

# Fungal sequencing and classification with the ITS Metagenomics Protocol

Perform rapid, reliable mycobiome sequencing with the demonstrated ITS Metagenomics Protocol.

## Introduction

The mycobiome plays an essential role in many biological processes such as symbiosis, decomposition, and disease progression—including skin mycosis. Understanding the role fungi play in these processes is critical to understanding microbial communities and the environment. However, due to culturing constraints, fungi can be difficult to assess and are often overlooked in metagenomics studies.

To address this challenge, many researchers have turned to culture-free fungal identification methods such as DNA sequencing for metagenomic studies (Figure 1). The internal transcribed spacer 1 (ITS1) region of the rRNA cistron is a commonly used DNA marker for identification of fungal species in metagenomic samples.<sup>1</sup> Amplification of the ITS1 region is a powerful method for culture-independent characterization and identification of fungal biota in complex and diverse environmental samples such as stool, saliva, skin, soil, and water.

To provide a rapid, next-generation sequencing (NGS)-based solution for ITS1 sequencing, Illumina offers the ITS Metagenomics Protocol for culture-free mycobiome characterization and analysis. This application note demonstrates the successful amplification and sequencing of the ITS1 amplicon as part of the ITS Metagenomics Workflow (Figure 2) using mock community and environmental samples sequenced on the MiSeq™ and iSeq™ 100 Systems.

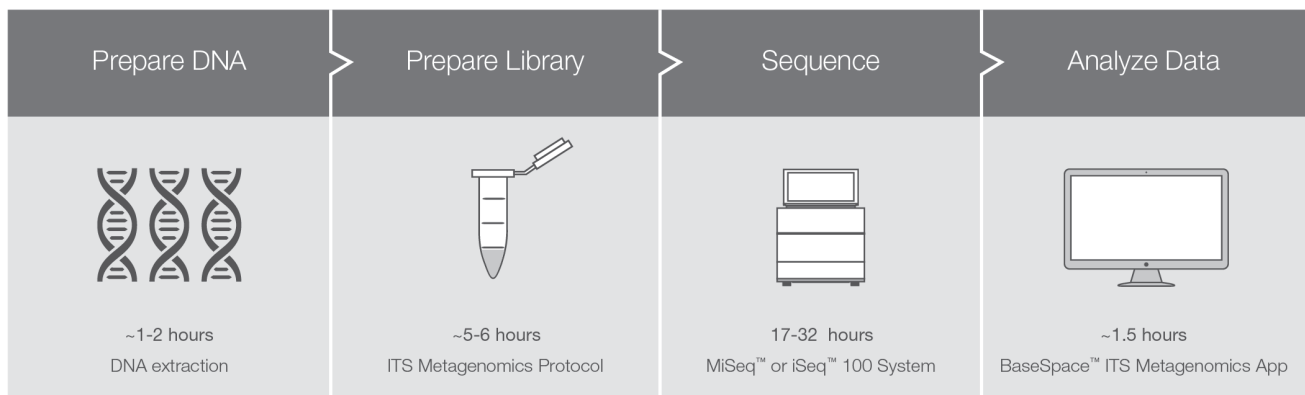


**Figure 1: Fungal metagenomics sequencing with the ITS Metagenomics Protocol**—The mycobiome, or fungal microbiome, is an important component in the environment and in human health. The human mycobiome may be important in a maintaining microbial community balance, metabolic function and immune system priming.<sup>2</sup>

## Methods

### Mock community samples

Two mock community samples of high-quality DNA were analyzed in this study. The first mock community was produced by combining five



**Figure 2: Fungal metagenomics sequencing with the ITS Metagenomics Workflow**—The ITS Metagenomics Workflow is part of an integrated NGS workflow that includes library preparation of the ITS1 amplicon, robust Illumina sequencing, and user-friendly data analysis with the BaseSpace™ ITS Metagenomics App.

