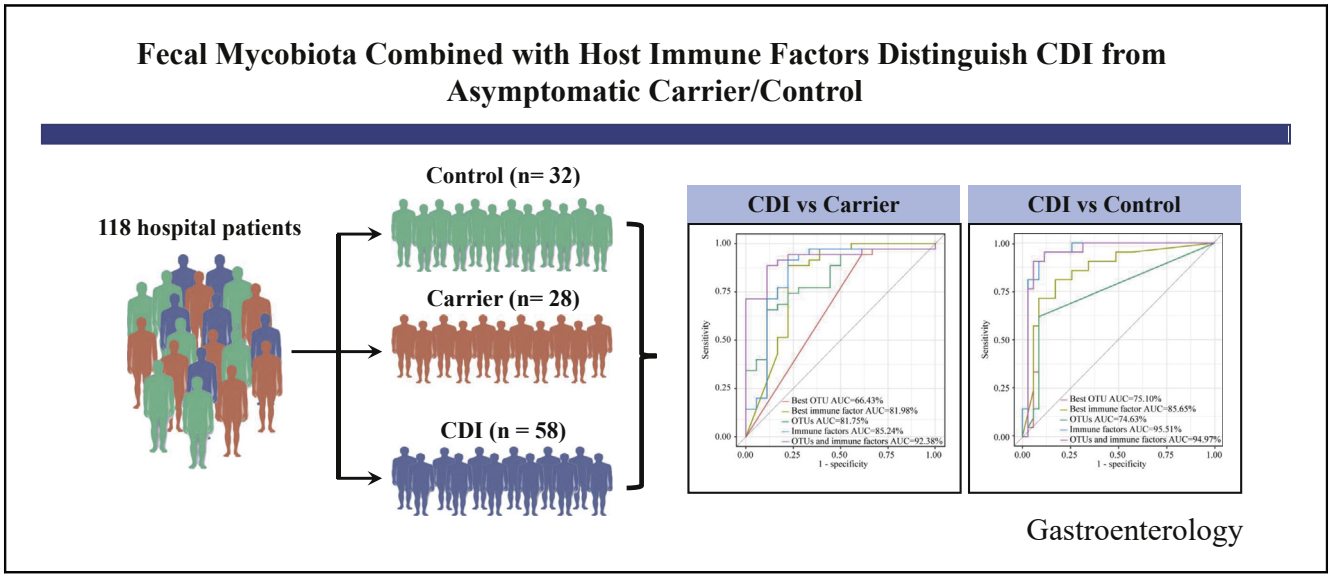
CLINICAL AT

Fecal Mycobiota Combined With Host Immune Factors Distinguish *Clostridioides difficile* Infection From Asymptomatic Carriage

Yangchun Cao,[1,2,*] **Lamei Wang**,[1,2,*] **Shanlin Ke**,[3,*] Javier A. Villafuerte Gálvez,[2] Nira R. Pollock,[4,5] Caitlin Barrett,[2] Rebecca Sprague,[2] Kaitlyn Daugherty,[2] Hua Xu,[2] Qianyun Lin,[2] Junhu Yao,[1] Yulin Chen,[1] Ciarán P. Kelly,[2] Yang-Yu Liu,[3] and Xinhua Chen[2]

[1]College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi Province, China; [2]Division of Gastroenterology, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts; [3]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; [4]Division of Infectious Diseases, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts; and [5]Department of Laboratory Medicine, Boston Children's Hospital, Boston, Massachusetts



BACKGROUND & AIMS: Although the role of gut microbiota in *Clostridioides difficile* infection (CDI) has been well established, little is known about the role of mycobiota in CDI. Here, we performed mycobiome data analysis in a well-characterized human cohort to evaluate the potential of using gut mycobiota features for CDI diagnosis. **METHODS:** Stool samples were collected from 118 hospital patients, divided into 3 groups: CDI (n = 58), asymptomatic carriers (Carrier, n = 28), and Control (n = 32). The nuclear ribosomal DNA internal transcribed spacer 2 was sequenced using the Illumina HiSeq platform to assess the fungal composition. Downstream statistical analyses (including Alpha diversity analysis, ordination analysis, differential abundance analysis, fungal correlation network analysis, and classification analysis) were then performed. **RESULTS:** Significant differences were observed in alpha and beta diversity between patients with CDI and Carrier ($P < .05$). Differential abundance analysis identified 2 genera (*Cladosporium* and *Aspergillus*) enriched in Carrier. The ratio of Ascomycota to Basidiomycota was dramatically higher in patients with CDI than in Carrier and Control ($P < .05$). Correlations between host immune factors and mycobiota features were weaker in patients with CDI than in Carrier. Using 4 fungal operational taxonomic units combined with 6 host immune markers in the random forest classifier can achieve very high performance (area under the curve ~92.38%) in distinguishing patients with CDI from Carrier. **CONCLUSIONS:** Our study provides specific markers of stool fungi combined with host immune factors to distinguish patients with CDI from Carrier. It highlights the importance of gut mycobiome in CDI, which may have been underestimated. Further studies on the diagnostic applications and therapeutic potentials of these findings are warranted.

Keywords: *C difficile*; Diagnostics; Gut Mycobiome; Immune Response.

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

It is an existing challenge for clinicians who care for patients with *Clostridioides difficile* infection to distinguish active infection from *C difficile* carriage. Little is known about the role of mycobiome in *C difficile* infection.

