Short title: Sequence-based classification and identification

Sequence-based classification and identification of Fungi

David Hibbett[1]

Biology Department, Clark University, Worcester, Massachusetts 01610

Kessy Abarenkov

Urmas Kõljalg

Maarja Öpik

Institute of Ecology and Earth Sciences, University of Tartu, 40 Lai St, Tartu 51005, Estonia

Benli Chai

James Cole

Qiong Wang

Department of Plant, Soil, and Microbial Sciences, Michigan State University, Plant and Soil Sciences Building, 1066 Bogue St. Room 540, East Lansing, Michigan 48824

Pedro Crous

Vincent Robert

Centraalbureau voor Schimmelcultures Fungal Biodiversity Centre (CBS-KNAW), 3508 AD, Utrecht, the Netherlands

Thorunn Helgason

Department of Biology, University of York, York YO10 5DD, United Kingdom

Joshua R. Herr

Department of Plant Pathology and Center for Plant Science Innovation, University of Nebraska, Lincoln, Nebraska 68503

Paul Kirk

Biodiversity Informatics and Spatial Analysis, Royal Botanic Gardens, Kew, Surrey TW9 3AF,

United Kingdom

Shiloh Lueschow

Kerry O'Donnell

NCAUR ARS USDA, 1815 N University St., Peoria, Illinois 61604

R. Henrik Nilsson

University of Gothenburg, Department of Biological and Environmental Sciences, Box 461, 405

30 Göteborg, Sweden

Ryoko Oono

Department of Ecology, Evolution, and Marine Biology, University of California Santa Barbara,

Santa Barbara, California 93106

Conrad Schoch

National Center for Biotechnology Information, National Library of Medicine, National

Institutes of Health, Bethesda, Maryland 20892

Christopher Smyth

Department of Plant Pathology and Environmental Microbiology, 121 Buckhout Laboratory,

Penn State University, University Park, Pennsylvania 16802

Donald M. Walker

Department of Biology, Tennessee Technological University, 1100 N. Dixie Ave., Cookeville,

Tennessee 38505

Andrea Porras-Alfaro

Department of Biological Sciences, Western Illinois University, Waggoner Hall 372, 1

University Circle Macomb, Illinois 61455

John W. Taylor

University of California, Department of Plant and Microbial Biology, 111 Koshland Hall,

Berkeley, California 94720

David M. Geiser

Department of Plant Pathology and Environmental Microbiology, 121 Buckhout Laboratory,

Penn State University, University Park, Pennsylvania 16802

Abstract: Fungal taxonomy and ecology have been revolutionized by the application of molecular methods and both have increasing connections to genomics and functional biology. However, data streams from traditional specimen- and culture-based systematics are not yet fully integrated with those from metagenomic and metatranscriptomic studies, which limits understanding of the taxonomic diversity and metabolic properties of fungal communities. This article reviews current resources, needs, and opportunities for sequence-based classification and identification (SBCI) in fungi as well as related efforts in prokaryotes. To realize the full potential of fungal SBCI it will be necessary to make advances in multiple areas. Improvements in sequencing methods, including long-read and single-cell technologies, will empower fungal molecular ecologists to look beyond ITS and current shotgun metagenomics approaches. Data quality and accessibility will be enhanced by attention to data and metadata standards and rigorous enforcement of policies for deposition of data and workflows. Taxonomic communities will need to develop best practices for molecular characterization in their focal clades, while also contributing to globally useful datasets including ITS. Changes to nomenclatural rules are needed to enable valid publication of sequence-based taxon descriptions. Finally, cultural shifts are necessary to promote adoption of SBCI and to accord professional credit to individuals who contribute to community resources.

INTRODUCTION

Fungi make up an underdescribed, poorly documented clade of eukaryotes that have immense ecological and economic impacts. Many fungi are microscopic or have cryptic life cycles that make detection difficult. Approximately 135 000 species of fungi have been described, but the actual diversity of the group is likely to be in the millions of species (Blackwell 2011, Taylor et al. 2014). Investigations into fungal diversity have traditionally been based on fruiting bodies or cultures, but an increasing number of studies that obtain DNA and RNA sequences directly from "environmental" sources—such as soil, water, air, or tissues of other organisms—are revealing potentially new fungal species at dramatically accelerated rates (Hibbett et al. 2011, Lindahl et al. 2013). For example, the Cryptomycota (Rozellomycota, rozellida) (Lara et al. 2010, Jones et al. 2011, Corsaro et al. 2014) and Archaeorhizomycetes are major clades of Fungi that are known almost entirely from environmental DNA sequences (Rosling et al. 2011, James and Berbee 2012). Many environmental sequences can only be identified to the level of a phylum or simply "fungi" (Nilsson et al. 2016), even in sophisticated analyses that use rigorous phylogenetic methods and that consider ribosomal RNA (rRNA) secondary structure (Glass et al. 2013). Thus, it is likely that other ancient clades are waiting to be described. Recent global or continental-scale analyses of patterns in fungal biodiversity have been based entirely on environmental DNA data (Amend et al. 2010, Talbot et al. 2014, Tedersoo et al. 2014, Davison et al. 2015).

Analyses of environmental DNA and RNA sequences may involve two complementary but distinct activities: Sequence-based classification (SBC) and sequence-based identification (SBI) (Herr et al. 2015). The goals of SBC are to discover, name, and classify fungal species

according to their phylogenetic relationships. In contrast, SBI uses the products of taxonomy to identify individuals and determine the composition of fungal communities with reference to existing classifications. SBI is a central element of ecological studies, including metatranscriptomic studies of community-level metabolic processes. Collectively, sequence-based classification and identification (SBCI) denotes the full range of activities required to detect and characterize fungi in the environment based on nucleic acid sequences (TABLE I).

New resources for SBCI are required to fully exploit the staggering volume of data flowing from fungal molecular ecology studies using high-throughput sequencing technologies. Huge numbers of undescribed taxa known only from environmental sequences need to be classified and linked to phenotypic, ecological, and functional traits. This article aims to: (i) envision the potential of SBCI and identify its conceptual challenges; (ii) survey current resources for SBCI in fungi and assess their strengths, limitations, and opportunities for enhancement; and (iii) consider options for integrating sequence-based fungal species into taxonomic systems based on specimens and cultures.

GOALS AND CONCEPTUAL CHALLENGES OF SBCI

In the ideal model of SBCI it would be possible to submit sequences of any nucleic acids from specimens or environmental samples to appropriate databases and retrieve lists of taxa with information on their relative abundance, phylogenetic relationships, ecological roles, and metabolic properties. The reference databases themselves would become richer as the results from each new study were integrated, creating new knowledge about fungal diversity, biogeography, population structure, and functional biology (FIG. 1). However, current methods of SBCI are based almost entirely on analyses of PCR-amplified nuclear rRNA genes, particularly the internal transcribed spacer (ITS) region, and they draw on incomplete taxonomic

and functional databases. New environmental data are not systematically integrated with existing resources. Here, we list six general challenges to achieving the model of SBCI described above; subsequent sections describe these challenges and the actions required to overcome them: (i) develop community standards for taxon recognition based on sequence data; (ii) create and curate sequence databases and analytical tools for SBCI; (iii) link sequence data to phenotypic data, including data from type specimens; (iv) achieve reproducibility in studies utilizing SBCI; (v) encourage the scientific and lay communities to adopt SBCI; and (vi) accord professional credit for contributing resources for SBCI.

SURVEYING THE LANDSCAPE: CURRENT RESOURCES AND NEEDS FOR SBCI IN FUNGI

Some of the earliest applications of comparative molecular data in fungi were to identify environmental samples that lacked the morphological characters necessary for traditional taxonomic identification (Gardes et al. 1991). Since then, many new web-accessible tools have been designed specifically for SBCI. URLs for resources described here are listed (TABLE II). All of these tools attempt to deal with the problem of misidentified or otherwise misleading sequences (Bridge et al. 2003, Bidartondo 2008) and insufficiently identified sequences (Ryberg et al. 2009). These are the so-called "dark taxa" (Parr et al. 2012, p 2013) that reside in the International Nucleotide Sequence Database Collaboration (INSDC), with its three partners: GenBank at the National Center for Biotechnology Information (NCBI), the European Nucleotide Sequence Archive of the European Molecular Biology Laboratory (EMBL), and the DNA Data Bank of Japan (DDBJ) (Cochrane et al. 2016).

Specialized databases and general resources for SBCI.—A number of databases and tools have been developed to enable high-throughput analyses and address the problems caused by misidentified or insufficiently identified sequences. One of the first such efforts was the

Ribosomal Database Project (RDP) (Larsen et al. 1993, Maidak et al. 1994), which initially

consisted of nuclear small subunit (SSU) rRNA gene sequences from Archaea, Bacteria, and

Eukarya. The RDP has evolved since its inception, providing alignments, trees, and similarity

assessment tools, based on its core database of curated sequences. Recent additions to RDP

include new databases for fungal ITS and large subunit (LSU) rRNA genes (Cole et al. 2014).

An innovation of the RDP is the naïve Bayesian classifier, which has been implemented

for both the bacterial and archaeal small subunit (16S) rRNA genes (Wang et al. 2007) and the

fungal LSU rRNA gene (Liu et al. 2012) and ITS (Porras-Alfaro et al. 2014). The RDP classifier

uses taxonomically grouped sequences as a training set and it attempts to place query sequences

(analyzed singly or in batch mode) at the lowest (least inclusive) taxonomic level possible in the

hierarchy. The RDP classifier has been implemented in automated microbial ecology pipelines

such as QIIME (Caporaso et al. 2010) and Mothur (Schloss et al. 2009). The latest release of

RDP now includes an LSU alignment of 108 901 sequences and an updated training set

(http://rdp.cme.msu.edu) (Cole et al. 2014). In addition, RDP has made available versions of the

Classifier trained on the UNITE database (described below) including an updated training set

named to honor the Australasian mycologist J.H. Warcup (Parberry and Robertson 1999). The

Warcup training set (2.0) is available from the RDP website and contains 8551 species and 17

878 ITS sequences organized into taxonomic classifications. Both sets performed well in testing

(Deshpande et al. 2015).

The UNITE database was initially developed for ITS sequences of ectomycorrhizal fungi

(Kõljalg et al. 2005), but it has since been expanded to represent all fungi (Abarenkov et al.

2010, Kõljalg et al. 2013). UNITE is the product of a consortium of fungal ecologists,

taxonomists, and bioinformaticians. Queries can be performed with reference to the custom-

curated UNITE database, which includes many sequences from specimens that were collected and deposited by taxonomic specialists specifically for the purpose of building the database or all INSDC sequences. Only about 0.5% of the 657813 sequences in UNITE have yet to be submitted to the INSDC databases but all can be searched on-line.

UNITE groups sequences into "Species Hypotheses" (SH), which are generated by a two-tier clustering process, first at the subgeneric/generic level then at a species level (Kõljalg et al. 2013). Sequences are grouped into SHs based on similarity to a reference sequence at a particular similarity cutoff (e.g. 97%, 98%, etc.). All SHs have a unique digital object identifier (DOI) to promote unambiguous communication. The UNITE database has also been adapted for use in QIIME and Mothur and a range of other applications (https://unite.ut.ee/repository.php).

The Barcode of Life Database (BOLD) and its European mirror, EUBOLD, provides sequence based identification tools (Ratnasingham and Hebert 2007). However, BOLD retains a strong focus on animal identification. Although it contains more than 117 000 ITS sequences with just under 100 000 sequences tied to voucher data, more than 84% of these were obtained from the public INSDC databases. Coverage for "Fungi and other organisms" is just over 16 000 species, compared to approximately 154 000 species for Animals and 58 000 for Plants. Currently, NCBI will only assign a BARCODE keyword to sequences that meet all BOLD criteria, including the deposition of original sequence trace files. Direct submissions to BOLD allow for trace submission and although it is technically feasible at NCBI as well, no fungal sequences have yet been tagged with the official BARCODE keyword in the INSDC. At the same time, the term "DNA barcode" has been widely applied as a generic descriptor in mycology for an ITS sequence tied to voucher data.

