the rationale and methodology used to infer them. This has implications for identification, as no single approach can be used to identify all organisms.

In prokaryotes, a species can be defined as “... a category that circumscribes monophyletic, and genomically and phenotypically coherent populations of individuals that can be clearly discriminated from other such entities by means of standardized parameters.” (Rosselló-Móra and Amann 2015). Monophyletic populations are populations that share a most recent common ancestor, which can be inferred using phylogenetic analyses (Rosselló-Móra and Amann 2015). The standardized parameters commonly used to delimit prokaryote species include 16S ribosomal RNA gene (16S rDNA) sequence identity and overall genome relatedness indices (OGRI), such as average nucleotide identity (ANI) and digital DNA-DNA hybridization (dDDH) (Meier-Kolthoff et al. 2013; Rosselló-Móra and Amann 2015). An organism is generally assigned to a new species if its ANI, dDDH, and 16S rDNA identity with respect to any named species is below 95, 70, and 98.7%, respectively. However, there are many examples of species that are more narrowly defined, including in *Mycobacterium* and *Streptomyces* where ANI between some species exceeds 99.9% (Ciufo et al. 2018), and in the *Bacillus cereus* group, where several species have identical 16S rDNA sequences (Liu et al. 2017). Once a new species is discovered it may be published, which requires designation of a nomenclatural type, a species name, and description of phenotypic and molecular characteristics (Fig. 1). A nomenclatural type is an organism voucher, usually a culture or dried specimen referred to as type material or type for short, to which a species name is permanently attached (Turland 2019). In order for a new species name to be valid and officially recognized, the rules of the *International Code of Nomenclature of Prokaryotes (ICNP)* must be followed (Parker et al. 2019). Prokaryote species names follow the binomial system adopted by Linnaeus (1753). In the binomial system, species names consist of two parts. The capitalized name of the genus, followed by the specific epithet in lowercase, as in *Escherichia coli*. Species names become

valid and officially recognized upon publication in the *International Journal of Systematic and Evolutionary Microbiology (IJSEM)*, whereas species names published in other journals are only considered effectively, not validly published (Parker et al. 2019). In order to be validated and officially recognized, effective names have to be vetted and published on one of the Validation Lists that are regularly compiled by the *IJSEM* (Oren et al. 2018; Parker et al. 2019; Trujillo et al. 2018). An appendix to the *ICNP* also lays out the steps for publication of provisional names for species that have not been cultured and have some other descriptive characteristics. These preliminary names are recognizable by the word *Candidatus* followed by a binomial or a single word, which is not italicized, e.g., ‘*Candidatus Liberibacter asiaticum*’ or ‘*Candidatus intracellularis*’. *Candidatus* names have no standing under the *ICNP* (Parker et al. 2019). Examples for new prokaryote species publications can be found in the literature (Albert et al. 2019; Johnson and Dunlap 2019; Tagini et al. 2019), including a recent *IJSEM* Validation List (Oren and Garrity 2019), and a *Candidatus* publication (Quaglino et al. 2013).

A commonly used species concept for fungi states that a species corresponds to “... the smallest aggregation of populations with a common lineage that share unique, diagnosable phenotypic characters.” (Harrington and Rizzo 1999). Finding populations with a common lineage frequently involves application of the genealogical concordance phylogenetic species recognition (GCPSR) approach (Taylor et al. 2000). GCPSR assigns species rank to groups that are supported, or at least not contradicted, by any of the phylogenetic trees from individual genes that are included in the study (Dettman et al. 2003). In practice, instead of comparing the topologies of individual gene trees, a single tree is frequently inferred from the combined data of multiple loci. Species-level status is then, in general, assigned to phylogenetic groups that share a most recent common ancestor and are thus monophyletic, and which differ phenotypically from their closest relatives. Statistical support of species-level monophyletic groups is based on the bootstrap support and Bayesian posterior probability metrics, with equal or greater than 70 and 95%, respectively, generally considered significant (Alfaro et al. 2003; Hillis and Bull 1993; Reeb et al. 2004). As for prokaryotes, publication of a new fungal species is more precisely the publication of a new species name, designation of a type and a description (Fig. 1). For the names of new fungal species to be valid, publication must adhere to the rules in the *International Code of Nomenclature for algae, fungi and plants (ICNafp)* (Turland et al. 2018). Similarly to prokaryotes, fungal species names are binomials. There are differences between the rules of the *ICNafp* and the ones of the *ICNP*. For instance, the names of new fungal species can be validly published in any journal, as long as all *ICNafp* requirements are satisfied. This includes registration of the new species names in one of the recognized fungal nomenclature databases. Examples for new fungal species publications can be found in the literature (Pereira et al. 2018; Valenzuela-Lopez et al. 2018; Yurkov and Kurtzman 2019).

Fig. 1. Overview of key terminology related to discovery, naming, and identification of prokaryote and fungal species. A species is named if it contains type material, type for short, to which a species name is permanently attached. Species names become valid if they meet the requirements of the respective codes for valid publication, which includes the designation of a type and preparation of a description. Species boundaries are drawn following best practices of taxonomy and are not regulated by the codes. An organism is identified if it falls within the boundaries of a species circumscribed by a type.

WHY NAMES CHANGE: ON SYNONYMS AND PRIORITY

With respect to species, a synonym is one of two or more species names that apply to the same species (Turland 2019). There are two kinds of synonyms, homotypic and heterotypic synonyms. Homotypic synonyms are names that are attached to the same type. They arise when the name of a genus changes. The names of the species in the affected genus are updated by combining the species epithets with the new genus name. For instance, the name *Bacillus subtilis* was published by Cohn in 1872, to replace the name *Vibrio*

subtilis published by Ehrenberg in 1835 (<https://www.ncbi.nlm.nih.gov/taxonomy>) (Sayers et al. 2019a). The name *B. subtilis* is attached to the earlier designated type for *Vibrio subtilis* and is therefore a later homotypic synonym of *Vibrio subtilis*. Only *B. subtilis* is the correct name for the species that contains the type of *V. subtilis* (Fig. 1).

As opposed to homotypic synonyms, heterotypic synonyms are species names attached to different types. Heterotypic synonyms arise when the same species is published more than once under different names, whereby each name is attached to a different type. The correct name among heterotypic synonyms is generally the oldest validly published name, which has priority over later names (Parker et al. 2019; Turland et al. 2018). For instance, the name *Saccharomyces capensis* was published by Van der Walt & Tschuschner in 1956, and is a later heterotypic synonym of the name *S. cerevisiae*, published by Meyen in 1838 (www.indexfungorum.org) (Index Fungorum Partnership 2019). Both *ICNP* and *ICNafp* have provisions to conserve names to protect them from earlier synonyms, or to reject validly published names for a variety of reasons, e.g., for being ambiguous or confusing. Requests for conservation and rejection of names are published in the *IJSEM* for prokaryotes and in *IMA Fungus* for fungi, and become binding upon acceptance by committees associated with the *ICNP* and *ICNafp*, respectively (Parker et al. 2019; Turland et al. 2018).

For many fungi, there is another kind of heterotypic synonym. These are the different names that were published for the same species depending on whether it displayed the sexual (teleomorph) or asexual (anamorph) morphological state (Hibbett and Taylor 2013). For example, the sexual state of the rice blast pathogen *Pyricularia oryzae* was referred to as *Magnaporthe oryzae*. This practice was abolished in 2013 (Hawksworth 2011), and an effort is underway to select one correct name for each species, e.g., *P. oryzae* over *Magnaporthe oryzae* (Zhang et al. 2016). The correct name among conspecific sexual and asexual names is not necessarily the oldest one (Hawksworth 2011).

THE IMPORTANCE OF USING DNA SEQUENCE DATA FROM TYPE MATERIAL FOR IDENTIFICATION

Type material or type for short, refers to organisms that are designated when new species names are published, and to which species names are permanently attached (Fig. 1) (Turland 2019). Types are preserved in culture collections or herbaria (Parker et al. 2019; Turland et al. 2018), from where they can be retrieved and sequenced. DNA sequence data from types (Federhen 2015; O'Leary et al. 2016) is integrated into GenBank (Sayers et al. 2019b) and can be downloaded and used as references for species identification. Almost all prokaryote species are represented in GenBank by a near full-length 16S rDNA sequence (Chun et al. 2018), which are associated with just over 20,000 validly or effectively published species names. Genome assemblies are available for approximately 57% of all species names, and 28% of species names have genome assemblies from types. Several projects are expected to increase the number of genomes from types in the near future (Whitman et al. 2019; Wu and Ma 2019). In contrast, the proportion of the approximately 120,000 named fungal species (Hawksworth and Luecking 2018) with at least some DNA sequence data, is less than half of all named species. Sequence data of the internal transcribed spacer ribosomal DNA (ITS rDNA) region from types exists for only approximately 12,400 species names, and whole genome sequencing data for currently fewer than 600 species names. The DNA sequence databases containing types at GenBank are continuously expanding as new data are being captured, and GenBank and RefSeq staff are actively involved in increasing the

scope and reliability of type-derived DNA sequence data in the GenBank and RefSeq databases (Ciufu et al. 2018; Robbertse et al. 2017).

Other databases contain DNA sequence data that is not derived from types and sometimes used for identification. These include Greengenes (McDonald et al. 2012), RDP training sets (Cole et al. 2013), SILVA (Yilmaz et al. 2013), UNITE (Kõljalg et al. 2013) and the Warcup Training Set (Deshpande et al. 2016). In fungi, non-types sometimes serve as type replacements in taxonomic studies (Chen et al. 2017), or on identification websites, including for *Fusarium* (O'Donnell et al. 2015). In general, non-types are commonly misidentified (Ciufu et al. 2018; Edgar 2018b; Nilsson et al. 2006) and DNA sequence data from non-types should only be used for identification if endorsed by an expert taxonomist.

