

Review

Open Access



Yeast metagenomics: analytical challenges in the analysis of the eukaryotic microbiome

Sonia Renzi[#] , Stefano Nenciarini[#] , Giovanni Bacci , Duccio Cavalieri 

Department of Biology, University of Florence, Sesto Fiorentino 50019, Italy.

[#]Authors contributed equally.

Correspondence to: Prof. Duccio Cavalieri, Department of Biology, University of Florence, Via Madonna del Piano 6, Sesto Fiorentino 50019, Italy. E-mail: duccio.cavalieri@unifi.it

How to cite this article: Renzi S, Nenciarini S, Bacci G, Cavalieri D. Yeast metagenomics: analytical challenges in the analysis of the eukaryotic microbiome. *Microbiome Res Rep* 2024;3:2. <https://dx.doi.org/10.20517/mrr.2023.27>

Received: 14 Apr 2023 **First Decision:** 12 Sep 2023 **Revised:** 9 Oct 2023 **Accepted:** 17 Oct 2023 **Published:** 23 Oct 2023

Academic Editor: Gabriele Andrea Lugli **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Even if their impact is often underestimated, yeasts and yeast-like fungi represent the most prevalent eukaryotic members of microbial communities on Earth. They play numerous roles in natural ecosystems and in association with their hosts. They are involved in the food industry and pharmaceutical production, but they can also cause diseases in other organisms, making the understanding of their biology mandatory. The ongoing loss of biodiversity due to overexploitation of environmental resources is a growing concern in many countries. Therefore, it becomes crucial to understand the ecology and evolutionary history of these organisms to systematically classify them. To achieve this, it is essential that our knowledge of the mycobiota reaches a level similar to that of the bacterial communities. To overcome the existing challenges in the study of fungal communities, the first step should be the establishment of standardized techniques for the correct identification of species, even from complex matrices, both in wet lab practices and in bioinformatic tools.

Keywords: Yeasts, fungi, microbiome, microbial eukaryotes, eukaryome, ngs, metagenomics, taxonomy

INTRODUCTION

In natural microbial systems, including host-associated microbiomes, microbial eukaryotes coexist with bacteria, archaea, and viruses, acting as decomposers, predators, parasites, and producers^[1]. Theoretically,



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



any ecosystem on Earth hosts eukaryotic microorganisms, from extremophiles in geothermal vents to endophytic fungi in plants to parasites or commensals with the gastrointestinal tracts of animals. In host-associated microbiomes, microbial eukaryotes implement complex interactions with their hosts: in plants, they defend the host against herbivorous organisms and enhance nutrients assimilation^[2]; in animals, they can metabolize plant compounds in the host's gastrointestinal systems^[3]. However, both plants and animals can also be afflicted by microbial eukaryotes^[4,5]. In humans, microbial eukaryotes interact with the host immune system in intricate ways. The low diversity in microbiomes from industrialized countries reflects the “extinction” reported for bacterial communities, which is a result of globalization^[6-8]. Beyond host interactions, microbial eukaryotes are essential to the ecology of aquatic and soil ecosystems, where they serve as primary producers, symbiotic partners, decomposers, and predators^[9,10].

Fungi constitute the group of eukaryotes with the highest diversity and global distribution. Thanks to a wide range of morphological, physiological, and ecological features, these organisms have evolved to colonize the most diverse ecosystems^[11]. Within the fungal kingdom, yeasts are not strictly identified, as the term refers to a unicellular lifestyle that has evolved several times rather than a taxonomic unit^[12]. Yeasts and yeast-like fungi are the most prevalent eukaryotic components of the microbiota due to their ubiquity, yet their abundance and influence are frequently underestimated.

Despite their relevance, eukaryotic microorganisms are generally largely neglected in microbiome investigations^[13]. Traditionally, culture-based techniques have been employed to explore and study microbial diversity and to obtain a representative set of isolates based on physicochemical variation. However, due to intrinsic methodological limitations, this approach has been progressively replaced by culture-independent ones, although it has been rediscovered and subjected to various refinements in recent years to enable the capture of a broader spectrum of microorganisms^[14,15]. Following the advent of Sanger sequencing, the use of DNA for the identification of microorganisms has become standard practice, revolutionizing microbial genotyping and taxonomy^[16,17]. The most recent rise of second- and third-generation sequencing approaches has facilitated the advancement of eukaryotic-specific amplicon sequencing, which is revolutionizing our understanding of the eukaryotic diversity in host-associated and environmental microbiomes^[18-22]. Like all amplicon-based techniques, this approach can suffer from poor taxonomic precision and difficulty discriminating between closely related species^[23,24]. In contrast, whole metagenome sequencing captures DNA from the entire pool of species present in a microbiome, including eukaryotes, without the need for experimental selection. Whole metagenome sequencing data are becoming predominant in microbiome research because they can be used to assemble unknown genomes, classify strains, and assess the presence or absence of genes and pathways^[25]. These methods are useful for identifying bacteria and archaea, but microbiome-associated eukaryotes, such as yeasts, are still difficult to detect, especially in large metagenome sequencing datasets. One of the main reasons for this issue is that despite being part of one of the largest branches of the “Tree of Life”, the number of high-quality fungal target sequences or genomes in curated databases is still significantly lower than that of available bacterial ones, severely limiting the possibility of properly investigating these organisms.

The aim of this review is to outline the current state of research regarding the techniques and experimental pipelines for the study of yeast metagenomics, focusing on the currently unresolved methodological challenges as well as the pros and cons of each different approach.

MYCOBIOME: FOCUS ON YEASTS

The term “mycobiome”, coined in 2009^[26] for a study of fungal communities on salt marsh plants using molecular fingerprinting, was then used in 2010 to refer to the human oral mycobiome^[27]. Now, it is used to

indicate the fungal component of every microbial ecosystem. Within the fungal kingdom, the term “yeast” is used to describe any fungus that reproduces asexually by budding or fission, produces single-cell stages, and has sexual structures that are not enclosed in a fruiting body^[28]. This broad description is frequently used to encompass dimorphic lineages that produce mycelial growth in their sexual phases, as well as biotrophic diseases and black yeasts. As a result, they do not constitute a taxonomic unit but rather a lifestyle shared by multiple distinct lineages, even though there are several exceptions and comments to the labile border between yeasts and dimorphic filamentous fungi that produce yeast-like stages, along with yeast lineages that grow solely as filamentous, are outlined^[29].

Yeasts occur in the division Ascomycota, mainly in the subdivisions Saccharomycotina (so-called budding yeasts) and Taphrinomycotina (that also includes so-called fission yeasts), as well as in three subdivisions of Basidiomycota, namely Ustilaginomycotina, Pucciniomycotina, and Agaricomycotina^[30].

These unicellular organisms have become popular in a various applications, including baking, brewing, winemaking, distilling, and an assortment of other conventional and non-conventional fermentations. They also serve as a versatile tool in biotechnology^[31], encompassing some of the most widely used model species (e.g., *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans*). The rapid expansion of scientific understanding of yeast diversity is attributed to the uncovering of new species in nature and the use of specific identification tools like nutritional tests, biochemical and molecular characterizations, and DNA barcode technology. As a result of this technological advancement, previously identified fungal species are continuously reevaluated, and the concept of yeast species itself is evolving^[32].

According to existing estimates, only a small fraction (about 5%-10%, depending on the environment) of the entire variety of fungi has been identified^[33,34]. It is estimated that Earth hosts between 2.2 and 3.8 million fungal species^[35], yet only about 4% of these are cataloged^[36]. This situation likely holds true for yeast as well. Out of the approximately 150,000 fungal species described so far^[37], only around 2,000 are yeasts. The mycobiome is often neglected, both due to its lower abundance compared to bacteria and the methodological challenges associated with its detection^[38].

The high incidence of cryptic and hybrid species hampers efforts to accurately quantify species diversity. These issues have long been acknowledged, but the advent of whole-genome sequencing has brought them to the forefront^[39]. In fact, when speaking about genomes, fungi exhibit more complex genetic features compared to bacteria, including multiple chromosomes, expanded repeated regions, and larger genome sizes, all of which introduce inaccuracies during sequence classification. Therefore, there is a need for comprehensive benchmarking of both classification algorithms and databases to optimize identification pipelines for the fungal kingdom.