The RefSeq Targeted Loci Project is a relatively new resource at NCBI. This is part of the curated microbial resources in the Reference Sequences databases (O'Leary et al. 2015, Tatusova et al. 2015). RefSeq ITS highlights significant sequences that do not meet all the strict DNA barcode requirements but that could act as references for SBCI. RefSeq ITS accessions are selected from GenBank records that pass a rigorous set of standards and can be curated by third parties (Schoch et al. 2014). Some entries have already been shared between RefSeq and UNITE and a formal pipeline to facilitate this process is in development.

RefSeq Targeted Loci will remain focused on sequences from type material, with a few exceptions for sequences from verified voucher material for economically important species or poorly represented groups. It currently contains data from the three most commonly used ribosomal regions. The most comprehensive RefSeq dataset contains 4779 ITS accessions. It is intended that this resource will increase representation and expand into other commonly used marker sequences with continued involvement of other databases and the research community. Whole genome sequence data from type material is one area of expansion and is already being applied for bacteria at NCBI. This is incorporated in a broader effort to add type material identifiers to the NCBI Taxonomy database, which allows use of BLAST and other tools to search multiple databases for reliable entries that are specifically tied to type material (Federhen 2015).

The projects described above provide diverse tools for analyzing environmental sequences, and new resources are being produced regularly. For example, a database of ITS sequences for medically important species was recently released by the International Society for Human and Animal Mycology (ISHAM) (Irinyi et al. 2015) and is now linked in NCBI and UNITE.

SBCI using alternative markers.—ITS has emerged as the primary marker for fungal barcoding (Schoch et al. 2012), but it is not the only region used for SBCI and it is not always the best choice. In fact, one of the first databases for SBCI in fungi used mitochondrial LSU rRNA gene sequences to identify ectomycorrhizal fungal taxa (Bruns et al. 1998). A current resource using alternative markers is the Maarj*AM* database (Öpik et al. 2010) of the arbuscular mycorrhizal (AM) Glomeromycota, which emphasizes SSU rRNA gene sequences. Maarj*AM* applies "virtual taxon" (VT) identifiers connected to curated "type sequences" and it currently contains ca. 22 000 accessions of SSU rRNA sequences linked to ca. 400 VT, 2000 ITS, and 2000 LSU sequences (Öpik et al. 2014). Maarj*AM* also includes curated information on sequence origin (ecological metadata), which is increasingly used to study AM fungal distribution patterns (Ohsowski et al. 2014). The inclusion of SSU rRNA gene sequences in Maarj*AM* reflects a historical preference for this region by AM fungal systematists (Schussler 2001) and is validated by comparison to ITS and LSU sequences within and among the same strains (Thiéry et al. 2016). As long-read sequencing technologies improve (Koren et al. 2013), it will become routine to obtain contiguous sequences containing ITS and its flanking LSU and SSU regions, which will facilitate integration of current datasets based on single markers.

Intragenomic heterogeneity among ITS copies poses challenges for SBCI, although an analysis using 454 sequencing of ITS amplicons in 99 diverse fungal isolates suggested that it may not be a major source of error in estimates of species richness (Lindner et al. 2013). Nevertheless, it is necessary to look beyond ITS in groups that are known to suffer from high levels of intragenomic heterogeneity, such as the polypore genus *Laetiporus* (Lindner and Banik 2011, Lindner et al. 2013). Another group in which ITS is a poor choice for SBCI is *Fusarium*,

which possesses highly divergent non-orthologous copies of the ITS2 (O'Donnell et al. 1998). ITS generally fails to resolve species boundaries in *Fusarium* that are discernible using other markers (O'Donnell et al. 2015). This difficulty has led to the recommended use of multiple protein-coding genes for species-level classification in *Fusarium*, with the intron-rich 5′ end of translation elongation factor 1-alpha gene (*tef1*) being favored as the primary locus for species identification, followed by partial nucleotide sequences of DNA-directed RNA polymerase II largest (*RPB1*) and second largest subunit (*RPB2*). Accordingly, the FUSARIUM-ID and *Fusarium MLST* databases were created to facilitate identification in *Fusarium* via BLAST (Geiser et al. 2004, O'Donnell et al. 2010). The current versions of FUSARIUM-ID and *Fusarium MLST* contain multilocus data from 1366 isolates, with data from ten different protein-coding and ribosomal gene regions (Park et al. 2010). Stand-alone multilocus databases have been developed for several other fungal groups, including TrichoBLAST for *Trichoderma* (Kopchinskiy et al. 2005) and multiple databases maintained at the CBS Biodiversity Centre focusing on *Aspergillus*, *Morchella*, *Russula*, dermatophytes, and indoor fungi.

Markers such as *tef1* are superior to ITS for SBCI in some groups (Stielow et al. 2015). Nevertheless, to develop a comparative sequence database with the broadest possible taxonomic coverage, it will be important for taxonomists to continue to include ITS among their target loci whenever possible. This was the philosophy followed by the AFTOL project, which focused on multigene analyses of higher-level fungal phylogeny but also generated ITS sequences to facilitate molecular ecology (Lutzoni et al. 2004). Conversely, ecologists using single-locus approaches should understand that ITS will enable only coarse identification in some clades, and they should be prepared to follow up with targeted analyses of other regions when species-level identification is critical.

Increasingly, data for SBCI are coming from whole genome sequencing studies, including those performed under the auspices of the 1000 Fungal Genomes Project (1KFG). Unfortunately, because of their repetitive nature, rRNA gene sequences are usually left unassembled in the sequence read archives (Schoch et al. 2014). This issue is getting some attention, as UNITE has begun recovering ITS sequences from genome sequence projects (Bengtsson-Palme et al. 2013). Where marker sequences can be extracted from whole genome sequences of sufficient quality and provenance, they can also be incorporated into the RefSeq system. The Fungal Genomes Program of the DOE Joint Genome Institute (Grigoriev et al. 2014) is now requiring ITS sequences as part of the metadata associated with samples for genome and transcriptome analyses. This information is requested to avoid sequencing of contaminants, but it could also be made available to link genomes with metagenomic data.

Barriers, inefficiencies, and unexploited opportunities.—The resources described above have evolved independently, resulting in a diversity of useful but largely unintegrated tools, with sequences and their metadata imported from a variety of sources. Greater attention to data standards and formats will facilitate data integration and SBCI (Tedersoo et al. 2015). At present, many individual databases require unique input formats, different data fields, and a great deal of manual curation. There is also inconsistency in the format and quality of metadata provided for sequence-based diversity studies, whether the sequences are derived from specimens or directly from the environment.

Ambitious attempts are underway to integrate and standardize phenotype descriptions across all biology (Deans et al. 2015). Fortunately, standards for formatting and minimum information have been developed and are available for adoption in SBCI. For example, the BIOM format (McDonald et al. 2012a) is widely used for transferring metadata from one source

to another, while the MIMARKS (minimum information about a marker gene sequence) standard (Yilmaz et al. 2011) establishes core annotation items for metagenomics/microbiome studies. These core metadata are key for SBCI, because they provide information regarding the origin and quality of sequences. RDP has produced a MIMARKS-compliant Google Sheet for metadata management that can be exported into WebIN and Sequin accessible formats, facilitating submission to EMBL and GenBank. With a few minimal adjustments, these tools should suffice for the development of a recommended format and standards for sequence data and metadata for SBCI that would also be compatible with formats for metagenomic analysis pipelines, submission to sequence repositories, and many other uses.

Linkages between environmental sequences and metabolic and phenotypic traits are needed to make predictions about the biological properties of organisms that have not been observed directly. Functional predictions for entire communities and individual species could be made by comparing environmental sequences to whole genomes. However, phenotypic data pose special challenges because they are diverse and are not archived in centralized databases. Apart from a few pioneering efforts, including MycoBank and the MycoPortal of the Macrofungi Collection Consortium, phenotypic information is generally scattered. An example of the sort of tools that are needed is the recently developed FUNGuild database and its associated bioinformatics resources (Nguyen et al. 2016), which make it possible to parse large numbers of environmental sequences into broad trophic guilds (e.g. saprotrophs vs. mycorrhizal fungi). Resources like FUNGuild have the potential to integrate SBCI with phenotypic and ecological traits, such as enzymatic activity (Talbot et al. 2015).

Building phenotypic databases will require an increased focus on specimen and culture annotations using a standardized format that can be traced across multiple databases and

electronically available publications. To take advantage of SBCI, these databases must be extended to include information on geographic distribution and habitat gained from environmental sequencing. Where possible, databases should use the Darwin Core data standards, which promote interoperability among biodiversity information resources (Wieczorek et al. 2012).

The lack of linkages among sequences, alignments, and phylogenies is also a limiting factor for SBCI. Phylogenetic inference is laborious and requires expert decisions about data inclusion and analytical settings. Delimitation of taxa by systematists is based on tree topologies and often involves consideration of morphological characters and other evidence such as genealogical exclusivity (Taylor et al. 2000), not merely sequence similarity. In contrast, SBCI as it is practiced now, is usually based only on pairwise sequence comparisons, with uniform (but adjustable) clustering criteria applied across all taxa. Phylogeny-based approaches to OTU delimitation have promise, such as the EPA-PTP (evolutionary placement algorithm-poisson tree process) method, which incorporates tree inference using RAxML (Zhang et al. 2013).

SBCI could be enhanced if phylogenetic trees and tree-based taxon concepts were incorporated into the pipelines used to analyze environmental sequences. The basic tools for tree-based SBCI already exist, such as lowest common ancestor (LCA) algorithms that can determine clade contents, as implemented in the (now defunct) *mor* pipeline for automated phylogenetic taxonomy of Agaricomycetes (Hibbett et al. 2005). Unfortunately, the input data for such analyses, alignments, and phylogenies are largely unavailable. An analysis by the Open Tree of Life Project found that only about 17% of published fungal phylogenies are available in electronic format (Drew et al. 2013), even though many mycological journals have stated requirements that datasets and trees be submitted to TreeBASE or Dryad. One reason that

submissions have lagged is that data submission is tedious, but lack of editorial oversight is also to blame. To facilitate the future development of tree-based SBCI that can take advantage of "gold standard" taxon delimitations by expert systematists, it will be necessary to create user-friendly tools for uploading phylogenetic trees to databases and for journals to enforce existing policies regarding data deposition. Development of informatics resources for tree storage is underway as part of the Open Tree of Life Project (Hinchliff et al. 2015).

LESSONS FROM PROKARYOTES

Microbiologists have developed rich databases and sophisticated informatics tools for SBCI, as well as broadly accepted standards for molecular taxon description. This section describes some of the major initiatives in SBCI of prokaryotes, including successful efforts that mycologists would be wise to emulate and pitfalls to avoid.

Prokaryotic taxonomy, informatics, and nomenclature.—Under the International Code of Nomenclature of Prokaryotes (ICNP) (Parker et al. 2015) 16S rRNA gene sequences are required for the description of new species, with a defined cutoff of <97% similarity between species (Wayne et al. 1987, Stackebrandt et al. 2002, Tindall et al. 2010). DNA-DNA Hybridization (DDH) or an equivalent technique is necessary if two strains show >97% similarity for the 16S rRNA region, although DDH has been criticized for its complexity, low reproducibility, and low throughput (Mende et al. 2013). Deposition of a type culture is also required.