HOW TO IDENTIFY PROKARYOTES AND FUNGI USING DNA

Species identification is a multistep process that involves classifying an organism among types and selecting the correct name among competing synonyms. Since the widespread adoption of DNA sequencing for the study of phylogenetic relationships in prokaryotes and fungi in the 1990s (Bruns et al. 1991; Stackebrandt and Goebel 1994), rDNA has played a crucial role for species discovery and identification. Most frequently used are 16S rDNA for prokaryotes and ITS rDNA for fungi. Whole genome sequencing data provides better resolution (Rosselló-Móra and Amann 2015; Stielow et al. 2015) and is now employed routinely for prokaryotes. Most fungal genomes are much larger than prokaryote genomes (Stajich 2017; Trevors 1996) and whole genome sequencing is not yet commonly used in fungi for species discovery and identification. The short 16S and ITS rDNA sequence reads from PCR-based metagenomic studies are not reliable for species identification (Edgar 2018a).

The following provides a general description of methods and thresholds for identification of prokaryotes and fungi to species. Additional details on reference databases and identification tools are provided in subsequent sections.

Prokaryotes. Due to the low cost of high throughput DNA sequencing, prokaryote identification is now commonly performed with DNA sequencing data from whole genome assemblies. A typical workflow consists of classifying an organism based on 16S rDNA identity, followed by assessment of overall genome relatedness and phylogenetic analyses (Table 1). An organism is identified if it falls into a species that contains a single type (Fig. 2). If there are two or more types, the names attached to the types are heterotypic synonyms, and the earliest heterotypic synonym is generally the correct name for the organism. For instance, several species of the *Mycobacterium tuberculosis* complex are synonyms of *Mycobacterium tuberculosis*, as their ANI values are >99% (Riojas et al. 2017), and therefore, only the name *Mycobacterium tuberculosis* has to be considered for species identification. For some species, there may be heterotypic synonyms that have not formally been synonymized. In those situations, instead of selecting the earliest synonym, it may be worthwhile to consult with an expert taxonomist.

Assessment of 16S rDNA identity is a convenient way for initial classification. A threshold of 98.7% identity across the near complete 16S rDNA is commonly regarded as species boundary (Stackebrandt and Ebers 2006). However, the 16S rDNA often has poor discriminatory power, may not reflect overall relatedness, and can be inaccurate for identification (Janda and Abbott 2007). For those reasons, even in the case of a single 16S rDNA hit of 98.7% or higher, it is advisable to perform additional analyses such as ANI,

dDDH, or other OGRI (Chun and Rainey 2014; Meier-Kolthoff et al. 2013; Rosselló-Móra and Amann 2015) (Table 1). Among OGRIs, ANI and dDDH are most commonly used for species delineation. An ANI threshold of 95% across approximately 65% of the genome is generally recognized as the species boundary (Chun et al. 2018; Parks et al. 2019). ANI as low as 93% is acceptable if confirmed by a knowledgeable taxonomist (Rosselló-Móra and Amann 2015). For dDDH, the species boundary threshold is 70% (Meier-Kolthoff et al. 2013). Classification using identity-based approaches such as ANI and dDDH is a phenetic approach that may not reflect evolutionary history (Sokal 1986). Therefore, OGRI results should be confirmed by phylogenetic analyses. The robustness of phylogenetic topology can be assessed with statistical metrics including bootstrap support, where 70% branch support translates to approximately 95% accuracy (Hillis and Bull 1993), or Bayesian posterior probability, where 95% branch support corresponds to approximately 95% accuracy (Alfaro et al. 2003; Reeb et al. 2004; Simmons et al. 2004). Species delimitation using phylogenetic analyses without consideration of OGRIs is not common practice in prokaryotes (Rosselló-Móra and Amann 2015).

Fungi. Identification of fungi differs from prokaryote identification in several important ways. The 18S rRNA gene (18S rDNA) of fungi, which is the equivalent of the prokaryote 16S rDNA, is too conserved and not useful to distinguish species (White et al. 1990). In fungi there are no DNA-based equivalents of the prokaryote ANI or other OGRI thresholds to establish species limits (Table 2), and whole genome sequencing data does not yet play a major role in species identification (Raja et al. 2017). As in prokaryotes, a fungal

organism is identified if it falls into a species that contains a single type, and if there are synonyms, only one of them can be the correct name for the species (Fig. 2).

Initial classification in fungi is commonly based on DNA sequence identity of the ITS rDNA, 28S rRNA (28S rDNA), and protein-coding genes (Schoch et al. 2012; Stielow et al. 2015). The ITS rDNA is most commonly used for identification in fungi and is the official fungal barcode (Schoch et al. 2012). Secondary barcodes include 28S rDNA and protein-coding gene regions depending on the group (Hoang et al. 2019; Stockinger et al. 2010; Tekpinar and Kalmer 2019). Secondary barcodes are required as in many fungal groups, rDNA does not offer sufficient resolution. In a study involving approximately 10,000 fungal species, nearly 20% could not be differentiated by ITS or 28S rDNA (Vu et al. 2016, 2019). These include species in important genera such as *Alternaria* (Woudenberg et al. 2013), *Aspergillus* (Samson et al. 2014), *Cladosporium* (Bensch et al. 2015), *Colletotrichum* (Damm et al. 2012), *Fusarium* (O'Donnell et al. 2015), and *Trichoderma* (Robbertse et al. 2017). In *Fusarium*, sequences from protein-coding genes, including translation elongation factor 1- α (*TEF1*), DNA-directed RNA polymerase II largest (*RPB1*) and second largest subunit (*RPB2*), are required to resolve species, and 100% identity across 99 to 100% of the reference sequence indicates conspecificity in most cases (O'Donnell et al. 2015). For groups where barcode diversity is comparably well studied such as *Fusarium* (O'Donnell et al. 2015) or *Verticillium* (Inderbitzin et al. 2011), identification can be performed reliably based on DNA sequence identity. For less well studied fungi, DNA identity-based identification is not suitable as the intraspecific

TABLE 1
Summary of analyses and criteria used to identify prokaryotes to species with DNA^a

Analysis ^b	Input	Identification threshold	Types above threshold	Conclusion
16S rRNA gene identity assessment	Full length 16S rDNA sequences	≥98.7% identity ^c	None	Unidentified
			≥1	Perform ANI or dDDH
Average nucleotide identity (ANI) assessment ^d	Genome assemblies	≥95% ANI ^e	None	Unidentified ^f
			≥1	Perform phylogenetic analyses
Digital DNA-DNA hybridization (dDDH) analyses ^d	Genome assemblies	≥70% dDDH ^g	None	Unidentified ^f
			≥1	Perform phylogenetic analyses
Core genome phylogenetic analyses ^d	Genome assemblies	≥70% bootstrap support or ≥95% Bayesian posterior probability ^h	0	Repeat analyses ⁱ
			1	Identified
			>1	Potentially identified ^j

^a Identification is typically initiated by assessing 16S rDNA identity between an organism to be identified and a type database, followed by calculation of ANI, dDDH, or other overall genome relatedness index (OGRI), and inference of evolutionary history with phylogenetic analyses.

^b Analyses require reference databases of 16S rDNA sequences and genome assemblies derived from types.

^c Identification using 16S rDNA identity is based on the following empirical criteria employed for species delimitation. If two near full length 16S rDNA sequences differ by >1.3% they represent separate species; if two near full length 16S rDNA sequences differ by ≤1.3% they may still represent separate species provided their overall genome relatedness index (OGRI) values are below the thresholds for conspecificity (dDDH < 70% or ANI < 95%) (Chun and Rainey 2014; Meier-Kolthoff et al. 2013; Rosselló-Móra and Amann 2015; Stackebrandt and Ebers 2006).

^d Taxon selection is generally based on 16S rDNA identity and may, e.g., be limited to all species at and above the 98.7% 16S rDNA identity threshold, or to all species of a genus.

^e An ANI threshold of ≥95% across ≥65% of the query genome is generally recognized as species boundary (Chun et al. 2018; Parks et al. 2019), with ANI as low as 93% acceptable if confirmed by a taxonomist (Rosselló-Móra and Amann 2015).

^f If organism cannot be identified, it is worth checking whether relevant species names, e.g., species names with high 16S rDNA identity to the organism to be identified, lack whole genome sequencing data and are thus missing from the analyses.

^g An empirical dDDH threshold of ≥70% is recognized as species boundary (Meier-Kolthoff et al. 2013).

^h Statistical support of species-level monophyletic groups is based on the bootstrap support and Bayesian posterior probability metrics, with ≥70 and ≥95%, respectively, generally considered significant (Alfaro et al. 2003; Hillis and Bull 1993; Reeb et al. 2004).

ⁱ It is expected that there is ≥1 type above the identification threshold in phylogenetic analyses since there is ≥1 type above at least one of the OGRI thresholds. If there is no underlying issue with the analyses, it may be worth contacting a taxonomist.

^j There are contending heterotypic synonyms and organism can only be identified if one of the synonyms is designated as the correct name for that species in a publication or database, e.g., the NCBI Taxonomy Browser (<https://www.ncbi.nlm.nih.gov/taxonomy>) (Sayers et al. 2019a).

variation is unknown, and phylogenetic analyses are recommended instead. With phylogenetic analyses, an organism is identified if it has a single closest relative among types with $\geq 70\%$ bootstrap support or $\geq 95\%$ Bayesian posterior probability. Multilocus phylogenetic analyses are more reliable and provide a better signal than single locus analyses (Santos et al. 2017).

In fungi, ANI or other OGRIs are not in common use. Phenotypic data, often morphological and microscopic (Kirk et al. 2008), can sometimes be useful to evaluate whether an organism matches the description of the species to which it was assigned using DNA (Fig. 2). However, morphological characters are often unstable and unreliable in culture (Slepecky and Starmer 2009), and there are no rules for how much morphological variation is allowed within a species.

Many fungal species lack DNA sequence data from types, and in those situations, DNA sequence data validated by expert taxonomists and published in fungal taxonomy papers and databases (Chen et al. 2017; O'Donnell et al. 2015) can be used for approximate identification. Genome-based identification is not yet feasible for most fungi due to the small number of sequenced types and the few high-quality whole genome assemblies that are currently publicly available. Additional details on fungal species identification are in Raja et al. (2017).