Table 1: Composition of ATCC five-pool mixture

Name	Percentage of mixture	Catalog No.
<i>Aspergillus brasiliensis</i> WLRI 034	20%	16404D-2
<i>Blastomyces dermatitidis</i>	20%	26199D-2
<i>Cryptococcus neoformans</i>	20%	208821D-2
<i>Paracoccidioides brasiliensis</i>	20%	MYA-826D-2
<i>Saccharomyces cerevisiae</i>	20%	201389D-5

Table 2: Composition of ZymoBIOMICS Microbial Community Standard

Strain Name	Percentage of mixture	Organism type
<i>Pseudomonas aeruginosa</i>	12%	bacterial
<i>Escherichia coli</i>	12%	bacterial
<i>Salmonella enterica</i>	12%	bacterial
<i>Lactobacillus fermentum</i>	12%	bacterial
<i>Enterococcus faecalis</i>	12%	bacterial
<i>Staphylococcus aureus</i>	12%	bacterial
<i>Listeria monocytogenes</i>	12%	bacterial
<i>Bacillus subtilis</i>	12%	bacterial
<i>Saccharomyces cerevisiae</i>	2%	fungus
<i>Cryptococcus neoformans</i>	2%	fungus

high-quality fungal isolates from the American Type Culture Collection (Table 1). The fungal DNA samples were quantified with a fluorometric method and pooled with equal DNA inputs from each isolate. The second mock community, the ZymoBIOMICS Microbial Community Standard (Zymo Research, Catalog No. D6306) mimics a mixed microbial community with a low percentage of fungal DNA. The ZymoBIOMICS standard is a mixture of genomic DNA isolated from pure cultures of eight bacterial strains (12% each) and two fungal strains (2% each) including *Saccharomyces cerevisiae* and *Cryptococcus neoformans* (Table 2).

**Environmental samples**

In addition to testing reference standards, three real world samples provided by an academic collaborator were tested using the ITS Metagenomics Workflow. The samples were from several sources with varying levels of fungal composition including a soil sample, a pine tree soil sample, and a skin swab sample.

**Library Preparation**

Libraries were prepared following the ITS Metagenomics Protocol which includes the use of indexes from the Nextera XT Index Kit v2 (Illumina, Catalog No. FC-131-2001).

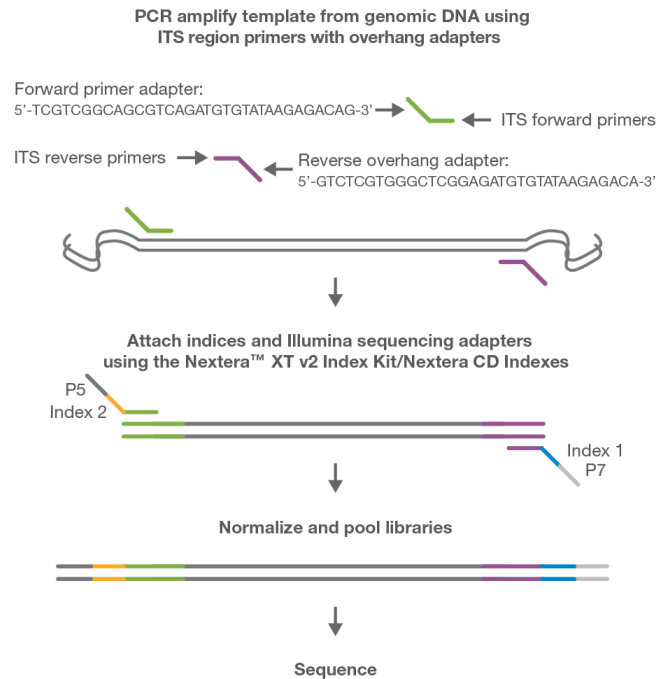
**Validated Primer Sequences**

Library preparation with the ITS Metagenomics Protocol includes two amplification steps (Figure 3). The first amplification was performed with primers modified from a known primer pair of the ITS1 region used widely in fungal barcoding (Table 3).<sup>3</sup> To increase the taxonomic coverage, these modified primers were optimized through probe hit analysis to address gaps in coverage. Sequence alignments for the modified ITS1 primers were performed against the SILVA<sup>4</sup> database and the Warcup Training Set.<sup>5</sup> Because the Warcup Training Set did not contain the forward primer,