The adoption of 16S rRNA genes as the primary taxonomic marker for Bacteria and Archaea greatly facilitated the discovery and documentation of prokaryotic diversity and resulted in a dramatic increase in the number of newly described species. Molecular documentation of described taxa is nearly complete, with about 99% of culturable prokaryotic species (including about 11 900 type strains) represented by 16S rRNA gene sequences (Yarza et al. 2008, Yarza et

al. 2010, Chun and Rainey 2014). Projects such as the sequencing orphan species (SOS) initiative have filled in the gaps for prokaryotes that had valid published names but lacked 16S rRNA genes sequences (Yarza et al. 2013). A similar effort to target orphan fungal species is needed, taking advantage of biodiversity collections networks such as the Global Biodiversity Information Facilty (GBIF), iDigBio and the World Federation for Culture Collections to identify sources of material.

Microbiologists have created comprehensive resources for analyses of rRNA gene sequences. For example, the All-Species Living Tree Project provides updated databases, alignments, and phylogenetic trees for about 11 900 species with sequenced type strains (Yarza et al. 2008, Munoz et al. 2011). Other curated databases contain sequences derived from both cultures and environmental samples. The latest release of the RDP includes 3.2 million aligned and annotated bacterial and archaeal sequences, of which about 85% come from environmental samples (Cole et al. 2014). RDP and other projects, such as SILVA (Quast et al. 2013), GreenGenes (DeSantis et al. 2006, McDonald et al. 2012b), Mothur (Schloss et al. 2009) and QIIME (Caporaso et al. 2010) provide additional tools and reference datasets to facilitate high throughput analyses of prokaryotic sequence data, and several now include reference datasets and tools for fungi from UNITE and other sources. The fungal community stands to benefit from continued collaboration with the microbial informatics community, with the aim of developing tools for fungal SBCI comparable to those already available for prokaryotes.

Despite the advances described above, there remains a huge gap between the number of described prokaryotes and the number of uncultured species that have already been discovered (Hedlund et al. 2015). Eighty-eight per cent of cultures belong to only four phyla, and more than half of the approximately 60 major prokaryotic lineages (phyla and divisions) are currently

represented only by sequence data (Rinke et al. 2013). Even when cultures are available, the minimal standards for descriptions of new taxa (http://www.bacterio.net/-minimalstandards.html) adopted by prokaryotic researchers, requiring phenotypic, chemotaxonomic, and genotypic data, too often represent insurmountable barriers to formal classification of much of the diversity of prokaryotes (Schleifer 2009).

Given the restrictions on naming uncultured species, microbiologists have made wide use of informal names for new species, as well as phyla and divisions (Brown et al. 2015, Spang et al. 2015). The *Candidatus* concept was proposed as a nomenclatural device for assigning provisional names to uncultured prokaryotes (Murray and Schleifer 1994, Murray and Stackebrandt 1995). However, sequence data alone are not sufficient to propose a *Candidatus* taxon; other information, such as phenotypic and ecological characters or in situ visualization, must also be provided (Schleifer 2009). The ad hoc committee for Systematic Bacteriology endorsed the use of *Candidatus* status for uncultured prokaryotes (Stackebrandt et al. 2002), but currently there are only 361 names under the *Candidatus* category in the List of Prokaryotic names with Standing in Nomenclature (LPSN) (Parte 2014), which represents a minuscule fraction of uncultured prokaryotic diversity. Several proposals have now been made to replace DNA–DNA hybridization for circumscription of species with genome sequences (Richter and Rosselló-Móra 2009, Chun and Rainey 2014) and even use it as type material (Whitman 2015, 2016).

Prokaryotic genomics and large-scale environmental sampling.—The number of prokaryotic genomes has increased exponentially in recent years (Land et al. 2015). In 2015 the GOLD database (http://www.genomesonline.org) reported 36 824 bacterial and 851 archaeal whole genome sequencing projects (Reddy et al. 2015) and over 30 000 prokaryotic genomes are

currently available through Ensembl (http://ensemblgenomes.org/). The explosive growth in prokaryote genomics has been driven in large part by coordinated efforts, such as the Genomic Encyclopedia of Bacteria and Archaea (GEBA). Nevertheless, there is still a need for many more genomes that represent the diversity of species in culture collections as well as unculturable taxa (Rinke et al. 2013). The NCBI Assembly database currently contains 4525 genomes from type strains, which represents 28% of the total number of deposited prokaryote with valid species names (15 559 Bacteria, 598 Archaea) (Federhen 2015). The use of genome wide comparisons has already made it possible to consider the streamlining of taxonomic identifications using metrics such as average nucleotide identity (ANI) and k-mer scores (frequencies of sequences of length k) (Federhen et al. 2016). Currently, RefSeq contains 18 454 16S rRNA gene sequences from type strains from 13 314 prokaryotic species (O'Leary et al. 2015).

The mycological counterpart to GEBA is the 1000 Fungal Genomes (1KFG) project (Grigoriev et al. 2014), which is providing a platform for sharing protocols and for community networking and contributing to the transformation of fungal biology into a genome-enabled discipline (Hibbett et al. 2013). Genomic data from 1KFG and other projects may not have a major impact on species-level identification, but they will be invaluable for predicting community function. Anyone can nominate a species for the 1KFG project, which aims to sequence two species per family. The emerging resources are publicly accessible and phylogenetically balanced, which will enhance their utility for SBCI. The 1KFG project is not complete, but mycologists should already be looking ahead to the collaborations that will develop the next generation of genomic resources, perhaps modeling their efforts on collaborative ventures associated with GEBA, such as the GEBA Type Strain (Kyrpides et al.

2014) and GEBA Microbial Dark Matter projects (http://microbialdarkmatter.org, http://standardsingenomics.org/index.php/sigen/article/view/sigs.5068949).

Prokaryotic genomics is set to expand further through the application of new technologies such as single-cell genomics (Hofer 2013, Rinke et al. 2013), which will provide data from previously inaccessible taxa and hybrid genome assembly, which will increase sequencing coverage and accuracy (Chun and Rainey 2014). Metagenomic binning makes it possible to assemble complete or near-complete genomes from unculturable taxa (Albertsen et al. 2013), while improved methods for isolating and growing previously uncultured microbes within their natural soil environments (Nichols et al. 2010) are opening possibilities to important new discoveries (Ling et al. 2015).

New analytical approaches are also facilitating comparisons of genomes for SBCI in prokaryotes. For example, new methods for multilocus sequence analysis have been developed (Mende et al. 2013) that could increase accuracy and ameliorate the impact of horizontal gene transfer, which is of particular concern in prokaryotes but also occurs in fungi (Kämpfer and Rosselló-Mora 2004, Chun and Rainey 2014). Nevertheless, even with whole genomes, prokaryotic SBCI is still challenged to define appropriate cut-off levels for species discrimination (Fox et al. 1992, Kämpfer and Rosselló-Mora 2004, Fraser et al. 2009).

Prokaryotes are focal taxa in global and local culture-based and environmental mega sequencing projects such as NEON (http://www.neoninc.org), the Earth Microbiome (http://www.earthmicrobiome.org/), Terragenome (http://www.terragenome.org/), the Global Microbial Identifier (http://www.globalmicrobialidentifier.org), and Human Microbiome Project (http://www.hmpdacc.org). Mycologists need to be engaged in these efforts, both to promote technology transfer and to ensure that the projects generate data that are useful for fungal SBCI.

INTEGRATING SBCI AND FUNGAL TAXONOMY AND NOMENCLATURE

In contrast to the prokaryotic ICNB, the International Code of Nomenclature for algae, fungi and plants (the Code) does not require sequence data to describe taxa, but it does mandate that a type specimen be indicated (an illustration or a culture may also be used for fungi under some conditions; Arts. 8.1, 8.4, 40.5) (McNeill et al. 2012). Consequently, there are many validly named species that lack sequences (FIG. 2), as well as species inferred only from environmental sequences that lack names (Hibbett et al. 2011, Öpik et al. 2014). This disconnect presents a barrier to SBCI and it limits understanding of fungal diversity among nonspecialists. To integrate fungal taxonomy and molecular ecology, it will be necessary to expand reference sequence databases based on specimens and consider formally naming species based only on sequences. Sequence data should be obtained not only from new collections but from existing specimens and isolates in fungaria and culture collections, which have been shown to house substantial unrecognized biodiversity (Brock et al. 2009, Nagy et al. 2011).

Growing sequence databases for described species.—From 1999 to 2009, only about 26% of newly described species of fungi had sequences of any locus deposited in GenBank (Hibbett et al. 2011), but from 2010 to 2015 the proportion of new species with sequences increased to 50% overall (60% in 2015, FIG. 2, listed in SUPPLEMENTARY INFORMATION). One way to grow the number of sequenced species even further would be to modify the Code to require sequence data as part of taxon descriptions. However, this would prevent the naming of fossils and other challenging materials, such as obligate biotrophs and other unculturable fungi that lack macroscopic structures, and could result in the needless destruction of specimens. For example, 64 species of Laboulbeniales (minute obligate insect symbionts) were described in 2010–2016

(so far), but only two have sequence data (see SUPPLEMENTARY INFORMATION). Even in cases where organisms grow well in culture or produce large reproductive structures, some scientists lack resources for molecular work, particularly in the developing world, where much of the undescribed fungal diversity resides. Increased collaboration between fungal molecular systematists and field mycologists could reduce the number of species that are described without sequence data.

Sequence data should not be required for taxon description under the Code, but there should be general standards, enforced by journal policies and reviewers, that molecular data and metadata accompany all new taxon descriptions when it is reasonable to expect them. Communities of taxonomic specialists will need to determine the appropriate loci to be sequenced, which should include ITS and other markers that are commonly used for SBCI in each group (e.g. *tef*1 in *Fusarium* and SSU in Glomeromycota). Organizations such as the International Commission on the Taxonomy of Fungi (ICTF; http://www.fungaltaxonomy.org/) could play a role in promoting best practices. Collaborative efforts to sequence representatives of described species, as in the prokaryotic SOS initiative, should be encouraged. Whenever possible, type materials should be sequenced, but this is often difficult. In such cases, systematists should consider epitypification of names with sequence data from a recent collection, from which a culture and/or nucleic acid sample can be derived (Ariyawansa et al. 2014). This approach is particularly important because many important fungal names are not associated with a physical type specimen.

Toward sequence-based species description.—The steps outlined above will increase the representation of described species in sequence databases, but they will not solve the problem posed by species known only from environmental sequences. For example, at present UNITE

contains 3412 SHs identified only to phylum (Nilsson et al. 2016). Similarly, Maarj*AM* currently

contains 292 Glomeromycota VTs composed of only unidentified sequences and 60 named VTs.

In natural communities, only 30–50% of AM fungal VTs are named (Ohsowski et al. 2014).

Groups such as Archaeorhizomycetes and Cryptomycota are known almost entirely from

environmental sequences (Hibbett 2016). Sequence-based taxonomic platforms such as UNITE

and MaarjAM make it possible to group sequences into species hypotheses (SH), virtual taxa

(VT), and other MOTUs (molecular operational taxonomic units) (Blaxter et al. 2005). These

resources permit fungal ecologists and evolutionary biologists to communicate about taxa known

only from sequences and to conduct repeatable analyses (because MOTUs are delimited based

on explicit algorithms). However, MOTUs of any sort are obscure concepts to nonspecialists,

and they are not included in names-based taxonomic databases such as the Catalogue of Life or

Global Biodiversity Information Facility. To facilitate communication and maximize awareness

of fungal biodiversity among scientists and the general public, it would be helpful to assign

Linnaean binomials to species based solely on sequences (Hibbett et al. 2011).