QUALITY CONTROL, SUBMISSION TO PUBLIC REPOSITORIES AND LINKING DNA DATA TO CULTURES AND HERBARIUM SPECIMENS

Adherence to high quality standards for genome and single locus DNA sequence data are crucial for reliable and accurate identification. Whole genome assemblies should fulfill the proposed standards for coverage, completeness and contamination (Chun et al. 2018; Chuvochina et al. 2019), and fungal ITS and 28S rDNA sequences should be trimmed to remove non-ITS and non-28S regions, respectively, to avoid bias in analyses (Schroeder et al. 2013). DNA data should be deposited in GenBank (Sayers et al. 2019b), or in one of its partner repositories (Karsch-Mizrachi et al. 2017). Representative cultures are best submitted to public culture collections, for instance the German Collection of Microorganisms and Cell Cultures (DSMZ) for prokaryotes, and the Westerdijk Fungal Biodiversity Institute CBS-KNAW collection for fungi. Additional collections are listed on the World Data Centre for Microorganisms website (<http://www.wdcm.org/>). Submitters to culture collections should be aware of international agreements that govern access to genetic resources and benefits derived thereof, including the Nagoya Protocol (Kariyawasam and Tsai 2018; McCluskey et al. 2017). To preserve the link between cultures, DNA and other data, GenBank records can accommodate strain and voucher identifiers and the collection codes of the facilities that store them. Collection codes of herbaria and culture collections are curated in the National Center for Biotechnology Information (NCBI) BioCollections Database (Sharma et al. 2018) that uses a controlled vocabulary.

WHEN IDENTIFICATION FAILS: SOME ADVICE FOR PUBLISHING NEW SPECIES

New species of prokaryotes and fungi are commonly found as the vast majority of microbial biodiversity remains undiscovered (Hawksworth and Luecking 2018; Louca et al. 2019). There are two major aspects to publishing new species. The first one relates to taxonomy, and focuses on classification and description of the organisms that comprise the new species. The second aspect concerns nomenclature and is about the rules of the respective code

that must be followed for valid publication. The rules specify requirements related to designation of types, selection of species names, preparation of descriptions and places of publication. Detailed rules are in the *ICNP* for prokaryotes, and the *ICNafp* for fungi (Parker et al. 2019; Turland et al. 2018). Both codes are periodically updated, every 6 years for the *ICNafp*, and irregularly for the *ICNP*, and the latest edition supersedes all previous editions. Neither code currently allows valid publication solely based on DNA (Hibbett et al. 2016; Whitman 2016). Prokaryote species not published in the *IJSEM* are only effectively published, and must be vetted by the *IJSEM* nomenclature editors. If in compliance with the *ICNP*, the new names are published on a Validation List, and are then valid and protected (Trujillo et al. 2018). It is the responsibility of the authors to notify the *IJSEM* about new names. Recent examples of new prokaryote species publications can be found in the *IJSEM*, which follows the highest standards for phenotypic characterization and description, or in other journals (Boxberger et al. 2019). For fungi, any published name that fulfills the nomenclatural rules of the *ICNafp* is automatically valid, and is entered into the Index Fungorum database (www.indexfungorum.org) (Index Fungorum Partnership 2019), and the MycoBank database (www.mycobank.org) (Robert et al. 2013). For valid publication of prokaryotes, cultures have to be deposited in at least two collections in different countries, whereas preserving cultures is merely encouraged for fungi (Parker et al. 2019; Turland et al. 2018). Culture, herbarium voucher, and institution identifiers should be added to GenBank records as described above (Sharma et al. 2018), and GenBank should be informed once a name is validly published (Schoch et al. 2017). Failure to follow taxonomic traditions and nomenclatural rules may result in rejection of the manuscript or invalidly published names. For assistance with new species descriptions,

Fig. 2. Different identification outcomes for prokaryotes and fungi. Shown are schematics of phylogenetic trees with types and organism to be identified. Ellipses represent species boundaries, and colors represent different species. Relevant support values are by the branches. Species boundaries are inferred using best taxonomic practices as explained in the text. **A**, Organism X belongs to the species circumscribed by Type 1. **B**, Organism X belongs to the species circumscribed by Type 2; Type 1 represents a later heterotypic synonym. **C**, Organism X represents a new species. **D**, Applies to fungi only. Unusually long branch (red) suggests Organism X may differ phenotypically from Type 1 and phenotypic analyses are warranted.

it is therefore advisable to collaborate with your friendly neighborhood taxonomist.

IDENTIFICATION RESOURCES

A summary of the methods and thresholds for identifying prokaryote and fungal species using DNA is presented in Tables 1 and 2. The following section provides practical advice on resources and tools for species identification. Many of the software packages require some knowledge of the Unix shell that is taught by excellent and free online resources, including the ones by Software Carpentry (<https://swcarpentry.github.io/shell-novice/>) (Wilson et al. 2019).

TYPE-DERIVED AND OTHER DNA DATABASES FOR IDENTIFICATION

DNA sequence data from types is stored in GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) (Sayers et al. 2019b). In order to access a particular set of DNA sequence data, such as all bacterial 16S rDNA sequences from types, dedicated queries (Table 3) can be performed at GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>). The queries take advantage of various filters and curated RefSeq targeted loci databases, including the prokaryote 16S, fungal ITS, and fungal 28S rDNA sequence databases (O’Leary et al. 2016). The sequence data can be downloaded and used for identification. Individual RefSeq targeted loci databases can be used as BLAST databases (Schoch et al. 2014) on the BLAST homepage, by selecting the database option ‘rRNA/ITS databases’ (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome). BLAST queries against the Nucleotide collection database (Sayers et al.

2019a), which is the default non-redundant BLAST database, can be restricted to DNA sequence data from types by selecting the ‘Sequences from type material’ option at the GenBank BLAST portal (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). The default Nucleotide collection database is not taxonomically validated for all records, and can be unreliable for species identification. However, the Nucleotide collection database does provide an overview of the nucleotide data available in the public repositories, and can be useful for general taxonomic context of an unidentified organism. Updates on databases, BLAST and other NCBI resources, are published regularly (Sayers et al. 2019a, b). Prokaryote names that are only effectively published, are still included on GenBank records, and their status is indicated in double quotes in the NCBI Taxonomy Browser (<https://www.ncbi.nlm.nih.gov/taxonomy>) (Sayers et al. 2019a). In addition to the above data, GenBank provides taxonomy, type info, and other data as complete dump files, updated every hour. The full NCBI Taxonomy is available at <ftp.ncbi.nlm.nih.gov/pub/taxonomy/>. In addition to the types in the general dump files, a complete set of annotated type-derived strains can be found at ftp://ftp.ncbi.nlm.nih.gov/genomes/ASSEMBLY_REPORTS/prokaryote_type_strain_report.txt.

GenBank type-derived data are also integrated in other databases commonly used for identification, including EzBioCloud (Yoon et al. 2017), GTDB (Parks et al. 2018, 2019) and the TYGS database (Meier-Kolthoff and Göker 2019) for prokaryotes, and MycoBank (Robert et al. 2013) and UNITE (Nilsson et al. 2018) for fungi. UNITE uses what are called species hypotheses, which correspond to groups of ITS rDNA sequences clustered at 97 to 100% identity thresholds. Species hypotheses that contain ITS rDNA sequences with conflicting taxonomic information are manually curated by experts. Since ITS rDNA provides insufficient

TABLE 2
Summary of analyses and criteria used to identify fungi to species with DNA^a

Analysis ^b	Input	Identification threshold	Types above threshold	Conclusion
ITS rDNA, 28S rDNA, or protein-coding gene identity assessment	ITS rDNA, 28S rDNA, or protein-coding sequence	Approximately >99% identity ^c	None	Perform phylogenetic analyses
			1	Confirm with phylogenetic analyses ^d
			>1	Perform phylogenetic analyses
Single locus phylogenetic analyses	ITS rDNA, 28S rDNA, or protein-coding sequence	≥70% bootstrap support or ≥95% Bayesian posterior probability ^e	None	Unidentified
			1	Identified ^{f,g}
			>1	Perform multilocus phylogenetic analyses
Multilocus phylogenetic analyses	DNA sequence data from multiple loci	≥70% bootstrap support or ≥95% Bayesian posterior probability ^e	None	Unidentified
			1 hit	Identified ^g
			>1 hit	Potentially identified ^{g,h}

^a Identification is typically initiated by assessing ITS rDNA, 28S rDNA or protein-coding gene identity between an organism to be identified and a type database, followed by inference of evolutionary history with phylogenetic analyses.

^b Analyses require reference databases with DNA data derived from types.

^c Thresholds for identity-based species identification in fungi, also known as barcoding, are group-dependent, but protein-coding genes provide more resolution than ribosomal genes (Stielow et al. 2015). Identity-based identification is only advisable for groups where intra and interspecific diversity is well characterized.

^d Phylogenetic analyses may have to be performed as confirmation, see footnote c.

^e Statistical support of species-level monophyletic groups is based on the bootstrap support and Bayesian posterior probability metrics, with equal or greater than 70 and 95%, respectively, generally considered significant (Alfaro et al. 2003; Hillis and Bull 1993; Reeb et al. 2004).

^f It is advisable to perform multilocus phylogenetic analyses as confirmation, as multilocus phylogenetic analyses are more reliable than single locus phylogenetic analyses (Santos et al. 2017).

^g The organism’s phenotype, generally morphological, should agree with the description of the species, with the caveat that phenotypic characters may be variable in culture and are often unreliable for identification.

^h There are contending heterotypic synonyms and organism is only identified if one of the synonyms is designated as the correct name for that species in a publication or database, e.g., the NCBI Taxonomy Browser (<https://www.ncbi.nlm.nih.gov/taxonomy>) (Sayers et al. 2019a). Alternatively, including additional loci in phylogenetic analyses may improve resolution (Santos et al. 2017).

species-level resolution for a large proportion of fungi (Vu et al. 2016, 2019), species hypotheses in many cases correspond to species clusters that require additional DNA data for separation into individual species (Robbertse et al. 2017).

