MICROBIOME

Mycobiome diversity: high-throughput sequencing and identification of fungi

R. Henrik Nilsson[1], Sten Anslan[2], Mohammad Bahram[3], Christian Wurzbacher[4], Petr Baldrian[5] and Leho Tedersoo[6]*

Abstract | Fungi are major ecological players in both terrestrial and aquatic environments by cycling organic matter and channelling nutrients across trophic levels. High-throughput sequencing (HTS) studies of fungal communities are redrawing the map of the fungal kingdom by hinting at its enormous — and largely uncharted — taxonomic and functional diversity. However, HTS approaches come with a range of pitfalls and potential biases, cautioning against unwary application and interpretation of HTS technologies and results. In this Review, we provide an overview and practical recommendations for aspects of HTS studies ranging from sampling and laboratory practices to data processing and analysis. We also discuss upcoming trends and techniques in the field and summarize recent and noteworthy results from HTS studies targeting fungal communities and guilds. Our Review highlights the need for reproducibility and public data availability in the study of fungal communities. If the associated challenges and conceptual barriers are overcome, HTS offers immense possibilities in mycology and elsewhere.

[1]*Gothenburg Global Biodiversity Centre, Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden.*

[2]*Zoological Institute, Braunschweig University of Technology, Braunschweig, Germany.*

[3]*Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden.*

[4]*Chair of Urban Water Systems Engineering, Technical University of Munich, Garching, Germany.*

[5]*Laboratory of Environmental Microbiology, Institute of Microbiology of the Czech Academy of Sciences, Praha, Czech Republic.*

[6]*Natural History Museum of Tartu University, Tartu, Estonia.*

*e-mail: henrik.nilsson@bioenv.gu.se

https://doi.org/10.1038/s41579-018-0116-y

The fungal kingdom, recently estimated to contain up to 3.8 million species, presents an immense diversity of life forms, nutritional strategies and associations with other organisms[1]. The origin of fungi can be traced to single-celled marine ancestors in the Mesoproterozoic era more than 1 billion years ago[2]. Fungi have since conquered not only land but also almost every potential habitat and substrate[3]. Although all fungi are heterotrophs, the fungal kingdom comprises a wide range of life strategies ranging from saprotrophy through mutualism to parasitism. Fungi fundamentally influence and shape ecosystems by producing and channelling nutrients across trophic levels and the food web at large[4].

The early molecular identification studies of the 1990s revealed a range of often uncultivable fungal species, which were unknown from fruiting bodies and other structures and whose nutritional modes and ecological associations regularly defied characterization[5]. As DNA sequencing technologies progressed from sequencing single specimens to parallel Sanger sequencing in the early 2000s, it became plainly clear that the unseen — and perhaps unseeable — mycobiota outnumbers the diversity observable through fruiting bodies and cultivation[6]. The second-generation sequencing methods were developed in the second half of the 2000s and marked the beginning of high-throughput sequencing (HTS) analyses of fungal communities. In their wake, fungal ecology caught fire[7,8]. The third-generation sequencing platforms operate at the level of single molecules and offer much higher read lengths than the earlier generations. Although they have been around for a number of years, the read quality has only recently improved enough to make them useful in metabarcoding and community analysis.

High-throughput analysis of fungal communities currently draws from the second and third generation of HTS technologies. The variety of HTS platforms (BOX 1) is matched by an even greater diversity of laboratory protocols and analysis pipelines. Staying abreast of the rapid development in these fields is not easy for individual researchers but, in this Review, we provide a comprehensive overview of commonly used cutting-edge molecular methods in the characterization of fungal communities, notably second-generation and third-generation sequencing tools but also related omics techniques. Our primary focus is high-throughput identification at the species level.

We begin by providing a brief overview and practical recommendations on the workflow of typical HTS metabarcoding studies ranging all the way from sampling and sample preparation through sequence processing to taxonomic and functional assignment. After that, we turn our attention to emerging techniques and methods. We then zoom out to survey how HTS has been applied to study fungal communities across guilds and habitats, highlighting the major findings, breakthroughs and challenges. We conclude with an outlook on the role of fungal community analysis in mycology and the conceptual barriers that must be overcome to fully take advantage of the immense possibilities offered by HTS.

Box 1 | **Overview of high-throughput sequencing platforms**

454 pyrosequencing (Roche). Discontinued in 2016, this comparatively expensive second-generation platform was the primary workhorse in fungal community high-throughput sequencing (HTS) studies from 2008 to around 2014. Although this platform suffered from insertions and deletions in homopolymer-rich DNA regions, it could generate up to 1.2 million reads of 800–1,000 bases per run. It left an indelible mark in fungal HTS software development and shed light onto the enormous diversity of the soil microbiota.

Ion Torrent PGM and GeneStudio (Thermo Fisher Scientific). This relatively inexpensive platform offers read lengths of up to 400–600 bases and a throughput of up to 5–12 million reads per run. Struggling with homopolymer-rich regions and read lengths, this platform never saw wide use in mycology, falling between the cracks of pyrosequencing and Illumina sequencing.

Illumina MiSeq, HiSeq and NovaSeq (Illumina). This platform is the major player among the second-generation HTS platforms. Although offering shorter single-end read lengths of up to 300 bases, its high read quality, massive throughput (20, 300 and 6,000 gigabases, respectively) and the possibility to combine reads in a paired-end approach translate into high-quality assemblies of full internal transcribed spacer subregion 1 (ITS1) or full ITS2 fungal sequences for most taxa at unprecedented sequencing depth. At present, MiSeq Illumina sequencing is the default choice for metabarcoding studies of fungi and other organisms, and the paired-end approach covers amplicons of up to ~550 bases in length (MiSeq 2 × 300). It is also the principal platform for conducting metagenomic and metatranscriptomic analyses on the basis of fragmented DNA or complementary DNA.

PacBio RSII and Sequel (Pacific Biosciences). This third-generation HTS platform produces long reads averaging 20–25 kilobases (up to 100 kilobases). The DNA template is circularized via hairpin adaptors and recorded multiple times to offer read qualities comparable to that of traditional Sanger sequencing. The new Sequel instrument generates ~400,000 reads per SMRT cell (3.2 million reads per run), which is much fewer reads than produced by Illumina platforms, but it produces high-quality sequences of amplicons or DNA fragments of ~3 kilobases in length. This makes the platform ideal for sequencing short-to-medium length amplicons such as the full ITS region and perhaps its flanking conserved genes for precise phylogenetic placement.

Oxford Nanopore MinION, GridION and PromethION (Oxford Nanopore Technologies). These third-generation nanopore technology-based instruments are relatively inexpensive, reaching read lengths of an unprecedented 2.4 megabases[165]. Remarkably, the US$1,000 MinION device can be run on powerful battery-powered laptops, removing the need for sending samples away for sequencing and enabling DNA sequencing in situ during field trips. Because of the high average error rate of 6–12%, nanopore sequencing has seen little use for microbial identification, although several research teams are experimenting with the inexpensive MinION sequencer. Methods for multiple consensus sequencing and improving read quality would greatly improve the applicability of Oxford Nanopore sequencing in ecology, beyond genome sequencing and linking genomic scaffolds.

Throughout, we emphasize the need for reproducibility and open data in the study of fungal communities.

Critical steps in HTS studies of fungi

This section gives a brief overview and practical recommendations on the major steps in a typical HTS metabarcoding effort, that is, sampling, DNA extraction, marker-based PCR amplification, DNA sequencing, sequence processing and data analysis (FIG. 1).

Sample preparation. Experimental design is a fundamental step that determines the analytical explanatory power and representativeness of a study[9,10]. Although highly sensitive HTS analyses enable recovering differences at the community level on the basis of as few as three replicates, the use of higher numbers of biological replicates is feasible for testing community-level differences without additional cost of sequencing. However, researchers should be prepared to lose up to 10% of the samples owing to technical failure or otherwise overly low numbers of sequences[11,12].

Care should be taken to avoid contamination and other sources of bias during sampling and processing[13]. Sampling equipment should be cleaned carefully between sampling rounds. For DNA analyses, samples should be frozen or dried within ~12 h to avoid overgrowth by fast-growing moulds. RNA analyses require rapid sample fixation in liquid nitrogen to ensure RNA preservation; commercial RNA preservation buffers may fail for complex substrates[14] and should be avoided or evaluated for efficacy beforehand. As RNA extraction is a sensitive process, only a small number of samples should be processed simultaneously, and it is recommended to construct cDNA libraries soon after extraction. DNA extraction usually involves mechanical cell lysis, and the extraction method should be suitable for fungi, as their recalcitrant chitinous cell walls and, sometimes, secondary metabolites may interfere with DNA extraction. Fungi that are particularly rich in polysaccharides, notably encapsulated yeasts, may require special treatment[15,16]. Optimization for particular substrates — notably soil, plant tissue and water — is often needed. Exclusively detecting living fungi requires additional steps (BOX 2).

Selection of markers and primers, and PCR. Selection of genetic markers, primers and amplification conditions is a critical step in HTS studies. Sequencing the internal transcribed spacer (ITS) region of the nuclear ribosomal RNA (rRNA) operon is a mainstay in Sanger sequencing-based species identification and HTS-based metabarcoding[17] and represents the primary fungal DNA barcode[18]. For several groups of important plant pathogens and endophytes, the ITS region provides insufficient resolution for species-level assignment[19], but there are currently no implementations of better markers for HTS-based separation of these lineages. Studies addressing aquatic fungi and arbuscular mycorrhizal fungi commonly use the small subunit (SSU) (18S) and large subunit (LSU) (28S) nuclear rRNA genes, but for ascomycetes and basidiomycetes, these markers are usually informative only on taxonomic levels above species (and sometimes genera) because there may be no or too little variation in SSU and LSU sequences between species in these fungal groups to enable robust determination. The main benefit of SSU, LSU and protein-coding genes such as the RNA polymerase gene *RPB2* (REF.[20]) is alignability across fungal phyla and hence assessment of large-scale phylogenetic affiliations at, for example, the phylum and order levels, which is a task that the ITS region is typically not in a position to address in the absence of reasonably similar reference sequences. Because the ITS region typically spans 500–700 bases, most HTS-based studies focus on either the ITS1 or ITS2 subregion of typically 250–400 bases. Advantages of the ITS2 subregion include lower length variation and more universal primer sites, resulting in less taxonomic bias than ITS1 (REF.[21]). Although the ITS1 and ITS2 subregions are suitable for second-generation

Heterotrophs
Organisms that cannot produce their own food (as most plants can through photosynthesis), relying instead on intake of nutrition and energy from external sources of organic carbon.

Saprotrophy
The process of obtaining nutrients and energy from decomposing non-living organic matter such as dead wood, detritus and fallen leaves.

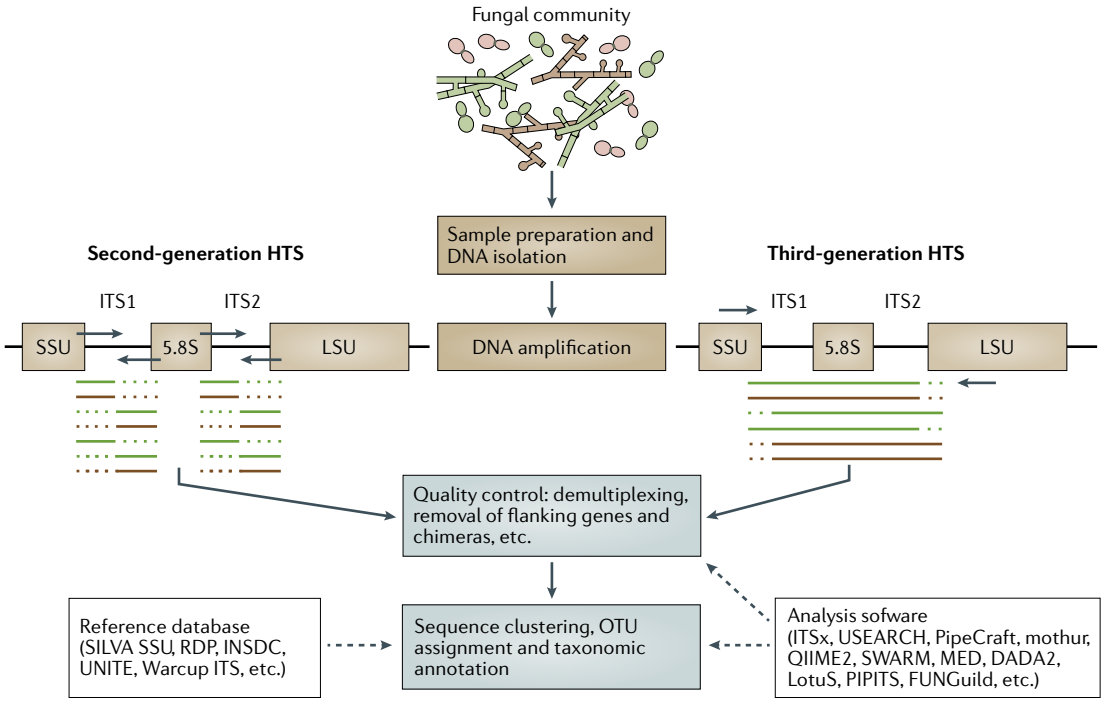

Fig. 1 | **The main steps in a fungal metabarcoding project.** Fungal diversity can be studied by sequencing genetic markers such as the internal transcribed spacer (ITS) region, small subunit (SSU) and large subunit (LSU). Such a study includes sample preparation and DNA extraction, PCR amplification, sequence processing and quality control, and sequence analysis. The fact that a non-trivial number of software tools and resources is usually needed in the analysis process is hinted at in the figure. Not shown are the subsequent statistical analysis of the resulting operational taxonomic unit (OTU) tables and associated metadata — an important step whose explanatory power is largely determined by the sampling strategy and that requires planning of the experimental design in a statistically aware way. HTS, high-throughput sequencing; INSDC, International Nucleotide Sequence Database Collaboration.