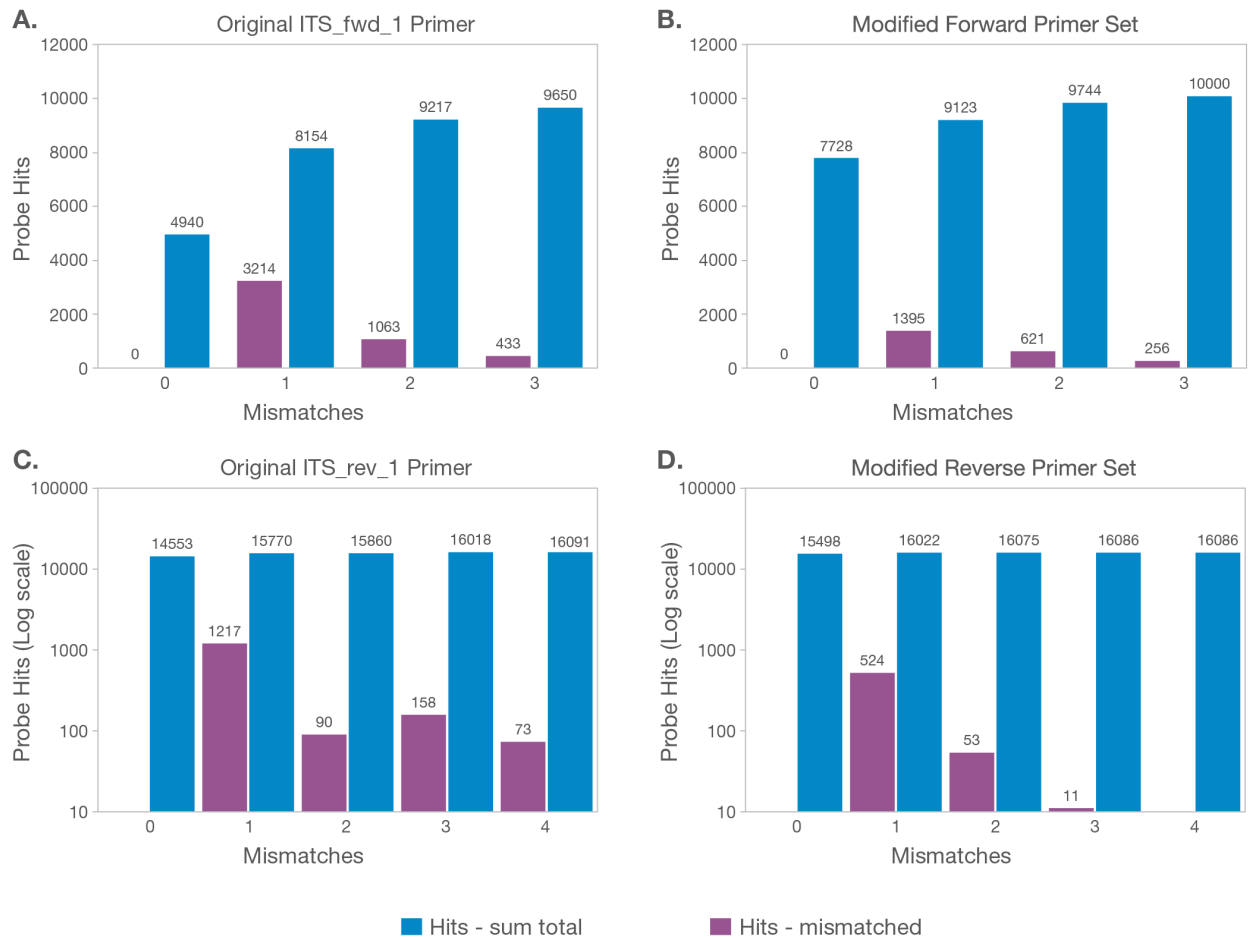
Table 3: ITS1 primer sequences and Illumina adapter sequences