　　　Mycologists were among the first to advocate for the use of DNA sequences in species

diagnoses (Reynolds and Taylor 1991, Renner 2016). Nevertheless, there is still no consensus

regarding the desirability of formally naming species of fungi based only on sequence data. This

may reflect confusion about the difference between taxonomy and nomenclature, which are

discrete but closely linked disciplines, and the role of the Code. Nomenclature is based on a set

of rules and conventions that determine whether names are "valid" (i.e. they have been published

correctly and are properly formed and documented), whereas taxonomy is the science concerned

with delimiting groups of organisms based on inferred phylogenetic relationships. Objections to

naming sequence-based species are largely rooted in concerns about taxonomic error, such as

that due to gene tree/species tree conflict or intragenomic heterogeneity. Concerns have also been raised that that some species described on the basis of sequences alone will prove to be synonyms of already described species that lack sequence data (Nagy et al. 2011). However, if estimates of the extant diversity of fungi are anywhere near correct (Blackwell 2011, Taylor et al. 2014), then it is likely that most species discovered with sequence data will be novel (Hibbett et al. 2011). In any case, it is not the purpose of the Code to certify taxonomic hypotheses; its rules do not address the scientific evidence required to justify taxonomic decisions, such as species delimitation. Nonetheless, the Code plays an important gatekeeper role in taxonomy, because only valid names are considered to be "correct" and have the protection of priority.

The next opportunity to change the Code will come in 2017, at the International Botanical Congress XIX in Shenzhen, China, and any changes will become effective on 1 Jan 2018. A proposal to modify the Code to allow sequence-based species description in fungi has recently been published and will be voted on at IBC XIX (Hawksworth et al. 2016). The proposal would apply to species "where data were obtained from voucherless environmental sequencing techniques and no individual material is available to serve as the type of a name of a new taxon." Proposed recommendations would suggest that: (i) new taxa based on sequences should be described with reference to a published phylogenetic analysis; (ii) sequences representing the new taxon should have been detected in multiple independent studies; and (iii) sequences used for taxon description should be drawn from regions deemed appropriate by the relevant taxonomic communities (Hawksworth et al. 2016). Other measures that would strengthen sequenced-based species description would include requiring that all appropriate reference databases have been searched and closely related sequences have been included in analyses and that appropriate tests for sequence quality, including tests for chimeric sequences,

have been performed (Nilsson et al. 2012, Hyde et al. 2013, Lindahl et al. 2013, Hart et al. 2015, Selosse et al. 2016). Again, organizations such as the ICTF could play an important role in developing and disseminating best practices, and journals could help by adopting and enforcing these criteria.

It remains to be seen if the proposal to allow sequences to serve as types (Hawksworth et al. 2016) will be approved. However, mycologists do not need to wait until 2018 to begin describing species based on environmental sequences. In the interim, demonstration papers could be produced that present worked examples of sequence-based species descriptions. To some extent, this is already happening. For example, a new species of Neocallimastigomycota, *Piromyces cryptodigmaticus*, was described based only on sequence data, although a technical voucher sample was obtained (Kirk 2012). Similarly, 48 species of *Archaeorhizomyces* that have been detected with environmental sequences in at least two independent studies were identified, although they were not given Linnaean binomials (Menkis et al. 2014).

A fine example of SBC in action is provided by *Hawksworthiomyces sequentia* ENAS, which was described by de Beer et al. (2016) based on two independent sequences, one from Canada and the other from Sweden. The ENAS suffix indicates that this taxon was described based on environmental sequences (TABLE I; Taylor 2011). In describing *H. sequentia*, de Beer et al. made a serious effort to obtain all relevant cultures, and they performed thorough phylogenetic analyses of sequences from GenBank and UNITE. The study was peer-reviewed and a diagnostic alignment was provided as supplementary material. Thus, the description of *H. sequentia* provides a model for other mycologists who wish to describe ENAS.

Until the Code is changed sequence-based species names will not have the protection of priority. If taxonomists working with physical specimens rediscover a species that was described

based on sequence data, they should consider validating the original sequence-based name, ideally in collaboration with the author of the sequence-based name. To do otherwise would cause confusion and create redundant names, as well as deprive individuals of credit for their discoveries. To promote recognition of sequence-based species, databases of names ("nomenclators") such as Index Fungorum, MycoBank, and NCBI should be encouraged to take up sequence-based species names, perhaps labeled as "candidate species" (Hibbett et al. 2011), or with a suffix to indicate their provisional status (nom. prov.) or their nature as sequence-based taxa (MOTU, ENAS), even if they are not compliant with the Code.

ENABLING THE COMMUNITY TO PARTICIPATE IN SBCI

Mycologists were early adopters of SBCI (Gardes et al. 1991, Reynolds and Taylor 1991, Bruns et al. 1998), molecular phylogenetics (White et al. 1990, Bruns et al. 1991) and comparative genomics (Birren et al. 2002), but there is still a long way to go before fungal molecular ecology, taxonomy, and functional biology are fully integrated (FIG. 1). Among the most important challenges facing fungal biology are to encourage and enable more mycologists to adopt SBCI, and to contribute molecular and phenotypic information to publicly accessible databases.

Promoting SBCI.—Probably the most effective means of promoting SBCI is to publish illustrative examples of research using community resources (TABLE II). Excellent models already exist for some areas, such as ecological studies of indoor air (Adams et al. 2013), human gut and skin microbes (Findley et al. 2013, Hoffmann et al. 2013), AM fungi (Davison et al. 2015), or forest soils (Talbot et al. 2014, Tedersoo et al. 2014). Innovative publications can suggest the way forward but without efforts to define and disseminate best practices, broad adoption will be slow. Critical analyses of published examples and focused meetings held in conjunction with regional and international mycology and microbiology meetings could help to

identify the optimal approaches for each group of fungi and research problem. To teach the best practices, tutorials can be developed for workflows and posted on sites such as Wikipedia or YouTube, with links to mycological sites such as MycoBank and UNITE. Mycological sites should catalog these resources, with Joseph Felsenstein's compilation of phylogeny programs as a model (http://evolution.genetics.washington.edu/phylip/software.html). Practical workshops providing intensive training in both wet-bench and computational methods would also promote adoption of best practices in SBCI.

Academic mycologists have been the leaders in developing SBCI, but to reach the full potential of these methods they must reach out to other groups, including scientists in industry and government who are focused on agriculture, biotechnology, and medicine. These professionals may be the best equipped to develop workflows and teach workshops that reach large numbers of practitioners (e.g. pathologists, quarantine agents, and industrial microbiologists). Liaison with professional societies, such as the American Phytopathological Society and others, will be key to involving these professionals in teaching SBCI.

Another group that will be important to the general acceptance of SBCI is non-professional mycologists, who often have a better knowledge of fungi in the environment than many professional mycologists. Outreach to non-professional groups, such as the North American Mycological Association (NAMA) and regional mycological clubs will help broaden awareness of SBCI. A good example is the North American Mycoflora Project (Bruns 2012), which seeks to conduct a continental-scale survey of fungal diversity. This ambitious undertaking relies on the distributed knowledge of fungal habitats and geographic distributions that resides in the amateur mycology community and was launched with support of NAMA. Tutorials and articles about SBCI in Web resources such as Mushroom Observer, Wikipedia, and

the Encyclopedia of Life could also reach large numbers of amateur mycologists. Citizen scientists should also be engaged in SBCI. Indeed, there are already examples of studies using SBCI with the support of citizen scientists, whose ability to sample environments from diverse locations has enabled ecological sampling on a broad scale (Amend et al. 2010; Barberan et al. 2015).

Recruiting citizen scientists calls for a multipronged approach, targeting K–12 schools, universities, and the general public, with social media playing a major role. Reaching students will require inclusion of SBCI in curricula and textbooks and direct outreach to organizations such as the National Association of Biology Teachers. Organizations involved in conservation and restoration biology can also help to promote SBCI. For this to occur, fungi must be explicitly included among the target organisms for conservation efforts, and there must be enhanced understanding of the role of SBCI for monitoring fungal populations and assessing their risk of extinction (Veresoglou et al. 2015).

Maintenance and sustainability of databases and collections.—The value of databases increases as they grow, and that growth depends on scientists depositing data and analytical workflows in publicly accessible repositories. To simplify this process, it would be helpful if researchers would document their workflows in formats that are ready to upload, e.g. an IPython Notebook. As noted previously, journals and granting agencies have an important role to play in enforcing good practices in research workflows and archiving.

The growth of SBCI also highlights the importance of fungaria and culture collections. Specimens can provide clues to the morphology of organisms known only from sequences, and cultures may provide sufficient DNA for genomic analysis and enable experimental studies that complement environmental observations. Direct sampling of nucleic acids in the environment

facilitates large-scale discovery of taxa and complements studies relying on culturing or direct observation of individual organisms. A brilliant example of this complementarity is provided by the aforementioned Archaeorhizomycetes, which was an unnamed and widely distributed clade of phenotypically mysterious fungi known only from environmental LSU sequences (Schadt et al. 2003). The connection of this DNA-based observation to its source organisms was made when researchers obtained rRNA gene sequences from previously cultivated fungi that matched the environmental DNA clade (Rosling et al. 2011). Culture collections and fungaria are storehouses of known fungal diversity, but the majority of these resources also remain uncharacterized at the sequence level. To facilitate SBCI, it is essential that culture collections and fungaria receive support, both for SBCI and for their continued maintenance. More generally, all community resources, including databases, require plans for long-term financial sustainability, possibly involving user fees as well as institutional support (Bell et al. 2010, Reiser et al. 2016).

According credit for tending the commons.—SBCI becomes more powerful as new data are added to public repositories. However, career advancement is traditionally based on publications, not unattributed contributions to databases and collections. An argument can be made that the most concrete legacy of any scientist is the data and materials that they have made available to the field (McNutt 2014). In the case of biology based on "big data," this argument is clear and to foster the science we need to move from a reward system that strongly emphasizes publications to one that also values "quantum contributions" (Maddison et al. 2012), such as DNA sequences and other data, specimens, and cultures, as well as curation and identification services, software, databases, and websites.

Technological as well as cultural shifts will be needed to assess and accord credit for contributions to community resources (McDade et al. 2011). Automated systems to track usage of data resources could provide quantitative measures of their impact, but they need to include unique identifiers for the workers who provided the data. New online publications such as Biodiversity Data Journal (http://biodiversitydatajournal.com/) and Research Ideas and Outcomes (http://riojournal.com/) allow researchers to publish datasets and workflows, allowing them to be cited by other researchers who use those resources.

Behavioral changes are also needed. For example, studies using data from GenBank/INSDC should not only include accession numbers but should also consider citing the articles in which the sequences were first reported, which will accord merit to the data providers and give them an incentive to update accessions with publication information (Seifert et al. 2008). In general, increased communication between data providers and data users, possibly leading to co-authorships, can only improve the quality of analyses. Researchers should include sections in their curricula vitae describing contributions to community resources, as is traditionally done by systematists in sections for "taxonomic novelties". Senior scientists and reviewers have a particular responsibility to train administrators and tenure-and-promotion committees to appreciate the value of contributions to resources for SBCI and their impact on understanding of fungal diversity.

species" took place at the International Mycological Congress, 5 Aug 2014, Bangkok, Thailand. The authors thank Sam Kovaka, who performed the query of GenBank for species names (FIG. 2, SUPPLEMENTARY INFORMATION), P. Brandon Matheny, and two anonymous reviewers for helpful comments. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the Victorian government or the US Department of Agriculture. USDA is an equal opportunity provider and employer. CLS acknowledges support from the Intramural Research Program of the National Institutes of Health, National Library of Medicine. MÖ acknowledges support from the Estonian Research Council (grant No. IUT20-28) and European Regional Development Fund (Centre of Excellence EcolChange). DSH acknowledges support of NSF award No. DEB-1208719.