There are various DNA databases dedicated to the identification of particular groups of species, particularly in fungi. These include arbuscular mycorrhizal fungi (Öpik et al. 2010), *Fusarium* (O'Donnell et al. 2015), medically important fungi (Hoang et al. 2019), and several additional databases listed among the MycoBank Polyphasic Identifications Databases (<http://www.mycobank.org/Defaultinfo.aspx?Page=PolyphasicID>).

GENERATION OF MOLECULAR DATA

Whole genome sequencing data plays a central role in prokaryote species delineation, and the field is moving toward single molecule,

long read sequencing that can potentially deliver a finished genome within a few hours (Land et al. 2015). Various workflows for assembly and annotation of different kinds of genome sequencing data from pure culture and environmental samples can be found in publications (Del Angel et al. 2018; Stewart et al. 2018).

Sanger sequencing has become less important for prokaryotes but is still the most commonly used sequencing method when delimiting fungal species. Near complete prokaryote 16S rDNA can be PCR amplified with primers 27F and 1492R, and Sanger sequenced with the PCR primers and internal 515F and 806R (Caporaso et al. 2011; Lane 1991; Turner et al. 1999; Weisburg et al. 1991). Primer sequences can also be found on several websites, including a site by the Northern Arizona University Environmental Genetics and Genomics Laboratory (<https://in.nau.edu/enggen/pcr-primers/>). Fungal ITS rDNA is commonly PCR amplified and Sanger sequenced with primers ITS1, ITS1-F or ITS5, paired with ITS4 (Gardes and Bruns

TABLE 3
Example queries to retrieve type-derived NCBI RefSeq (O'Leary et al. 2016) and International Nucleotide Sequence Database (INSD) (Karsch-Mizrachi et al. 2017) assembly and nucleotide records from the Assembly and Nucleotide collection databases (Sayers et al. 2019a)

DNA data type	Taxonomic group	NCBI database ^a	Source	Entrez query ^b
16S rDNA	<i>Archaea</i>	Nucleotide	NCBI RefSeq	PRJNA33317 [BioProject] ^c
16S rDNA	<i>Bacteria</i>	Nucleotide	NCBI RefSeq	PRJNA33175 [BioProject] ^c
16S rDNA	Prokaryotes	Nucleotide	NCBI RefSeq	PRJNA33175 [BioProject] OR PRJNA33317 [BioProject] ^c
16S rDNA	<i>Pseudomonas syringae</i>	Nucleotide	NCBI RefSeq	PRJNA33175 [BioProject] AND <i>Pseudomonas syringae</i> [orgn] ^c
ITS rDNA	<i>Fungi</i>	Nucleotide	NCBI RefSeq	PRJNA177353 [BioProject] ^c
ITS rDNA	<i>Oomycetes</i>	Nucleotide	INSD	Oomycetes [orgn] AND ("sequence from type" [filter] OR "sequence from synonym type" [filter]) AND (5.8S OR 5.8 S OR 5.8 OR internal [title] OR (ITS1 [title] AND ITS2 [title])) NOT WGS [filter] NOT mRNA [filter] NOT mitochondrion [filter] NOT RefSeq [filter] ^{d,e}
ITS rDNA	<i>Verticillium</i> spp.	Nucleotide	NCBI RefSeq	PRJNA177353 [BioProject] AND <i>Verticillium</i> [orgn] ^c
28S rDNA	<i>Fungi</i>	Nucleotide	NCBI RefSeq	PRJNA51803 [BioProject] ^c
28S rDNA	<i>Oomycetes</i>	Nucleotide	INSD	28S [ti] AND Oomycetes [orgn] AND ("sequence from type" [filter] OR "sequence from synonym type" [filter]) ^{d,e}
28S rDNA	<i>Mucoromycota</i>	Nucleotide	NCBI RefSeq	PRJNA51803 [BioProject] AND <i>Mucoromycota</i> [orgn] ^c
Any nucleotide data	<i>Verticillium dahliae</i>	Nucleotide	INSD	<i>Verticillium dahliae</i> [orgn] AND ("sequence from type" [filter] OR "sequence from synonym type" [filter]) ^{d,e}
Genome assembly	<i>Archaea</i>	Assembly	NCBI RefSeq	<i>Archaea</i> [orgn] AND "from type" [properties] AND "latest refseq" [filter] ^{f,g}
Genome assembly	<i>Bacteria</i>	Assembly	NCBI RefSeq	<i>Bacteria</i> [orgn] AND "from type" [properties] AND "latest refseq" [filter] ^{f,g}
Genome assembly	Prokaryotes	Assembly	NCBI RefSeq	Prokaryotes [orgn] AND "from type" [properties] AND "latest refseq" [filter] ^{f,g}
Genome assembly	<i>Firmicutes</i>	Assembly	NCBI RefSeq	<i>Firmicutes</i> [orgn] AND "from type" [properties] AND "latest refseq" [filter] ^{f,g}
Genome assembly	<i>Fungi</i>	Assembly	INSD	<i>Fungi</i> [orgn] AND "from type" [properties] AND "latest" [filter] ^{f,h}

^a The NCBI Nucleotide collection database is at <https://www.ncbi.nlm.nih.gov/nucleotide/>, the NCBI Assembly collection database is at <https://www.ncbi.nlm.nih.gov/assembly/>.

^b Entrez query commands can be pasted into the search windows on the sites given in footnote a. The Assembly and Nucleotide Advanced Search Builder sites are useful for building queries (<https://www.ncbi.nlm.nih.gov/assembly/advanced>; <https://www.ncbi.nlm.nih.gov/nucleotide/advanced>).

^c This BioProject contains a set of curated sequences referred to as RefSeq targeted loci database, the sequences are generally from type material (O'Leary et al. 2016).

^d The "sequence from type" [filter] command restricts queries to records derived from type material (Federhen 2015).

^e The "sequence from synonym type" [filter] command restricts queries to records derived from synonym types. Examples of synonym types are the types of later heterotypic synonyms (Federhen 2015).

^f The "from type" [properties] command restricts queries to all different types for assemblies.

^g The "latest refseq" [filter] command restricts queries to the most recent NCBI RefSeq assemblies.

^h The "latest" [filter] command restricts queries to the most recent INSD assemblies.

1993; White et al. 1990), and the ITS plus partial 28S rDNA with an ITS rDNA forward primer and one of various 28S rDNA reverse primers, such as TW13 (Taylor and Bruns 1999). Primer maps and sequences can be found on the Bruns Lab website (<https://nature.berkeley.edu/brunslab/tour/primers.html>) and the Vilgalys Lab website (https://sites.duke.edu/vilgalyslab/rDNA_primers_for_fungi/). One option to assemble forward and reverse Sanger sequencing reads is Geneious (Drummond et al. 2010). It is also possible to PCR amplify and sequence the complete ribosomal operon in fungi using third generation sequencing methods (Wurzbacher et al. 2019). For all DNA sequence data, proper quality control is imperative as mentioned above (Chun et al. 2018; Chuvochina et al. 2019; Schroeder et al. 2013), and DNA sequences should be submitted to GenBank or to one of its partner repositories (Karsch-Mizrachi et al. 2017).

SOFTWARE AND ONLINE PLATFORMS FOR IDENTIFICATION

Automated identification of prokaryotes. Due to the large proportion of known species with genome sequencing data, prokaryote identification is particularly well-suited for automation. EzBioCloud (Yoon et al. 2017) and TYGS (Meier-Kolthoff and Göker 2019) are examples of online platforms for prokaryote identification where users can upload DNA sequence data, and species identification is performed automatically using type-derived reference data and computation of phylogenetic trees and OGRI. GTDB-Tk is a software package for whole genome-based identification of prokaryotes (Chaumeil et al. 2019; Parks et al. 2019). Similar identification tools do not currently exist for fungi.

Extracting ribosomal and other genes from genomes. Ribosomal DNA and other regions for identity-based identification and species selection for ANI or phylogenetic analyses, can be extracted from whole genome assemblies. However, contaminating, nontarget sequences and other errors in genome assemblies (Acuña-Amador et al. 2018; Kryukov and Imanishi 2016) can lead to misleading results. Particularly when describing new species, the integrity of extracted regions should be confirmed by PCR and Sanger sequencing (Chun et al. 2018). There are different tools to assess the level of contamination in genome assemblies (Laetsch and Blaxter 2017; Lee et al. 2017). Prokaryote 16S rDNA can be extracted from genome assemblies with barrnap (Seemann 2019), the fungal ITS rDNA with ITSx (Bengtsson-Palme et al. 2013) and the fungal 28S rDNA with Infernal (Kalvari et al. 2018). BLASTn is useful for extraction of protein-coding gene sequences from fungal genomes using query sequences from close relatives. An example command to extract the highest identity hits to a query sequence from a genome assembly, and saving results in tabular format as a text file, is as follows: `blastn -query query_sequence.fasta -subject genome_assembly.fasta -outfmt '6 qseqid sseqid sseq' > results.txt`. Arguments are explained on the BLAST Command Line Applications User Manual website (<https://www.ncbi.nlm.nih.gov/books/NBK279690/>) (National Center for Biotechnology Information 2019).

Assessing DNA sequence identity between a query and reference sequences. DNA sequence identity can be used for identification in some situations, and as a starting point for OGRI assessment and phylogenetic analyses. For these purposes, a query sequence needs to be compared across its entire length to one or more reference sequences, which can be done by performing global alignments. Global alignments between two nucleotide sequences using the Needleman-Wunsch algorithm can be performed from the BLAST page (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Pairwise global alignments between a query sequence and a reference database can be performed by USEARCH (Edgar 2010), with the

following command: `usearch -usearch_global query_sequence.fasta -db database.fasta -id 0.8 -userout pairwise_alignment_results.txt -strand both -maxaccepts 0 -maxrejects 0 -userfields query+target+id+difs+ql+tl+evalue`. Arguments are explained on the Usearch website (<https://drive5.com/usearch/>). Global alignments assess pairwise identity across the entire length of the query sequence, and in this sense are different from the local alignments that are performed by BLAST (Altschul et al. 1997).

