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. 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.

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 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 maintaining microbial community balance, metabolic function and immune system priming.

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
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). 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 database and the Warcup Training Set. Because the Warcup Training Set did not contain the forward primer,
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).

![Figure 4: Probe hits in database increase after addition of modified primers](image)

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.
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 (Figure 5). The ITS Metagenomics App performs taxonomic classification of amplicon reads using UNITE, a curated fungal database. Genus read counts were visualized using JMP software.

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 Megenomics 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. Fast-growing yeasts such as Saccharomyces cerevisiae tend to have higher copy numbers than slow-growing yeasts such as Paracoccidioides brasiliensis. GC-bias, amplicon length, taxonomic primer bias, and DNA extraction efficiency can also be confounding factors for quantitative fungal community profiling. 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.
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.  

- Basidiobolus is easily isolated from soil or leaf-litter while *Delfinachytrium* is found in soil and aquatic habitats.
- *Exophiala* and *Cryptococcus* are also known to inhabit soil.

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*. 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.  

- *Cladosporium* and *Penicillium* are found in soil while *Mylitinioid* are found on wood. 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. These results illustrate that the ITS Metagenomics Workflow can detect fungal genera in low-diversity environments.
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

<table>
<thead>
<tr>
<th>Product Name</th>
<th>Catalog number</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiSeq System</td>
<td>SY-410-1003</td>
</tr>
<tr>
<td>MiSeq Reagent Kit v3 (150-cycle)</td>
<td>MS-102-3001</td>
</tr>
<tr>
<td>iSeq 100 System</td>
<td>20021532</td>
</tr>
<tr>
<td>iSeq 100 s1 Reagent Kit (318-cycle)</td>
<td>20021533</td>
</tr>
<tr>
<td>Nextera XT Index Kit v2 Set A (96 indexes, 384 samples)</td>
<td>FC-131-2001</td>
</tr>
<tr>
<td>PhiX Control v3</td>
<td>FC-110-3001</td>
</tr>
</tbody>
</table>

Learn More

- www.illumina.com/systems/sequencing-platforms/iseq.html
- www.illumina.com/systems/sequencing-platforms/miseq.html

ITS Metagenomics Protocol

References