LITERATURE CITED

Abarenkov K, Henrik Nilsson R, Larsson KH, Alexander IJ, Eberhardt U, Erland S, Hoiland K, Kjoller R, Larsson E, Pennanen T, Sen R, Taylor AF, Tedersoo L, Ursing BM, Vralstad T, Liimatainen K, Peintner U, Koljalg U. 2010. The UNITE database for molecular identification of fungi—recent updates and future perspectives. New Phytol 186:281–285.

Adams RI, Miletto M, Taylor JW, Bruns TD. 2013. The diversity and distribution of fungi on residential surfaces. PLoS ONE 8:e78866.

Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat Biotechnol 31:533–538.

Amend AS, Seifert KA, Samson R, Bruns TD. 2010. Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. Proc Natl Acad Sci USA 107:13748–13753.

Ariyawansa HA, Hawksworth DL, Hyde KD, Jones EBG, Maharachchikumbura SSN, Manamgoda DS, Thambugala KM, Udayanga D, Camporesi E, Daranagama A, Jayawardena R, Liu J-K, McKenzie EHC, Phookamsak R, Senanayake IC, Shivas RG, Tian Q, Xu J-C. 2014. Epitypification and neotypification: guidelines with appropriate and inappropriate examples. Fungal Divers 69:57–91.

Barberan A, Ladau J, Leff JW, Pollard KS, Menninger HL, Dunn RR, Fierer N. 2015. Continental-scale distributions of dust-associated bacteria and fungi. Proc Natl Acad Sci USA 112:5756–5761.

Bell J, Masaoka J, Zimmerman S. 2010. Nonprofit sustainability: making strategic decisions for financial viability. San Francisco: John Wiley & Sons. 208 p.

Bengtsson-Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, De Wit P, Sánchez-García M, Ebersberger I, de Sousa F, Amend A, Jumpponen A, Unterseher M, Kristiansson E, Abarenkov K, Bertrand YJK, Sanli K, Eriksson KM, Vik U, Veldre V, Nilsson RH. 2013. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods Ecol Evol 4:914–919.

Bidartondo MI. 2008. Preserving accuracy in GenBank. Science 319:1616.

Birren B, Fink G, Lander E. 2002. Fungal genome initiative. Cambridge, Massachusetts: Whitehead Institute Center for Genome Research. 26 p.

Blackwell M. 2011. The fungi: 1, 2, 3 ... 5.1 million species? Am J Bot 98:426–438.

Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E. 2005. Defining operational taxonomic units using DNA barcode data. Phil Trans R Soc B 360:1935–1943.

Bridge PD, Roberts PJ, Spooner BM, Panchal G. 2003. On the unreliability of published DNA sequences. New Phytol 160:43–48.

Brock PM, Döring H, Bidartondo MI. 2009. How to know unknown fungi: the role of a herbarium. New Phytol 181:719–724.

Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, Wilkins MJ, Wrighton KC, Williams KH, Banfield JF. 2015. Unusual biology across a group comprising more than 15% of domain Bacteria. Nature 523:208–211.

Bruns TD. 2012. The North American Mycoflora Project—the first steps on a long journey. New Phytol 192:972–974.

———, Szaro TM, Gardes M, Cullings KW, Pan JJ, Taylor DL, Horton TR, Krentzer A, Garbelotto M, Li Y. 1998. A sequence database for the identification of ectomycorrhizal basidiomycetes by phylogenetic analysis. Molec Ecol Resour 7:257–272.

———, White T, Taylor J. 1991. Fungal molecular systematics. Ann Rev Ecol Syst 22:525–564.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336.

Chun J, Rainey FA. 2014. Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. Int J Syst Evol Micr 64:316–324.

Cochrane G, Karsch-Mizrachi I, Takagi T, Sequence Database Collaboration IN. 2016. The International Nucleotide Sequence Database Collaboration. Nucleic Acids Res 44:D48–D50.

Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res 42:D633–D642.

Corsaro D, Walochnik J, Venditti D, Steinmann J, Muller KD, Michel R. 2014. Microsporidia-like parasites of amoebae belong to the early fungal lineage Rozellomycota. Parasitol Res 113:1909–1918.

Davison J, Moora M, Öpik M, Adholeya A, Ainsaar L, Bâ A, Burla S, Diedhiou AG, Hiiesalu I, Jairus T, Johnson NC, Kane A, Koorem K, Kochar M, Ndiaye C, Pärtel M, Reier Ü, Saks Ü, Singh R, Vasar M, Zobel M. 2015. Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. Science 349:970–973.

Deans AR, Lewis SE, Huala E, Anzaldo SS, Ashburner M, Balhoff JP, Blackburn DC, Blake JA, Burleigh JG, Chanet B, Cooper LD, Courtot M, Csˆsz Sn, Cui H, Dahdul W, Das S, Dececchi TA, Dettai A, Diogo R, Druzinsky RE, Dumontier M, Franz NM, Friedrich F, Gkoutos GV, Haendel M, Harmon LJ, Hayamizu TF, He Y, Hines HM, Ibrahim N, Jackson LM, Jaiswal P, James-Zorn C, Kˆhler S, Lecointre G, Lapp H, Lawrence CJ, Le NovËre N, Lundberg JG, Macklin J, Mast AR, Midford PE, MikÛ In, Mungall CJ, Oellrich A, Osumi-Sutherland D, Parkinson H, RamÌrez MJ, Richter S, Robinson PN, Ruttenberg A, Schulz KS, Segerdell E, Seltmann KC, Sharkey MJ, Smith AD, Smith B, Specht CD, Squires RB, Thacker RW, Thessen A, Fernandez-Triana J, Vihinen M, Vize PD, Vogt L, Wall CE, Walls RL, Westerfeld M, Wharton RA, Wirkner CS, Woolley JB, Yoder MJ, Zorn AM, Mabee P. 2015. Finding our way through phenotypes. PLoS Biol 13:e1002033.

de Beer ZW, Marincowitz S, Duong TA, Kim J-J, Rodrigues A, Wingfield MW. 2016. *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, http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1016/j.funbio.2016.07.004

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microb 72:5069–5072.

Deshpande V, Wang Q, Greenfield P, Charleston M, Porras-Alfaro A, Kuske CR, Cole JR, Midgley DJ, Tran-Dinh N. 2015. Fungal identification using a Bayesian classifier and the Warcup training set of internal transcribed spacer sequences. Mycologia 108:1–5.

Drew BT, Gazis R, Cabezas P, Swithers KS, Deng J, Rodriguez R, Katz LA, Crandall KA, Hibbett DS, Soltis DE. 2013. Lost branches on the tree of life. PLoS Biology 11:e1001636.

Federhen S. 2015. Type material in the NCBI taxonomy database. Nucleic Acids Res 43:D1086–D1098.

———, Rossello-Mora R, Klenk H-P, Tindall BJ, Konstantinidis KT, Whitman WB, Brown D, Labeda D, Ussery D, Garrity GM, Colwell RR, Hasan N, Graf J, Parte A, Yarza P, Goldberg B, Sichtig H, Karsch-Mizrachi I, Clark K, McVeigh R, Pruitt KD, Tatusova T, Falk R, Turner S, Madden T, Kitts P, Kimchi A, Klimke W, Agarwala R, DiCuccio M, Ostell J. 2016. Meeting report: GenBank microbial genomic taxonomy workshop (12–13 May 2015). Stand Genomic Sci 11:1–8.

Findley K, Oh J, Yang J, Conlan S, Deming C, Meyer JA, Schoenfeld D, Nomicos E, Park M, Kong HH, Segre JA, Comp NIHISC. 2013. Topographic diversity of fungal and bacterial communities in human skin. Nature 498:367–370.

Fox GE, Wisotzkey JD, Jurtshuk P. 1992. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int J Syst Evol Micr 42:166–170.

Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP. 2009. The bacterial species challenge: making sense of genetic and ecological diversity. Science 323:741–746.

Gardes M, Mueller GM, Fortin JA, Kropp BR. 1991. Mitochondrial DNA polymorphisms in *Laccaria bicolor*, *L. laccata*, *L. proxima* and *L. amethystina*. Mycol Res 95:206–216.

Geiser DM, del Mar Jiménez-Gasco M, Kang S, Makalowska I, Veeraraghavan N, Ward TJ, Zhang N, Kuldau GA, O'Donnell K. 2004. FUSARIUM-ID 1.0: a DNA sequence database for identifying *Fusarium*. In: Mulè G, Bailey JA, Cooke BM, Logrieco A, eds. Molecular diversity and PCR-detection of toxigenic *Fusarium* species and ochratoxigenic fungi: under the aegis of COST Action 835 'Agriculturally Important Toxigenic Fungi 1998–2003', EU project (QLK1-CT-1998-01380) and the ISPP 'Fusarium Committee'. Dordrecht, the Netherlands: Springer. p 473–479.

Glass DJ, Takebayashi N, Olson LE, Taylor DL. 2013. Evaluation of the authenticity of a highly novel environmental sequence from boreal forest soil using ribosomal RNA secondary structure modeling. Mol Phylognet Evol 67:234–245.

Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I. 2014. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res 42:D699–D704.

Hart MM, Aleklett K, Chagnon P-L, Egan C, Ghignone S, Helgason T, Lekberg Y, Öpik M, Pickles BJ, Waller L. 2015. Navigating the labyrinth: a guide to sequence-based, community ecology of arbuscular mycorrhizal fungi. New Phytol 207:235–247.

Hawksworth DL, Hibbett D, Kirk PM, Lücking R. 2016. (308–310) Proposals to amend Art. 8 to permit the use of DNA sequences as types of names of fungi. Taxon 65:899–900.

Hedlund BP, Dodsworth JA, Staley JT. 2015. The changing landscape of microbial biodiversity exploration and its implications for systematics. Syst Appl Microbiol 38:231–236.

Herr JR, Öpik M, Hibbett D. 2015. Towards the unification of sequenced-based classification and sequenced-based identification of host-associated microorganisms. New Phytol 205:27–31.

Hibbett D. 2016. The invisible dimension of fungal diversity. Science 351:1150–1151.

———, Nilsson R, Snyder M, Fonseca M, Costanzo J, Shonfeld M. 2005. Automated phylogenetic taxonomy: an example in the Homobasidiomycetes (mushroom-forming Fungi). Syst Biol 54:660–668.

———, Ohman A, Glotzer D, Nuhn M, Kirk P, Nilsson RH. 2011. Progress in molecular and morphological taxon discovery in Fungi and options for formal classification of environmental sequences. Fungal Biol Rev 25:38–47.

———, Stajich JE, Spatafora JW. 2013. Toward genome-enabled mycology. Mycologia 105:1339–1349.

Hinchliff CE, Smith SA, Allman JF, Burleigh JG, Chaudhary R, Coghill LM, Crandall KA, Deng J, Drew BT, Gazis R, Gude K, Hibbett DS, Katz LA, Laughinghouse HDI, McTavish EJ, Midford PE, Owen CL, Ree RH, Rees JA, Soltis DE, Williams T, Cranston KA. 2015. Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proc Natl Acad Sci USA 112:12764–12769.

Hofer U. 2013. Environmental microbiology: Exploring diversity with single-cell genomics. Nat Rev Microbiol 11:598–599.

Hoffmann C, Dollive S, Grunberg S, Chen J, Li HZ, Wu GD, Lewis JD, Bushman FD. 2013. Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. PLoS One 8:e66019.