Generating DNA sequence alignments. Aligning DNA sequence data is a prerequisite for phylogenetic analyses and there are various options for this task, including MAFFT (Katoh and Standley 2013) and MUSCLE (Edgar 2004), which are also implemented in Mesquite (Maddison and Maddison 2011) and Geneious (Drummond et al. 2010). The latter two packages can also be used to concatenate alignments of individual loci for multilocus phylogenetic analyses. Some online platforms align DNA sequence data, including Phylogeny.fr (<http://www.phylogeny.fr/>) (Dereeper et al. 2008). TreeBASE is a repository for alignments used in publications (<https://www.treebase.org/treebase-web>) (Piel et al. 2002), and T-BAS contains reference alignments for some groups of fungi (Carbone et al. 2019).

Building and visualizing phylogenetic trees. Phylogenetic trees play a central role in DNA-based identification and many excellent resources explain the topic at different levels of detail (Baldauf 2003; Baum and Smith 2013; Hall 2011; Lemey et al. 2009). Among the most widely used methods and software for inferring phylogenetic trees are MrBayes (Ronquist and Huelsenbeck 2003), PAUP (Swofford 2002), and RAxML (Stamatakis 2014). Phylogeny.fr is a site that performs a complete phylogenetic workflow online (<http://www.phylogeny.fr/>) (Dereeper et al. 2008). There is also the CIPRES Science Gateway for inference of large phylogenetic trees (<http://www.phylo.org/index.php/>) (Miller et al. 2015). For prokaryotes, several packages infer phylogenetic trees from whole genome assemblies, including GTDB-Tk (Chaumeil et al. 2019; Parks et al. 2019), GToTree (Lee 2019), and GET_PHYLOMARKERS (Vinuesa et al. 2018). Options for editing and visualizing phylogenetic trees include Dendroscope (Huson and Scornavacca 2012), iTol (Letunic and Bork 2006) and the NCBI Tree Viewer (<https://www.ncbi.nlm.nih.gov/tools/treewiewer/>). More capabilities are available as part of the NCBI Genome Workbench (<https://www.ncbi.nlm.nih.gov/tools/gbench/>) (Sayers et al. 2019b) and a complete set of all NCBI tools are available under the Tools tab at <https://www.ncbi.nlm.nih.gov/guide/all/>.

Assessing overall genome relatedness. OGRI such as ANI and dDDH are central to identification of prokaryotes and can be performed with different online tools, including EzBioCloud (Yoon et al. 2017) and JSpeciesWS (Richter and Rosselló-Móra 2009) for ANI, and the TYGS portal for dDDH (Meier-Kolthoff and Göker 2019). ANI can be implemented locally with pyani (Pritchard et al. 2016), or with the faster, BLAST-free fastANI (Jain et al. 2018), and dDDH as described (Meier-Kolthoff et al. 2013). NCBI will release their ANI analysis tool as both a standalone and as part of the Prokaryotic Genome Annotation Pipeline (PGAP) (Ciufo et al. 2018; Haft et al. 2018; Tatusova et al. 2016).

DATABASES OF PROKARYOTE AND FUNGAL NAMES AND SYNONYMS

In order to check the validity of a species name, several online databases can be consulted. These include the List of Prokaryotic Names with Standing in Nomenclature (LPSN) (<http://www.bacterio.net>), which contains validly published prokaryote names and literature references to effective and valid publications (Parte 2018). According to the LPSN website, LPSN is being migrated into the

DSMZ Prokaryotic Nomenclature Up-to-date (PNU) database (<https://www.dsmz.de/services/online-tools/prokaryotic-nomenclature-up-to-date>), and will no longer be updated. *BacDive* contains similar information as well as phenotypic data (<https://bacdiv.dsmz.de/>) (Reimer et al. 2018). For fungi, species names and references are compiled in Index Fungorum (<http://www.indexfungorum.org>) (Index Fungorum Partnership 2019) and MycoBank (<http://www.mycobank.org>) (Robert et al. 2013). The NCBI Taxonomy Browser also provides information about species names, including lineage and type information (<https://www.ncbi.nlm.nih.gov/taxonomy>) (Sayers et al. 2019a) with recent updates linking directly to records from types (<https://ncbiinsights.ncbi.nlm.nih.gov/2020/01/27/improving-the-display-of-type-material-in-the-ncbi-taxbrowser/>).

LITERATURE RESOURCES FOR IDENTIFICATION

Fungi may require verification of morphological or other phenotypic information to confirm DNA-based identification. Phenotypic data are part of the description that is provided when new species names are published (Fig. 1). Literature references to species names can be found via name search on Index Fungorum (<http://www.indexfungorum.org/Names/Names.asp>) (Index Fungorum Partnership 2019) or MycoBank (<http://www.mycobank.org/quick-search.aspx>) (Robert et al. 2013). Index Fungorum also links directly to literature for some of the species names. The NCBI Taxonomy Browser lists relevant literature references for species names (<https://www.ncbi.nlm.nih.gov/taxonomy>) (Sayers et al. 2019a), and references are also included in the downloadable NCBI Taxonomy dump files (<ftp.ncbi.nih.gov/pub/taxonomy/>).

FUTURE OF IDENTIFICATION

Accurate species identification of prokaryotes and fungi is crucial for many sectors of the economy including academia, agriculture, government, health care, and industry and can be challenging even for experts. There are encouraging signs that fully automated, DNA-based identification could be the norm in the near future, at least for prokaryotes. Platforms that use DNA for species identification in prokaryotes already exist (Meier-Kolthoff and Göker 2019; Yoon et al. 2017), and genome sequences have been generated for nearly half of all known species (Chun et al. 2018). Also, there may soon be a way to validly name uncultured species based on DNA (Chuvochina et al. 2019; Hibbett et al. 2016; Whitman 2016). In order to discover most of the still largely unknown microbial biodiversity (Hawksworth and Luecking 2018; Louca et al. 2019), the circumscription and naming of new species must be fully automated using universally recognized criteria (Bobay and Ochman 2017; Matute and Sepúlveda 2019). This, together with portable DNA sequencing platforms, will enable the real-time identification across all of life, and have profound implications on many aspects of science and society.

ACKNOWLEDGMENTS

The following individuals provided insightful comments that substantially improved the manuscript: three anonymous reviewers; Carolyn Young, Editor-in-Chief and Senior Editor, *Phytobiomes Journal*; and Daniela Pignatta and Charley Hubbard, Indigo Ag. Thank you all for your time and effort!

LITERATURE CITED

- long-read resequencing of *Porphyromonas gingivalis* reference strains. *BMC Genomics* 19:54.
- Albert, R. A., McGuire, M., Pavlons, S. C., Roecker, J., Bruess, J., Mossman, S., Sun, S., King, M., Hong, S., Farrance, C. E., Danner, J., Joung, Y., Shapiro, N., Whitman, W. B., and Busse, H.-J. 2019. *Bosea psychrotolerans* sp. nov., a psychrotrophic alphaproteobacterium isolated from Lake Michigan water. *Int. J. Syst. Evol. Microbiol.* 69:1376-1383.
- Alfaro, M. E., Zoller, S., and Lutzoni, F. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Mol. Biol. Evol.* 20:255-266.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.
- Bai, Y., Müller, D. B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M., Dombrowski, N., Münch, P. C., Spaepen, S., Remus-Emsermann, M., Hüttel, B., McHardy, A. C., Vorholt, J. A., and Schulze-Lefert, P. 2015. Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature* 528:364-369.
- Baldauf, S. L. 2003. Phylogeny for the faint of heart: A tutorial. *Trends Genet.* 19:345-351.
- Baum, D. A., and Smith, S. D. 2013. *Tree Thinking: An Introduction to Phylogenetic Biology*. Roberts, Greenwood Village, CO.
- Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., et al. 2013. Improved software detection and extraction of ITS1 and ITS 2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods Ecol. Evol.* 4:914-919.
- Bensch, K., Groenewald, J. Z., Braun, U., Dijksterhuis, J., de Jesús Yáñez-Morales, M., and Crous, P. W. 2015. Common but different: The expanding realm of *Cladosporium*. *Stud. Mycol.* 82:23-74.
- Bobay, L.-M., and Ochman, H. 2017. Biological species are universal across life's domains. *Genome Biol. Evol.* 9:491-501.
- Boxberger, M., Anani, H., and La Scola, B. 2019. Genome sequence and description of *Alterileibacterium massiliense* gen. nov., sp. nov., a new bacterium isolated from human ileum of a patient with Crohn's disease. *New Microbes New Infect.* 30:100533.
- Bruns, T. D., White, T. J., and Taylor, J. W. 1991. Fungal molecular systematics. *Annu. Rev. Ecol. Syst.* 22:525-564.
- Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., and Knight, R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci.* 108:4516-4522.
- Carbone, I., White, J. B., Miadlikowska, J., Arnold, A. E., Miller, M. A., Magain, N., U'Ren, J. M., and Lutzoni, F. 2019. T-BAS Version 2.1: Tree-based selector toolkit for evolutionary placement of DNA sequences and viewing alignments and specimen metadata on curated and custom trees. *C. A. Cuomo, ed. Microbiol. Resour. Announc.* 8:e00328-e19.
- Cavalier-Smith, T. 2001. What are Fungi? Pages 3-37 in: *The Mycota. VII. Part A. Systematics and Evolution*. D. J. McLaughlin, E. G. McLaughlin, and P. A. Lemke, eds. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-10376-0_1
- Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. 2019. GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* btz848.
- Chen, Q., Hou, L. W., Duan, W. J., Crous, P. W., and Cai, L. 2017. *Didymellaceae* revisited. *Stud. Mycol.* 87:105-159.
- Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahall, D. R., da Costa, M. S., Rooney, A. P., Yi, H., Xu, X.-W., De Meyer, S., and Trujillo, M. E. 2018. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 68:461-466.
- Chun, J., and Rainey, F. A. 2014. Integrating genomics into the taxonomy and systematics of the *Bacteria* and *Archaea*. *Int. J. Syst. Evol. Microbiol.* 64:316-324.
- Chuvochina, M., Rinke, C., Parks, D. H., Rappé, M. S., Tyson, G. W., Yilmaz, P., Whitman, W. B., and Hugenholtz, P. 2019. The importance of designating type material for uncultured taxa. *Syst. Appl. Microbiol.* 42:15-21.
- Ciufo, S., Kannan, S., Sharma, S., Badretin, A., Clark, K., Turner, S., Brover, S., Schoch, C. L., Kimchi, A., and DiCuccio, M. 2018. Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *Int. J. Syst. Evol. Microbiol.* 68:2386-2392.
- Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., Brown, C. T., Porras-Alfaro, A., Kuske, C. R., and Tiedje, J. M. 2013. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 42:D633-D642.
- Cornut, J., De Respinis, S., Tonolla, M., Petrini, O., Bärlocher, F., Chauvet, E., and Bruder, A. 2019. Rapid characterization of aquatic hyphomycetes by