HTS, the third-generation techniques using Pacific Biosciences (PacBio) and Oxford Nanopore platforms enable targeting the full ITS region, as well as parts of, or even the entire, flanking rRNA genes. The benefits of targeting the full ITS region over any of its subregions include greater taxonomic resolution and reduced amplification of dead organisms, but a disadvantage is the poorer performance on low-quality material such as legacy herbarium specimens, in which the DNA has degraded to the point at which full-length ITS DNA sequencing is no longer possible[22].

The choice of primers dictates what fungi will be recovered from the sample, and we recommend spending substantial time evaluating and choosing primers. An overview of primarily fungus-specific, HTS-oriented primers for amplification of targets in the SSU–ITS–LSU region is given in Supplementary Fig. 1. Only a few of these primers amplify >90% of all fungal groups, and primer choice necessitates careful consideration of the target taxa[21]. Regarding the ITS region, we recommend targeting the ITS2 subregion or the full ITS region by using the degenerate forward primers gITS7ngs and ITS9MUNngs, respectively, in combination with the reverse primer ITS4ngs owing to their superior coverage of the fungal kingdom (Supplementary Table 1). For living plant tissue, the inclusion of at least one primer that preferentially amplifies fungi over plants (for example, ITSOF, ITS1catta, LF-402 or LR5F; Supplementary Table 1) is warranted. It is important to recognize that the ITS1 primers endorsed by several large-scale microbiome projects (that is, ITS1F and ITS2) suffer strongly from primer biases and the presence of an intron that is common in several fungal groups, which can lead to biased amplification[23].

In HTS-based studies, samples are usually amplified with primer variants equipped with unique identifier indices (also known as tags, molecular identifiers or barcodes) to recognize the samples bioinformatically. These indices should be longer than six bases and differ from each other by at least three (four in the case of PacBio) nucleotides to prevent technical cross-contamination; unique indices should be used for each sample on both the forward and the reverse primer to recognize index-switching artefacts[10]. The primers may also contain a spacer (1–7 bases to prevent index ends from acting as primers) and sequencing adaptors for particular platforms. On Illumina systems, length variation of spacers facilitates an increased yield by phasing[24]. Platform-specific adaptors make the primers more expensive and reduce PCR efficiency but allow determination of sequencing direction. Indices for Illumina sequencing can be added to the products of conventional PCR primers through a short PCR step, which, however, is costlier and more laborious when targeting hundreds of samples or preparing multiple libraries.

A high-fidelity proofreading PCR polymerase with low GC content bias is required to produce fully

Mutualism
A symbiotic and mutually beneficial interaction between organisms, such as mycorrhizal relationships between fungi and plants.

Parasitism
A relationship between two organisms in which one organism, the parasite, obtains an advantage at the expense of the other organism.

Metabarcoding
A rapid method of PCR-based biodiversity assessment powered by high-throughput DNA sequencing.

Endophytes
Organisms that live inside a plant for at least a part of their life cycle without specialized nutrient-exchange structures or symptoms of apparent harm or disease.

Box 2 | **Detecting active fungi**

Only a fraction of the fungi present in any given environment are metabolically active. A recent soil-oriented study indicated that relic DNA may account for >40% of the internal transcribed spacer (ITS) sequences recovered[166]. These sequences from dead organisms inflate biodiversity estimates and obscure the analysis of short-term changes. Similarly, for dead wood, it is known that DNA of endophytic fungi is still present after years of decomposition, although the RNA of those fungi can no longer be detected[167]. This highlights the importance of identifying the living component of the total fungal community. The most common approach uses amplicon sequencing of the fungal ITS region derived from total RNA. The transcribed precursor of the fungal ribosomal RNA (rRNA) operon contains not only rRNA but also its internal spacers, which are spliced and degraded when rRNA molecules are assembled into ribosomes. The analysis of the RNA precursor ITS thus targets the transcripts produced by active fungi in the process of synthesizing their ribosomes[168] and may be used for assessment of metabolically active fungi[169]. The RNA–DNA ratio can also be used as a measure of transcription of individual genes in litter, such as cellulase genes[169]. Another option is to use longer DNA barcodes because the vast majority of relic DNA fragments are <200 bases in length[170].

Although the activity of the microbiota can be estimated by monitoring total transcription[63] or metaproteome composition[171], these alternative methods are not suitable for fine-level taxonomic classification of a broad range of fungi, as the classification of fungal metatranscriptomic contigs is typically reliable only at the phylum level[78]. If individual environmental processes are targeted, amplicon analysis of those microbial members that are labelled following addition of a stable isotope-containing substrate (stable isotope probing) can be applied[172]. The metabolic activity of individual fungal taxa is best explored through real-time quantitative PCR analysis after total RNA extraction[173].

extended amplicons and reduce random errors and chimaeras that accumulate with cycles. Diluted DNA (if feasible), a low number of PCR cycles and multiple parallel PCR reactions per sample yield fewer compromised sequences[25,26].

Controls and technical replication. To understand the abundance and nature of contamination and technical artefacts — error accumulation, chimaera formation and index switching (BOX 3) — we recommend the use of a negative control (no sample), a positive control (known species unlikely to be found in the samples) and a mock community such as SynMock[27]. Negative control samples from DNA extraction and subsequent steps indicate sources of potential contamination and can be used for blank correction of low-level contamination that might occur, for example, in buffers[28]. Positive control and mock community samples allow assessment and quantification of index switching, chimaera formation, operational taxonomic unit (OTU) inference stringency and abundance shifts[26,29]. To mitigate stochastic variation, technical PCR and DNA extraction replicates (for pooling) are recommended.

Quality-filtering of HTS data. Optimal ways of demultiplexing HTS samples and examining reads for quality differ between markers and sequencing platforms. Owing to the lack of large-scale alignability and the length polymorphism of the ITS region, bioinformatics analysis of fungal ITS data sets poses additional challenges and requires tailored software solutions that are not always readily available. Demultiplexing of samples based on indices in both ends will remove most index-switching artefacts[30]. When excluding potentially low-quality sequences, researchers should consider that the length of complete ITS1 and ITS2 sequences may be as short as 50 bases and that homopolymers >10 bases are common. The Illumina platform offers paired-end sequences, and we recommend using paired sequences for short fragments such as ITS1 or ITS2. Single ends can be used for longer reads such as the full ITS region, for which ends would not be expected to meet given present Illumina read lengths.

It is important to remove any substantial portions of the genes flanking the ITS1 and ITS2 subregions (those encoding SSU, 5.8S and LSU) to refine taxonomic resolution[31]. We recommend using ITSx[32] with relaxed threshold values for removal of flanking genes. USEARCH[33] and VSEARCH[34] work well for chimaera filtering of all organisms and markers we have examined.

Sequence clustering and operational taxonomic units. Clustering of sequences into roughly species-level OTUs is the most common approach in microbial HTS studies. Typically, ITS sequence similarity thresholds of 97.0–98.5% are used to strike a balance between species-level taxonomic resolution and undesired accumulation of sequencing errors. Barcoding thresholds can be determined for major taxonomic groups separately to account for taxon-specific marker conservation[19]. For fungal ITS data sets, single-linkage clustering methods (in which a new sequence is added to a nascent OTU if it satisfies the matching condition for at least one sequence in that OTU) perform the best, as single-linkage clustering inflates the number of OTUs less than does complete-linkage clustering (in which a new sequence has to satisfy the matching condition for all sequences in that OTU). Furthermore, single-linkage clustering does not depend on 'seed' sequences — the same OTU will always be obtained, regardless of which sequence is chosen to define that OTU[10]. Single-linkage clustering is compute-intensive and thus requires initial dereplication or other pre-processing steps[5]. Before committing to any particular clustering solution, we recommend users to explore that tool on a subset of the target sequences. The default settings of clustering tools are often tailored for 16S rRNA sequences, in which gaps in the alignment less frequently occur than in ITS alignments. Reducing gap penalties (and gap extension penalties) may help if unexpected clustering behaviour happens. Popular OTU inference methods include algorithms implemented in USEARCH, VSEARCH and mothur, for example[35]. A growing number of innovative clustering and clustering-free OTU approaches do not require specific designations of direct similarity thresholds, including single-linkage based SWARM[36] and the data-filtering based tools MED[37] and DADA2 (REF.[38]), which have been used mostly for bacterial SSU data sets. DADA2 supports analysis of distinct sequence variants, which may not be optimal for the fungal ITS region, given that it is sometimes present in multiple distinct copies per genome[39]. Detailed documentation of the used clustering approach and deposition of OTU tables and standards-compliant metadata (as a supplementary item or in the Dryad Digital Repository) will increase scientific reproducibility and comparability across studies.

Arbuscular mycorrhizal fungi
Fungi of the phylum Glomeromycota (and the Endogonales of the Mucoromycota) that establish mutualistic symbioses with primarily herbaceous plants; plant cell walls are penetrated, and the fungi produce arbuscules and sometimes vesicles inside the plant cells.

Technical cross-contamination
Mutations that turn distinct indices into indices used for other samples will lead to errors in sample assignment.

Phasing
Barcoded amplicon sequences are determined in different sequencing phases by adding spacers of different lengths to the primer sequences.

Box 3 | **High-throughput sequencing biases**

Extraction bias. Fungi differ greatly in the presence and composition of their cell walls. Similarly, the biochemical and physical properties of their habitat vary, notably in the content of humic acids and saccharides that interfere with molecular analyses. Therefore, there is no universal DNA or RNA extraction method that will work well for all organisms and environments. Methods should be selected carefully on the basis of both substrate and target organisms[10], and further purification might be needed.

Marker bias. DNA and RNA markers differ substantially in length, taxonomic resolution, copy number and alignability. Therefore, they also differ in ease of amplification, power to distinguish between species and phylogenetic applicability[5]. Because fungal ribosomal RNA operon copy numbers vary, there is a strong bias towards organisms with more copies. Although the internal transcribed spacer (ITS) region represents the formal fungal barcode, it sometimes provides insufficient resolution to distinguish species, necessitating the use of additional or alternative markers[174].

Primer bias. Primers differ in their melting temperature, binding specificity and binding position and thus in amplicon length and coverage of the targeted organisms. Primers should be chosen carefully to have similar melting temperatures and to match the targeted taxa and the optimal amplicon length for the respective sequencing platform. Primer–template mismatches in the 3′ end are detrimental to elongation, and single near-terminal mismatches may virtually eliminate species from detection[21]. We recommend the use of universal eukaryote primers, except in cases with living host material or scanty fungal biomass[175]. A more expensive and resource-demanding alternative is to avoid the PCR step altogether.

PCR bias. Polymerases differentially amplify DNA molecules of variable size and nucleotide contents. Because shorter fragments are preferentially amplified, markers containing introns should be avoided. In addition, polymerases sometimes make mistakes in nucleotide incorporation, and some extensions are interrupted, resulting in chimaera formation, especially at high numbers of PCR cycles. To reduce chimaera formation and sequencing errors, diluted DNA samples should be used (if feasible), and the number of amplification cycles should be kept low[25,176]. High-fidelity proofreading polymerases with low GC bias should be used[177].

Library preparation bias. Previous results with index and primer variants should be considered, and indices with A + T:G + C ratios <30% and >70% should be avoided. Short linkers not matching the template or indices with the same leading nucleotide may improve the results. If feasible, different primer pairs could be included in different libraries. Alternatively, the use of long primers already carrying the sequencing adaptor can be considered, but these are more costly, difficult to amplify and useable with only some platforms.

Sequencing bias. Amplicons of smaller size and regular A + T:G + C ratios are often preferentially sequenced, but these differences depend strongly on the platform and sequencing chemistry[175]. Third-generation sequencing methods have lower sequencing biases despite greater error rates.