CHARACTERISING THE MYCOBIOME: IDENTIFICATION AND TECHNOLOGICAL ISSUES

As mentioned above, many questions regarding mycobiota remain to be addressed. Several methodologies commonly applied for the investigation of the bacteriome are not consistent when used for studying the fungal community. Consequently, non-standardized techniques, technical challenges, restricted availability of reference data, and other issues have emerged^[40]. Therefore, it is crucial to enhance our knowledge and expand the spectrum of available technologies in order to address the challenges posed by the fungal communities inhabiting the environmental ecosystem and our bodies.

Culture-dependent approaches

Traditionally, culture-dependent approaches have been employed to investigate microorganisms' diversity, including fungi. However, these techniques have well-known limitations. For instance, many species remain undetected because appropriate culture conditions are either unknown or challenging to reproduce^[41]. Moreover, culture methods are time-consuming and hardly suitable for high-throughput analysis. Culturomic approaches offer undeniable benefit as they provide access to the fungus itself, allowing for the assessment of its viability, metabolites, phenotypical and functional characterization, and other host-adaptation features^[42]. In recent years, the integration between culture-dependent and culture-independent approaches has increased, thanks to molecular techniques. Sequencing of large portions or entire microbial genomes has provided the necessary information for fine-tuning the growth conditions of even those microorganisms considered “unculturable” until a few years ago^[43,44]. As a result, culture-dependent approaches remain useful and of great interest^[45,46]. This is especially important given that some fungal strains cannot be accurately identified by a culture-independent method. This underrepresentation of some species might result from factors such as cell wall structure or the inadequacy of the chosen PCR primers and/or barcode sequence^[47]. However, the identification process for isolated fungal strains is not yet complete and requires further steps, often involving culture-independent approaches.

Culture-independent approaches

The use of DNA as an identifying marker in culture-independent approaches avoids some of the aforementioned issues. However, this method strongly relies on the choice and efficiency of DNA recovery methods, and it also introduces new limits and hurdles [Figure 1]. Fungi, unlike bacteria, have a strong and complex cell wall rich in glucans and chitin^[48-51]. Consequently, the efficient destruction of the fungal cell wall is crucial for genomic DNA extraction. Several bead-beating stages followed by enzymatic cell lysis are required for successful mycobiota analysis of any sample matrix^[47]. Following DNA extraction, different approaches can be used to detect and identify fungi. These methods may include PCR^[52], metabarcoding sequencing analysis, or whole genome sequencing (WGS) metagenomics.

Amplicon-based sequencing: a matter of target

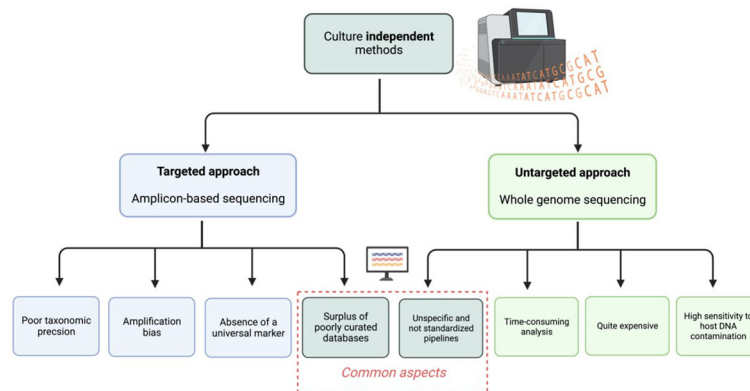
While amplicon sequencing techniques have successfully revealed the microbiome of a plethora of organisms^[53-56], the choice of the marker to use is crucial as it drastically affects the type of organisms that can be detected. In the micro-eukaryotic world, mainly composed of fungi, protists, algae, and other microorganisms known to inhabit almost all ecological niches explored on Earth, the selection of “universal” targets is limited [Table 1]. Only a few available pipelines are available to cope with markers different from the well-known bacterial 16S rRNA gene^[57-60].

Similarly to bacterial metabarcoding, the usual fungal barcode is the rRNA gene locus, which includes the genes for 18S rRNA, 5.8S rRNA, and 28S rRNA, separated by the internal transcribed spacers (ITS1 and ITS2). This approach seems to discriminate better at higher taxonomic ranks than the 16S rRNA gene^[61]. After exploring fungal rRNA genes, Schoch *et al.* in 2012 identified the ITS as the possible universal DNA barcode identifier for fungi^[24,62], although currently, it is still not clear which of the two ITS components has the better resolution in strain prediction. Recent findings have shown that both regions suffer from amplification biases, resulting in an uneven representation of synthetic fungal communities^[63-65]: ITS1-based PCR appears to favor Basidiomycota, whereas Ascomycota seems to be favored by ITS2-based PCR^[66-68], although this consideration should not be generalized. In fact, there are known ascomycetes species (such as the ones belonging to the genera *Saccharomyces* and *Komagataella*) that are discriminated with greater resolution by employing the ITS1 marker^[69]. Hoggard *et al.* recommend the selection of the ITS2 region in human mycobiota investigation after comparing four sets of primers targeting the small subunit (SSU) rRNA (18S), ITS1, ITS2, and large subunit (LSU) rRNA (26S) genomic regions^[70]. In yeast, the D1/D2

Table 1. List of barcode loci for fungal taxonomic identification

Genomic locus	Proposed by	Ref.
ITS1-4 (whole region)	Schoch <i>et al.</i>	[24]
ITS1 as preferred for Basidiomycota identification	Bellemain <i>et al.</i>	[68]
ITS2 as preferred for Ascomycota identification	Bellemain <i>et al.</i>	[68]
ITS2 as preferred for human mycobiota identification	Hoggard <i>et al.</i>	[70]
ITS2 subregion	Nilsson <i>et al.</i>	[62]
TEF1 α	James <i>et al.</i>	[77]
TOP1	Stielow <i>et al.</i>	[76]
PGK	Stielow <i>et al.</i>	[76]
RPB1	Matheny <i>et al.</i>	[78]
RPB2 for environmental fungal communities	Větrovský <i>et al.</i>	[80]
IGS	Morrison <i>et al.</i>	[81]
β -tubulin	Geiser <i>et al.</i>	[82]
LSU (D1/D2 region)	Kurtzman and Robnett	[72]

IGS: Intergenic spacer; LSU: large subunit; PGK: phosphoglycerate kinase; TOP1: topoisomerase I.

**Figure 1.** Schematic representation of current limitations in culture-independent methods.

region of the LSU gene cluster within the ribosomal DNA (rDNA) has been a longstanding and effective tool for species identification and strain differentiation, pre-dating the conceptualization of DNA barcoding^[71,72]. In addition, Nilsson *et al.* propose a set of fungus-specific primers with superior coverage of the fungal kingdom, targeting the ITS2 sub-region with degenerate forward primers gITS7ngs and a reverse primer ITS4ng^[61]. Besides primers' choice, length variation among ITS sequences from fungal species, spanning from 200 to 800 bp, has a strong impact on PCR efficiency as well as sequencing technologies^[73,74]. Moreover, not only is this region present in multiple copies within one species^[75], but intragenomic variation within a single species, such as numerous paralogous or non-orthologous copies, may lead to an overestimation of global fungal diversity^[24]. Since the ITS copy number has highly interspecific variation, an accurate determination of fungal abundance is hard to reach, and quantitative comparisons between diverse species in mixed populations must be made with caution. The lack of universal taxonomic resolution and the potential presence of non-homologous ITS copies in the genome made the identification of supplementary molecular markers necessary. Using *in silico* pipelines, Stielow *et al.*^[76] confirmed the already known TEF1 α ^[77] as a secondary barcoding marker for fungi and proposed the genes topoisomerase I (TOP1) and phosphoglycerate kinase (PGK) as promising ascomycetes identifiers based on the analysis of complete sequenced genomes^[76]. Other suggested secondary markers for fungal DNA amplification are the

intergenic spacer (IGS), RNA polymerase II (RPB1 and RPB2), β -tubulin II (TUB2), and the minichromosome maintenance complex component 7 (MCM7) protein^[78-82]. The selection of one or more reference genes is crucial for standardization and promotion of large-scale investigations, but in some cases, primer bias in targeted sequencing can be overcome by opting for the shotgun metagenomic approach.