Forward Primer Set	Sequence (5')
ITS_fwd_1 <sup>a</sup>	CTTGGTCATTTAGAGGAAGTAA
ITS_fwd_2	CTCGGTCATTTAGAGGAAGTAA
ITS_fwd_3	CTTGGTCATTTAGAGGAAGTAA
ITS_fwd_4	CCC GGTCATTTAGAGGAAGTAA
ITS_fwd_5	CTAGGTCATTTAGAGGAAGTAA
ITS_fwd_6	CTTAGTTATTTAGAGGAAGTAA
ITS_fwd_7	CTACGTCATTTAGAGGAAGTAA
ITS_fwd_8	CTTGGTCATTTAGAGGTCGTAA
Reverse Primer Set	
ITS_rev_1 <sup>b</sup>	GCTGCGTTCCTTCATCGATGC
ITS_rev_2	GCTGCGTTCCTTCATCGATGG
ITS_rev_3	GCTACGTTCTTCATCGATGC
ITS_rev_4	GCTGCGTTCCTTCATCGATGT
ITS_rev_5	ACTGTGTTCTTCATCGATGT
ITS_rev_6	GCTGCGTTCCTTCATCGTTGC
ITS_rev_7	GCGTTCTTCATCGATGC
Overhang Adapter Sequences	
Forward overhang: 5'	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[locus-specific sequence]
Reverse overhang: 5'	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-[locus-specific sequence]

- a. The same sequence as the original ITS1-F primer sequence in the Bellemain paper.
  - b. The same sequence as the original ITS2 primer sequence in the Bellemain paper.
- Note: In this Illumina study, this sequence was used as the reverse primer for ITS1 region amplification.



**Figure 3: ITS1 amplicon chemistry**-ITS1 amplicon chemistry includes ITS1 amplification, PCR clean-up steps, second stage PCR and PCR clean-up steps, followed by library quantification, normalization, and pooling.



**Figure 4: Probe hits in database increase after addition of modified primers** -Charts show results from primer sequence alignments to the SILVA or Warcup databases. X-axis indicates the number of mismatched nucleotides in alignments to database sequences. Blue bars represent the total number of hits including hits with mismatched nucleotides. Purple bars represent hits with mismatched nucleotides. (A) The original ITS\_fwd\_1 primer and the (B) modified forward primer set aligned against the SILVA database. (C) The original ITS\_rev\_1 primer and the (D) modified reverse primer set aligned against the Warcup Training Set. Note: Due to the large number of probe hits in the Warcup Training Set, the y-axis is displayed in logarithmic scale.

Because the Warcup Training Set does not contain the forward primer, the SILVA database was used to test the forward primer set and the Warcup Training Set was used to test the reverse primer set. Inclusion of these modified primers increased the number of hits to fungal ITS1 sequences (Figure 4).

Following probe hit analysis, the forward and reverse primers complementary to the upstream and downstream sequences of the ITS1 region were designed with overhang adapters and used to amplify templates from genomic DNA. A subsequent limited-cycle amplification step was performed to add Illumina sequencing adapters and dual-index barcodes from the Nextera XT DNA Index Kit to the amplicons. Libraries were quantified, hand-normalized, and pooled according to the ITS Metagenomics Protocol.

### Sequencing

Libraries were sequenced on the MiSeq System with the MiSeq Reagent Kit v3 (Illumina, Catalog No. MS-102-3003). The library prepared from the ATCC five-pool sample was also sequenced on the iSeq 100 System using the iSeq 100 i1 Reagent Kit (Illumina, Catalog No. 20021533). All MiSeq and iSeq 100 System runs were performed with a run configuration of 2 x 151 cycles. To enable calculation of error-rate metrics and to increase nucleotide base diversity for more accurate base-calling, all libraries were spiked with 5% PhiX Control v3 (Illumina, Catalog No. FC-110-3001).

**Data Analysis**

Sequence data can be transferred, analyzed, and stored securely in BaseSpace Sequence Hub, the Illumina genomics computing environment. FASTQ sequence files were generated in BaseSpace Sequence Hub followed by sequence alignment and phylogenetic classification with the BaseSpace ITS Metagenomics App<sup>6</sup> (Figure 5). The ITS Metagenomics App performs taxonomic classification of amplicon reads using UNITE,<sup>7</sup> a curated fungal database. Genus read counts were visualized using JMP software.<sup>8</sup>