Bioinformatics biases. Insufficient knowledge of the properties of the genetic marker may result in poor choices of bioinformatics solutions. The length of the ITS subregion 1 (ITS1) and ITS2 markers may vary from 50 bases to several kilobases, and several fungi have ITS regions with homopolymer regions >10 bases. Some divergent fungi are also lost with the default parameters of ITSx, although relaxation of the e-value threshold might help[175]. Incorrect classification may lead to the inclusion or exclusion of false positive (non-fungal) or valid operational taxonomic units (OTUs), respectively. During chimaera filtering, false positives are common, as some species have a variable ITS1 and relatively conserved ITS2 region or vice versa. Discussion with fungal molecular ecologists might help to make informed choices.

Index switching. During the library preparation and sequencing steps, secondary chimeric molecules may form at low abundance. These differ from PCR chimaeras by cross-hybridization across samples (self-switching may also occur), which results in technical cross-contamination between samples[178]. Such errors can be recognized and quantified on the basis of positive and negative controls and removal of rare occurrences of abundant species from samples.

Poor clustering. All known clustering algorithms produce more OTUs than expected owing to accumulation of rare sequencing errors, which leads to formation of satellite OTUs that overestimate richness and reinforce community-level tests. Such errors can generate strong artificial positive co-occurrence patterns and may overestimate inter-species facilitative interactions. Stringent denoising, removal of rare OTUs and the use of several clustering rounds mitigate this problem[50]. No single sequence similarity threshold will reflect the species level across the fungal kingdom, which is a problem for analyses resting on the species level. An alternative is to adopt dynamic similarity thresholds or sequence variants[179].

Unequal sequencing depth. Having different numbers of sequences between samples complicates tests of taxonomic diversity because OTU accumulation curves fail to reach a plateau in large composite samples. Although rarefying to the lowest common number of sequences is widely used, too much information may be lost[11]. Accounting for library size in downstream analyses is an alternative to rarefaction, except when there are great differences in alpha diversity between samples (for example, soil samples have more OTUs than root samples[180]). In theory, evenness in sequencing depth can be promoted by adding sequencing adaptors to amplicons, mixing equal amounts of amplicons and using barcodes starting with the same nucleotide.

Sequence-based taxonomic identification and taxon communication. The taxonomic annotation of OTUs largely relies on sequence similarity searches in reference databases. The most commonly used tool for sequence similarity comparison is BLAST[40], although speedier approaches are available[33]. Regarding reference databases in the context of the fungal kingdom and SSU and LSU sequences, we recommend the SILVA SSU Ref NR and the RDP databases, respectively. We advise that users targeting arbuscular mycorrhizal

fungi through the SSU gene consider Maarja-AM[41]. However, these SSU and LSU databases might need to be complemented with recent sequences and lineages not belonging to the ascomycetes and basidiomycetes by recourse to the International Nucleotide Sequence Database Collaboration (INSDC). Three main reference ITS databases spanning the fungal kingdom are available: INSDC, UNITE[42] and Warcup ITS[43]. We recommend UNITE, as it represents a middle ground between including the very latest sequences and offering detailed taxonomic annotation, and it also includes the many fungal species-level lineages for which no resolved names are available. We recommend a subsequent manual BLAST search in INSDC or UNITE of the most abundant OTUs to potentially improve taxonomic annotations. Deciding at which taxonomic level to annotate ITS sequences is notoriously difficult but, as an approximation, ≥97.0–98.5% similarity (in a pairwise alignment covering ≥90% of the query sequence) corresponds to the species level, ≥90% similarity corresponds to the genus level, ≥85% corresponds to the family level, ≥80% corresponds to the order level, ≥75% corresponds to the class level and ≥70% corresponds to the phylum level[12]. In our experience, OTUs that produce only partial matches to fungal reference sequences may or may not represent true fungi, and many of them may, on closer inspection, be found to belong to Chlorophyta, Rhizaria or Alveolata instead. Comparison of the OTUs to the 'most wanted fungi' ITS release of UNITE will reveal whether the data contain any of these particularly enigmatic fungal lineages[44]. Users are well advised to classify their OTUs against the UNITE digitally archived species hypotheses, as this enables tagging of their OTUs with the digital object identifier (DOI) of any highly similar species hypothesis. This classification provides a standardized documentation of which taxa were found, enables knowledge to be assembled on those taxa and makes it possible to compare taxonomic profiles across studies even in the total absence of known taxonomic affiliations[45]. We recommend FUNGuild[29] for assignment of functional guilds to taxa, although many genera contain more than one trophic strategy, which necessitates manual evaluation. Similarly, insufficient guild data are available for many fungal groups, cautioning against community-wide conclusions on data sets for which only a limited proportion of the OTUs could be sorted into well-defined guilds.

Data processing and analysis pipelines. A panoply of software pipelines exists for processing and analysing HTS metabarcode data (TABLE 1), with common choices including mothur, USEARCH and QIIME. HTS pipelines include multiple steps of the analysis procedure, ranging from filtering of raw data to generating sample-by-OTU abundance tables, but there is variation in which steps are included in individual pipelines. Owing to the lack of inclusion of tools that extract the ITS subregions (for example, ITSx) in these pipelines and/or the fact that fungal ITS sequences cannot be robustly aligned even across some genera (which is particularly problematic when using clustering approaches in mothur), users need to incorporate additional tools into the workflow. Software packages that cover all steps necessary for ITS amplicon analysis and accommodate many of the considerations above include PipeCraft[46], LotuS[47] and PIPITS[48]. In addition, the Amplicon toolkit (AMPtk[27]) accurately accounts for amplicons of variable length, but ITS region extraction is not implemented. These software packages are easy to install, work well with fungal ITS sequences and produce quantitatively comparable results[49]. However, there is at present no support for PacBio data in either PIPITS or the otherwise much often used QIIME pipeline. LotuS and PipeCraft work with raw multiplexed data (reads are split into separate files based on indices), whereas PIPITS requires demultiplexed data (a single file containing all sequences from one run). PipeCraft provides an easy-to-use graphical user interface and multiple options for data analysis. PIPITS offers a limited number of tools, but data analysis is easily performed with a straightforward pipeline. LotuS is a rapid tool that is run through the command line and whose developers sought to minimize computational time and memory requirements. One of the benefits of AMPtk is automatic calculation of the index-switching rate (BOX 3). With the exception of AMPtk, these pipelines support extraction of the ITS1 and ITS2 subregions with ITSx to discard non-target sequences and to increase taxonomic resolution. For Illumina data, we recommend PipeCraft, PIPITS, LotuS and AMPtk. For PacBio data, we recommend PipeCraft, as it bundles tools that are particularly well suited for analysing long reads[49]. The latest version of DADA2 (v1.9.1) can also handle PacBio data. However, none of the existing pipelines fully filter out all errors that accumulate during sample preparation and sequencing[27,49]. Semi-automated curation of the results, particularly of the OTU tables, is highly recommended[50]. Manual validation of at least the major OTUs by examining the underlying sequence alignments and the proposed taxonomic affiliations (for example, through a BLAST search in GenBank or UNITE) will often reveal putative pipeline glitches and lead to refined taxonomic resolution. Comparing and combining the results from several taxonomy assignment tools may be helpful[51].

Emerging methods

In this section, we peek around the corner and consider emerging methods and new combinations of existing methods for studying fungal communities. We discuss quantification of fungal taxa, arrays and microarrays, metagenomics, metatranscriptomics and approaches to identify fungal strains and even single cells (FIG. 2).

Quantification. Quantitative PCR (qPCR) is a well-established method to quantify fungal biomass based on the content of DNA markers[52]. However, fungal qPCR methods typically target the rRNA genetic markers, and because fungal species differ markedly in number of rRNA operon copies per genome, the performance of qPCR is limited[20]. Furthermore, many of the fungus-specific primers that are used for qPCR are in fact not fungus-specific or do not cover all fungi. Although it is possible to estimate fungal biomass on the

Species hypotheses
A species hypothesis is a group of similar sequences that is tentatively delimited at the species level.

Table 1 | **List of commonly used tools for metabarcoding data analysis**

Name	Description and link	Refs
DADA2	Amplicon sequence variant analysis pipeline • https://benjjneb.github.io/dada2/	38
Galaxy	Web-based platform, including various analytical tools • https://usegalaxy.org/	183
LotuS	Full pipeline for amplicon data • http://psbweb05.psb.ugent.be/lotus/index.html	47
mothur	Versatile software suite (designed mostly for 16S rRNA) • https://www.mothur.org	35
AMPtk	Full pipeline for amplicon data • http://amptk.readthedocs.io	27
OBITools	Versatile software package • https://git.metabarcoding.org/obitools	184
PipeCraft	Full pipeline for amplicon data (with graphical user interface) • https://plutof.ut.ee/#/datacite/10.15156%2FBIO%2F587450	46
PIPITS	Full pipeline for fungal ITS amplicon data (only for Illumina data) • https://github.com/hsgweon/pipits	48
QIIME	Full pipeline for amplicon data (designed mostly for 16S rRNA) • https://qiime2.org	185
SEED2	Full pipeline for amplicon data (with graphical user interface; on Windows) • http://www.biomed.cas.cz/mbu/lbwrf/seed	186
Microbiology.se	Tools, including ITSx and Metaxa2, for processing ITS, SSU and LSU data • http://microbiology.se	32,187
USEARCH	Versatile software package • https://www.drive5.com/usearch	33
VSEARCH	Versatile software package • https://github.com/torognes/vsearch	34

ITS, internal transcribed spacer; LSU, large subunit; rRNA, ribosomal RNA; SSU, small subunit.

basis of qPCR instead of using other methods, such as quantification of phospholipid fatty acids or ergosterol, there seems to be much higher variation with qPCR than with the other methods[53]. Fungal qPCR provides a decent average measure of biomass together with qPCR of bacterial rRNAs, an approach that has been used to assess the relative abundance of fungi and bacteria in soil[54]. Taking the mentioned shortcomings into account, qPCR methods can be tailored to study specific guilds of fungi, for example, arbuscular mycorrhizal fungi[55]. For individual species[56], PCR-based abundance estimates are a solid indicator of fungal biomass because the number of rDNA copies varies little in many species[57], although there are exceptions[58]. Moreover, in special situations such as during sporulation, nuclear density may cause problems. The first qPCR methods had low throughput, but recently developed microchips enable quantification of >5,000 samples simultaneously[59].

The use of droplet digital PCR will probably increase, as it provides more precise estimations of copy numbers and does not require comparison with external standards[60].

Droplet digital PCR
A PCR approach in which the PCR solution is divided into smaller oil-covered droplets in which the PCRs are then carried out.

Ectomycorrhizal fungi
Fungi that form mutualistic symbioses between fungi and various species of primarily woody plants at the root tips of the plants, although the fungi do not penetrate the plant cell walls. The ability to form ectomycorrhiza is primarily found in the fungal phyla Ascomycota and Basidiomycota and has evolved and been lost multiple times independently.

Total DNA and RNA sequencing, that is, metagenomics and metatranscriptomics, are more expensive, alternative approaches that allow relative quantification of DNA and gene expression of any organism. By assigning individual reads to taxa, these approaches also provide relative abundances of higher-level fungal taxa that are unbiased by PCR[61,62]. The relative abundance of fungal transcripts in the total metatranscriptome can function as a proxy for the contribution of fungi to the activity, rather than the abundance, of the total microbiome[63].

As an alternative to direct PCR-based quantification, DNA spiking protocols can be used for absolute quantification of targeted organisms, including prokaryotes, eukaryotes and fungi[64]. Spiking involves adding a known quantity of artificial DNA molecules with specific priming sites to samples before DNA extraction[65]. These are then quantified relative to targeted organisms on the basis of HTS read abundance. Although this method has lower PCR costs, it cannot be reliably used for pre-extracted DNA samples owing to variation in DNA extraction efficiency and sample chemistry across aliquots and batches. Spiking may also lead to the loss of many of the targeted sequences when DNA content varies greatly between samples.

Arrays and microarrays. Hybridization-based array technologies emerged in the late 1990s and gained popularity in clinical microbiology and bacteriology but not in mycology. Although microarrays for select human pathogenic fungi[66] and for ectomycorrhizal fungi[67] exist, the scientific community did not widely adopt them. The widely used Geochip detects soil functions: it probes functional genes, including some fungal genes, in DNA or RNA samples[68]. Arrays may contain any short molecular sequence, including rRNA markers and protein-coding genes. For both fungi and bacteria, however, the main issue with microarrays lies in the closed format, which means that only pre-defined taxa and functions are detected and there is a risk of obtaining mixed signals from a target taxon and closely related but untargeted taxa[69]. Therefore, in the study of fungal communities, HTS is more popular than arrays despite the capacity of the latter to provide absolute taxon quantification compared with the relative quantification provided by HTS.