Metagenomic whole genome sequencing

Shotgun metagenomic sequencing allows for a higher taxonomic resolution as it sequences most of the genomes of every organism present within a sample^[83]. This capability not only identifies the organism but also characterizes extended profiles, including antimicrobial resistance, genetic subtypes, metabolism, and virulence^[84]. Despite being a highly effective method for describing pathways and discovering novel functions, shotgun metagenomics is significantly more expensive and computationally more intensive than amplicon sequencing, depending on sequencing depth^[85].

Moreover, due to its non-specificity, WGS is the most unbiased technique but also the most sensitive to host DNA contamination, especially in soft tissues and biological fluid samples where host DNA can dominate the sequenced reads^[86]. This sensitivity is a significant concern for the study of mycobiota since fungi represent only a small fraction of the total microbial biomass. Achieving adequate sequencing depth is required to perform the analysis. Currently, it appears that low fungal abundance in human samples is impeding the broad use of metagenomic WGS in human samples, a finding that is unrelated to DNA extraction techniques and reflects really low total in vivo fungal abundance^[87].

The development of high-throughput sequencing techniques has greatly benefited our understanding of microbial ecology. Nevertheless, the most common methods currently in use, which produce short reads, often suffer from limited species-level resolution and identification uncertainty. Fortunately, recent developments in long-read sequencing technologies by PacBio and Oxford Nanopore are enabling the reconstruction of more complete fungal genomes. These long reads, often exceeding 10 kb in length, can cover critical genomic regions, including highly repetitive ones^[84,88-91].

Using long-read sequencers, researchers have successfully generated whole genomes of major pathogenic fungi, often in combination with short-read sequencing, a technique known as hybrid assemblies^[92-99].

Bioinformatics

In metagenomics and metabarcoding analyses, data interpretation is a significant challenge. While these approaches enhance the objectivity of fungal phylogeny and subsequent accurate identification, they simultaneously generate ever-growing amounts of sequencing data. Addressing the prompt delivery of the enormous amount of sequence data available to end user introduces a new challenge.

Databases: need for unification

Thanks to advancements in computational technology and bioinformatics tools, large volumes of data can now be easily stored, annotated, and accessed remotely with relative ease. As a result, a surplus of nucleotide sequence databases for fungal studies was created^[23]. The strategic value of a database is based on its accessibility, through which end users may deposit, save, annotate, and retrieve data. It must be considered that every database has an intrinsic proclivity to become outdated over time. To maintain useful and relevant databases for diagnostics and research, a dedicated group of trained professionals is required to carry out an ongoing and systematic curation. Over the last decade, many online fungal databases have been established for the mycology research community. However, not all of them have a dedicated team of curators or an updated maintenance system. Some of the most widely used repositories [Table 2], such as

Table 2. Principal genomic databases described according to their ability to discriminate fungal sequences

Database	Ref.	Description	Taxonomic discriminative potential
AspGD	[100]	AspGD focuses on the genomes of <i>Aspergillus</i> species. It provides detailed genomic data, including gene annotations, functional information, and comparative genomics. Enables the identification of both the species and strain levels.	High within the genus <i>Aspergillus</i>
BOLD	[101]	BOLD is a comprehensive online platform primarily dedicated to DNA barcoding and biodiversity research. While it is a valuable resource, its primary focus is on animal barcoding. As a result, its fungal taxonomic discriminative potential is limited compared to databases specifically tailored for fungi.	Limited
Broad Insite Database	http://www.broadinstitute.org/scientific-community/data/	The Broad Institute has contributed extensively to fungal genomics. It offers genomic data for a variety of fungal species, with an emphasis on pathogenic fungi.	High
CGD	[102]	CGD is dedicated to <i>Candida</i> species, and it offers genomic sequences, gene annotations, and pathogenicity-related information, supporting research on the genus <i>Candida</i> .	High within the genus <i>Candida</i>
CYGD	[103]	CYGD offers comprehensive genome annotation and functional data primarily for <i>Saccharomyces cerevisiae</i> . While it provides essential information for yeast research, its taxonomic scope is restricted to this species.	Limited to species <i>S. cerevisiae</i>
Ensembl Fungi	https://fungi.ensembl.org	Ensembl Fungi is a component of the Ensembl project, offering genomic data and tools for various fungal species. While it covers a range of fungi, it may be more comprehensive for some taxa than others.	Moderate
FungiDB	[104]	FungiDB is a genomic database focused on fungal pathogens. It includes a diverse set of fungal genomes, with an emphasis on medically important species.	Moderate
FUNGIpath	[105]	FUNGIpath is a resource for fungal pathogen genomics. It provides genomic sequences and annotations for pathogenic fungi, with relevance to disease research.	Moderate
Fusarium-ID	[106]	Fusarium-ID is a specialized database for <i>Fusarium</i> species identification and classification. It provides detailed molecular and phenotypic data for various <i>Fusarium</i> species, including pathogenic strains.	High within the genus <i>Fusarium</i>
Fusarium MLST	[107]	Fusarium MLST is a database that focuses on sequence-based typing for <i>Fusarium</i> species. It allows researchers to differentiate between closely related <i>Fusarium</i> isolates by analyzing multiple gene loci. This database is particularly useful for studying genetic diversity within the genus.	High within the genus <i>Fusarium</i>
ISHAM-ITS	[108]	ISHAM-ITS database is designed to aid in the identification and classification of medically important fungi using the fungal Internal Transcribed Spacer (ITS) region of ribosomal DNA. Its taxonomic discriminative potential is high within the context of identifying and characterizing fungi relevant to human and animal health.	High within the medical mycology
ISHAM-MLST	http://mlst.mycologylab.org/	ISHAM-MLST is dedicated to the Multilocus Sequence Typing of medically important fungi, particularly those associated with human and animal mycoses. It has a higher taxonomic discriminative potential for distinguishing between closely related strains within a species.	Very high within the medical mycology
JGI MycoCosm	[109]	MycoCosm, hosted by the JGI, provides access to a diverse collection of fungal genomes, including those from various taxonomic groups, making it suitable for discriminative research.	High
NCBI GenBank	https://www.ncbi.nlm.nih.gov/genbank/	NCBI GenBank is a comprehensive and widely used repository for genomic data. It covers a wide taxonomic range, including fungi, but the level of detail and annotation quality can vary.	Moderate to high
NCBI RefSeq	http://www.ncbi.nlm.nih.gov/refseq/	NCBI RefSeq offers high-quality genomic annotations and reference sequences, making it a preferred choice for researchers seeking accurate taxonomic and functional information for well-studied fungal species.	High
PomBase	[110]	PomBase is primarily focused on <i>Schizosaccharomyces pombe</i> . It provides detailed genomic and functional information for this species, making it an excellent resource for <i>S. pombe</i> research. However, its taxonomic scope is limited to this species.	Limited to species <i>S. pombe</i>
SGD	[111]	SGD is dedicated to <i>Saccharomyces cerevisiae</i> and is a comprehensive resource. While its primary focus is <i>S. cerevisiae</i> , it contains extensive genomic and functional data that can support the study of other <i>Saccharomyces</i> species as well.	High within the genus <i>Saccharomyces</i>
UNITE	[112]	UNITE provides a comprehensive collection of fungal ITS sequences, covering a broad range of fungal taxa, from common and well-studied species to rare and less-known fungi.	High

AspGD: Aspergillus genome database; BOLD: barcode of life data systems; CGD: candida genome database; CYGD: comprehensive yeast genome database; ISHAM-ITS: international society for human and animal mycology-internal transcribed spacer; ISHAM-MLST: international society for human and animal mycology - multilocus sequence typing; JGI: joint genome institute; MLST: multilocus sequence typing; SGD: saccharomyces genome database.