**Results**

**Fungal sequencing with mock microbial communities**

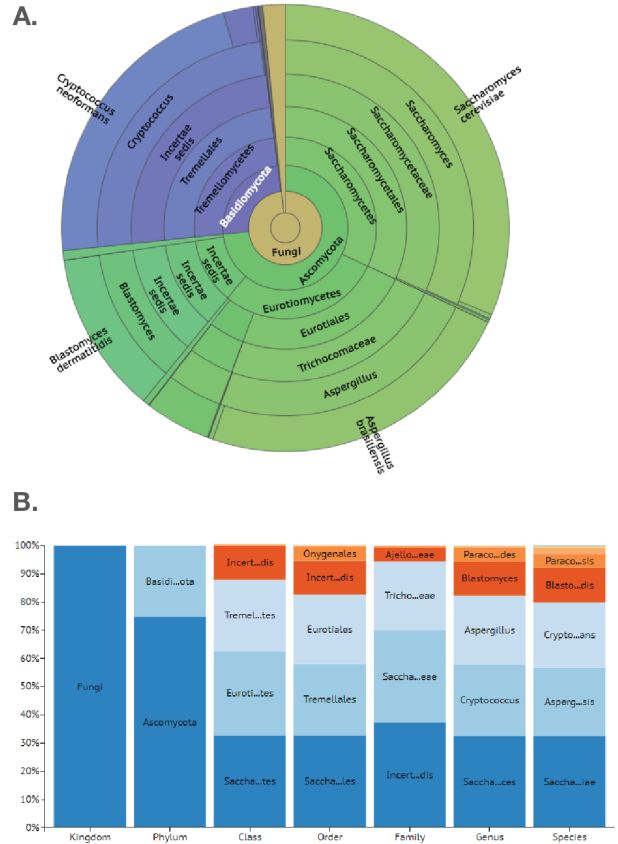
To assess the performance of the ITS Metagenomics Workflow from library preparation to data analysis, two mock community samples were analyzed: the ATCC five-pool mixture and the ZymoBIOMICS Microbial Community Standard. Mock community samples can be used to assess bias or errors associated with library preparation, sequencing, and data analysis in the NGS workflow. The validated community mixtures can also be used as a standard for quality control purposes throughout the workflow.

**Accurate classification of fungal isolates in mock community with even mixture of isolates**

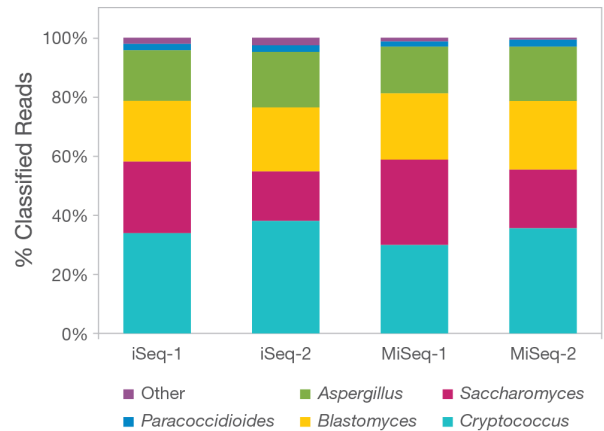
The ATCC five-pool mock community, which contains equal proportions of fungal DNA, was used to compare sequencing data between the MiSeq and iSeq 100 Systems with the ITS Metagenomics Workflow. The results showed accurate detection and classification for all five isolates (Figure 6). Consistency between technical replicates and between sequencing platforms was also demonstrated. We observed an imbalance in the abundance representation of each isolate, however this is likely due to varying rRNA cistron copy number between isolates which is commonly seen in the field.<sup>9</sup> Fast-growing yeasts such as *Saccharomyces cerevisiae* tend to have higher copy numbers than slow-growing yeasts such as *Paracoccidioides lutzii*. GC-bias, amplicon length, taxonomic primer bias, and DNA extraction efficiency can also be confounding factors for quantitative fungal community profiling.<sup>9,10</sup> Due to these constraints, the workflow should be considered limited as a quantitative profiling method.