Hyde KD, Udayanga D, Manamgoda DS, Tedersoo L, Larsson E, Abarenkov K, Bertrand YJK, Oxelman B, Hartmann M, Kauserud H, Ryberg M, Kristiansson E, Nilsson RH. 2013. Incorporating molecular data in fungal systematics: a guide for aspiring researchers. Curr Res Environ Appl Mycol J Fungal Biol 3:32.

Irinyi L, Serena C, Garcia-Hermoso D, Arabatzis M, Desnos-Ollivier M, Vu D, Cardinali G, Arthur I, Normand A-C, Giraldo A, da Cunha KC, Sandoval-Denis M, Hendrickx M, Nishikaku AS, de Azevedo Melo AS, Merseguel KB, Khan A, Parente Rocha JA, Sampaio P, da Silva Briones MR, e Ferreira RC, de Medeiros Muniz M, Castañón-

Olivares LR, Estrada-Barcenas D, Cassagne C, Mary C, Duan SY, Kong F, Sun AY, Zeng X, Zhao Z, Gantois N, Botterel F, Robbertse B, Schoch C, Gams W, Ellis D, Halliday C, Chen S, Sorrell TC, Piarroux R, Colombo AL, Pais C, de Hoog S, Zancopé-Oliveira RM, Taylor ML, Toriello C, de Almeida Soares CM, Delhaes L, Stubbe D, Dromer F, Ranque S, Guarro J, Cano-Lira JF, Robert V, Velegraki A, Meyer W. 2015. International Society of Human and Animal Mycology (ISHAM)-ITS reference DNA barcoding database—the quality controlled standard tool for routine identification of human and animal pathogenic fungi. Med Mycol 53:313–337.

James TY, Berbee ML. 2012. No jacket required--new fungal lineage defies dress code: recently described zoosporic fungi lack a cell wall during trophic phase. Bioessays 34:94–102.

Jones MDM, Forn I, Gadelha C, Egan MJ, Bass D, Massana R, Richards TA. 2011. Discovery of novel intermediate forms redefines the fungal tree of life. Nature 474:200–203.

Kämpfer P, Rosselló-Mora R. 2004. The species concept for prokaryotic microorganisms—an obstacle for describing diversity? Poiesis Prax 3:62–72.

Kirk PM. 2012. Nomenclatural novelties. Index Fungorum 1:1–1.

Kõljalg U, Larsson K-H, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, Erland S, Høiland K, Kjøller R, Larsson E, Pennanen T, Sen R, Taylor AFS, Tedersoo L, Vrålstad T. 2005. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. New Phytol 166:1063–1068.

———, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, Bates ST, Bruns TD, Bengtsson-Palme J, Callaghan TM, Douglas B, Drenkhan T, Eberhardt U, Dueñas M, Grebenc T, Griffith GW, Hartmann M, Kirk PM, Kohout P, Larsson E, Lindahl BD, Lücking R, Martín MP, Matheny PB, Nguyen NH, Niskanen T, Oja J, Peay KG, Peintner U, Peterson M, Põldmaa K, Saag L, Saar I, Schüßler A, Scott JA, Senés C, Smith ME, Suija A, Taylor DL, Telleria MT, Weiss M, Larsson K-H. 2013. Towards a unified paradigm for sequence-based identification of fungi. Mol Ecol 22:5271–5277.

Kopchinskiy A, Komon M, Kubicek CP, Druzhinina IS. 2005. Trichoblast: a multilocus database for *Trichoderma* and *Hypocrea* identifications. Mycol Res 109:658–660.

Koren S, Harhay GP, Smith TP, Bono JL, Harhay DM, Mcvey SD, Radune D, Bergman NH, Phillippy AM. 2013. Reducing assembly complexity of microbial genomes with single-molecule sequencing. Genome Biol 14:1.

Kyrpides NC, Hugenholtz P, Eisen JA, Woyke T, Gˆker M, Parker CT, Amann R, Beck BJ, Chain PSG, Chun J, Colwell RR, Danchin A, Dawyndt P, Dedeurwaerdere T, DeLong EF, Detter JC, De Vos P, Donohue TJ, Dong X-Z, Ehrlich DS, Fraser C, Gibbs R, Gilbert J, Gilna P, Göker FO, Jansson JK, Keasling JD, Knight R, Labeda D, Lapidus A, Lee J-S, Li W-J, Ma J, Markowitz V, Moore ERB, Morrison M, Meyer F, Nelson KE, Ohkuma M, Ouzounis CA, Pace N, Parkhill J, Qin N, Rossello-Mora R, Sikorski J, Smith D, Sogin M, Stevens R, Stingl U, Suzuki K-i, Taylor D, Tiedje JM, Tindall B, Wagner M, Weinstock G, Weissenbach J, White O, Wang J, Zhang L, Zhou Y-G, Field D, Whitman WB, Garrity GM, Klenk H-P. 2014. Genomic Encyclopedia of Bacteria and Archaea: sequencing a myriad of type strains. PLoS Biol 12:e1001920.

Land M, Hauser L, Jun SR, Nookaew I, Leuze MR, Ahn TH, Karpinets T, Lund O, Kora G, Wassenaar T, Poudel S, Ussery DW. 2015. Insights from 20 years of bacterial genome sequencing. Funct Integr Genomics 15:141–161.

Lara E, Moreira D, López-García P. 2010. The environmental clade LKM11 and *Rozella* form the deepest branching clade of Fungi. Protist 161:116–121.

Larsen N, Olsen GJ, Maidak BL, McCaughey MJ, Overbeek R, Macke TJ, Marsh TL, Woese CR. 1993. The ribosomal database project. Nucleic Acids Res 21:3021–3023.

Lindahl BD, Nilsson RH, Tedersoo L, Abarenkov K, Carlsen T, Kjøller R, Kõljalg U, Pennanen T, Rosendahl S, Stenlid J, Kauserud H. 2013. Fungal community analysis by high-throughput sequencing of amplified markers—a user's guide. New Phytol 199:288–299.

Lindner DL, Banik MT. 2011. Intra-genomic variation in the ITS rDNA region obscures phylogenetic relationships and inflates estimates of operational taxonomic units in genus *Laetiporus*. Mycologia 103:731–740.

———, Carlsen T, Henrik Nilsson R, Davey M, Schumacher T, Kauserud H. 2013. Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. Ecol Evol 3:1751–1764.

Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP, Mueller A, Schaberle TF, Hughes DE, Epstein S, Jones M, Lazarides L, Steadman VA, Cohen DR, Felix CR, Fetterman KA, Millett WP, Nitti AG, Zullo AM, Chen C, Lewis K. 2015. A new antibiotic kills pathogens without detectable resistance. Nature 517:455–459.

Liu K-L, Porras-Alfaro A, Kuske CR, Eichorst SA, Xie G. 2012. Accurate, rapid taxonomic classification of fungal large-subunit rRNA genes. Appl Environ Microb 78:1523–1533.

Lutzoni F, Kauff F, Cox CJ, McLaughlin D, Celio G, Dentinger B, Padamsee M, Hibbett D, James TY, Baloch E, Grube M, Reeb V, Hofstetter V, Schoch C, Arnold AE, Miadlikowska J, Spatafora J, Johnson D, Hambleton S, Crockett M, Shoemaker R, Sung G-H, Lücking R, Lumbsch T, O'Donnell K, Binder M, Diederich P, Ertz D, Gueidan C, Hansen K, Harris RC, Hosaka K, Lim Y-W, Matheny B, Nishida H, Pfister D, Rogers J, Rossman A, Schmitt I, Sipman H, Stone J, Sugiyama J, Yahr R, Vilgalys R. 2004. Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits. Am J Bot 91:1446–1480.

Maddison DR, Guralnick R, Hill A, Reysenbach AL, McDade LA. 2012. Ramping up biodiversity discovery via online quantum contributions. Trends Ecol Evol 27:72–77.

Maidak BL, Larsen N, McCaughey MJ, Overbeek R, Olsen GJ, Fogel K, Blandy J, Woese CR. 1994. The ribosomal database project. Nucleic Acids Res 22:3485–3487.

McDade LA, Maddison DR, Guralnick R, Piwowar HA, Jameson ML, Helgen KM, Herendeen PS, Hill A, Vis ML. 2011. Biology needs a modern assessment system for professional productivity. Bioscience 61:619–625.

McDonald D, Clemente JC, Kuczynski J, Rideout JR, Stombaugh J, Wendel D, Wilke A, Huse S, Hufnagle J, Meyer F, Knight R, Caporaso JG. 2012a. The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. Gigascience 1:1–6.

———, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. 2012b. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J 6:610–618.

McNeill J, Barrie FR, Buck WR, Demoulin V, Greuter W, Hawksworth DL, Herendeen PS, Knapp S, Marhold K, Prado J, Prud'homme van Reine WF, Smith GF, Wiersema JH, Turland NJ. 2012. International Code of Nomenclature for algae, fungi, and plants (Melbourne Code), Adopted by the Eighteenth International Botanical Congress Melbourne, Australia, Jul 2011. Köenigstein, Germany: Koeltz Scientific Books. 240 p.

McNutt M. 2014. The measure of research merit. Science 346:1155–1155.

Mende DR, Sunagawa S, Zeller G, Bork P. 2013. Accurate and universal delineation of prokaryotic species. Nat Methods 10:881–884.

Menkis A, Urbina H, James TY, Rosling A. 2014. Archaeorhizomyces borealis sp. nov. and a sequence-based classification of related soil fungal species. Fungal Biol 118:943–955.

Munoz R, Yarza P, Ludwig W, Euzéby J, Amann R, Schleifer K-H, Oliver Glöckner F, Rosselló-Móra R. 2011. Release LTPs104 of the All-Species Living Tree. Syst Appl Microbiol 34:169–170.

Murray R, Schleifer K. 1994. Taxonomic notes: a proposal for recording the properties of putative taxa of procaryotes. Int J Syst Bacteriol 44:174–176.

———, Stackebrandt E. 1995. Taxonomic note: implementation of the provisional status Candidatus for incompletely described procaryotes. Int J Syst Bacteriol 45:186–187.

Nagy L, Petkovits T, Kovács G, Voigt K, Vágvölgyi C, Papp T. 2011. Where is the unseen fungal diversity hidden? A study of *Mortierella* reveals a large contribution of reference collections to the identification of fungal environmental sequences. New Phytol 191:789–794.

Nguyen NH, Song Z, Bates ST, Branco S, Tedersoo L, Menke J, Schilling JS, Kennedy PG. 2016. FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. Fungal Ecol 20:241–248.

Nichols D, Cahoon N, Trakhtenberg EM, Pham L, Mehta A, Belanger A, Kanigan T, Lewis K, Epstein SS. 2010. Use of Ichip for high-throughput in situ cultivation of "uncultivable" microbial species. Appl Environ Microb 76:2445–2450.

Nilsson RH, Tedersoo L, Abarenkov K, Ryberg M, Kristiansson E, Hartmann M, Schoch CL, Nylander JAA, Bergsten J, Porter TM, Jumpponen A, Vaishampayan P, Ovaskainen O, Hallenberg N, Bengtsson-Palme J, Eriksson KM, Larsson K-H, Larsson E, Kõljalg U. 2012. Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences. MycoKeys 4:37–63.

———, Wurzbacher C, Bahram M, R. M. Coimbra V, Larsson E, Tedersoo L, Eriksson J, Duarte C, Svantesson S, Sánchez-García M, Ryberg MK, Kristiansson E, Abarenkov K. 2016. Top 50 most wanted fungi. MycoKeys 12:29–40.

O'Donnell K, Cigelnik E, Nirenberg HI. 1998. Molecular systematics and phylogeography of the *Gibberella fujikuroi* species complex. Mycologia 90:465–493.