Aspergillus Genome Database (AspGD)^[100], Barcode of Life Data Systems (BOLD)^[101], Broad Institute databases (<http://www.broadinstitute.org/scientific-community/data/>), Candida Genome Database (CGD)^[102], Comprehensive Yeast Genome Database (CYGD)^[103], Ensembl Fungi (<https://fungi.ensembl.org>), FungiDB^[104], FUNGIpath^[105], Fusarium-ID^[106], Fusarium Multilocus Sequence Typing (MLST)^[107], International Society for Human and Animal Mycology-Internal Transcribed Spacer (ISHAM-ITS)^[108], International Society for Human and Animal Mycology - MultiLocus Sequence Typing (ISHAM-MLST) (<http://mlst.mycologylab.org/>), JGI MycoCosm^[109], NCBI GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), NCBI RefSeq (<http://www.ncbi.nlm.nih.gov/refseq/>), PomBase^[110], Saccharomyces Genome Database (SGD)^[111], and UNITE^[112] have been resumed and extensively classified by Prakash *et al.*^[113]. To avoid the hampering issues of comprehensive data management, they suggest a cloud-based, dynamic network platform based on the integration of particular focused-group databases with maximum access and functional characteristics for the user community.

One of the most concerning analytic challenges in mycobiota investigations is the inadequate curation of fungal databases. This deficiency in high-quality fungal sequences within curated databases results in a substantial number of unclassified reads. Addressing this issue may involve producing additional high-quality metagenomic and whole-fungal genome assemblies^[87]. Furthermore, sequencing data are frequently devoid of any biologically relevant information, such as the substrate of origin or details on the technology used. Thus, well-curated fungal databases with accurate sequence data play a pivotal role in further research and diagnostics in the field of mycology. The current fungal databases only poorly represent the diversity of the fungal kingdom, limiting their analytical power.

Pipelines

The bioinformatics analysis workflow for amplicon data can be summarized into four main steps: (i) pre-processing; (ii) “grouping” of amplicon sequences; (iii) taxonomic classification; and (iv) visualization and statistical analysis^[114]. While various tools can be used in each of these steps, producing slightly different results, the second step, in particular, is crucial. Amplicon sequences can be clustered based on their similarity^[115-119], akin to classical clustering techniques such as k-mean clustering or agglomerative clustering - or based on single nucleotide differences across them, an approach currently known as sequence variant inference^[60]. Methods falling into the first category profile bacterial communities by grouping similar sequences into Operational Taxonomic Units (OTUs), but the definition of a similarity threshold has always been empirical. As a consequence, these methods tend to produce a large number of OTUs that are not always biologically relevant, an issue that goes by the name of “OTU inflation”^[120]. This massive production of OTUs may lead to wrong conclusions and/or to the generation of huge datasets, which can be difficult to analyze. Tackling this issue is not trivial, and a series of novel approaches have been proposed. These approaches rely on the definition of sequence variants from single nucleotide differences in the amplicon reconstruction, trying to profile microbial communities based on “real” differences instead of sequence similarity. Nowadays, the research communities are gradually moving to the new concept of Amplicons Sequence Variants (ASVs) or Exact Sequences Variants (ESVs)^[121] for profiling bacterial communities, and it should also be recommended for yeasts and yeast-like organisms. These approaches generate an error model for each sequencing run, which enables discriminating between a true sequence variant (i.e., one

sequence with a single SNP with respect to another) from sequencing errors^[60]. Since these processes rely on the single nucleotide variation of amplicons for defining taxonomy, they usually lead to an increased estimation of alpha diversity, mainly due to their higher sensitivity with respect to identity-based approaches. One of the greatest assumptions of these methods is that the amplicon sequence should not vary in length, and ITS sequences from fungi do not share this assumption. This may lead to biases in the discriminatory potential of these methods, even if, at present, no extensive survey has been performed^[122]. To reduce these biases, a number of ITS sequencing-based systems have been created to identify different fungal species. Some of them are able to examine both 16S rRNA (from bacteria) and ITS (from fungi), such as Kraken^[123], Mothur^[115], Qiime^[119,124], Vsearch^[117], and DADA2^[60]; others are specialized on fungi only, such as Plutof^[125], Clotu^[126], PIPITS^[116], CloVR-ITS^[127], MICCA^[128], and BioMaS^[129]. Despite these well-known issues, standardized pipelines are still to come, leaving the choice of the analysis method in the hands of researchers. This situation opens a whole new scenario where researchers are responsible for the pipeline they used (which, in most cases, is published and freely available), and this choice may alter the research outcomes^[130], paving the way for contrasting conclusions. Although pipelines based on the bacterial 16S gene (or part of it) have been extensively used in the last three decades, the “yeast world” remains largely unexplored, and the effect of one pipeline compared to another is unpredictable. A summary of the main pipelines available is reported in [Table 3](#)^[60,115-117,119,123,124,126-129,131,132].

In the context of metagenomic WGS, two primary strategies are commonly employed to analyze raw data: the alignment-based approach and the assembly-based approach. The first one involves mapping individual sequencing reads to a reference database or a reference genome. On the other hand, the second approach assembles reads de novo to form contigs, which are then clustered into so-called genome bins during a binning process. Combining both approaches is frequently advocated for result accuracy^[84]. By now, many bioinformatic tools are available. Alignment-based tools are strong in taxonomic profiling and identifying known microorganisms. They include a step of fragment recruitment in order to map all the reads to one or more selected references. Among taxonomic profilers, MetaPlhAn2^[133], Kraken2^[134], and DIAMOND^[135] stand out for different skills. If you need high specificity and rapid analysis, MetaPhlAn2 might be a good choice. For comprehensive database coverage and strain-level resolution, Kraken 2 is valuable. DIAMOND allows customization and offers fast alignment capabilities, but it requires additional steps for taxonomic profiling. Assembly-based tools, instead, are essential for discovering novel organisms and in-depth functional analysis within metagenomic communities. Their workflow includes an assembler^[136] that is well suited for the reconstruction of long contigs and a genome binner to cluster such sequences from the same organism^[137]. When selecting an assembler for WGS data, the type of sequencing technology used, the genome size, the desired level of assembly completeness, and the availability of computational resources should be taken into consideration. MetaSPAdes^[138], MegaHit^[139], and IDBA-UD^[140] are the most popular metagenome assemblers, also for fungal genomes. As well as for assemblers, there is no binning tool designed exclusively for fungal sequences, so general metagenomic bidders are being used, like METABAT2^[141], CONCOCT^[142], MaxBin 2.0^[143] and MetaWrap^[144] to name a few of the most efficient. Many researchers also employ hybrid assembly strategies that combine short-read and long-read data to achieve more accurate and complete genome assemblies^[95]. To delve deeper into the metagenomic data beyond taxonomic composition, functional annotation becomes necessary. Fragment recruitment, as previously described, involves leveraging a database of functionally annotated genes or proteins. This approach provides a straightforward means to achieve functional annotation. Subsequently, annotations showing a specific level of coverage can be linked to various aspects, such as metabolic pathways, with tools like KEGG^[145]. Metagenomic WGS of fungi offers valuable insights into complex fungal communities, but it also comes with several drawbacks and challenges. Bioinformatic complexity, functional annotation, short-read sequencing, not standardized pipelines, data volume and processing are probably the main ones. Addressing these drawbacks often requires a combination of improved sequencing technologies, more comprehensive

Table 3. List of currently available pipelines for meta-barcoding

Name	Clustering algorithm	Yeast-specific	Ref.
Clotu	Identity-based clustering	Yes	[126]
PIPITS	Identity-based clustering	Yes	[116]
CloVR-ITS	Identity-based clustering	Yes	[127]
BioMaS	Reference-based	No	[129]
Kraken	Reference-based	No	[123]
Mothur	Mixed	No	[115]
Qiime (1 & 2)	Mixed	No	[119,124]
MICCA	Mixed	No	[128]
Vsearch	Identity-based	No	[117]
Uparse	Identity-based	No	[131]
Unoise (1 & 2)	Variant-based	No	[132]
DADA2	Variant-based	No	[60]

Clustering algorithms were divided into: (1) Identity-based, those relying on an empirical percentage of identity between two sequences for grouping them into a single cluster; (2) Reference-based, algorithms which group sequences into taxonomic bins according to their identities; (3) Variant-base, those defining sequence variants according to the presence of SNPs or mutations; (4) Mixed, pipelines which contain different algorithms for clustering.

reference databases, advances in bioinformatics methods, and careful experimental design to mitigate potential biases and methodological limitations.