**Sensitive detection of fungal isolates in a mock community with low fungal composition**

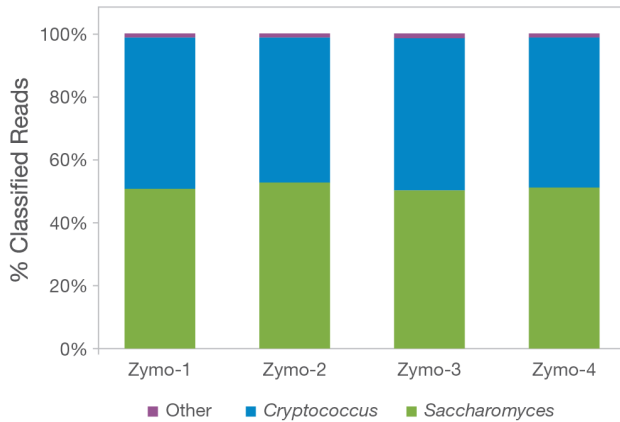
To test the sensitivity of the ITS Metagenomics Workflow, the ZymoBIOMICS Standard, which contains a low percentage of fungal gDNA, was tested. The ZymoBIOMICS Standard contains 96% bacterial, 2% *Saccharomyces*, and 2% *Cryptococcus* gDNA. The results consistently demonstrated >99% correct classification with four technical replicates (Figure 7). Furthermore, these results were achieved with input as low as 1 ng gDNA, demonstrating the robust performance of the ITS Metagenomics Workflow.



**Figure 5: BaseSpace ITS Metagenomics App** - The ITS Metagenomics App performs metagenomic analysis including read alignment to the UNITE database and read classification. The app generates reports and charts including dendrograms, principle coordinate analysis, pie charts, (A) sunburst charts, and (B) stacked column charts. The ITS Metagenomics App is freely accessible through BaseSpace Sequence Hub.



**Figure 6: Accurate classification of fungal genera in ATCC five-pool mock community** - The ITS Metagenomic Workflow successfully detected all five fungal isolates in a mock community mixture. High concordance was observed between iSeq and MiSeq System data.



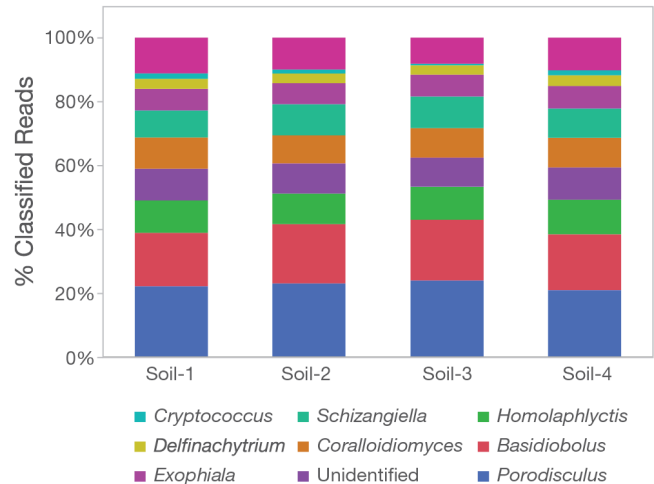
**Figure 7: Sensitive detection of fungal isolates in ZymoBIOMICS mock community** - The ITS Metagenomics Workflow successfully determined the correct identity and balance of two fungal genera in a mock community standard with extremely low fungal content. Data also show high concordance between replicates.

### Fungal sequencing with real-world environmental samples

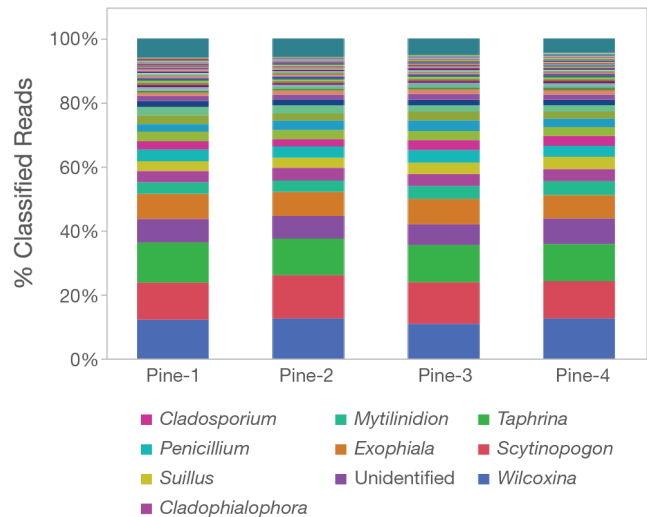
The ITS Metagenomics Workflow was also evaluated with three real-world environmental samples. The first sample, a soil sample, was assayed with 12.5 ng input DNA. Results showed identification of eight abundant fungal taxa previously identified in soil (Figure 8). *Porodisculus* is common to wood substrate and freshly dead branches.<sup>11</sup> *Basidiobolus* is easily isolated from soil or leaf-litter while *Delfinachytrium* is found in soil and aquatic habitats.<sup>12,13</sup> *Exophiala* and *Cryptococcus* are also known to inhabit soil.<sup>14</sup>