Acuña-Amador, L., Primot, A., Cadieu, E., Roulet, A., and Barloy-Hubler, F. 2018. Genomic repeats, misassembly and reannotation: A case study with

- matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Mycologia* 111:177-189.
- Cross, K. L., Campbell, J. H., Balachandran, M., Campbell, A. G., Cooper, S. J., Griffen, A., Heaton, M., Joshi, S., Klingeman, D., Leys, E., Yang, Z., Parks, J. M., and Podar, M. 2019. Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. *Nat. Biotechnol.* 37:1314-1321.
- Crous, P. W., Groenewald, J. Z., Slippers, B., and Wingfield, M. J. 2016. Global food and fibre security threatened by current inefficiencies in fungal identification. *Philos. Trans. R. Soc. B Biol. Sci.* 371:20160024.
- Damm, U., Cannon, P., Woudenberg, J., and Crous, P. 2012. The *Colletotrichum acutatum* species complex. *Stud. Mycol.* 73:37-113.
- Darwin, C. 1859. On the Origin of Species. John Murray, London. <http://darwin-online.org.uk/Variorum/1859/1859-52-dns.html>
- Del Angel, V. D., Hjerde, E., Sterck, L., Capella-Gutierrez, S., Notredame, C., Pettersson, O. V., et al. 2018. Ten steps to get started in genome assembly and annotation. *F1000 Res.* 7:148.
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.-M., and Gascuel, O. 2008. Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36:W465-W469.
- Deshpande, V., Wang, Q., Greenfield, P., Charleston, M., Porras-Alfaro, A., Kuske, C. R., Cole, J. R., Midgley, D. J., and Tran-Dinh, N. 2016. Fungal identification using a Bayesian classifier and the Warcup training set of internal transcribed spacer sequences. *Mycologia* 108:1-5.
- Dettman, J. R., Jacobson, D. J., and Taylor, J. W. 2003. A multilocus genealogical approach to phylogenetic species recognition in the model eukaryote *Neurospora*. *Evolution* 57:2703-2720.
- Drummond, A. J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Heled, J., et al. 2010. Geneious v4.8.5. <https://www.geneious.com/>
- Edgar, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792-1797.
- Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460-2461.
- Edgar, R. C. 2018a. Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences. *PeerJ* 6:e4652.
- Edgar, R. C. 2018b. Taxonomy annotation and guide tree errors in 16S rRNA databases. *PeerJ* 6:e5030.
- Eme, L., Spang, A., Lombard, J., Stairs, C. W., and Ettema, T. J. G. 2017. *Archaea* and the origin of eukaryotes. *Nat. Rev. Microbiol.* 15:711-723.
- Epstein, S. S., Sizova, M., and Hazen, A. 2013. New approaches to cultivation of human microbiota. Pages 303-314 in: *The Human Microbiota*. John Wiley & Sons, New York.
- Federhen, S. 2015. Type material in the NCBI Taxonomy Database. *Nucleic Acids Res.* 43:D1086-D1098.
- Gardes, M., and Bruns, T. D. 1993. ITS primers with enhanced specificity of basidiomycetes: Application to the identification of mycorrhizae and rusts. *Mol. Ecol.* 2:113-118.
- Haft, D. H., DiCuccio, M., Badretdin, A., Brover, V., Chetvernin, V., O'Neill, K., et al. 2018. RefSeq: An update on prokaryotic genome annotation and curation. *Nucleic Acids Res.* 46:D851-D860.
- Hall, B. G. 2011. *Phylogenetic Trees Made Easy: A How-To Manual*. Sinauer Associates, Inc., Sunderland, MA.
- Harrington, T. C., and Rizzo, D. M. 1999. Defining species in the fungi. Pages 43-71 in: *Structure and Dynamics of Fungal Populations*. J. J. Worrall, ed. Kluwer Press, Dordrecht, The Netherlands.
- Hassani, M. A., Durán, P., and Hacquard, S. 2018. Microbial interactions within the plant holobiont. *Microbiome* 6:58.
- Hawksworth, D. 2011. A new dawn for the naming of fungi: Impacts of decisions made in Melbourne in July 2011 on the future publication and regulation of fungal names. *Mycoskeys* 1:7-20.
- Hawksworth, D. L., and Luecking, R. 2018. Fungal diversity revisited: 2.2 to 3.8 million species. *The Fungal Kingdom*. J. Heitman, P. Crous, E. Stukenbrock, B. Howlett, and N. Gow, eds. ASM Press, Washington, DC.
- Hibbett, D., Abarenkov, K., Kõljalg, U., Öpik, M., Chai, B., Cole, J., et al. 2016. Sequence-based classification and identification of fungi. *Mycologia* 108:1049-1068.
- Hibbett, D. S., and Taylor, J. W. 2013. Fungal systematics: Is a new age of enlightenment at hand? *Nat. Rev. Microbiol.* 11:129-133.
- Hillis, D. M., and Bull, J. J. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* 42:182-192.
- Hoang, M. T. V., Irinyi, L., Chen, S. C., Sorrell, T. C., and Meyer, W. 2019. Dual DNA barcoding for the molecular identification of the agents of invasive fungal infections. *Front. Microbiol.* 10:1647.
- Huson, D. H., and Scornavacca, C. 2012. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* 61:1061-1067.
- Imachi, H., Nobu, M. K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., et al. 2020. Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* 577:519-525.
- Inderbitzin, P., Bostock, R. M., Davis, R. M., Usami, T., Platt, H. W., and Subbarao, K. V. 2011. Phylogenetics and taxonomy of the fungal vascular wilt pathogen *Verticillium*, with the descriptions of five new species. *PLoS One* 6:e28341.
- Index Fungorum Partnership. 2019. Index Fungorum. <http://www.indexfungorum.org/Names/Names.asp>
- Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114.
- Janda, J. M., and Abbott, S. L. 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *J. Clin. Microbiol.* 45:2761-2764.
- Johnson, E. T., and Dunlap, C. A. 2019. Phylogenomic analysis of the *Brevibacillus brevis* clade: A proposal for three new *Brevibacillus* species, *Brevibacillus fortis* sp. nov., *Brevibacillus porteri* sp. nov. and *Brevibacillus schisleri* sp. nov. *Antonie van Leeuwenhoek* 112:991-999.
- Kalvari, I., Nawrocki, E. P., Argasinska, J., Quinones-Olvera, N., Finn, R. D., Bateman, A., and Petrov, A. I. 2018. Non-coding RNA analysis using the Rfam database. *Curr. Protoc. Bioinformatics* 62:e51.
- Kariyawasam, K., and Tsai, M. 2018. Access to genetic resources and benefit sharing: Implications of Nagoya Protocol on providers and users. *J. World Intellect. Prop.* 21:289-305.
- Karsch-Mizrahi, I., Takagi, T., Cochrane, G., and the International Nucleotide Sequence Database Collaboration. 2017. The international nucleotide sequence database collaboration. *Nucleic Acids Res.* 46:D48-D51.
- Katoh, K., and Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30:772-780.
- Kirk, P. M., Cannon, P. F., Minter, D. W., and Stalpers, J. A. 2008. *Ainsworth & Bisby's Dictionary of the Fungi*. 10th ed. CABI, Wallingford, Oxon. <https://doi.org/10.1079/9780851998268.0000>
- Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F. S., Bahram, M., et al. 2013. Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* 22:5271-5277.
- Kryukov, K., and Imanishi, T. 2016. Human contamination in public genome assemblies. *PLoS One* 11:e0162424.
- Laetsch, D. R., and Blaxter, M. L. 2017. BlobTools: Interrogation of genome assemblies. *F1000 Res.* 6:1287.
- Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M. R., Ahn, T.-H., Karpins, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S., and Ussery, D. W. 2015. Insights from 20 years of bacterial genome sequencing. *Funct. Integr. Genomics* 15:141-161.
- Lane, D. J. 1991. 16S/23S rRNA sequencing. Pages 115-175 in: *Nucleic Acid Techniques in Bacterial Systematics*. E. Stackebrandt and M. Goodfellow, eds. John Wiley & Sons, New York.
- Lee, I., Chaita, M., Ha, S.-M., Na, S.-I., Yoon, S.-H., and Chun, J. 2017. ContEst16S: An algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences. *Int. J. Syst. Evol. Microbiol.* 67:2053-2057.
- Lee, M. D. 2019. Applications and considerations of GToTree: A user-friendly workflow for phylogenomics. *Evol. Bioinform.* 15:1176934319862245.
- Lemey, P., Salemi, M., and Vandamme, A.-M. 2009. *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*. Cambridge University Press, Cambridge, U.K. <https://doi.org/10.1017/CBO9780511819049>
- Letunic, I., and Bork, P. 2006. Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* 23:127-128.
- Linnaeus, C. 1753. *Species plantarum*. Laurentii Salvii, Stockholm. <http://www.botanicus.org/title/b12069590>
- Liu, Y., Du, J., Lai, Q., Zeng, R., Ye, D., Xu, J., and Shao, Z. 2017. Proposal of nine novel species of the *Bacillus cereus* group. *Int. J. Syst. Evol. Microbiol.* 67:2499-2508.
- Louca, S., Mazel, F., Doebeli, M., and Parfrey, L. W. 2019. A census-based estimate of Earth's bacterial and archaeal diversity. *PLoS Biol.* 17:e3000106.
- Maddison, W. P., and Maddison, D. R. 2011. Mesquite: A modular system for evolutionary analysis. Version 2. <http://www.mesquiteproject.org/>
- Matute, D. R., and Sepúlveda, V. E. 2019. Fungal species boundaries in the genomics era. *Fungal Genet. Biol.* 131:103249.
- McCluskey, K., Barker, K. B., Barton, H. A., Boundy-Mills, K., Brown, D. R., Coddington, J. A., et al. 2017. The U.S. Culture Collection Network responding to the requirements of the Nagoya Protocol on access and benefit sharing. *MBio* 8:e00982-e17.

- McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., Andersen, G. L., Knight, R., and Hugenholtz, P. 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of *Bacteria* and *Archaea*. *ISME J.* 6:610-618.
- Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., and Göker, M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60.
- Meier-Kolthoff, J. P., and Göker, M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* 10:2182.
- Miller, M. A., Schwartz, T., Pickett, B. E., He, S., Klem, E. B., Scheuermann, R. H., Passarotti, M., Kaufman, S., and O'Leary, M. A. 2015. A RESTful API for access to phylogenetic tools via the CIPRES science gateway. *Evol. Bioinform.* 11:EBO-S21501.
- Monteiro, A. C. M., Fortaleza, C. M. C. B., Ferreira, A. M., de Souza Cavalcante, R., Mondelli, A. L., Bagagli, E., et al. 2016. Comparison of methods for the identification of microorganisms isolated from blood cultures. *Ann. Clin. Microbiol. Antimicrob.* 15:45.
- National Center for Biotechnology Information. 2019. BLAST® Command Line Applications User Manual. National Center for Biotechnology Information (US), Bethesda, MD. <https://www.ncbi.nlm.nih.gov/books/NBK279690/>
- Nelkner, J., Henke, C., Lin, T. W., Pätzold, W., Hassa, J., Jaenicke, S., Grosch, R., Puhler, A., Sczyrba, A., and Schluter, A. 2019. Effect of long-term farming practices on agricultural soil microbiome members represented by metagenomically assembled genomes (MAGs) and their predicted plant-beneficial genes. *Genes (Basel)* 10:424.
- Nilsson, H. R., Ryberg, M., Kristiansson, E., Abarenkov, K., Larsson, K.-H., and Koljag, U. 2006. Taxonomic reliability of DNA sequences in public sequence databases: A fungal perspective. *PLoS One* 1:e59.
- Nilsson, R. H., Larsson, K.-H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., et al. 2018. The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* 47:D259-D264.
- O'Donnell, K., Ward, T. J., Robert, V. A. R. G., Crous, P. W., Geiser, D. M., and Kang, S. 2015. DNA sequence-based identification of *Fusarium*: Current status and future directions. *Phytoparasitica* 43:583-595.
- O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufu, S., Haddad, D., McVeigh, R., et al. 2016. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44:D733-D745.
- Öpik, M., Vanatoa, A., Vanatoa, E., Moora, M., Davison, J., Kalwij, J., Reier, U., and Zobel, M. 2010. The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (*Glomeromycota*). *New Phytol.* 188:223-241.
- Orellana, L. H., Chee-Sanford, J. C., Sanford, R. A., Löffler, F. E., and Konstantinidis, K. T. 2018. Year-round shotgun metagenomes reveal stable microbial communities in agricultural soils and novel ammonia oxidizers responding to fertilization. *Appl. Environ. Microbiol.* 84:e01646-e17.
- Oren, A., and Garrity, G. M. 2019. List of new names and new combinations previously effectively, but not validly, published. *Int. J. Syst. Evol. Microbiol.* 69:597-599.
- Oren, A., Garrity, G. M., and Parte, A. C. 2018. Why are so many effectively published names of prokaryotic taxa never validated? *Int. J. Syst. Evol. Microbiol.* 68:2125-2129.
- Parker, C. T., Tindall, B. J., and Garrity, G. M. 2019. International Code of Nomenclature of Prokaryotes. *Int. J. Syst. Evol. Microbiol.* 69:S1-S111.
- Parks, D. H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A. J., and Hugenholtz, P. 2019. Selection of representative genomes for 24,706 bacterial and archaeal species clusters provide a complete genome-based taxonomy. *bioRxiv* 771964. <https://www.biorxiv.org/content/early/2019/09/18/771964>
- Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., and Hugenholtz, P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36:996-1004.
- Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B. J., Evans, P. N., Hugenholtz, P., and Tyson, G. W. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* 2:1533-1542.
- Parte, A. C. 2018. LPSN-List of prokaryotic names with standing in nomenclature (bacterio.net), 20 years on. *Int. J. Syst. Evol. Microbiol.* 68: 1825-1829.
- Pendleton, K. M., Erb-Downward, J. R., Bao, Y., Branton, W. R., Falkowski, N. R., Newton, D. W., Huffnagle, G. B., and Dickson, R. P. 2017. Rapid pathogen identification in bacterial pneumonia using real-time metagenomics. *Am. J. Respir. Crit. Care Med.* 196:1610-1612.
- Pereira, C. B., Ward, T. J., Tessmann, D. J., Del Ponte, E. M., Laraba, I., Vaughan, M. M., McCormick, S. P., Busman, M., Kelly, A., Proctor, R. H., and O'Donnell, K. 2018. *Fusarium subtropicale*, sp. nov., a novel nivalenol mycotoxin-producing species isolated from barley (*Hordeum vulgare*) in Brazil and sister to *F. praegraminearum*. *Mycologia* 110:860-871.
- Piel, W. H., Donoghue, M. J., and Sanderson, M. J. 2002. TreeBASE: A database of phylogenetic knowledge. Pages 41-47 in: To the Interoperable "Catalog of Life"—With Partners Species 2000 Asia Oceania. J. Shimura, K. L. Wilson, and D. Gordon, eds. Research Report from the National Institute for Environmental Studies No. 171, Tsukuba, Japan.
- Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., and Toth, I. K. 2016. Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8:12-24.
- Quaglino, F., Zhao, Y., Casati, P., Bulgari, D., Bianco, P. A., Wei, W., and Davis, R. E. 2013. '*Candidatus* Phytoplasma solani', a novel taxon associated with stolbur-and bois noir-related diseases of plants. *Int. J. Syst. Evol. Microbiol.* 63:2879-2894.
- Raja, H. A., Miller, A. N., Pearce, C. J., and Oberlies, N. H. 2017. Fungal identification using molecular tools: A primer for the natural products research community. *J. Nat. Prod.* 80:756-770.
- Reeb, V., Lutzoni, F., and Roux, C. 2004. Contribution of *RPB2* to multilocus phylogenetic studies of the euascomycetes (*Pezizomycotina*, *Fungi*) with special emphasis on the lichen-forming *Acarosporaceae* and evolution of polyspory. *Mol. Phylogenet. Evol.* 32:1036-1060.
- Reimer, L. C., Vetcinova, A., Carbasse, J. S., Söhngen, C., Gleim, D., Ebeling, C., and Overmann, J. 2018. BacDive in 2019: Bacterial phenotypic data for high-throughput biodiversity analysis. *Nucleic Acids Res.* 47: D631-D636.
- Reydon, T. A., and Kunz, W. 2019. Species as natural entities, instrumental units and ranked taxa: New perspectives on the grouping and ranking problems. *Biol. J. Linn. Soc.* 126:623-636.
- Richter, M., and Rosselló-Móra, R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci.* 106:19126-19131.
- Riojas, M. A., McGough, K. J., Rider-Riojas, C. J., Rastogi, N., and Hazbón, M. H. 2017. Phylogenomic analysis of the species of the *Mycobacterium tuberculosis* complex demonstrates that *Mycobacterium africanum*, *Mycobacterium bovis*, *Mycobacterium caprae*, *Mycobacterium microti* and *Mycobacterium pinnipedii* are later heterotypic synonyms of *Mycobacterium tuberculosis*. *Int. J. Syst. Evol. Microbiol.* 68:324-332.
- Robertse, B., Strophe, P. K., Chaverri, P., Gazis, R., Ciufu, S., Domrachev, M., and Schoch, C. L. 2017. Improving taxonomic accuracy for fungi in public sequence databases: Applying 'one name one species' in well-defined genera with *Trichoderma/Hypocrea* as a test case. *Database (Oxford)* 2017:bax072.
- Robert, V., Vu, D., Amor, A. B. H., van de Wiele, N., Brouwer, C., Jabas, B., et al. 2013. MycoBank gearing up for new horizons. *IMA Fungus* 4:371-379.
- Ronquist, F., and Huelsenbeck, J. P. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.
- Rosselló-Móra, R., and Amann, R. 2015. Past and future species definitions for *Bacteria* and *Archaea*. *Syst. Appl. Microbiol.* 38:209-216.
- Samson, R. A., Visagie, C. M., Houbraken, J., Hong, S.-B., Hubka, V., Klaassen, C. H. W., Perrone, G., Seifert, K. A., Susca, A., Tanney, J. B., Varga, J., Kocsube, S., Szigeti, G., Yaguchi, T., and Frisvad, J. C. 2014. Phylogeny, identification and nomenclature of the genus *Aspergillus*. *Stud. Mycol.* 78:141-173.
- Santos, L., Alves, A., and Alves, R. 2017. Evaluating multi-locus phylogenies for species boundaries determination in the genus *Diaporthe*. *PeerJ* 5:e3120.
- Sayers, E. W., Beck, J., Brister, J. R., Bolton, E. E., Canese, K., Comeau, D. C., et al. 2019a. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 48:D9-D16.
- Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., and Karsch-Mizrachi, I. 2019b. GenBank. *Nucleic Acids Res.* 48:D84-D86.
- Schoch, C. L., Aime, M. C., De Beer, W., Crous, P. W., Hyde, K. D., Penev, L., Seifert, K. A., Stadler, M., Zhang, N., and Miller, A. N. 2017. Using standard keywords in publications to facilitate updates of new fungal taxonomic names. *IMA Fungus* 8:70-73.
- Schoch, C. L., Robertse, B., Robert, V., Vu, D., Cardinali, G., Irinyi, L., et al. 2014. Finding needles in haystacks: Linking scientific names, reference specimens and molecular data for Fungi. *Database (Oxford)* 2014:bau061. <https://doi.org/10.1093/database/bau061>
- Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., Chen, W., and Fungal Barcoding Consortium. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci. USA* 109:6241-6246.
- Schroeder, K. L., Martin, F. N., de Cock, A. W. A. M., Lévesque, C. A., Spies, C. F. J., Okubara, P. A., and Paulitz, T. C. 2013. Molecular detection and