CONCLUSION

In conclusion, fungi play a pivotal role in shaping diverse ecosystems, and while our understanding of their importance has grown considerably, there remain numerous avenues for exploration within the fungal kingdom. The advent of DNA-based classification methods has ushered in a transformative era in mycology, revolutionizing traditional taxonomic approaches while also providing robust validation of species identities. Despite significant progress, challenges persist in the field of fungal genomics. Sequencing techniques have revealed biases and limitations, particularly in fungal markers amplification. Recent innovations like long-range amplification and long-read sequencing hold promise for more accurate fungal classifications. The increasing availability of whole-genome shotgun sequencing and expanding genome databases offer opportunities to map newly generated fungal DNA sequences directly to comprehensive references.

Advancements in sequencing technologies are complemented by the development of taxonomic classification algorithms, but critical gaps remain. Benchmarking long-read sequencing strategies for fungal communities lags behind bacterial community studies. Similar disparities exist in the relative maturity of bioinformatic platforms and databases. Fungi's unique complexities, such as multiple chromosomes, extended repeat regions, and larger genome sizes, add to the challenges.

The intricacies of fungal taxonomy further complicate identification efforts. The absence of standardized pipelines for sequencing data analysis remains a significant hurdle in mycobiota investigations. Given these challenges and opportunities, it's evident that fungal research continues to rapidly evolve. Future progress will hinge on collaborative efforts to address existing gaps, harmonize methodologies, and advance our understanding of these essential and enigmatic organisms in the intricate network of global ecosystems.

DECLARATIONS

Authors' contributions

Conceived the work: Cavalieri D

Gathered data from the literature: Renzi S, Nenciarini S

Wrote and revised the manuscript: Cavalieri D, Renzi S, Nenciarini S, Bacci G

All authors contributed to the article and approved the submitted version.

Availability of data and materials

Not applicable.

Financial support and sponsorship

This work was supported by (1) FishEU Trust project Horizon-CL6-2021-FARM2FORK01 (grant n. 101060712); (2) FNS-Cloud WP3: Standardization (<https://www.fns-cloud.eu/>), which has received funding from the European Union's Horizon 2020 Research and Innovation program (H2020-EU.3.2.2.3.-A sustainable and competitive agri-food industry) under Grant Agreement No. 863059; and (3) Bando Salute 2018 RISKCROHNBIOM project (grant number G84I18000160002), by the Italian Ministry of Agriculture, Food, and Forestry Policies (MiPAAF), within the trans-national project INTIMIC-Knowledge Platform on food, diet, intestinal microbiomics.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2023.

REFERENCES

1. Bik HM, Porazinska DL, Creer S, Caporaso JG, Knight R, Thomas WK. Sequencing our way towards understanding global eukaryotic biodiversity. *Trends Ecol Evol* 2012;27:233-43. [DOI](#) [PubMed](#) [PMC](#)
2. Rodriguez RJ, White JF Jr, Arnold AE, Redman RS. Fungal endophytes: diversity and functional roles. *New Phytol* 2009;182:314-30. [DOI](#) [PubMed](#)
3. Akin DE, Borneman WS. Role of rumen fungi in fiber degradation. *J Dairy Sci* 1990;73:3023-32. [DOI](#) [PubMed](#)
4. Kamoun S, Furzer O, Jones JD, et al. The Top 10 oomycete pathogens in molecular plant pathology. *Mol Plant Pathol* 2015;16:413-34. [DOI](#) [PubMed](#) [PMC](#)
5. Haque R. Human intestinal parasites. *J Health Popul Nutr* 2007;25:387-91. [PubMed](#) [PMC](#)
6. Laforest-Lapointe I, Arrieta MC. Microbial eukaryotes: a missing link in gut microbiome studies. *mSystems* 2018;3:e00201-17. [DOI](#) [PubMed](#) [PMC](#)
7. Parfrey LW, Walters WA, Lauber CL, et al. Communities of microbial eukaryotes in the mammalian gut within the context of environmental eukaryotic diversity. *Front Microbiol* 2014;5:298. [DOI](#) [PubMed](#) [PMC](#)
8. Sonnenburg ED, Smits SA, Tikhonov M, Higginbottom SK, Wingreen NS, Sonnenburg JL. Diet-induced extinctions in the gut microbiota compound over generations. *Nature* 2016;529:212-5. [DOI](#) [PubMed](#) [PMC](#)
9. Caron DA, Alexander H, Allen AE, et al. Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nat Rev Microbiol* 2017;15:6-20. [DOI](#)
10. Brussaard L, de Ruyter PC, Brown GG. Soil biodiversity for agricultural sustainability. *Agr Ecosyst Environ* 2007;121:233-44. [DOI](#)
11. James TY, Stajich JE, Hittinger CT, Rokas A. Toward a fully resolved fungal tree of life. *Annu Rev Microbiol* 2020;74:291-313. [DOI](#)
12. Shen XX, Oplente DA, Kominek J, et al. Tempo and mode of genome evolution in the budding yeast subphylum. *Cell*