Metagenomics and metatranscriptomics. Although amplicon sequencing is currently the most popular method to target marker genes, the recent development and cost-effectiveness of shotgun HTS enables targeting a suite of genes. Compared with amplicon sequencing, omics approaches have the potential to detect taxonomic composition and relative abundances more accurately by avoiding biases inherent to PCR and primer choice[70,71] (BOX 1). Omics data can be useful for designing and evaluating new primers[72] and thereby aid the discovery of new fungal clades by amplicon sequencing. In addition, omics also has the potential to improve standardization and comparison of relative abundances of phyla across studies in the future, as — compared with metabarcoding, for which relative abundance strongly depends on sequencing depth and amplification biases — relative

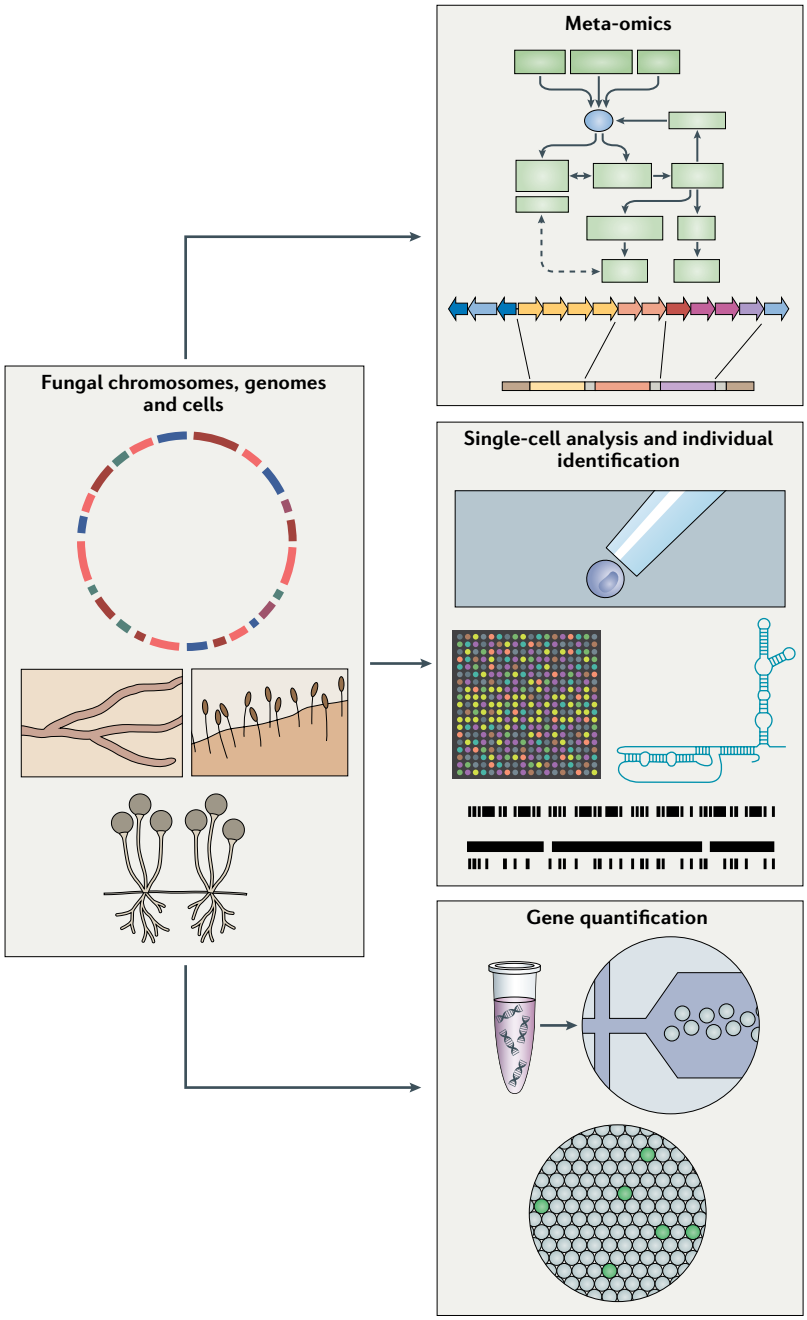

Fig. 2 | **Emerging sequencing techniques.** Metatranscriptomics represents the many emerging omics techniques, giving researchers the opportunity to target expressed genes and thus functional aspects of the fungal community at hand. Identification of fungal populations, individual fungal strains and even single cells promises hitherto unprecedented resolution in the pursuit of fungal diversity and interactions. Finally, refined approaches to gene quantification offer a much-needed quantitative aspect in the study of fungal communities and are expected to become routine in the foreseeable future.

abundance in omics samples constitutes the fraction of all reads mapping a given gene to those mapping all other genes in the community genomes and transcriptomes[70]. In addition, omics approaches enable determining the functional potential by characterizing the prevalence of metabolic genes[61,73], which ultimately could be used for genome-scale metabolic models[74].

Poly(A) tail
A stretch of mRNA that has only adenine (A) bases; it is important for the nuclear export, translation and stability of mRNA.

Metagenomics makes it possible to identify fungal taxa together with prokaryotes[75] and has highlighted their relative role in microbiomes[61]. However, fungal DNA, unlike bacterial DNA, contains many non-coding regions, and fungal genes contain both exons and introns[76], which complicates fungal metagenomics. In addition, fungal genome databases are sparsely populated at present and, in particular, compared with bacterial genome databases[77]. As a result, the proportion of sequences identified as fungal is low even in metagenomes with high fungal abundance, such as topsoil metagenomes[61]. Because of this low sequence abundance, it is difficult to perform OTU analyses from metagenomic data. An alternative is to map short metagenomics reads to reference sequences of known taxonomic identity and then cluster sequences according to taxonomic identities (that is, the MiTag approach[70]), such as was done in a recent study[61] at the level of class and phylum. Metagenomics has nevertheless underscored the importance of fungi for the decomposition of plant biomass[78]. Furthermore, another study has used metagenomics to evaluate amplification biases of universal and specific primers in fungal community profiling[21]. In general, with further development of long-read HTS technologies, metagenomic-based taxonomic profiles need to be compared with profiles from amplicon-based full-length ITS sequences, as has been done for 16S rRNA-based profiles[71].

Fungal metatranscriptomics has the benefit that expressed genes lack introns and can thus be annotated efficiently. Fungal transcripts represented >50% of all transcripts in forest litter[63], which testified to the fungal dominance in the decomposition of plant biopolymers in forest soils[78]. Metatranscriptomics has also been used to determine gene expression of fungal and bacterial communities in response to nitrogen deposition[79]. Unfortunately, but similarly to fungal metagenomics, the limited number of available fungal genomes does not permit taxonomic classification of fungal transcripts much beyond the phylum level. To enrich mRNA and make analysis feasible, rRNA can be removed with the Illumina RiboZero rRNA depletion kit[79,80]. An alternative option is to purify mRNA by hybridization targeting of its poly(A) tail, as exemplified in an expression study of mycorrhizal pine symbionts[81], although this requires that the RNA is highly intact. An interesting new option for HTS is native RNA sequencing[82], in which RNA is sequenced directly and which has recently been implemented by Oxford Nanopore Technologies. RNA-based methods provide a straightforward way to quantify active fungal communities and their functionality in an unbiased way.

HTS methods for identification of strains and individuals. When larger numbers of nuclear genes are targeted, HTS methods become powerful at delimiting even recently evolved species and species complexes. By also considering SNPs within a gene, the resolution increases further, facilitating genotyping of strains and even distinct fungal individuals. Several genotyping methods have been developed over the years, particularly for pathogenic fungi[83]. For example, an assay exists

for identifying individual clades of the amphibian pathogen *Batrachochytrium dendrobatidis* from swab samples[84]. Genotyping can be expanded by whole-genome re-sequencing[85], which depends on the availability of fungal reference genomes. The 1000 Fungal Genomes Project[86] and similar initiatives provide important reference data for future research using comparative genomics. With the help of genome-wide association studies, it is now possible to distinguish phenotypes, for example, of opportunistic pathogens such as *Cryptococcus neoformans*, and to identify the respective virulence factors[87]. Furthermore, it is now possible to compare genomes of monokaryotic and dikaryotic strains of a single species, which enables examination of differential transcription patterns[88]. HTS coupled with single-cell techniques and whole-genome amplification makes it possible to analyse the genomes of uncultivable fungi derived from microscopic material such as single hyphae, conidia and spores. A recent whole-genome HTS study even extended to the subcellular level and revealed the heterogeneity of nuclei from a single arbuscular mycorrhizal spore[89].

HTS studies of fungal communities

This section provides an overview of recent and noteworthy results from second-generation and third-generation HTS studies that investigated fungal communities and specific fungal guilds (FIG. 3).

Overall fungal diversity. Wide-reaching HTS studies that encompass whole continents or even the whole planet often recover tens of thousands of OTUs depending on the targeted substrate. A substantial proportion of these OTUs defies identification at various taxonomic levels (often even at the phylum level), which highlights the importance of fungal metabarcoding to help quantify and ultimately describe fungal biodiversity. A global study of soil fungi found 44,563 non-singleton OTUs[12], ~6% of which could not be assigned to any known fungal class. The authors highlighted 14 large groups of OTUs as potential undescribed classes of fungi, which were further characterized through extended ribosomal DNA sequencing in a subsequent study[22]. Another study looked at indoor settled dust and recovered 38,473 OTUs from 1,000 swab samples that were collected across the continental USA[90].

No single factor determines the global distribution and abundance of all fungi, although patterns can be discerned for specific taxonomic and functional groups of fungi, as well as at smaller spatial scales. Globally, soil fungal diversity is decoupled from plant diversity, although relationships between plant and fungal diversity can be strong locally at the level of species richness[91] and in phylogenetic diversity[92]. Climate, soil chemistry and location are the strongest predictors of fungal richness and community composition globally. Most but not all fungal groups and guilds follow the traditionally predicted gradient of increasing diversity towards the equator[12]. By contrast, ectomycorrhizal fungi, which are mainly associated with forest trees, are the most diverse in temperate biomes, whereas the relative abundance and diversity of agriculturally important arbuscular mycorrhizal fungi are higher in grasslands than forests, with unclear latitudinal patterns[93]. Other studies have pointed at the influence of soil organic carbon content[94], temperature[95] and pH[96]. It is also important to understand the distribution of ecophysiological traits in an environmental context.

Saprotrophic fungi. Saprotrophic fungi have key roles in nutrient cycling and pedogenesis[97]. Several groups of saprotrophic fungi are subjects of longstanding taxonomic scrutiny, notably wood-decaying fungi. HTS-based studies of saprotrophs have nevertheless offered their share of surprises in terms of taxonomy, community structure and the timing of colonization and decomposition processes. Studies of senescent oak leaves revealed colonization by a range of cellulolytic fungi well before leaf abscission, and after shedding, there was a rapid succession of fungal lineages. Ascomycetes dominated the early stages of decomposition, whereas basidiomycetes increased gradually over time, presumably through a combination of changes in litter nutrient contents, litter chemistry, order of colonization and other fungal interactions[98]. These results are compatible with the concept of three phases of plant litter decomposition: early (targeting hemicelluloses and soluble compounds), intermediate (targeting cellulose) and late (targeting lignin)[99]. Studies of fungal wood decomposition following forest and tree perturbation indicated dramatic changes in fungal biomass, community composition and wood decay rates, questioning whether current wood decomposition models account for fungi in an adequate way[100,101]. Studies of soil communities revealed strong vertical partitioning of the fungal community, with a decrease of diversity and an increase of species of unclear taxonomic placement with increasing soil depth[102]. A comparison of fungal communities between soil and fallen, decaying wood found marked qualitative taxonomic differences, although the communities converged over time[103]. Decaying logs were found to influence the composition of the soil community below, which together with the recovery of ectomycorrhizal species from inside the dead wood hinted at incompletely understood processes and interactions between species. Saprotrophic fungi were also found in less conventional habitats, for example, in deep-sea hydrothermal vents[104], which deepens our understanding of decomposition processes driven by fungi.

Mycorrhizal fungi. HTS enables exploration of mycorrhizal fungi from not only plant roots but also soil, wood and air[105]. Studies of soil and litter provide information about the extraradical phase and recover greater diversity than analyses of only root tips, as they also include hyphae originating from unsampled roots as well as germinating and dormant spores. The diversity of ectomycorrhizal fungi in soil increases with stand age[106], tree richness[107] and productivity[108], peaking at intermediate soil moisture[109]. Ectomycorrhizal hyphae also colonize decomposing wood and may become abundant at late stages, when energy sources are depleted[103]. Ectomycorrhizal fungal diversity declines rapidly in response to tree cutting[110], which together

Hyphae
The branching filaments that collectively make up the mycelium of a fungus.

Conidia
Asexual, non-motile fungal spores typically produced on specialized stalked cells (conidiophores) for survival and dispersal.