The second soil sample was collected in proximity to Pine trees and assayed with 11.5 ng input. The pine tree soil sample produced read classification of more than 17 fungal taxa with >1% composition within the sample (Figure 9). The highest number of reads in the sample were aligned to *Wilcoxina*, a genera of mycorrhizal fungi known to infect a variety of conifers including *Pinus*, *Betula*, and *Quercus*.<sup>15</sup> Other fungal taxa identified have been found in soil and wooded environments. *Taphrina* is a fungal genus within the *Ascomycota* that causes leaf and catkin curl diseases.<sup>16</sup> *Cladophialophora* are found in soil while *Mytilinidion* are found on wood.<sup>17,18</sup> The pine tree soil sample demonstrates the successful detection and classification of fungal biota in a highly diverse microbial community.

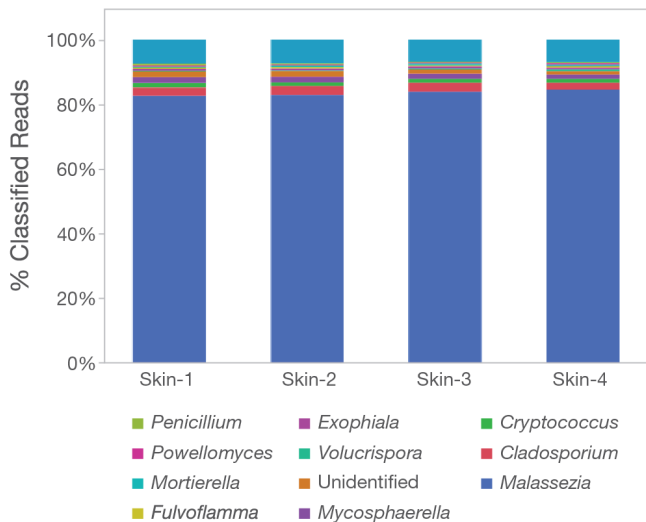
The third environmental sample, a skin swab collected from an individual with a fungal skin infection, was assayed with only 2.3 ng due to sample volume constraints. The classification results illustrate a low-diversity environment dominated by one fungal taxon (Figure 10). This taxon, *Malassezia*, is a common fungus that can cause skin rashes and infections.<sup>19</sup> These results illustrate that the ITS Metagenomics Workflow can detect fungal genera in low-diversity environments.



**Figure 8: Identification of eight fungal genera known to inhabit soil** - The ITS Metagenomics Workflow successfully detected and identified eight fungal genera known to inhabit soil in the environmental soil sample. Four replicates demonstrated high concordance. Note: The "Unidentified" category is taxon category of UNITE database.



**Figure 9: Identification of more than 17 fungal genera in highly diverse pine tree soil sample** - The ITS Metagenomics Workflow successfully detected and identified a variety of fungal taxa known to inhabit soil environments, and for some genera, known to inhabit soil samples near pine trees. The read classification results also demonstrate high concordance between the four replicates.



**Figure 10: Identification of fungal genera in a low-diversity skin sample** - The skin swab sample from an individual with a fungal skin infection represents an environment dominated by a single taxon. The ITS Metagenomics Workflow successfully detected and identified the *Malassezia* taxa, a common cause of skin mycosis.

## Summary

The ITS Metagenomics Workflow outlines an integrated, comprehensive NGS workflow for identification of fungal biota for a variety of sample types. This application note demonstrated successful amplification and sequencing of the ITS1 biomarker region in metagenomics samples, and subsequent classification with the BaseSpace ITS Metagenomics App. Additionally, the accuracy of genera identification, and concordance between the MiSeq and iSeq Systems was demonstrated using reference standards in mock community samples. Numerous fungal genera were then detected and identified in real-world environmental samples of varied diversity levels. With the ITS Metagenomics Workflow, microbiologists can leverage the power of next-generation sequencing to advance their research and increase our understanding of the mycobiome.