- 2018;175:1533-45.e20. [DOI](#) [PubMed](#) [PMC](#)
13. Hernández-Santos N, Klein BS. Through the scope darkly: the gut mycobiome comes into focus. *Cell Host Microbe* 2017;22:728-9. [DOI](#) [PubMed](#) [PMC](#)
 14. Alou M, Naud S, Khelaifia S, Bonnet M, Lagier JC, Raoult D. State of the art in the culture of the human microbiota: new interests and strategies. *Clin Microbiol Rev* 2020;34:e00129-19. [DOI](#) [PubMed](#) [PMC](#)
 15. Vu D, Groenewald M, Szöke S, et al. DNA barcoding analysis of more than 9000 yeast isolates contributes to quantitative thresholds for yeast species and genera delimitation. *Stud Mycol* 2016;85:91-105. [DOI](#) [PubMed](#) [PMC](#)
 16. Makimura K. Species identification system for dermatophytes based on the DNA sequences of nuclear ribosomal internal transcribed spacer 1. *Nihon Ishinkin Gakkai Zasshi* 2001;42:61-7. [DOI](#) [PubMed](#)
 17. Leaw SN, Chang HC, Sun HF, Barton R, Bouchara JP, Chang TC. Identification of medically important yeast species by sequence analysis of the internal transcribed spacer regions. *J Clin Microbiol* 2006;44:693-9. [DOI](#) [PubMed](#) [PMC](#)
 18. Del Campo J, Pons MJ, Herranz M, et al. Validation of a universal set of primers to study animal-associated microeukaryotic communities. *Environ Microbiol* 2019;21:3855-61. [DOI](#) [PubMed](#)
 19. del Campo J, Bass D, Keeling PJ, Bennett A. The eukaryome: diversity and role of microeukaryotic organisms associated with animal hosts. *Functional Ecology* 2020;34:2045-54. [DOI](#)
 20. Parfrey LW, Walters WA, Knight R. Microbial eukaryotes in the human microbiome: ecology, evolution, and future directions. *Front Microbiol* 2011;2:153. [DOI](#) [PubMed](#) [PMC](#)
 21. Andersen LO, Vedel Nielsen H, Stensvold CR. Waiting for the human intestinal Eukaryotome. *ISME J* 2013;7:1253-5. [DOI](#) [PubMed](#) [PMC](#)
 22. Franco-Duarte R, Mendes I, Gomes AC, Santos MA, de Sousa B, Schuller D. Genotyping of *Saccharomyces cerevisiae* strains by interdelta sequence typing using automated microfluidics. *Electrophoresis* 2011;32:1447-55. [DOI](#) [PubMed](#)
 23. Lücking R, Aime MC, Robbertse B, et al. Unambiguous identification of fungi: where do we stand and how accurate and precise is fungal DNA barcoding? *IMA Fungus* 2020;11:14. [DOI](#) [PubMed](#) [PMC](#)
 24. Schoch CL, Seifert KA, Huhndorf S, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc Natl Acad Sci U S A* 2012;109:6241-6. [DOI](#) [PubMed](#) [PMC](#)
 25. Knight R, Vrbnac A, Taylor BC, et al. Best practices for analysing microbiomes. *Nat Rev Microbiol* 2018;16:410-22. [DOI](#)
 26. Gillevet PM, Sikaroodi M, Torzilli AP. Analyzing salt-marsh fungal diversity: comparing ARISA fingerprinting with clone sequencing and pyrosequencing. *Fungal Ecology* 2009;2:160-7. [DOI](#)
 27. Ghannoum MA, Jurevic RJ, Mukherjee PK, et al. Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLoS Pathog* 2010;6:e1000713. [DOI](#) [PubMed](#) [PMC](#)
 28. Kurtzman CP, Sugiyama J. 1 Saccharomycotina and taphrinomycotina: the yeasts and yeastlike fungi of the ascomycota. In: Mclaughlin DJ, Spatafora JW, editors. Systematics and Evolution. Berlin: Springer Berlin Heidelberg; 2015. p. 3-33. [DOI](#)
 29. Kurtzman CP, Fell JW, Boekhout T. Chapter 1 - Definition, classification and nomenclature of the yeasts. In: The Yeasts. Elsevier; 2011. p. 3-5. [DOI](#)
 30. Li Y, Steenwyk JL, Chang Y, et al. A genome-scale phylogeny of the kingdom fungi. *Curr Biol* 2021;31:1653-65.e5. [DOI](#) [PubMed](#) [PMC](#)
 31. Żymaniak-duda E, Brzezińska-rodak M, Klimek-ochab M, Duda M, Zerka A. Yeast as a versatile tool in biotechnology. In: Morata A, Loira I, editors. Yeast - Industrial Applications. InTech; 2017. [DOI](#)
 32. Boekhout T, Aime MC, Begerow D, et al. The evolving species concepts used for yeasts: from phenotypes and genomes to speciation networks. *Fungal Divers* 2021;109:27-55. [DOI](#) [PubMed](#) [PMC](#)
 33. Hawksworth DL. The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycol Res* 2001;105:1422-32. [DOI](#)
 34. Blackwell M. The fungi: 1, 2, 3 ... 5.1 million species? *Am J Bot* 2011;98:426-38. [DOI](#) [PubMed](#)
 35. Hawksworth DL, Lücking R. Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiol Spectr* 2017;5. [DOI](#) [PubMed](#)
 36. Cheek M, Nic Lughadha E, Kirk P, et al. New scientific discoveries: plants and fungi. *Plants People Planet* 2020;2:371-88. [DOI](#)
 37. Lücking R, Aime MC, Robbertse B, et al. Fungal taxonomy and sequence-based nomenclature. *Nat Microbiol* 2021;6:540-8. [DOI](#) [PubMed](#) [PMC](#)
 38. Huseyin CE, O'Toole PW, Cotter PD, Scanlan PD. Forgotten fungi-the gut mycobiome in human health and disease. *FEMS Microbiol Rev* 2017;41:479-511. [DOI](#) [PubMed](#)
 39. Naranjo-Ortiz MA, Gabaldón T. Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. *Biol Rev Camb Philos Soc* 2019;94:2101-37. [DOI](#) [PubMed](#) [PMC](#)
 40. Suhr MJ, Hallen-Adams HE. The human gut mycobiome: pitfalls and potentials - a mycologist's perspective. *Mycologia* 2015;107:1057-73. [PubMed](#)
 41. Hinsu A, Dumadiya A, Joshi A, et al. To culture or not to culture: a snapshot of culture-dependent and culture-independent bacterial diversity from peanut rhizosphere. *PeerJ* 2021;9:e12035. [DOI](#) [PubMed](#) [PMC](#)
 42. Strati F, Di Paola M, Stefanini I, et al. Age and gender affect the composition of fungal population of the human gastrointestinal tract. *Front Microbiol* 2016;7:1227. [DOI](#) [PubMed](#) [PMC](#)
 43. Browne HP, Forster SC, Anonye BO, et al. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* 2016;533:543-6. [DOI](#) [PubMed](#) [PMC](#)
 44. Gutleben J, Chaib De Mares M, van Elsas JD, Smidt H, Overmann J, Sipkema D. The multi-omics promise in context: from sequence

- to microbial isolate. *Crit Rev Microbiol* 2018;44:212-29. DOI PubMed
45. Borges FM, de Paula TO, Sarmiento MRA, et al. Fungal diversity of human gut microbiota among eutrophic, overweight, and obese individuals based on aerobic culture-dependent approach. *Curr Microbiol* 2018;75:726-35. DOI
 46. Hamad I, Ranque S, Azhar EI, et al. Culturomics and amplicon-based metagenomic approaches for the study of fungal population in human gut microbiota. *Sci Rep* 2017;7:16788. DOI PubMed PMC
 47. Huseyin CE, Rubio RC, O'Sullivan O, Cotter PD, Scanlan PD. The fungal frontier: a comparative analysis of methods used in the study of the human gut mycobiome. *Front Microbiol* 2017;8:1432. DOI PubMed PMC
 48. Aïmanianda V, Clavaud C, Simenel C, Fontaine T, Delepierre M, Latgé JP. Cell wall beta-(1,6)-glucan of *Saccharomyces cerevisiae*: structural characterization and in situ synthesis. *J Biol Chem* 2009;284:13401-12. DOI PubMed PMC
 49. Valiante V, Macheleidt J, Föge M, Brakhage AA. The *Aspergillus fumigatus* cell wall integrity signaling pathway: drug target, compensatory pathways, and virulence. *Front Microbiol* 2015;6:325. DOI PubMed PMC
 50. Gow NAR, Latge JP, Munro CA. The fungal cell wall: structure, biosynthesis, and function. *Microbiol Spectr* 2017;5. DOI PubMed
 51. Machová E, Kvapilová K, Kogan G, Sandula J. Effect of ultrasonic treatment on the molecular weight of carboxymethylated chitin-glucan complex from *Aspergillus niger*. *Ultrason Sonochem* 1999;5:169-72. DOI PubMed
 52. Mendonça A, Carvalho-Pereira J, Franco-Duarte R, Sampaio P. Correction to: optimization of a quantitative PCR methodology for detection of *Aspergillus* spp. and *Rhizopus arrhizus*. *Mol Diagn Ther* 2022;26:527. DOI PubMed PMC
 53. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature* 2007;449:804-10. DOI PubMed
 54. Stefanini I, Dapporto L, Legras JL, et al. Role of social wasps in *Saccharomyces cerevisiae* ecology and evolution. *Proc Natl Acad Sci U S A* 2012;109:13398-403. DOI PubMed PMC
 55. Abdelrhman KF, Bacci G, Mancusi C, Mengoni A, Serena F, Ugolini A. A first insight into the gut microbiota of the sea turtle caretta caretta. *Front Microbiol* 2016;7:1060. DOI PubMed PMC
 56. Abdelrhman KF, Bacci G, Marras B, et al. Exploring the bacterial gut microbiota of supralittoral talitrid amphipods. *Res Microbiol* 2017;168:74-84. DOI
 57. Ramazzotti M, Bacci G. Chapter 5 - 16S rRNA-based taxonomy profiling in the metagenomics era. In: Nagarajan M, editor. *Metagenomics*. Academic Press; 2018. p. 103-19. DOI
 58. Arranz V, Pearman WS, Aguirre JD, Liggins L. MARES, a replicable pipeline and curated reference database for marine eukaryote metabarcoding. *Sci Data* 2020;7:209. DOI PubMed PMC
 59. Frøslev TG, Kjølner R, Bruun HH, et al. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nat Commun* 2017;8:1188. DOI PubMed PMC
 60. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016;13:581-3. DOI PubMed PMC
 61. Nilsson RH, Anslan S, Bahram M, Wurzbacher C, Baldrian P, Tedersoo L. Mycobiome diversity: high-throughput sequencing and identification of fungi. *Nat Rev Microbiol* 2019;17:95-109. DOI PubMed
 62. Nilsson RH, Kristiansson E, Ryberg M, Hallenberg N, Larsson KH. Intraspecific ITS variability in the kingdom fungi as expressed in the international sequence databases and its implications for molecular species identification. *Evol Bioinform Online* 2008;4:193-201. DOI PubMed PMC
 63. Ali NABM, Mac Aogáin M, Morales RF, Tiew PY, Chotirmall SH. Optimisation and benchmarking of targeted amplicon sequencing for mycobiome analysis of respiratory specimens. *Int J Mol Sci* 2019;20:4991. DOI PubMed PMC
 64. Bokulich NA, Mills DA. Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. *Appl Environ Microbiol* 2013;79:2519-26. DOI PubMed PMC
 65. Tedersoo L, Lindahl B. Fungal identification biases in microbiome projects. *Environ Microbiol Rep* 2016;8:774-9. DOI PubMed
 66. Franco-Duarte R, Fernandes I, Gulis V, Cássio F, Pascoal C. ITS rDNA barcodes clarify molecular diversity of aquatic hyphomycetes. *Microorganisms* 2022;10:1569. DOI PubMed PMC
 67. Bradshaw MJ, Aime MC, Rokas A, et al. Extensive intragenomic variation in the internal transcribed spacer region of fungi. *iScience* 2023;26:107317. DOI PubMed PMC
 68. Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kausserud H. ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC Microbiol* 2010;10:189. DOI PubMed PMC
 69. Mbareche H, Veillette M, Bilodeau G, Duchaine C. Comparison of the performance of ITS1 and ITS2 as barcodes in amplicon-based sequencing of bioaerosols. *PeerJ* 2020;8:e8523. DOI PubMed PMC
 70. Hoggard M, Vesty A, Wong G, et al. Characterizing the human mycobiota: a comparison of small subunit rRNA, ITS1, ITS2, and large subunit rRNA genomic targets. *Front Microbiol* 2018;9:2208. DOI PubMed PMC
 71. Peterson SW, Kurtzman CP. Ribosomal RNA sequence divergence among sibling species of yeasts. *Syst Appl Microbiol* 1991;14:124-9. DOI
 72. Kurtzman CP, Robnett CJ. Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Antonie Van Leeuwenhoek* 1998;73:331-71. DOI PubMed
 73. Tang J, Iliev ID, Brown J, Underhill DM, Funari VA. Mycobiome: approaches to analysis of intestinal fungi. *J Immunol Methods* 2015;421:112-21. DOI PubMed PMC
 74. Filippis F, Laiola M, Blaiotta G, Ercolini D. Different amplicon targets for sequencing-based studies of fungal diversity. *Appl*

