Supplemental Methods

Sequence Analysis

Sequence data was processed using QIIME [1]. Sequences were quality trimmed and assigned to their respective sample based on their barcodes. Sequences were binned into de novo Operational Taxonomic Units (OTUs) using CDHIT [2]. with a 97% minimum sequence identity threshold for 16S and a 95.2% minimum sequence identity threshold for ITS. The most abundant sequence from each OTU was selected as the representative sequence for that OTU.

For the ITS analysis, spacer sequences of lengths 130 to 1000 nt were used for the analysis. In a previous study [3], a longer minimum was used, which resulted in loss of the unusually short Cyberlindnera jadinii (aka Pichia jadinii and Candida utilis), which is only 142 nt in length. We also note that length variation in the ITS region may have resulted in under-estimation of the abundance of Saccharomyces cerevisiae, which has a relatively long ITS region (368 nt).

Taxonomy was assigned to the representative sequences using Ribosomal Database Project (RDP) for 16S [4] and the UNITE database for the ITS. The bacterial sequences were then NAST aligned using the Greengenes' reference database [5] and used to build a phylogenetic tree using the using the FastTree algorithm [6]. Bacterial community distances were calculated between all pair of samples using Unifrac [7]. UniFrac distances are based on the fraction of branch length shared in a phylogenetic tree between two samples' microbial communities. Weighted UniFrac incorporate the

relative abundances of each OTU [8]. For the fungal community distances, Jaccard and abundance-weighted Jaccard indices were calculated.

Principal Coordinate Analysis (PCoA) based on UniFrac distances was used to compare samples. Heatmaps that show the taxonomic distribution of each sample's sequences were created using Qiimer (http://cran.r-project.org/web/packages/qiimer). Diversity was assessed using the Shannon diversity metric. LEFSE [9] was used to identify taxa that differed between IBD and healthy samples.

Supplemental References.

1. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;**7**:335-6.
2. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 2006;**22**:1658-9.
3. Hoffmann C, Dollive S, Grunberg S, et al. Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. PLoS One. 2013;8:e66019
4. Cole JR, Wang Q, Cardenas E, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 2009;**37**:D141-5.
5. DeSantis TZ, Jr., Hugenholtz P, Keller K, et al. NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic Acids Res 2006;**34**:W394-9.
6. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 2009;**26**:1641-50.
7. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 2005;**71**:8228-35.
8. Lozupone CA, Hamady M, Kelley ST, et al. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Appl Environ Microbiol 2007;**73**:1576-85.
9. Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation. Genome Biol 2011;**12**:R60.