## Products Mentioned

Product Name	Catalog number
MiSeq System	SY-410-1003
MiSeq Reagent Kit v3 (150-cycle)	MS-102-3001
iSeq 100 System	20021532
iSeq 100 i1 Reagent Kit (318-cycle)	20021533
Nextera XT Index Kit v2 Set A (96 indexes, 384 samples)	FC-131-2001
PhiX Control v3	FC-110-3001

## Learn More

[www.illumina.com/systems/sequencing-platforms/iseq.html](http://www.illumina.com/systems/sequencing-platforms/iseq.html)

[www.illumina.com/systems/sequencing-platforms/miseq.html](http://www.illumina.com/systems/sequencing-platforms/miseq.html)

## ITS Metagenomics Protocol

illumina, Inc. • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

© 2018 illumina, Inc. All rights reserved. All trademarks are the property of illumina, Inc. or their respective owners. For specific trademark information, see [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html). Pub. No. 1270-2018-001-A QB #6724

For Research Use Only. Not for use in diagnostic procedures.

## References

- 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*. 2012; 109(16):6241-6.
- Cui L, Morris A, Ghedin E. The human mycobiome in health and disease. *Genome Med*. 2013;5 (7): 63.
- Bellemain E, Carlsen, T, Brochmann C, et al. ITS as an environmental DNA barcode for fungi: an *in silico* approach reveals potential PCR biases. *BMC Microbiol*. 2010;10: 189.
- SILVA database. [www.arb-silva.de/](http://www.arb-silva.de/). Accessed September 14, 2018.
- Deshpande V, Wang Q, Greenfield P, et al. Fungal identification using a Bayesian classifier and the Warcup training set of internal transcribed spacer sequences. *Mycologia*. 2015;108(1)1-5.
- ITS Metagenomics App. Accessed September 14, 2018.
- UNITE v01.12.2017. [unite.ut.ee/](http://unite.ut.ee/). Accessed September 14, 2018.
- JMP Software. [www.jmp.com/en\\_us/home.html](http://www.jmp.com/en_us/home.html). Accessed September 14, 2018.
- Taylor DL, Walters WA, Lennon NJ, et al. Accurate estimation of fungal diversity and abundance through improved lineage-specific primers optimized for Illumina amplicon sequencing. *Appl and Environ Microbiol*. 2016;82(24):7217-7226.
- Bokulich NA, Mills DA. Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. *Appl and Environ Microbiol*. 2013;79(8): 2519-2526.
- Murrill WA. Polyporaceae, Part 1. North American Flora. New York Botanical Garden. 1907;9(1):47.
- Drechsler C. Supplementary Developmental Stages of *Basidiobolus ranarum* and *Basidiobolus haptosporus*. *Mycologia*. 1956;48(5):655.
- Vélez CG, Letcher PM, Schultz S, et al. Three new genera in *Chytridiales* from aquatic habitats in Argentina. *Mycologia*. 2013;105(5):1251-65.
- Moll J, Hoppe B, König S, et al. Spatial distribution of fungal communities in an arable soil. *PLoS One*. 2016;11(2): e0148130.
- Hagedorn C, Nakas JP. *Biotechnology of Plant-Microbe Interactions*. New York: McGraw-Hill. 2013; 4(3): e00055-13.
- Ousmane HC, Almeida MGCFJ, Fonseca A, et al. Genome sequencing of the plant pathogen *Taphrina deformans*, the causal agent of peach leaf curl. *mBio*. 2013; 4(3): e00055-13.
- He J, Xu Z, Hughes J. Analyses of soil fungal communities in adjacent natural forest and hoop pine plantation ecosystems of subtropical Australia using molecular approaches based on 18S rRNA genes. *FEMS Microbiol Lett*. 2005;247(1):91-100.
- Cannon PF, Kirk PM Fungal Families of the World. Wallingford, UK, CAB International. 2007; 233-4.
- Inamadar AC, Palit A. The genus *Malassezia* and human disease. *Indian J Dermatol Venereol Leprol*. 2003;69(4):265-70.