- Environ Microbiol* 2017;83:e00905-17. DOI PubMed PMC
75. Kiss L. Limits of nuclear ribosomal DNA internal transcribed spacer (ITS) sequences as species barcodes for Fungi. *Proc Natl Acad Sci U S A* 2012;109:E1811; author reply E1812. DOI PubMed PMC
 76. Stielow JB, Lévesque CA, Seifert KA, et al. One fungus, which genes? Development and assessment of universal primers for potential secondary fungal DNA barcodes. *Persoonia* 2015;35:242-63. DOI
 77. James TY, Kauff F, Schoch CL, et al. Reconstructing the early evolution of fungi using a six-gene phylogeny. *Nature* 2006;443:818-22. DOI PubMed
 78. Matheny PB, Liu YJ, Ammirati JF, Hall BD. Using RPB1 sequences to improve phylogenetic inference among mushrooms (Inocybe, Agaricales). *Am J Bot* 2002;89:688-98. DOI PubMed
 79. Meyer W, Irinyi L, Hoang MTV, et al. Database establishment for the secondary fungal DNA barcode *translational elongation factor 1a* (TEF1a)¹. *Genome* 2019;62:160-9. DOI
 80. 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 Resour* 2016;16:388-401. DOI PubMed
 81. Morrison GA, Fu J, Lee GC, et al. Nanopore sequencing of the fungal intergenic spacer sequence as a potential rapid diagnostic assay. *J Clin Microbiol* 2020;58:e01972-20. DOI PubMed PMC
 82. Geiser DM, Frisvad JC, Taylor JW. Evolutionary relationships in *Aspergillus* section *Fumigati* inferred from partial β -tubulin and hydrophobin DNA sequences. *Mycologia* 1998;90:831-45. DOI
 83. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: an overview. *Hum Immunol* 2021;82:801-11. DOI PubMed
 84. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* 2017;35:833-44. DOI PubMed
 85. Morgan XC, Huttenhower C. Meta-omic analytic techniques for studying the intestinal microbiome. *Gastroenterology* 2014;146:1437-48.e1. DOI PubMed
 86. Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;486:207-14. DOI PubMed PMC
 87. Nash AK, Auchtung TA, Wong MC, et al. The gut mycobiome of the Human Microbiome Project healthy cohort. *Microbiome* 2017;5:153. DOI PubMed PMC
 88. Hoang MTV, Irinyi L, Hu Y, Schwessinger B, Meyer W. Long-reads-based metagenomics in clinical diagnosis with a special focus on fungal infections. *Front Microbiol* 2021;12:708550. DOI PubMed PMC
 89. Pollard MO, Gurdasani D, Mentzer AJ, Porter T, Sandhu MS. Long reads: their purpose and place. *Hum Mol Genet* 2018;27:R234-41. DOI PubMed PMC
 90. Mantere T, Kersten S, Hoischen A. Long-read sequencing emerging in medical genetics. *Front Genet* 2019;10:426. DOI PubMed PMC
 91. Sui Y, Wisniewski M, Droby S, Piombo E, Wu X, Yue J. Genome sequence, assembly, and characterization of the antagonistic yeast *Candida oleophila* used as a biocontrol agent against post-harvest diseases. *Front Microbiol* 2020;11:295. DOI PubMed PMC
 92. Cuomo CA, Shea T, Yang B, Rao R, Forche A. Whole genome sequence of the heterozygous clinical isolate *Candida krusei* 81-B-5. *G3* 2017;7:2883-9. DOI PubMed PMC
 93. Luo R, Zimin A, Workman R, et al. First draft genome sequence of the pathogenic fungus *Lomentospora prolificans* (Formerly *Scedosporium prolificans*). *G3* 2017;7:3831-6. DOI PubMed PMC
 94. Vale-Silva L, Beaudoin E, Tran VDT, Sanglard D. Comparative genomics of two sequential *Candida glabrata* clinical isolates. *G3* 2017;7:2413-26. DOI PubMed PMC
 95. Panthee S, Hamamoto H, Ishijima SA, Paudel A, Sekimizu K. Utilization of hybrid assembly approach to determine the genome of an opportunistic pathogenic fungus, *Candida albicans* TIMM 1768. *Genome Biol Evol* 2018;10:2017-22. DOI PubMed PMC
 96. Rhodes J, Abdolrasouli A, Farrer RA, et al. Genomic epidemiology of the UK outbreak of the emerging human fungal pathogen *Candida auris*. *Emerg Microbes Infect* 2018;7:43. DOI PubMed PMC
 97. Morand SC, Bertignac M, Iltis A, et al. Complete genome sequence of *Malassezia restricta* CBS 7877, an opportunist pathogen involved in dandruff and seborrheic dermatitis. *Microbiol Resour Announc* 2019;8:e01543-18. DOI PubMed PMC
 98. Schultzhau Z, Cuomo CA, Wang Z. Genome sequence of the black yeast *Exophiala lecanii-corni*. *Microbiol Resour Announc* 2019;8:e01709-18. DOI PubMed PMC
 99. Pehelin IM, Azarov DV, Churina MA, et al. Whole genome sequence of first *Candida auris* strain, isolated in Russia. *Med Mycol* 2020;58:414-6. DOI
 100. Arnaud MB, Chibucos MC, Costanzo MC, et al. The aspergillus genome database, a curated comparative genomics resource for gene, protein and sequence information for the Aspergillus research community. *Nucleic Acids Res* 2010;38:D420-7. DOI PubMed PMC
 101. Ratnasingham S, Hebert PD. bold: The barcode of life data system (<http://www.barcodinglife.org>). *Mol Ecol Notes* 2007;7:355-64. DOI PubMed PMC
 102. Inglis DO, Arnaud MB, Binkley J, et al. The *Candida* genome database incorporates multiple *Candida* species: multispecies search and analysis tools with curated gene and protein information for *Candida albicans* and *Candida glabrata*. *Nucleic Acids Res* 2012;40:D667-74. DOI PubMed PMC
 103. Güldener U, Münsterkötter M, Kastenmüller G, et al. CYGD: the comprehensive yeast genome database. *Nucleic Acids Res*