———, Sutton DA, Rinaldi MG, Sarver BAJ, Balajee SA, Schroers H-J, Summerbell RC, Robert VARG, Crous PW, Zhang N, Aoki T, Jung K, Park J, Lee Y-H, Kang S, Park B, Geiser DM. 2010. Internet-accessible DNA sequence database for identifying Fusaria from human and animal infections. J Clin Microbiol 48:3708–3718.

———, Ward TJ, Robert VARG, Crous PW, Geiser DM, Kang S. 2015. DNA sequence-based identification of *Fusarium*: current status and future directions. Phytoparasitica 43:583–595.

Ohsowski BM, Zaitsoff PD, Öpik M, Hart MM. 2014. Where the wild things are: looking for uncultured Glomeromycota. New Phytol 204:171–179.

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD. 2015. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res gkv1189v1.

Öpik M, Davison J, Moora M, Zobel M. 2014. DNA-based detection and identification of Glomeromycota: the virtual taxonomy of environmental sequences1. Botany 92:135–147.

———, Vanatoa A, Vanatoa E, Moora M, Davison J, Kalwij JM, Reier U, Zobel M. 2010. The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). New Phytol 188:223–241.

Page RD. 2013. BioNames: linking taxonomy, texts, and trees. PeerJ 1:e190.

Parberry DG, Robertson NF. 1999. Obituary: John Henry (Jack) Warcup: 29 October 1921–15 May 1998. Mycologist 13:48–49.

Park B, Park J, Cheong K-C, Choi J, Jung K, Kim D, Lee Y-H, Ward TJ, O'Donnell K, Geiser DM, Kang S. 2010. Cyber infrastructure for Fusarium: three integrated platforms supporting strain identification, phylogenetics, comparative genomics and knowledge sharing. Nucleic Acids Res 2010:1–7.

Parker CT, Tindall BJ, Garrity GM. 2015. International Code of Nomenclature of Prokaryotes. Int J Syst Evol Microbiol (In press), doi:10.1099/ijsem.0.000778

Parr CS, Guralnick R, Cellinese N, Page RDM. 2012. Evolutionary informatics: unifying knowledge about the diversity of life. Trends Ecol Evol 27:94–103.

Parte AC. 2014. LPSN—list of prokaryotic names with standing in nomenclature. Nucleic Acids Res 42:D613–D616.

Porras-Alfaro A, Liu K-L, Kuske CR, Xie G. 2014. From genus to phylum: large-subunit and internal transcribed spacer rRNA operon regions show similar classification accuracies influenced by database composition. Appl Environ Microb 80:829–840.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596.

Ratnasingham S, Hebert PDN. 2007. BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Mol Ecol Notes 7:355–364.

Reddy TB, Thomas AD, Stamatis D, Bertsch J, Isbandi M, Jansson J, Mallajosyula J, Pagani I, Lobos EA, Kyrpides NC. 2015. The genomes online database (GOLD) 5: a metadata management system based on a four level (meta)genome project classification. Nucleic Acids Res 43:D1099–1106.

Reiser L, Berardini TZ, Li D, Muller R, Strait EM, Li Q, Mezheritsky Y, Vetushko A, Huala E. 2016. Sustainable funding for biocuration: the Arabidopsis Information Resource (TAIR) as a case study of a subscription-based funding model. Database bav018.

Renner SS. 2016. A return to Linnaeus's focus on diagnosis, not description: the use of DNA characters in the formal naming of species. Syst Biol (In press), doi:10.1093/sysbio/syw032

Reynolds D, Taylor J. 1991. DNA specimens and the 'International code of botanical nomenclature'. Taxon 40:311–351.

Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci USA 106:19126–19131.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu W-T, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T. 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature 499:431–437.

Rosling A, Cox F, Cruz-Martinez K, Ihrmark K, Grelet G-A, Lindahl BD, Menkis A, James TY. 2011. Archaeorhizomycetes: unearthing an ancient class of ubiquitous soil fungi. Science 333:876–879.

Ryberg M, Kristiansson E, Sjökvist E, Nilsson RH. 2009. An outlook on the fungal internal transcribed spacer sequences in GenBank and the introduction of a web-based tool for the exploration of fungal diversity. New Phytol 181:471–477.

Schadt CW, Martin AP, Lipson DA, Schmidt SK. 2003. Seasonal dynamics of previously unknown fungal lineages in tundra soils. Science 301:1359–1361.

Schleifer KH. 2009. Classification of Bacteria and Archaea: past, present and future. Syst Appl Microbiol 32:533–542.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microb 75:7537–7541.

Schoch CL, Robbertse B, Robert V, Vu D, Cardinali G, Irinyi L, Meyer W, Nilsson RH, Hughes K, Miller AN, Kirk PM, Abarenkov K, Aime MC, Ariyawansa HA, Bidartondo M, Boekhout T, Buyck B, Cai Q, Chen J, Crespo A, Crous PW, Damm U, De Beer ZW, Dentinger BTM, Divakar PK, Dueñas M, Feau N, Fliegerova K, García MA, Ge Z-W, Griffith GW, Groenewald JZ, Groenewald M, Grube M, Gryzenhout M, Gueidan C, Guo L, Hambleton S, Hamelin R, Hansen K, Hofstetter V, Hong S-B, Houbraken J, Hyde KD, Inderbitzin P, Johnston PR, Karunarathna SC, Kõljalg U, Kovács GM, Kraichak E, Krizsan K, Kurtzman CP, Larsson K-H, Leavitt S, Letcher PM, Liimatainen K, Liu J-K, Lodge DJ, Jennifer Luangsa-ard J, Lumbsch HT, Maharachchikumbura SSN, Manamgoda D, Martín MP, Minnis AM, Moncalvo J-M, Mulè G, Nakasone KK, Niskanen T, Olariaga I, Papp T, Petkovits T, Pino-Bodas R, Powell MJ, Raja HA, Redecker D, Sarmiento-Ramirez JM, Seifert KA, Shrestha B, Stenroos S, Stielow B, Suh S-O, Tanaka K, Tedersoo L, Telleria MT, Udayanga D, Untereiner WA, Diéguez Uribeondo J, Subbarao KV, Vágvölgyi C, Visagie C, Voigt K, Walker DM, Weir BS, Weiß M, Wijayawardene NN, Wingfield MJ, Xu JP, Yang ZL, Zhang N, Zhuang W-Y, Federhen S. 2014. Finding needles in haystacks: linking scientific names, reference specimens and molecular data for Fungi. Database 2014.

———, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Bolchacova E, Voigt K, Crous PW, Miller AN, Wingfield MJ, Aime MC, An KD, Bai FY, Barreto RW, Begerow D, Bergeron MJ, Blackwell M,

Boekhout T, Bogale M, Boonyuen N, Burgaz AR, Buyck B, Cai L, Cai Q, Cardinali G, Chaverri P, Coppins BJ, Crespo A, Cubas P, Cummings C, Damm U, de Beer ZW, de Hoog GS, Del-Prado R, Dentinger B, Dieguez-Uribeondo J, Divakar PK, Douglas B, Dueñas M, Duong TA, Eberhardt U, Edwards JE, Elshahed MS, Fliegerova K, Furtado M, García MA, Ge ZW, Griffith GW, Griffiths K, Groenewald JZ, Groenewald M, Grube M, Gryzenhout M, Guo LD, Hagen F, Hambleton S, Hamelin RC, Hansen K, Harrold P, Heller G, Herrera G, Hirayama K, Hirooka Y, Ho HM, Hoffmann K, Hofstetter V, Hognabba F, Hollingsworth PM, Hong SB, Hosaka K, Houbraken J, Hughes K, Huhtinen S, Hyde KD, James T, Johnson EM, Johnson JE, Johnston PR, Jones EB, Kelly LJ, Kirk PM, Knapp DG, Kõljalg U, Kovács GM, Kurtzman CP, Landvik S, Leavitt SD, Liggenstoffer AS, Liimatainen K, Lombard L, Luangsa-Ard JJ, Lumbsch HT, Maganti H, Maharachchikumbura SS, Martín MP, May TW, McTaggart AR, Methven AS, Meyer W, Moncalvo JM, Mongkolsamrit S, Nagy LG, Nilsson RH, Niskanen T, Nyilasi I, Okada G, Okane I, Olariaga I, Otte J, Papp T, Park D, Petkovits T, Pino-Bodas R, Quaedvlieg W, Raja HA, Redecker D, Rintoul T, Ruibal C, Sarmiento-Ramirez JM, Schmitt I, Schussler A, Shearer C, Sotome K, Stefani FO, Stenroos S, Stielow B, Stockinger H, Suetrong S, Suh SO, Sung GH, Suzuki M, Tanaka K, Tedersoo L, Telleria MT, Tretter E, Untereiner WA, Urbina H, Vágvölgyi C, Vialle A, Vu TD, Walther G, Wang QM, Wang Y, Weir BS, Weiß M, White MM, Xu J, Yahr R, Yang ZL, Yurkov A, Zamora JC, Zhang N, Zhuang WY, Schindel D. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci USA 109:6241–6246.

Schussler A. 2001. A new fungal phylum, the Glomeromycota: phylogeny and evolution. Mycologia 105:1413–1421.

Seifert KA, Crous PW, Frisvad JC. 2008. Correcting the impact factors of taxonomic journals by appropriate citation of taxonomy. Persoonia 20:105–105.

Selosse MA, Vincenot L, Öpik M. 2016. Data processing can mask biology: towards better reporting of fungal barcoding data? New Phytol 210:1159–1164.

Spang A, Saw JH, Jorgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJG. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521:173–179.

Stackebrandt E, Frederiksen W, Garrity GM, Grimont PAD, Kampfer P, Maiden MCJ, Nesme X, Rossello-Mora R, Swings J, Truper HG, Vauterin L, Ward AC, Whitman WB. 2002. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Micr 52:1043–1047.

Stielow JB, Levesque CA, Seifert KA, Meyer W, Iriny L, Smits D, Renfurm R, Verkley GJ, Groenewald M, Chaduli D, Lomascolo A, Welti S, Lesage-Meessen L, Favel A, Al-Hatmi AM, Damm U, Yilmaz N, Houbraken J, Lombard L, Quaedvlieg W, Binder M, Vaas LA, Vu D, Yurkov A, Begerow D, Roehl O, Guerreiro M, Fonseca A, Samerpitak K, van Diepeningen AD, Dolatabadi S, Moreno LF, Casaregola S, Mallet S, Jacques N, Roscini L, Egidi E, Bizet C, Garcia-Hermoso D, Martin MP, Deng S, Groenewald JZ, Boekhout T, de Beer ZW, Barnes I, Duong TA, Wingfield MJ, de Hoog GS, Crous PW, Lewis CT, Hambleton S, Moussa TA, Al-Zahrani HS, Almaghrabi OA, Louis-Seize G, Assabgui R, McCormick W, Omer G, Dukik K, Cardinali G, Eberhardt U, de Vries M, Robert V. 2015. One fungus, which genes? Development and assessment of universal primers for potential secondary fungal DNA barcodes. Persoonia 35:242–263.

Talbot JM, Bruns TD, Taylor JW, Smith DP, Branco S, Glassman SI, Erlandson S, Vilgalys R, Liao H-L, Smith ME, Peay KG. 2014. Endemism and functional convergence across the North American soil mycobiome. Proc Natl Acad Sci USA 111:6341–6346.

———, Martin F, Kohler A, Henrissat B, Peay KG. 2015. Functional guild classification predicts the enzymatic role of fungi in litter and soil biogeochemistry. Soil Biol Biochem 88:441–456.

Tatusova T, Ciufo S, Federhen S, Fedorov B, McVeigh R, O'Neill K, Tolstoy I, Zaslavsky L. 2015. Update on RefSeq microbial genomes resources. Nucleic Acids Res 43:D599–D605.