- quantification of *Pythium* species: Evolving taxonomy, new tools, and challenges. *Plant Dis.* 97:4-20.
- Seemann, T. 2019. barnap 0.9: Rapid ribosomal RNA prediction. <https://github.com/tseemann/barnap>
- Sharma, S., Ciufu, S., Starchenko, E., Darji, D., Chlumsky, L., Karsch-Mizrachi, I., and Schoch, C. L. 2018. The NCBI BioCollections Database. Database (Oxford) 2018:bay006. <https://doi.org/10.1093/database/bay006>
- Simmons, M. P., Pickett, K. M., and Miya, M. 2004. How meaningful are Bayesian support values? *Mol. Biol. Evol.* 21:188-199.
- Slepecky, R. A., and Starmer, W. T. 2009. Phenotypic plasticity in fungi: A review with observations on *Aureobasidium pullulans*. *Mycologia* 101:823-832.
- Sokal, R. R. 1986. Phenetic taxonomy: Theory and methods. *Annu. Rev. Ecol. Syst.* 17:423-442.
- Stackebrandt, E., and Ebers, J. 2006. Taxonomic parameters revisited: Tarnished gold standards. *Microbiol. Today* 152-155.
- Stackebrandt, E., and Goebel, B. M. 1994. Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* 44:846-849.
- Stajich, J. E. 2017. Fungal genomes and insights into the evolution of the kingdom. *Microbiol. Spectr.* 5:4.
- Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312-1313.
- Stewart, R. D., Auffret, M. D., Warr, A., Wiser, A. H., Press, M. O., Langford, K. W., Liachko, I., Snelling, T. J., Dewhurst, R. J., Walker, A. W., Roche, R., and Watson, M. 2018. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat. Commun.* 9:870.
- Stielow, J. B., Levesque, C. A., Seifert, K. A., Meyer, W., Iriny, L., Smits, D., et al. 2015. One fungus, which genes? Development and assessment of universal primers for potential secondary fungal DNA barcodes. *Persoonia* 35:242-263.
- Stockinger, H., Krüger, M., and Schüßler, A. 2010. DNA barcoding of arbuscular mycorrhizal fungi. *New Phytol.* 187:461-474.
- Swofford, D. L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, MA.
- Tagini, F., Aeby, S., Bertelli, C., Droz, S., Casanova, C., Prod'homme, G., Jaton, K., and Greub, G. 2019. Phylogenomics reveal that *Mycobacterium kansasii* subtypes are species-level lineages. Description of *Mycobacterium pseudokansasii* sp. nov., *Mycobacterium innocens* sp. nov. and *Mycobacterium attenuatum* sp. nov. *Int. J. Syst. Evol. Microbiol.* 69:1696-1704.
- Tatusova, T., DiCuccio, M., Badretdin, A., Chetvermin, V., Nawrocki, E. P., Zaslavsky, L., Lomsadze, A., Pruitt, K. D., Borodovsky, M., and Ostell, J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44:6614-6624.
- Taylor, D. L., and Bruns, T. D. 1999. Community structure of ectomycorrhizal fungi in a *Pinus muricata* forest: Minimal overlap between the mature forest and resistant propagule communities. *Mol. Ecol.* 8:1837-1850.
- Taylor, J. W., Jacobson, D. J., Kroken, S., Kasuga, T., Geiser, D. M., Hibbett, D. S., and Fisher, M. C. 2000. Phylogenetic species recognition and species concepts in fungi. *Fungal Genet. Biol.* 31:21-32.
- Tekpinar, A. D., and Kalmer, A. 2019. Utility of various molecular markers in fungal identification and phylogeny. *Nova Hedwigia* 108:3-4.
- Toju, H., Peay, K. G., Yamamichi, M., Narisawa, K., Hiruma, K., Naito, K., et al. 2018. Core microbiomes for sustainable agroecosystems. *Nat. Plants* 4:247-257.
- Trevors, J. 1996. Genome size in bacteria. *Antonie van Leeuwenhoek* 69:293-303.
- Trujillo, M. E., Oren, A., and Garrity, G. M. 2018. Preparation of the Validation Lists and the role of the List Editors. *Int. J. Syst. Evol. Microbiol.* 69:3-4.
- Turland, N. 2019. The Code Decoded. A user's guide to the International Code of Nomenclature for algae, fungi, and plants, 2nd ed. Pensoft Publishers, Sofia, Bulgaria.
- Turland, N. J., Wiersema, J. H., Barrie, F. R., Greuter, W., Hawksworth, D., Herendeen, P. S., Knapp, S., Kusber, W.-H., Li, D.-Z., Marhold, K., May, T. W., McNeill, J., Monro, A. M., Prado, J., Price, M. J., and Smith, G. F. 2018. International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, Koeltz Botanical Books, China. <https://doi.org/10.12705/Code.2018>
- Turner, S., Pryer, K. M., Miao, V. P., and Palmer, J. D. 1999. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J. Eukaryot. Microbiol.* 46:327-338.
- Valenzuela-Lopez, N., Cano-Lira, J. F., Guarro, J., Sutton, D. A., Wiederhold, N., Crous, P. W., and Stchigel, A. M. 2018. Coelomycetous *Dothideomycetes* with emphasis on the families *Cucurbitariaceae* and *Didymellaceae*. *Stud. Mycol.* 90:1-69.
- Vinuesa, P., Ochoa-Sánchez, L. E., and Contreras-Moreira, B. 2018. GET_PHYLOMARKERS, a software package to select optimal orthologous clusters for phylogenomics and inferring pan-genome phylogenies, used for a critical geno-taxonomic revision of the genus *Stenotrophomonas*. *Front. Microbiol.* 9:771.
- Vu, D., Groenewald, M., de Vries, M., Gehrman, T., Stielow, B., Eberhardt, U., et al. 2019. Large-scale generation and analysis of filamentous fungal DNA barcodes boosts coverage for kingdom fungi and reveals thresholds for fungal species and higher taxon delimitation. *Stud. Mycol.* 92:135-154.
- Vu, D., Groenewald, M., Szöke, S., Cardinali, G., Eberhardt, U., Stielow, B., de Vries, M., Vekleij, G. J. M., Crous, P. W., Boekhout, T., and Robert, V. 2016. DNA barcoding analysis of more than 9,000 yeast isolates contributes to quantitative thresholds for yeast species and genera delimitation. *Stud. Mycol.* 85:91-105.
- Weisburg, W. G., Barns, S. M., Pelletier, D. A., and Lane, D. J. 1991. 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.* 173:697-703.
- White, T. J., Bruns, T. D., Lee, S. B., and Taylor, J. W. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pages 315-322 in: PCR Protocols. M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, eds. Academic Press, San Diego, CA.
- Whitman, W. B. 2016. Modest proposals to expand the type material for naming of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 66:2108-2112.
- Whitman, W. B., Klenk, H.-P., Arahall, D. R., Aznar, R., Garrity, G., Pester, M., et al. 2019. Genomic Encyclopedia of *Bacteria* and *Archaea* (GEBA) VI: Learning from type strains. *Microbiol. Aust.* 40:125-129.
- Wilson, G., Capes, G., Devenyi, G. A., Koch, C., Silva, R., Srinath, A., et al. 2019. swcarpentry/shell-novice: Software Carpentry: The UNIX shell, June 2019 (Version v2019.06.1). Zenodo. G. A. Devenyi, G. Capes, C. Morris, and W. Pitchers, eds. <http://doi.org/10.5281/zenodo.3266823>
- Woudenberg, J. H. C., Groenewald, J. Z., Binder, M., and Crous, P. W. 2013. *Alternaria* redefined. *Stud. Mycol.* 75:171-212.
- Wu, C. H., Bernard, S. M., Andersen, G. L., and Chen, W. 2009. Developing microbe-plant interactions for applications in plant-growth promotion and disease control, production of useful compounds, remediation and carbon sequestration. *Microbiol. Biotechnol.* 2:428-440.
- Wu, L., and Ma, J. 2019. The Global Catalogue of Microorganisms (GCM) 10K type strain sequencing project: Providing services to taxonomists for standard genome sequencing and annotation. *Int. J. Syst. Evol. Microbiol.* 69:895-898.
- Wurzbacher, C., Larsson, E., Bengtsson-Palme, J., Van den Wyngaert, S., Svantesson, S., Kristiansson, E., Kagami, M., and Nilsson, R. H. 2019. Introducing ribosomal tandem repeat barcoding for fungi. *Mol. Ecol. Resour.* 19:118-127.
- Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., and Glockner, F. O. 2013. The SILVA and "all-species living tree project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* 42:D643-D648.
- Yoon, S.-H., Ha, S.-M., Kwon, S., Lim, J., Kim, Y., Seo, H., and Chun, J. 2017. Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* 67:1613-1617.
- Yurkov, A. M., and Kurtzman, C. P. 2019. Three new species of *Tremellomyces* isolated from maize and northern wild rice. *FEMS Yeast Res.* 19:foz004.
- Zachos, F. E. 2016. Species Concepts in Biology: Historical Development, Theoretical Foundations and Practical Relevance. Springer International Publishing, Cham, Switzerland.
- Zhang, N., Luo, J., Rossman, A. Y., Aoki, T., Chuma, I., Crous, P. W., et al. 2016. Generic names in Magnaporthales. *IMA Fungus* 7:155-159.