- 2005;33:D364-8. [DOI PubMed PMC](#)
104. Stajich JE, Harris T, Brunk BP, et al. FungiDB: an integrated functional genomics database for fungi. *Nucleic Acids Res* 2012;40:D675-81. [DOI PubMed PMC](#)
105. Grossetête S, Labedan B, Lespinet O. FUNGIpath: a tool to assess fungal metabolic pathways predicted by orthology. *BMC Genomics* 2010;11:81. [DOI PubMed PMC](#)
106. Geiser DM, del Mar Jiménez-gasco M, Kang S, et al. FUSARIUM-ID v. 1.0: a DNA sequence database for identifying fusarium. *Eur J Plant Pathol* 2004;110:473-9. [DOI](#)
107. O'donnell K, Sutton DA, Rinaldi MG, et al. Internet-accessible DNA sequence database for identifying fusaria from human and animal infections. *J Clin Microbiol* 2010;48:3708-18. [DOI PubMed PMC](#)
108. Irinyi L, Serena C, Garcia-Hermoso D, 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* 2015;53:313-37. [DOI](#)
109. Ahrendt SR, Mondo SJ, Haridas S, Grigoriev IV. MycoCosm, the JGI's fungal genome portal for comparative genomic and multiomics data analyses. In: Martin F, Uroz S, editors. *Microbial Environmental Genomics (MEG)*. New York: Springer US; 2023. p. 271-91. [DOI](#)
110. Wood V, Harris MA, McDowall MD, et al. PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Res* 2012;40:D695-9. [DOI PubMed PMC](#)
111. Cherry JM, Hong EL, Amundsen C, et al. Saccharomyces genome database: the genomics resource of budding yeast. *Nucleic Acids Res* 2012;40:D700-5. [DOI PubMed PMC](#)
112. Abarenkov K, Henrik Nilsson R, Larsson KH, et al. The UNITE database for molecular identification of fungi - recent updates and future perspectives. *New Phytol* 2010;186:281-5. [DOI](#)
113. Prakash PY, Irinyi L, Halliday C, Chen S, Robert V, Meyer W. Online databases for taxonomy and identification of pathogenic fungi and proposal for a cloud-based dynamic data network platform. *J Clin Microbiol* 2017;55:1011-24. [DOI PubMed PMC](#)
114. Kuczynski J, Lauber CL, Walters WA, et al. Experimental and analytical tools for studying the human microbiome. *Nat Rev Genet* 2011;13:47-58. [DOI PubMed PMC](#)
115. Schloss PD, Westcott SL, Ryabin T, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;75:7537-41. [DOI PubMed PMC](#)
116. Gweon HS, Oliver A, Taylor J, et al. PIPITS: an automated pipeline for analyses of fungal internal transcribed spacer sequences from the Illumina sequencing platform. *Methods Ecol Evol* 2015;6:973-80. [DOI PubMed PMC](#)
117. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 2016;4:e2584. [DOI PubMed PMC](#)
118. Mysara M, Njima M, Leys N, Raes J, Monsieurs P. From reads to operational taxonomic units: an ensemble processing pipeline for MiSeq amplicon sequencing data. *Gigascience* 2017;6:1-10. [DOI PubMed PMC](#)
119. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019;37:852-7. [DOI PubMed PMC](#)
120. He Y, Caporaso JG, Jiang XT, et al. Stability of operational taxonomic units: an important but neglected property for analyzing microbial diversity. *Microbiome* 2015;3:20. [DOI PubMed PMC](#)
121. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J* 2017;11:2639-43. [DOI PubMed PMC](#)
122. Chiarello M, McCauley M, Villéger S, Jackson CR. Ranking the biases: the choice of OTUs vs. ASVs in 16S rRNA amplicon data analysis has stronger effects on diversity measures than rarefaction and OTU identity threshold. *PLoS One* 2022;17:e0264443. [DOI PubMed PMC](#)
123. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 2014;15:R46. [DOI PubMed PMC](#)
124. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7:335-6. [DOI PubMed PMC](#)
125. Abarenkov K, Tedersoo L, Nilsson RH, et al. PlutoF - a web based workbench for ecological and taxonomic research, with an online implementation for fungal ITS sequences. *Evol Bioinform Online* 2010;6:EBO.S6271. [DOI PMC](#)
126. Kumar S, Carlsen T, Mevik BH, et al. CLOTU: an online pipeline for processing and clustering of 454 amplicon reads into OTUs followed by taxonomic annotation. *BMC Bioinform* 2011;12:182. [DOI PubMed PMC](#)
127. White JR, Maddox C, White O, Angiuoli SV, Fricke WF. CloVR-ITS: automated internal transcribed spacer amplicon sequence analysis pipeline for the characterization of fungal microbiota. *Microbiome* 2013;1:6. [DOI PubMed PMC](#)
128. Albanese D, Fontana P, De Filippo C, Cavalieri D, Donati C. MICCA: a complete and accurate software for taxonomic profiling of metagenomic data. *Sci Rep* 2015;5:9743. [DOI PubMed PMC](#)
129. Fosso B, Santamaria M, Marzano M, et al. BioMaS: a modular pipeline for Bioinformatic analysis of Metagenomic AmpliconS. *BMC Bioinform* 2015;16:203. [DOI PubMed PMC](#)
130. Odom AR, Faits T, Castro-Nallar E, Crandall KA, Johnson WE. Metagenomic profiling pipelines improve taxonomic classification for 16S amplicon sequencing data. *Sci Rep* 2023;13:13957. [DOI PubMed PMC](#)
131. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* 2013;10:996-8. [DOI PubMed](#)

132. Edgar RC, Flyvbjerg H. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics* 2015;31:3476-82. [DOI](#) [PubMed](#)
133. Truong DT, Franzosa EA, Tickle TL, et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods* 2015;12:902-3. [DOI](#)
134. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019;20:257. [DOI](#) [PubMed](#) [PMC](#)
135. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2015;12:59-60. [DOI](#) [PubMed](#)
136. Olson ND, Treangen TJ, Hill CM, et al. Metagenomic assembly through the lens of validation: recent advances in assessing and improving the quality of genomes assembled from metagenomes. *Brief Bioinform* 2019;20:1140-50. [DOI](#) [PubMed](#) [PMC](#)
137. Jünemann S, Kleinbölting N, Jaenicke S, et al. Bioinformatics for NGS-based metagenomics and the application to biogas research. *J Biotechnol* 2017;261:10-23. [DOI](#)
138. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 2017;27:824-34. [DOI](#) [PubMed](#) [PMC](#)
139. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015;31:1674-6. [DOI](#)
140. Peng Y, Leung HC, Yiu SM, Chin FY. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 2012;28:1420-8. [DOI](#)
141. Kang DD, Li F, Kirton E, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019;7:e7359. [DOI](#) [PubMed](#) [PMC](#)
142. Alneberg J, Bjarnason BS, de Bruijn I, et al. Binning metagenomic contigs by coverage and composition. *Nat Methods* 2014;11:1144-6. [DOI](#)
143. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 2016;32:605-7. [DOI](#) [PubMed](#)
144. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 2018;6:158. [DOI](#) [PubMed](#) [PMC](#)
145. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28:27-30. [DOI](#) [PubMed](#) [PMC](#)