Spores
The fungal spore is the unit for sexual and asexual reproduction, as well as for dispersal and, at times, survival during unfavourable conditions through dormancy.

Saprotrophic fungi
Fungi deriving their energy and nutrients from decomposing non-living organic matter; these are found throughout the fungal tree of life and are often intermingled with species with other nutritional strategies in puzzling ways.

Pedogenesis
The process of soil formation as affected by the soil biota and the environment at large.

Extraradical phase
Scavenging fungal hyphae that emanate, for example, from ectomycorrhizal root tips.

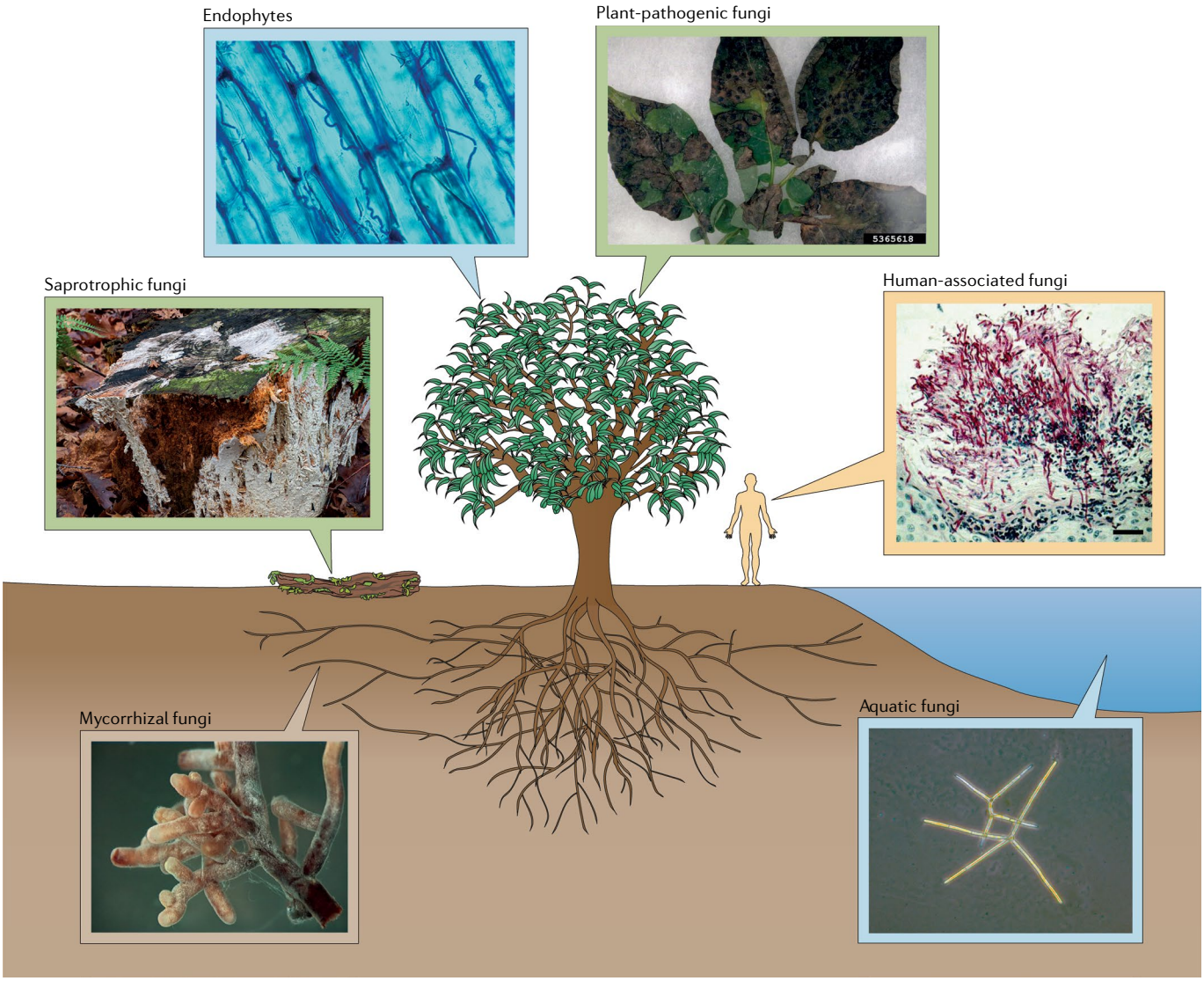

Fig. 3 | **Fungal diversity in different environments.** Fungi permeate all aspects of natural ecosystems and are crucial, yet largely overlooked, for nutrient recycling across trophic levels. Saprotrophic fungi mediate the decomposition of organic matter, as illustrated in the figure by decaying wood and a white-rotting fungus that has decomposed the lignin but largely left the cellulose intact. Mycorrhizal fungi associate with roots and are represented in the figure by ectomycorrhiza of the *Amanita* (fly agaric) type that form a hyphal sheath around the root tips; other types of mycorrhiza are less conspicuous. Plant pathogens are represented in the figure by a species of *Alternaria* decomposing a living leaf; *Alternaria* spp. are known as major plant pathogens throughout the world. By contrast, endophytes live inside plants and are rarely visible to the naked eye. They have a range of poorly understood nutritional strategies and taxonomic affiliations. Aquatic fungi comprise a mixture of early-diverging lineages and fungi that later transitioned back into the aquatic realm. Shown in the figure is an example of the latter, a so-called aquatic hyphomycete. Finally, human-associated fungi are represented by a *Candida* sp. in a microscope mount, dyed to make it easier to examine. *Candida* infections (candidiasis) range from merely annoying to life-threatening in humans, although it is primarily infants and patients with weakened immune systems that are at risk of serious infection. Adapted from REF.[4], Springer Nature Limited. Mycorrhizal fungi image reproduced from REF.[181], Springer Nature Limited. Saprotrophic fungi image courtesy of Arterra Picture Library/Alamy Stock Photo. Endophytes photo by Nick Hill, USDA Agricultural Research Service. Plant-pathogenic fungi image courtesy of Howard F. Schwartz, Colorado State University, Bugwood.org. Human-associated fungi image reproduced from REF.[182], Springer Nature Limited. Aquatic fungi image courtesy of George Barron, University of Guelph.

with interrupted belowground carbon flows hampers nutrient cycling in clear-felled habitats[111]. Arbuscular mycorrhizal fungal diversity correlates with plant richness locally[112], with a strong effect of soil pH[113]. Arbuscular mycorrhizal fungal communities cluster phylogenetically, indicating niche overlap between closely related taxa[114]. Ericaceous plants share their root ericoid mycorrhizal symbionts and putative endophytes with other plants, which complicates the identification of functionally active ericoid mycorrhizal fungi[115]. Long-fragment HTS provides additional resolution compared with the commonly used SSU gene for

arbuscular mycorrhizal fungi[116]. Co-occurring orchid species differ greatly in the specificity and composition of their mycorrhizal mycobionts, whereas members of the same species harbour distinct fungal communities across habitat types and geographic regions[117]. Soil fungal communities may determine the ability of certain orchid species to inhabit particular niches and promote coexistence of multiple taxa[118]. Meta-analyses of several sequencing studies revealed no evidence for systematic differences in spatiotemporal variability, host specificity or community network properties between mycorrhizal guilds, although arbuscular mycorrhizal fungal communities have been considered the least host-dependent[119,120].

Plant-pathogenic fungi. The use of HTS methods in mycopathology has somewhat lagged behind other fields of mycology despite the potential usefulness of HTS for surveillance. Much of the available information comes from studies that did not specifically address pathogens. Molecular diagnostic assays are available but typically do not involve HTS and are limited to a handful of key pathogens on select economically important hosts[121]. Spore trapping and metabarcoding have been recommended for monitoring the spread of fungal pathogens and invasive species[122]. Monitoring the persistence of the root rot biocontrol agent *Phlebiopsis gigantea* revealed its disappearance from infected stumps within 12 months, suggesting poor long-term performance as a biocontrol agent in natural conditions[123]. Rotten olives harboured a high proportion of pathogenic *Colletotrichum* spp. and a generally more diverse fungal community than undamaged olives[124]. Balsam poplar leaves harboured fewer pathogens the further north[125], consistent with strong latitudinal diversity gradients of fungal pathogens. Across fungi, the abundance of pathogens declined with decreasing anthropogenic influence on habitats[126] and soil dryness[109]. Fruitflies and bark beetles introduced with imported timber may disperse invasive plant-pathogenic fungi[127]. Early-diverging vascular plants are overlooked hosts for poorly characterized fungal pathogens[128]. The concept of disease-suppressive soils, in which the local microbiota suppresses disease development, is of interest for the biological management of plant pathogens[129]. A combination of metabarcoding and metaproteomic analyses may soon enable detailed investigations of the fungal proteins that are produced by pathogenic fungi and of interactions between plant and fungal proteins during health and disease.

Foliar endophytes. Endophytes inhabit all vegetative organs of plants and include (latent) saprotrophs, commensals, mutualists and even parasites. HTS studies often recover several hundred OTUs from single leaf extracts[130], which eclipses the diversity recovered by culturing. Although ascomycetes, and to a lesser extent basidiomycetes, are abundant in most leaf mycobiota, a high abundance of zygomycetes and unclassifiable OTUs was found in pine needles[131]. The interplay between foliar endophytes, the microorganisms living on the surface and their hosts is poorly understood. Endophytes can increase host fitness, notably through pathogen antagonism and nutrient exchange. Other endophytes have adverse effects on their hosts or may switch to parasitic phases as conditions change[132]. Abiotic factors, plant traits and competition are thought to shape leaf mycobiota[133], but disentangling individual, general factors has proved difficult.

In temperate plants, the host species usually strongly influences the composition of the leaf mycobiota, but this relationship is much weaker in tropical plants. The leaf mycobiota differed substantially between neighbouring tropical *Amorphophallus* and *Camellia* plants[134]. Furthermore, the number of OTUs increased with *Camellia* leaf age, and different plant tissues of *Amorphophallus* harboured different fungal communities. Interestingly, the foliar endophytes of tropical grasses showed no host specificity; instead, there was a strong spatial structure consistent with dispersal limitation[135]. Local habitat conditions, rather than host genotype, structured the composition of the fungal communities of Arctic white spruce needles[136]. The composition of the fungal phyllosphere in bur oaks changed drastically over time[137]. HTS can help to elucidate the role of foliar endophytes in plant health, the potential effects of leaf endophytes on other mycobiota of the plant and how foliar endophytes mediate resistance against pathogens[138].

Aquatic fungi. Aquatic habitats include oceans, hydrothermal systems, sediments, aquifers, glaciers and inland waters. Fungi evolved in the sea, and current aquatic fungi are a mixture of direct descendants from the earliest fungi and more recent lineages that transitioned back into the aquatic realm. This diversity necessitates the use of several marker regions, as any single marker will be either too conserved or so divergent that it lacks satisfactory reference sequences[139]. However, multimarker studies are still rare, and most studies of aquatic fungi are based on SSU sequences[140]. The first HTS studies confirmed an unprecedented diversity of unknown, early-diverging fungal lineages in aquatic environments[141]. Metastudies of HTS freshwater data sets predicted a central role of fungi in inland waters and confirmed that there is a lack of fungal reference data, even when using the SSU for phylum-level assignment[142]. Given the higher divergence rates of the LSU and ITS markers than the SSU, the classification success of ITS sequences in aquatic fungi is often low, sometimes with >60% unclassifiable sequences even at the phylum level[143]. The exceptions include samples derived from plants, such as submerged leaf litter in streams and macrophytes in the ocean, which are frequently colonized by mitosporic ascomycetes and basidiomycetes, for which ITS works well. HTS studies have widened the scope of aquatic mycology to include various aquatic habitats and biomes, such as microbial biofilms[144], sea-ice[145] and marine snow[146]. It is now possible to examine biogeographical structuring[147] and monitor seasonal patterns of aquatic fungi[148]. Reanalysis of Tara ocean data[73] revealed no latitudinal diversity gradients for fungi, but oomycetes were among the groups showing the strongest correlation

Ericoid mycorrhizal symbionts
Organisms that participate in mutualistic symbiosis formed between members of the plant family Ericaceae and a number of fungal lineages mainly of the ascomycetes; plant cell walls are penetrated, and fungal coils are found within the plant cells.

Mitosporic
Referring to fungi in their asexual state.

Marine snow
Organic matter falling from upper waters to the deep ocean; it is often the dominant external source of carbon in these nutrient-deprived systems.

with the abundance of bacterial antibiotic resistance genes, implying antagonism[61]. Given the predominance of early-diverging and formally undescribed fungal lineages, the enormous variety of habitats and the paucity of research, aquatic mycology remains in its infancy and has much to gain from a combination of sequencing long DNA stretches and visualization with fluorescent hybridization[149,150].