Taylor DL, Hollingsworth TN, McFarland JW, Lennon NJ, Nusbaum C, Ruess RW. 2014. A first comprehensive census of fungi in soil reveals both hyperdiversity and fine-scale niche partitioning. Ecol Monogr 84:3–20.

Taylor JW. 2011. One fungus = one name: DNA and fungal nomenclature twenty years after PCR. IMA Fungus 2:113–120.

———, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, Fisher MC. 2000. Phylogenetic species recognition and species concepts in fungi. Fungal Genet Biol 31:21–32.

Tedersoo L, Bahram M, Põlme S, Kõljalg U, Yorou NS, Wijesundera R, Ruiz LV, Vasco-Palacios AM, Thu PQ, Suija A, Smith ME, Sharp C, Saluveer E, Saitta A, Rosas M, Riit T, Ratkowsky D, Pritsch K, Põldmaa K, Piepenbring M, Phosri C, Peterson M, Parts K, Pärtel K, Otsing E, Nouhra E, Njouonkou AL, Nilsson RH, Morgado LN, Mayor J, May TW, Majuakim L, Lodge DJ, Lee SS, Larsson K-H, Kohout P, Hosaka K, Hiiesalu I, Henkel TW, Harend H, Guo L-d, Greslebin A, Grelet G, Geml J, Gates G, Dunstan W, Dunk C, Drenkhan R, Dearnaley J, De Kesel A, Dang T, Chen X, Buegger F, Brearley FQ, Bonito G, Anslan S, Abell S, Abarenkov K. 2014. Global diversity and geography of soil fungi. Science 346:1256688.

———, Ramirez KS, Nilsson RH, Kaljuvee A, Kõljalg U, Abarenkov K. 2015. Standardizing metadata and taxonomic identification in metabarcoding studies. GigaScience 4:1–4.

Thiéry O, Vasar M, Jairus T, Davison J, Roux C, Kivistik P-A, Metspalu A, Milani L, Saks Ü, Moora M, Zobel M, Öpik M. 2016. Sequence variation in nuclear ribosomal small subunit, internal transcribed spacer and large subunit regions of *Rhizophagus irregularis* and *Gigaspora margarita* is high and isolate-dependent. Mol Ecol 25:2816–2832.

Tindall BJ, Rossello-Mora R, Busse HJ, Ludwig W, Kampfer P. 2010. Notes on the characterization of prokaryote strains for taxonomic purposes. Int J Syst Evol Micr 60:249–266.

Veresoglou SD, Halley JM, Rillig MC. 2015. Extinction risk of soil biota. Nat Commun 6:8862.

Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microb 73:5261–5267.

Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, Krichevsky MI, Moore LH, Moore WEC, Murray RGE, Stackebrandt E, Starr MP, Truper HG. 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int J Syst Evol Micr 37:463–464.

White TJ, Bruns T, Lee S, Taylor J. 1990 Amplification and firect sequencing of fungal ribosomal RNA genes for phylogenies. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds. PCR protocols. San Diego, California: Academic Press. p 315–322.

Whitman WB. 2015. Genome sequences as the type material for taxonomic descriptions of prokaryotes. Syst Appl Microbiol 38:217–222.

———. 2016. Modest proposals to expand the type material for naming of prokaryotes. Int J Syst Evol Micr 66:2108–2112.

Wieczorek J, Bloom D, Guralnick R, Blum S, Dˆring M, Giovanni R, Robertson T, Vieglais D. 2012. Darwin Core: an evolving community-developed biodiversity data standard. PLoS ONE 7:e29715.

Yarza P, Ludwig W, Euzéby J, Amann R, Schleifer K-H, Glöckner FO, Rosselló-Móra R. 2010. Update of the all-species living tree project based on 16S and 23S rRNA sequence analyses. Syst Appl Microbiol 33:291–299.

———, Richter M, Peplies J, Euzeby J, Amann R, Schleifer K-H, Ludwig W, Glöckner FO, Rosselló-Móra R. 2008. The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. Syst Appl Microbiol 31:241–250.

———, Spröer C, Swiderski J, Mrotzek N, Spring S, Tindall BJ, Gronow S, Pukall R, Klenk H-P, Lang E, Verbarg S, Crouch A, Lilburn T, Beck B, Unosson C, Cardew S, Moore ERB, Gomila M, Nakagawa Y, Janssens D, De Vos P, Peiren J, Suttels T, Clermont D, Bizet C, Sakamoto M, Iida T, Kudo T, Kosako Y, Oshida Y, Ohkuma M, R. Arahal D, Spieck E, Pommerening Roeser A, Figge M, Park D, Buchanan P, Cifuentes A, Munoz R, Euzéby JP, Schleifer K-H, Ludwig W, Amann R, Glöckner FO, Rosselló-Móra R. 2013. Sequencing orphan species initiative (SOS): filling the gaps in the 16S rRNA gene sequence database for all species with validly published names. Syst Appl Microbiol 36:69–73.

Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, Gilbert JA, Karsch-Mizrachi I, Johnston A, Cochrane G, Vaughan R, Hunter C, Park J, Morrison N, Rocca-Serra P, Sterk P, Arumugam M, Bailey M, Baumgartner L, Birren BW, Blaser MJ, Bonazzi V, Booth T, Bork P, Bushman FD, Buttigieg PL, Chain PSG, Charlson E, Costello EK, Huot-Creasy H, Dawyndt P, DeSantis T, Fierer N, Fuhrman JA, Gallery RE, Gevers D, Gibbs RA, Gil IS, Gonzalez A, Gordon JI, Guralnick R, Hankeln W, Highlander S, Hugenholtz P, Jansson J, Kau AL, Kelley ST, Kennedy J, Knights D, Koren O, Kuczynski J, Kyrpides N, Larsen R, Lauber CL, Legg T, Ley RE, Lozupone CA, Ludwig W, Lyons D, Maguire E, Methe BA, Meyer F, Muegge B, Nakielny S, Nelson KE, Nemergut D, Neufeld JD, Newbold LK, Oliver AE, Pace NR, Palanisamy G, Peplies J, Petrosino J, Proctor L, Pruesse E, Quast C, Raes J, Ratnasingham S, Ravel J, Relman DA, Assunta-Sansone S, Schloss PD, Schriml L, Sinha R, Smith MI, Sodergren E, Spor A, Stombaugh J, Tiedje JM, Ward DV, Weinstock GM, Wendel D, White O, Whiteley A, Wilke A, Wortman JR, Yatsunenko T, Glockner FO. 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat Biotechnol 29:415–420.

Zhang J, Kapli P, Pavlidis P, Stamatakis A. 2013. A general species delimitation method with applications to phylogenetic placements. Bioinformatics 29:2869–2876.

LEGENDS

FIG. 1. Conceptual diagram of Sequence-Based Classification and Identification (SBCI) in fungi. The upper part of the figure shows traditional mycological data streams originating with cultures and collections (green boxes), whereas the lower part indicates environmental molecular sampling (blue boxes). SBCI is included in the integrating middle layer (red arrows), with outputs including synthetic understanding of community composition and functional biology and contributions to taxonomic and functional databases. Metadata not indicated in the diagram include temporal and geographic information or host information, which could contribute to fields such as ecological niche modeling, biogeography, and epidemiology. Metaproteomics and non-molecular aspects of functional biology (e.g. morphology and development) are also not shown, but could be integrated into SBCI and used to predict phenotypic properties of species detected with environmental sequences.

FIG. 2. New fungal species published from 2010 to 2015 based on Index Fungorum records, with and without sequence data of any locus in GenBank as of Jun 2016, as total (A) and proportional (B) values. The complete list of taxa with their higher-level classifications, publication information, and GenBank accession numbers (where these exist) are included (SUPPLEMENTARY INFORMATION).

FOOTNOTES

[1] Corresponding author. E-mail: dhibbett@clarku.edu

TABLE I. Terms and concepts for sequence-based classification and identification

Candidatus	A provisional taxonomic category for prokaryotes that lack a type culture
Environmental nucleic acid species (ENAS)	A species recognized solely with environmental molecular sequences
Environmental sequence	A DNA or RNA sequence obtained directly from a microbial community using amplicon or shotgun methods
Molecular operational taxonomic unit (MOTU)	An unranked taxonomic entity recognized with environmental sequences
Nomenclature	The set of rules detailed in the ICNAFP that determine the correct name for algae, fungi, and plants
Sequence-based classification (SBC)	The process by which species are discovered, named, and classified according to their phylogenetic relationships.
Sequence-based identification (SBI)	The process by which the products of taxonomy are used to identify individuals and determine the composition of communities with reference to existing classifications
Species hypotheses (SH)	A term coined to describe taxa whose ITS rRNA gene sequences cluster at user-defined cutoff levels i.e. 97–99%
Taxonomy	The branch of science focused on naming, describing, and classifying all forms of life
Virtual taxa (VT)	Phylogenetically defined sequence groups that roughly correspond to species

TABLE II. Databases and tools for sequence-based classification and identification

General identification tools and data repositories	
BOLD	http://www.boldsystems.org/
CBS-KNAW	http://www.cbs.knaw.nl/Collections/BioloMICSSequences.aspx—contains 30 BLASTn searchable databases
Dryad	http://datadryad.org/
FUSARIUM-ID	http://isolate.fusariumdb.org/
GreenGenes	http://greengenes.lbl.gov/cgi-bin/nph-index.cgi
Maarj*AM*	http://maarjam.botany.ut.ee/
Mothur	http://www.mothur.org/
Naïve Bayesian Classifier	http://aem.asm.org/content/73/16/5261.short?rss=1&ssource=mfc
Open Tree of Life	http://www.opentreeoflife.org/
QIIME	http://qiime.org/
RefSeq Targeted Loci	http://www.ncbi.nlm.nih.gov/refseq/targetedloci/
Ribosomal Database Project (RDP)	https://rdp.cme.msu.edu/
Silva	http://www.arb-silva.de/
TreeBASE	https://treebase.org/
TrichoBLAST	http://www.isth.info/tools/blast/
UNITE	https://unite.ut.ee/
Data standards	
BIOM	http://biom-format.org/
MIMARKS	http://www.nature.com/nbt/journal/v29/n5/full/nbt.1823.html
Darwin Core	http://rs.tdwg.org/dwc/
Genomics databases and tools	
1000 Fungal Genomes Project (1KFG)	http://1000.fungalgenomes.org/home/
FungiDB	http://fungidb.org/fungidb/
GEBA	http://jgi.doe.gov/our-science/science-programs/microbial-genomics/phylogenetic-diversity/
MycoCosm	http://genome.jgi.doe.gov/programs/fungi/index.jsf
Functional database	
FUNGuild	https://github.com/UMNFuN/FUNGuild

Nomenclature and nomenclatural databases and organizations	
Catalogue of Life (COL)	http://www.catalogueoflife.org/
Index Fungorum	http://www.indexfungorum.org/
International code of nomenclature for algae, fungi, and plants (ICNAFP)	http://www.iapt-taxon.org/nomen/main.php
International Commission on the Taxonomy of Fungi (ICTF)	http://www.fungaltaxonomy.org/
List of prokaryotic names with standing in nomenclature (LPSN)	http://www.bacterio.net/
MycoBank	http://www.mycobank.org/
Biodiversity collections databases	
Global Biodiversity Information Facility (GBIF)	http://www.gbif.org/
iDigBio	https://www.idigbio.org/
MycoPortal	http://mycoportal.org/portal/index.php
World Federation of Culture Collections (WFCC)	http://www.wfcc.info/
Citizen science resources	
Encyclopedia of Life	http://eol.org/
Mushroom Observer	http://mushroomobserver.org/