Human-associated fungi. The fungal component of the human microbiota was relatively overlooked until recently. Growing awareness of the roles of fungi in disease development and the increasing prevalence of fungal infections, as well as their potential health benefits and probiotic effects, have served to put the human mycobiota in the spotlight[151–153]. So far, the mycobiota of the oral cavity, skin, lung, gut and vagina have been studied[154]. Humans, in both health and disease, are hosts to a wide range of fungi from large parts of the fungal tree of life. Key genera include many that have been on the agenda of medical mycology for decades, notably *Candida*, *Malassezia*, *Cryptococcus* sensu lato and *Penicillium*. Others, such as *Alternaria*, *Mucor* and *Schizophyllum*, are perhaps less often thought of in medical contexts. Recent results indicate that fungi colonize human infants very early in life through the mother, other caregivers and the environment[155]. However, the methods used to study the human mycobiota can affect the results, cautioning against (premature) conclusions on whether core mycobiota exist for humans and for different organs and body parts[23]. Fungi play important roles in decomposition and nitrogen cycling of mammalian corpses, with many key species established from the soil community below[156]. HTS studies of the built (indoor) environment have revealed a far greater fungal diversity than previously assumed, forcing researchers to consider new definitions of terms and refine the ways in which the built mycobiota is sampled[157,158]. HTS methods are increasingly used to study food-borne fungi and fungi used in brewing and fermentation, which are now known to be complex, multi-organism processes[159]. Fungi also interact with non-human mammals and other animals in multiple biologically important ways as parasites, members of normal skin, nail and claw microbiota and fermenting gut symbionts[160,161].

Perspectives and conclusions

Rapid advances in DNA and RNA sequencing technologies now enable us to study fungal communities in an integrative way, including exploring the taxonomic profiles of fungal communities and their functional and ecological attributes. These data allow us to interpret communities in a richer way and to address questions related to ecosystem functioning. Similarly, there may be little reason to target individual fungal guilds such as ectomycorrhizal fungi in isolation when we now know that fungi interact across taxonomic groups and functional guilds, although the details of these interactions are still poorly understood. Finally, it may be misguided to sequence only fungi; it would be better to consider including other eukaryotes and prokaryotes, given that subsets of these groups often co-occur and interact[61]. As HTS data accumulate — more than a billion HTS-derived ITS reads are now publicly available[162] — it becomes increasingly important to use those data to glean clues to previously overlooked, as well as new, research questions, hypotheses and theories. We seem to have reached the stage at which imagination and theoretical understanding, rather than access to sufficient amounts of data, set the limit for what research hypotheses are pursued (Supplementary Fig. 2).

We would like to make a case for fungal metabarcoding as an interdisciplinary and essentially reproducible research strategy. Project teams should be assembled to comprise expertise in not only mycology but also ecology, the Earth sciences, bioinformatics, statistics and laboratory and analytical procedures as deemed relevant for the research questions at hand. As an often brushed aside field, mycology needs to ask grand scientific questions and target large-scale patterns and processes[163]. Adopting the view that any ongoing study is a step forward not only in one's career but also for mycology and science would be a good start. The key here is producing reproducible results by following applicable standards and protocols, providing ample detail on data processing and analysis and making all relevant data freely and openly available[164]. Failure to do so will serve to maintain the all-too-common view that fungi matter only to mycologists, a belief that has haunted mycology for far too long.

Published online 15 November 2018

1. Hawksworth, D. L. & Lücking, R. Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiol. Spectr.* **5**, 4 (2017).
2. Tedersoo, L. et al. High-level classification of the fungi and a tool for evolutionary ecological analyses. *Fung. Div.* **90**, 135–159 (2018).
3. Berbee, M. L., James, T. Y. & Strullu-Derrien, C. Early diverging fungi: diversity and impact at the dawn of terrestrial life. *Ann. Rev. Microbiol.* **71**, 41–60 (2017).
This eye-opening paper is a mycological must-read that substantially broadens the scope of what fungi are and which fungi should be included in both HTS-based efforts and general mycological parlance.
4. Peay, K. G., Kennedy, P. G. & Talbot, J. M. Dimensions of biodiversity in the Earth mycobiome. *Nat. Rev. Microbiol.* **14**, 434–447 (2016).
5. Tedersoo, L. & Nilsson, R. H. in *Molecular Mycorrhizal Symbiosis* (ed. Martin, F.) 301–322 (Wiley, Hoboken, 2016).
6. O'Brien, H. E., Parrent, J. L., Jackson, J. A., Moncalvo, J. M. & Vilgalys, R. Fungal community analysis by large-scale sequencing of environmental samples. *Appl. Environ. Microbiol.* **71**, 5544–5550 (2005).
7. Hibbett, D. S., Ohman, A. & Kirk, P. M. Fungal ecology catches fire. *New Phytol.* **184**, 279–282 (2009).
8. Hibbett, D. et al. Sequence-based classification and identification of fungi. *Mycologia* **108**, 1049–1068 (2016).
This thoughtful article discusses how to translate HTS results into mycological progress in the context of systematics, taxonomy, ecology and nomenclature.
9. Petersen, L., Minkkinen, P. & Esbensen, K. H. Representative sampling for reliable data analysis: theory of sampling. *Chemometr. Intell. Lab.* **77**, 261–277 (2005).
10. Lindahl, B. D. et al. Fungal community analysis by high-throughput sequencing of amplified markers — a user's guide. *New Phytol.* **199**, 288–299 (2013).
11. McMurdie, P. J. & Holmes, S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLOS Comput. Biol.* **10**, e1003531 (2014).
This somewhat controversial paper discusses the thorny issue of rarefication in a most informative way.
12. Tedersoo, L. et al. Global diversity and geography of soil fungi. *Science* **346**, 1256688 (2014).
This landmark paper presents the first comprehensive HTS-based overview of soil fungi and the patterns and processes governing their distribution.
13. Song, Z. et al. Effort versus reward: preparing samples for fungal community characterization in high-throughput sequencing surveys of soils. *PLOS ONE* **10**, e0127234 (2015).
14. Rissanen, A. J., Kurhela, E., Aho, T., Oittinen, T. & Tiirola, M. Storage of environmental samples for guaranteeing nucleic acid yields for molecular microbiological studies. *Appl. Microbiol. Biotechnol.* **88**, 977–984 (2010).
15. Bolano, A. et al. Rapid methods to extract DNA and RNA from *Cryptococcus neoformans*. *FEMS Yeast Res.* **1**, 221–224 (2001).
16. Huang, X. et al. CTAB-PEG DNA extraction from fungi with high contents of polysaccharides. *Mol. Biol.* **52**, 621–628 (2018).

17. Begerow, D., Nilsson, H., Unterseher, M. & Maier, W. Current state and perspectives of fungal DNA barcoding and rapid identification procedures. *Appl. Microbiol. Biot.* **87**, 99–108 (2010).
18. Schoch, C. L. et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proc. Natl Acad. Sci. USA* **109**, 6241–6246 (2012).
19. Vu, D. et al. Large-scale generation and analysis of filamentous fungal DNA barcodes boosts coverage for kingdom Fungi and reveals thresholds for fungal species and higher taxon delimitation. *Stud. Mycol.* **92**, 135–154 (2018).
20. Větrovský, T., Kolařík, M., Žifčáková, L., Zelenka, T. & Baldrian, P. The *rpb2* gene represents a viable alternative molecular marker for the analysis of environmental fungal communities. *Mol. Ecol. Res.* **16**, 388–401 (2016).
21. Tedersoo, L. et al. Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycoKeys* **10**, 1–43 (2015).
22. Tedersoo, L., Bahram, M., Puusepp, R., Nilsson, R. H. & James, T. Y. Novel soil-inhabiting clades fill gaps in the fungal tree of life. *Microbiome* **5**, 42 (2017).
This paper takes off where other HTS-based efforts wrap up by providing a method for extending partial ITS sequence reads into longer rRNA reads that are useful for phylogenetic assignment and ulterior species descriptions.
23. Tedersoo, L. & Lindahl, B. Fungal identification biases in microbiome projects. *Env. Microbiol. Rep.* **8**, 774–779 (2016).
24. Wu, L. et al. Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. *BMC Microbiol.* **15**, 125 (2015).
25. D'Amore, R. et al. A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC Genomics* **17**, 55 (2016).
This is a very useful, if 16S rRNA-oriented, overview of workflows and resources for HTS-oriented metabarcoding.
26. Bakker, M. G. A fungal mock community control for amplicon sequencing experiments. *Mol. Ecol. Res.* **18**, 541–556 (2018).
27. Palmer, J. M., Jusino, M. A., Banik, M. T. & Lindner, D. L. Non-biological synthetic spike-in controls and the AMPtk software pipeline improve mycobiome data. *PeerJ* **6**, e4925 (2018).
28. Salter, S. J. et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, 87 (2014).
29. Nguyen, N. H. et al. FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* **20**, 241–248 (2016).
This paper introduces a much-needed tool to estimate functional (guild) properties for taxa recovered in HTS studies, paving the way for ecologically informed analyses of fungal communities.
30. Carlsen, T. et al. Don't make a mista(g)ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies? *Fungal Ecol.* **5**, 747–749 (2012).
31. Nilsson, R. H. et al. Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences. *MycoKeys* **4**, 37–63 (2012).
32. Bengtsson-Palme, J. et al. 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 (2013).
33. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
34. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
35. Schloss, P. D. et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
36. Mahé, F., Rognes, T., Quince, C., de Vargas, C. & Dunthorn, M. Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ* **2**, e593 (2014).
37. Eren, A. M. et al. Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J.* **9**, 968 (2015).
38. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
39. Lindner, D. L. et al. Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. *Ecol. Evol.* **3**, 1751–1764 (2013).
40. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
41. Öpik, M. et al. The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). *New Phytol.* **188**, 223–241 (2010).
42. Kõljalg, U. et al. Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* **22**, 5271–5277 (2013).
43. Deshpande, V. et al. Fungal identification using a Bayesian classifier and the Warcup training set of internal transcribed spacer sequences. *Mycologia* **108**, 1–5 (2016).
44. Nilsson, R. H. et al. Top 50 most wanted fungi. *MycoKeys* **12**, 29–40 (2016).
This paper urges the mycological community to keep the potential of taxonomic feedback loops in HTS efforts in mind and introduces a software tool to that effect.
45. Kõljalg, U., Tedersoo, L., Nilsson, R. H. & Abarenkov, K. Digital identifiers for fungal species. *Science* **352**, 1182–1183 (2016).
46. Anslan, S., Bahram, M., Hiiesalu, I. & Tedersoo, L. PipeCraft: flexible open-source toolkit for bioinformatics analysis of custom high-throughput amplicon sequencing data. *Mol. Ecol. Res.* **17**, e234–e240 (2017).
47. Hildebrand, F., Tadeo, R., Voigt, A. Y., Bork, P. & Raes, J. LotuS: an efficient and user-friendly OTU processing pipeline. *Microbiome* **2**, 30 (2014).
48. Gweon, H. S. et al. PIPITS: an automated pipeline for analyses of fungal internal transcribed spacer sequences from the Illumina sequencing platform. *Meth. Ecol. Evol.* **6**, 973–980 (2015).
49. Anslan, S. et al. Great differences in performance and outcome of high-throughput sequencing data analysis platforms for fungal metabarcoding. *Mycokeys* **39**, 29–40 (2018).
50. Frøslev, T. G. et al. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nat. Commun.* **8**, 1188 (2017).
This article contributes a novel and much-needed approach for removing compromised OTUs from HTS community data.
51. Gdanetz, K., Benucci, G. M. N., Pol, N. V. & Bonito, G. CONSTAX: a tool for improved taxonomic resolution of environmental fungal ITS sequences. *BMC Bioinf.* **18**, 538 (2017).
52. Wallander, H. et al. Evaluation of methods to estimate production, biomass and turnover of ectomycorrhizal mycelium in forests soils — a review. *Soil Biol. Biochem.* **57**, 1034–1047 (2013).
53. Baldrian, P. et al. Estimation of fungal biomass in forest litter and soil. *Fungal Ecol.* **6**, 1–11 (2013).
54. Rousk, J. et al. Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J.* **4**, 1340–1351 (2010).
55. Voříšková, A. et al. Real-time PCR quantification of arbuscular mycorrhizal fungi: does the use of nuclear or mitochondrial markers make a difference? *Mycorrhiza* **27**, 577–585 (2017).
56. Yamaguchi, M. et al. A qPCR assay that specifically quantifies *Tricholoma matsutake* biomass in natural soil. *Mycorrhiza* **26**, 847–861 (2016).
57. Amend, A. S., Seifert, K. A., Samson, R. & Bruns, T. D. Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. *Proc. Natl Acad. Sci. USA* **107**, 13748–13753 (2010).
58. Liti, G. et al. Population genomics of domestic and wild yeasts. *Nature* **458**, 337–341 (2009).
59. Muurinen, J. et al. Influence of manure application on the environmental resistome under Finnish agricultural practice with restricted antibiotic use. *Environ. Sci. Technol.* **51**, 5989–5999 (2017).
60. Hindson, B. J. et al. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* **83**, 8604–8610 (2011).
61. Bahram, M. et al. Structure and function of the global topsoil microbiome. *Nature* **560**, 233–237 (2018).
By leveraging a large set of metagenomes, this study introduces the first comprehensive soil gene catalogue and reports on the global diversity of soil fungi and bacteria and their gene functions in an environmental context.
62. Karst, S. M. et al. Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. *Nat. Biotechnol.* **36**, 190–195 (2018).
63. Žifčáková, L., Větrovský, T., Howe, A. & Baldrian, P. Microbial activity in forest soil reflects the changes in ecosystem properties between summer and winter. *Environ. Microbiol.* **18**, 288–301 (2016).
64. Tkacz, A., Hortala, M. & Poole, P. S. Absolute quantitation of microbiota abundance in environmental samples. *Microbiome* **6**, 110 (2018).
65. Smets, W. et al. A method for simultaneous measurement of soil bacterial abundances and community composition via 16S rRNA gene sequencing. *Soil Biol. Biochem.* **96**, 145–151 (2016).
66. Leinberger, D. M., Schumacher, U., Autenrieth, I. B. & Bachmann, T. T. Development of a DNA microarray for detection and identification of fungal pathogens involved in invasive mycoses. *J. Clin. Microbiol.* **43**, 4943–4953 (2005).
67. Reich, M., Kohler, A., Martin, F. & Buée, M. Development and validation of an oligonucleotide microarray to characterize ectomycorrhizal fungal communities. *BMC Microbiol.* **9**, 241 (2009).
68. He, Z. et al. GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J.* **1**, 67–77 (2007).
69. Zhou, J., He, Z., Deng, Y., Tringe, S. G. & Alvarez-Cohen, L. High-throughput metagenomic technologies for complex microbial community analysis: open and closed formats. *mBio.* **6**, e02288–14 (2015).
70. Logares, R. et al. Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ. Microbiol.* **16**, 2659–2671 (2014).
71. Singer, E. et al. High-resolution phylogenetic microbial community profiling. *ISME J.* **10**, 2020–2032 (2016).
72. Klindworth, A. et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41**, e1 (2013).
73. Sunagawa, S. et al. Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
74. Witherden, E. A., Moyes, D. L., Bruce, K. D., Ehrlich, S. D. & Shoaie, S. Using systems biology approaches to elucidate cause and effect in host–microbiome interactions. *Curr. Opin. System. Biol.* **3**, 141–146 (2017).
75. Geisen, S. et al. Metatranscriptomic census of active protists in soils. *ISME J.* **9**, 2178–2190 (2015).
76. Martin, F. et al. Perigord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* **464**, 1033–1038 (2010).
77. Sokol, H. et al. Fungal microbiota dysbiosis in IBD. *Gut* **66**, 1039–1048 (2017).
78. Žifčáková, L. et al. Feed in summer, rest in winter: microbial carbon utilization in forest topsoil. *Microbiome* **5**, 122 (2017).
This paper shows the power of integrating metagenomics and metatranscriptomics to distinguish the genomic potential and real activity of fungi and bacteria in a fungus-dominated environment of forest soils.
79. Hesse, C. N. et al. Forest floor community metatranscriptomes identify fungal and bacterial responses to N deposition in two maple forests. *Front. Microbiol.* **6**, 337 (2015).
80. Kuske, C. R. et al. Prospects and challenges for fungal metatranscriptomics of complex communities. *Fungal Ecol.* **14**, 133–137 (2015).
81. Liao, H. L. et al. Metatranscriptomic analysis of ectomycorrhizal roots reveals genes associated with *Piloderma-Pinus* symbiosis: improved methodologies for assessing gene expression in situ. *Environ. Microbiol.* **16**, 3730–3742 (2014).
82. Garalde, D. R. et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* **15**, 201–206 (2018).
83. Grünwald, N. J., McDonald, B. A. & Milgroom, M. G. Population genomics of fungal and oomycete pathogens. *Annu. Rev. Phytopathol.* **54**, 323–346 (2016).
84. Byrne, A. Q. et al. Unlocking the story in the swab: a new genotyping assay for the amphibian chytrid fungus *Batrachochytrium dendrobatidis*. *Mol. Ecol. Res.* **17**, 1283–1292 (2017).
85. Fuentes-Pardo, A. P. & Ruzzante, D. E. Whole-genome sequencing approaches for conservation biology:

advantages, limitations, and practical recommendations. *Mol. Ecol.* **26**, 5369–5406 (2017).
86. Grigoriev, I. V. et al. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* **42**, D699–D704 (2013).
87. Desjardins, C. A. et al. Population genomics and the evolution of virulence in the fungal pathogen *Cryptococcus neoformans*. *Genome Res.* **27**, 1207–1219 (2017).
88. López, S. C. et al. Induction of genes encoding plant cell wall-degrading carbohydrate-active enzymes by lignocellulose-derived monosaccharides and cellobiose in the white-rot fungus *Dichomitus squalens*. *Appl. Environ. Microb.* **84**, e00403–e00418 (2018).
89. Lin, K. et al. Single nucleus genome sequencing reveals high similarity among nuclei of an endomycorrhizal fungus. *PLOS Genet.* **10**, e1004078 (2014).
90. Grantham, N. S. et al. Fungi identify the geographic origin of dust samples. *PLOS ONE* **10**, e0122605 (2015).
91. Yang, T. et al. Plant diversity and productivity drive soil fungal richness in natural grasslands of the Tibetan Plateau. *New Phytol.* **215**, 756–765 (2017).
92. Nguyen, N. H. et al. Ectomycorrhizal fungal diversity and saprotrophic fungal diversity are linked to different tree community attributes in a field-based tree experiment. *Mol. Ecol.* **25**, 4032–4046 (2016).
93. Davison, J. et al. Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science* **28**, 970–973 (2015).
94. Maestre, F. T. et al. Increasing aridity reduces soil microbial diversity and abundance in global drylands. *Proc. Natl Acad. Sci. USA* **112**, 15684–15689 (2015).
95. Newsham, K. K. et al. Relationship between soil fungal diversity and temperature in the maritime Antarctic. *Nat. Clim. Change* **6**, 182–186 (2016).
96. Lanzén, A. et al. The community structures of prokaryotes and fungi in mountain pasture soils are highly correlated and primarily influenced by pH. *Front. Microbiol.* **6**, 1321 (2015).
97. Boddy, L., Frankland, J. & Van West, P. *Ecology of Saprotrophic Basidiomycetes (British Mycological Society Symposia Series)* Vol. 28 (Academic Press, London, 2008).
98. Voříšková, J. & Baldrian, P. Fungal community on decomposing leaf litter undergoes rapid successional changes. *ISME J.* **7**, 477–486 (2013).
99. Šnajdr, J. et al. Transformation of *Quercus petraea* litter: successive changes in litter chemistry are reflected in differential enzyme activity and changes in the microbial community composition. *FEMS Microbiol. Ecol.* **75**, 291–303 (2011).
100. Štursová, M. et al. When the forest dies: the response of forest soil fungi to a bark beetle-induced tree dieback. *ISME J.* **8**, 1920–1931 (2014).
101. van der Wal, A., Ottosson, E. & de Boer, W. Neglected role of fungal community composition in explaining variation in wood decay rates. *Ecology* **96**, 124–133 (2015).
102. Jumpponen, A., Jones, K. L. & Blair, J. Vertical distribution of fungal communities in tallgrass prairie soil. *Mycologia* **102**, 1027–1041 (2010).
103. Mäkipää, R. et al. Interactions between soil-and dead wood-inhabiting fungal communities during the decay of Norway spruce logs. *ISME J.* **11**, 1964–1974 (2017).
104. Xu, W., Gong, L. F., Pang, K. L. & Luo, Z. H. Fungal diversity in deep-sea sediments of a hydrothermal vent system in the Southwest Indian Ridge. *Deep-Sea Res. I* **131**, 16–26 (2017).
105. Dickie, I. A. & John, M. G. St. in *Molecular Mycorrhizal Symbiosis* (ed. Matin, F.) 473–491 (John Wiley & Sons, 2016).
106. Kyaschenko, Y., Clemmensen, K., Hagenbo, A., Karltun, E. & Lindahl, B. Shift in fungal communities and associated enzyme activities along an age gradient of managed *Pinus sylvestris* stands. *ISME J.* **11**, 863–874 (2017).
107. Tedersoo, L. et al. Tree diversity and species identity effects on soil fungi, protists and animals are context-dependent. *ISME J.* **10**, 346–362 (2016).
108. Sterkenburg, E. et al. Changes in fungal communities along a boreal forest soil fertility gradient. *New Phytol.* **207**, 1145–1158 (2015).
109. Hiiesalu, I., Bahram, M. & Tedersoo, L. Plant species richness and productivity determine the diversity of soil fungal guilds in temperate coniferous forest and bog habitats. *Mol. Ecol.* **26**, 4846–4858 (2017).
110. Hartmann, M. et al. Significant and persistent impact of timber harvesting on soil microbial communities in northern coniferous forests. *ISME J.* **6**, 2199–2218 (2012).
111. Kohout, P. et al. Clearcutting alters decomposition processes and initiates complex restructuring of fungal communities in soil and tree roots. *ISME J.* **12**, 692–703 (2018).
112. Hiiesalu, I. et al. Species richness of arbuscular mycorrhizal fungi: associations with grassland plant richness and biomass. *New Phytol.* **203**, 233–234 (2014).
113. Dumbrell, A. J., Nelson, M., Helgason, T., Dytham, C. & Fitter, A. H. Relative roles of niche and neutral processes in structuring a soil microbial community. *ISME J.* **4**, 337–345 (2010).
114. Horn, S., Caruso, T., Verbruggen, E., Rillig, M. C. & Hempel, S. Arbuscular mycorrhizal fungal communities are phylogenetically clustered at small scales. *ISME J.* **8**, 2231–2242 (2014).
115. Kohout, P. in *Biogeography of Mycorrhizal Symbiosis* (ed. Tedersoo, L.) 179–193 (Springer, Cham, 2017).
116. Schlaeppi, K. et al. High-resolution community profiling of arbuscular mycorrhizal fungi. *New Phytol.* **212**, 780–791 (2016).
117. Jacquemyn, H. et al. Habitat-driven variation in mycorrhizal communities in the terrestrial orchid genus *Dactylorhiza*. *Sci. Rep.* **6**, 37182 (2016).
118. Waud, M. W., Busschaert, P., Lievens, B. & Jacquemyn, H. Specificity and localised distribution of mycorrhizal fungi in the soil may contribute to co-existence of orchid species. *Fung. Ecol.* **20**, 155–165 (2016).
119. Bahram, M., Peay, K. G. & Tedersoo, L. Local-scale biogeography and spatiotemporal variability in communities of mycorrhizal fungi. *New Phytol.* **205**, 1454–1463 (2015).
120. Pölme, S. et al. Host preference and network properties in biotrophic plant–fungal associations. *New Phytol.* **217**, 1230–1239 (2018).
121. Abdelfattah, A., Malacrinò, A., Wisniewski, M., Cacciola, S. O. & Schena, L. Metabarcoding: a powerful tool to investigate microbial communities and shape future plant protection strategies. *Biol. Control* **120**, 1–10 (2017).
122. Banchi, E. et al. DNA metabarcoding uncovers fungal diversity of mixed airborne samples in Italy. *PLOS ONE* **13**, e0194489 (2018).
123. Terhonen, E. et al. Effects of the use of biocontrol agent (*Phlebiopsis gigantea*) on fungal communities on the surface of *Picea abies* stumps. *Forest Ecol. Manag.* **310**, 428–433 (2013).
124. Abdelfattah, A., Nicosia, M. G. L. D., Cacciola, S. O., Droby, S. & Schena, L. Metabarcoding analysis of fungal diversity in the phyllosphere and carposphere of olive (*Olea europaea*). *PLOS ONE* **10**, e0131069 (2015).
125. Bälint, M. et al. Relocation, high-latitude warming and host genetic identity shape the foliar fungal microbiome of poplars. *Mol. Ecol.* **24**, 235–248 (2015).
126. Yan, D. et al. High-throughput eDNA monitoring of fungi to track functional recovery in ecological restoration. *Biol. Conserv.* **217**, 113–120 (2018).
127. Malacrino, A. et al. Fungal communities associated with bark and ambrosia beetles trapped at international harbours. *Fungal Ecol.* **28**, 44–52 (2017).
128. Rimington, W. R., Pressel, S., Duckett, J. G. & Bidartondo, M. I. Fungal associations of basal vascular plants: reopening a closed book? *New Phytol.* **205**, 1394–1398 (2015).
129. Sun, P., Otto-Hanson, L. K., Arenz, B. E., Ma, Q. & Kinkel, L. L. Molecular and functional characteristics of streptomycete communities in relation to soil factors and potato common scab. *Eur. J. Soil Biol.* **70**, 58–66 (2015).
130. Datlof, E. M. et al. Uncovering unseen fungal diversity from plant DNA banks. *PeerJ* **5**, e3730 (2017).
131. Peršoh, D. Factors shaping community structure of endophytic fungi — evidence from the *Pinus-Viscum*-system. *Fungal Divers.* **60**, 55–69 (2013).
132. Porras-Alfaro, A. & Bayman, P. Hidden fungi, emergent properties: endophytes and microbiomes. *Ann. Rev. Phytopathol.* **49**, 291–315 (2011).
133. van Bael, S., Estrada, C. & Arnold, A. E. in *The Fungal Community: its Organization and Role in the Ecosystem* (eds Dighton, J. & White, J. F.) 79–94 (CRC Press, Boca Raton, 2017).
134. Unterseher, M. et al. Mycobiota of sympatric *Amorphophallus albispathus* (Araceae) and *Camellia sinensis* (Theaceae) — a case study reveals clear tissue preferences and differences in diversity and composition. *Mycol. Prog.* **17**, 489–500 (2018).
135. Higgins, K. L., Arnold, A. E., Coley, P. D. & Kursar, T. A. Communities of fungal endophytes in tropical forest grasses: highly diverse host-and habitat generalists characterized by strong spatial structure. *Fungal Ecol.* **8**, 1–11 (2014).
136. Eusemann, P. et al. Habitat conditions and phenological tree traits overrule the influence of tree genotype in the needle mycobiome–*Picea glauca* system at an arctic treeline ecotone. *New Phytol.* **211**, 1221–1231 (2016).
137. Jumpponen, A., Jones, K. L., Mattox, J. D. & Yaege, C. Massively parallel 454-sequencing of fungal communities in *Quercus* spp. ectomycorrhizas indicates seasonal dynamics in urban and rural sites. *Mol. Ecol.* **19** (Suppl. 1), 41–53 (2010).
138. Busby, P. E., Peay, K. G. & Newcombe, G. Common foliar fungi of *Populus trichocarpa* modify *Melampsora* rust disease severity. *New Phytol.* **209**, 1681–1692 (2016).
139. Heeger, F. et al. Long-read DNA metabarcoding of ribosomal rRNA in the analysis of fungi from aquatic environments. *Mol. Ecol. Res.* https://doi.org/10.1111/1755-0998.12937 (2018).
140. Panzer, K. et al. Identification of habitat-specific biomes of aquatic fungal communities using a comprehensive nearly full-length 18S rRNA dataset enriched with contextual data. *PLOS ONE* **10**, e0134377 (2015).
141. Richards, T. A. et al. Molecular diversity and distribution of marine fungi across 130 European environmental samples. *Proc. R. Soc. B* **282**, 20152243 (2015).
142. Grossart, H. P., Wurzbacher, C., James, T. Y. & Kagami, M. Discovery of dark matter fungi in aquatic ecosystems demands a reappraisal of the phylogeny and ecology of zoosporic fungi. *Fungal Ecol.* **19**, 28–38 (2016).
143. Wahl, H. E. et al. What lies beneath? Fungal diversity at the bottom of Lake Michigan and Lake Superior. *J. Great Lakes Res.* **44**, 263–270 (2018).
144. Wurzbacher, C. et al. High habitat-specificity in fungal communities in oligo-mesotrophic, temperate Lake Stechlin (North-East Germany). *MycoKeys* **16**, 17–44 (2016).
145. Hassett, B. T., Ducluzeau, A. L. L., Collins, R. E. & Gradinger, R. Spatial distribution of aquatic marine fungi across the western Arctic and sub-arctic. *Environ. Microbiol.* **19**, 475–484 (2017).
146. Bochdansky, A. B., Clouse, M. A. & Herndl, G. J. Eukaryotic microbes, principally fungi and labyrinthulomycetes, dominate biomass on bathypelagic marine snow. *ISME J.* **11**, 362–373 (2017).
147. Khomich, M., Davey, M. L., Kauserud, H., Rasconi, S. & Andersen, T. Fungal communities in Scandinavian lakes along a longitudinal gradient. *Fungal Ecol.* **27**, 36–46 (2017).
148. Taylor, J. D. & Cunliffe, M. Multi-year assessment of coastal planktonic fungi reveals environmental drivers of diversity and abundance. *ISME J.* **10**, 2118–2128 (2016).
149. Jones, M. D. et al. Discovery of novel intermediate forms redefines the fungal tree of life. *Nature* **474**, 200–203 (2011).
150. Wurzbacher, C. et al. Introducing ribosomal tandem repeat barcoding for fungi. *Mol. Ecol. Res.* https://doi.org/10.1111/1755-0998.12944 (2018).
This forward-thinking methods article unlocks the explanatory power residing in HTS-based sequencing of full-length rRNA operons.
151. Casadevall, A. Fungi and the rise of mammals. *PLOS Pathog.* **8**, e1002808 (2012).
152. Irinyi, L. et al. 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 (2015).
153. Halwachs, B. et al. Critical issues in mycobiota analysis. *Front. Microbiol.* **8**, 180 (2017).
154. Kong, H. H. & Morris, A. The emerging importance and challenges of the human mycobiome. *Virulence* **8**, 310–213 (2017).
155. Ward, T. L. et al. Development of the human mycobiota over the first month of life and across body sites. *mSystems* **3**, e00140–17 (2018).
156. Metcalf, J. L. et al. Microbial community assembly and metabolic function during mammalian corpse decomposition. *Science* **351**, 158–162 (2016).
157. Abarenkov, K. et al. Annotating public fungal ITS sequences from the built environment according to the MIxS-built environment standard-a report from a May 23–24, 2016 workshop (Gothenburg, Sweden). *MycoKeys* **16**, 1–15 (2016).
158. Nilsson, R. H. et al. Taxonomic annotation of public fungal ITS sequences from the built

environment — a report from an April 10–11, 2017 workshop (Aberdeen, UK). *MycoKeys* **28**, 65–82 (2018).
159. Liang, H., Yin, L., Zhang, Y., Chang, C. & Zhang, W. Dynamics and diversity of a microbial community during the fermentation of industrialized *Qingcai paocai*, a traditional Chinese fermented vegetable food, as assessed by Illumina MiSeq sequencing, DGGE and qPCR assay. *Ann. Microbiol.* **68**, 111–122 (2018).
160. Callaghan, T. M. et al. *Buwchfawromyces eastonii* gen. nov., sp. nov.: a new anaerobic fungus (Neocallimastigomycota) isolated from buffalo faeces. *MycoKeys* **9**, 11–28 (2015).
161. Seyedmousavi, S. et al. Fungal infections in animals: a patchwork of different situations. *Med. Mycol.* **56**, S165–S187 (2018).
162. Lücking, R. & Hawksworth, D. L. Formal description of sequence-based voucherless Fungi: promises and pitfalls, and how to resolve them. *IMA Fungus* **9**, 143–166 (2018).
This provocative article makes a persuasive case for the formal description of fungal species known only from sequence data.
163. Pautasso, M. Fungal under-representation is (indeed) diminishing in the life sciences. *Fungal Ecol.* **6**, 460–463 (2013).
164. Tedersoo, L. et al. Standardizing metadata and taxonomic identification in metabarcoding studies. *Gigascience* **4**, 34 (2015).
165. Payne, A., Nadine, H., Vardhman, R. & Matthew, L. Whale watching with BulkVis: a graphical viewer for Oxford Nanopore bulk fast5 files. *Preprint at bioRxiv* https://doi.org/10.1101/312256 (2018).
166. Carini, P. et al. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat. Microbiol.* **2**, 16242 (2016).
167. Rajala, T., Peltoniemi, M., Hantula, J., Mäkipää, R. & Pennanen, T. RNA reveals a succession of active fungi during the decay of Norway spruce logs. *Fungal Ecol.* **4**, 437–448 (2011).
168. Anderson, I. C. & Parkin, P. I. Detection of active soil fungi by RT-PCR amplification of precursor rRNA molecules. *J. Microbiol. Methods* **68**, 248–253 (2007).
169. Baldrian, P. et al. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME J.* **6**, 248–258 (2012).
170. Allentoft, M. E. et al. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. Biol. Sci.* **279**, 4724–4733 (2012).
171. Hultman, J. et al. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* **521**, 208–212 (2015).
172. Hannula, S. E. et al. Shifts in rhizosphere fungal community during secondary succession following abandonment from agriculture. *ISME J.* **11**, 2294–2304 (2017).
173. Llanos, A., Francois, J. M. & Parrou, J. L. Tracking the best reference genes for RT-qPCR data normalization in filamentous fungi. *BMC Genomics* **16**, 71 (2013).
174. Stielow, B. et al. One fungus, which genes? Development and assessment of universal primers for potential secondary fungal DNA barcodes. *Persoonia* **35**, 242–263 (2015).
175. Tedersoo, L., Tooming-Klunderud, A. & Anslan, S. PacBio metabarcoding of fungi and other eukaryotes: biases and perspectives. *New Phytol.* **217**, 1370–1385 (2018).
176. Castle, S. C. et al. DNA template dilution impacts amplicon sequencing-based estimates of soil fungal diversity. *Phytobiomes J.* **2**, 100–107 (2018).
177. Gohl, D. M. et al. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat. Biotechnol.* **34**, 942–949 (2016).
178. Schnell, I. B., Bohmann, K. & Gilbert, M. T. P. Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. *Mol. Ecol. Res.* **15**, 1289–1303 (2015).
179. Callahan, B. J., McMurdie, P. J. & Holmes, S. P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* **11**, 2639–2643 (2017).
180. Bálint, M. et al. Millions of reads, thousands of taxa: microbial community structure and associations analyzed via marker genes. *FEMS Microbiol. Rev.* **40**, 686–700 (2016).
181. Nilsson, R. H., Kristiansson, E., Ryberg, M. & Larsson, K. H. Approaching the taxonomic affiliation of unidentified sequences in public databases — an example from the mycorrhizal fungi. *BMC Bioinformatics* **6**, 178 (2005).
182. Solis, N. V. & Filler, S. G. Mouse model of oropharyngeal candidiasis. *Nat. Protoc.* **7**, 637–642 (2012).
183. Afgan, E. et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* **44**, W3–W10 (2016).
184. Boyer, F. et al. OBITOOLS: a UNIX-inspired software package for DNA metabarcoding. *Mol. Ecol. Res.* **16**, 176–182 (2016).
185. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
186. Větrovský, T., Baldrian, P. & Morais, D. SEED 2: a user-friendly platform for amplicon high-throughput sequencing data analyses. *Bioinformatics* **34**, 2292–2294 (2018).
187. Bengtsson-Palme, J., Thorell, K., Wurzbacher, C., Sjöling, Å. & Nilsson, R. H. Metaxa2 diversity tools: easing microbial community analysis with Metaxa2. *Ecol. Inform.* **33**, 45–50 (2016).

Acknowledgements
M. Unterseher is acknowledged for valuable feedback on parts of the manuscript. R.H.N. and the UNITE community acknowledge support from the Alfred P. Sloan Foundation (grant no. G-2015-14062). M.B. received support from the Estonian Research Council (grant no. PUT1317).

Author contributions
All authors researched data for the article, made substantial contributions to discussions of the content, wrote the article and reviewed and edited the manuscript before submission.

Competing interests
The authors declare no competing interests.

Publisher's note

Reviewer information
Nature Reviews Microbiology thanks G. Giardinali, P. Kennedy, H. Toju and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Supplementary information
Supplementary information is available for this paper at https://doi.org/10.1038/s41579-018-0116-y.

RELATED LINKS
BLAST (a BLAST search interface to the sequence data in INSDC): https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch
Dryad Digital Repository (digital archive for scientific data): https://datadryad.org
Genomic Standards Consortium (set of metadata standards (notably MIxS and MIMARKS) for increased reproducibility in metabarcoding): http://gensc.org/mixs
International Nucleotide Sequence Database Collaboration (primary database for molecular data in science): https://www.ncbi.nlm.nih.gov/genbank
RDP (reference database for prokaryotic and eukaryotic SSU and fungal LSU sequences; also provides RDP Classifier training sets (16S rRNA, fungal LSU, Warcup fungal ITS and UNITE fungal ITS): http://rdp.cme.msu.edu
SILVA (reference database for prokaryotic and eukaryotic SSU and LSU sequences): https://www.arb-silva.de
UNITE (reference database for fungal ITS sequences and various HTS pipelines; provides DOI-tagged species hypotheses under the 'Resources' tab): https://unite.ut.ee