NEW FINDINGS

Mycobiota appears to be an important component of microbial dysbiosis associated with *C difficile* infection. Fungal operational taxonomic units combined with host immune factors provide high power for distinguishing *C difficile* infection from Carrier.

LIMITATIONS

This was a cross-sectional study of 58 patients with *C difficile* infection, 28 Carrier and 32 Control from a single hospital. Further studies are needed in other geographical regions and larger populations to validate these findings.

IMPACT

Fecal mycobiota combined with host immune factors may be useful biomarkers to distinguish *C difficile* infection from Carrier.

Clostridioides difficile remains the leading cause of health care–associated infectious diarrhea and is responsible for 500,000 illnesses and up to 30,000 deaths annually in the United States.[1] Exposure to *C difficile* can lead to asymptomatic carriage (presence of toxinogenic *C difficile* in the colon, but no symptoms), or *C difficile* infection (CDI) with a range of clinical presentations (ranging from mild diarrhea to severe colitis and/or death).[2] Asymptomatic *C difficile* colonization refers to the shedding of *C difficile* in stool but without diarrhea or other clinical symptoms.[3] Previous studies revealed that asymptomatic *C difficile*–colonized patients in the acute care setting may be protected from progression to infection because they can mount a humoral immune response to *C difficile* toxins.[4] Toxin-targeting treatments, such as vaccines and monoclonal antibodies, may protect against CDI recurrence but are unlikely to prevent asymptomatic colonization with *C difficile*.[5,6]

Currently, no diagnostic method can accurately differentiate CDI from *C difficile* colonization. This constitutes a critical unmet need in clinical care. Literature regarding colonized patients varies significantly in the patient inclusion criteria, tested material, and applied diagnostic and gold standard tests. In addition, various diagnostic screening tests have been used to detect *C difficile*, frequently divided into assays to recognize toxinogenic or nontoxinogenic strains.[3] Our previous study also revealed that neither stool toxin concentration nor nucleic acid amplification testing (NAAT) cycle threshold value can reliably distinguish a symptomatic CDI patient from a *C difficile*–colonized patient with diarrhea due to other causes.[2,7] Therefore, novel diagnostic markers for differentiation of CDI from asymptomatic carriers (Carrier) are urgently needed.

The human gastrointestinal tract harbors a complex and diverse community of commensal microorganisms, providing a variety of beneficial effects to the host. They contribute to the maintenance of intestinal homeostasis and epithelial integrity and exert anti-inflammatory effects by interacting with the mucosal immune system.[8] A healthy microbiome, composed of diverse communities of bacteria, viruses, fungi, protozoa, and archaea, offers colonization resistance against pathogens through various mechanisms.[9] Hence, disruption of the microbiome (also known as microbial dysbiosis) due to immunodeficiency, chemotherapy, antibiotic use, or other factors, is known to increase the risk of CDI by disrupting the gut microbiome's ability to resist pathogen colonization or by weakening the intestinal barrier.[10]

Although growing evidence supports the importance of the gut microbiota[11,12] and bacteriophages[13] in the pathogenesis of CDI, the potential role of the fungal component of the gut microbiota, namely the gut mycobiota, in CDI has long been overlooked. A few existing studies focused on the gut mycobiota comparison between patients with CDI and healthy Controls.[14] There is a paucity of literature studying the gut mycobiota difference between patients with CDI and Carrier. This represents a significant knowledge gap that warrants filling and can be essential for understanding the overall gut microbiota dysbiosis associated with CDI.

We hypothesize that the fecal mycobiota can serve for CDI diagnosis purposes. To test this hypothesis, we analyzed mycobiota data analysis of 118 hospitalized individuals that consist of patients with CDI (n = 58), antibiotic-exposed Carrier (n = 28), and antibiotic-exposed asymptomatic noncarriers (n = 32). In this study, we aimed to profile gut mycobiota using internal transcribed spacer 2 (ITS2) sequencing of stool samples from these individuals. Infection with *C difficile* leads to both adaptive and innate immune responses.[15] Our several previous studies revealed that adaptive immune responses to *C difficile* toxins have been associated with symptomless carriage.[16,17] Meanwhile, *C difficile* and its toxins are potent activators of innate

*** Authors share co-first authorship.**

Abbreviations used in this paper: **ANCOM, analysis of composition of microbiomes; AUC, area under the curve; Carrier, asymptomatic carriers; CDI, *Clostridioides difficile* infection; GCSF, granulocyte colony-stimulating factor; Ig, immunoglobulin; IL, interleukin; ITS2, internal transcribed spacer 2; MCP1, monocyte chemoattractant protein 1; NAAT, nucleic acid amplification testing; OTU, operational taxonomic unit; PERMANOVA, permutational multivariate analysis of variance; TNF, tumor necrosis factor.**

Most current article

0016-5085/$36.00
https://doi.org/10.1053/j.gastro.2021.02.069

CLINICAL AT

immune responses in vitro and in vivo.[18] Our previous study also showed that specific serum markers of innate and adaptive immunity can distinguish CDI from Carrier.[7] Hence, here we also aimed to determine whether these immune factors combined with specific fungal markers could further increase the discriminative power between CDI and Carrier.

Methods

Study Participants

The background and design of this cohort has been detailed in our previous study.[2,7] Concisely, all individuals were adults (age >18 years). Patients with CDI were inpatients with positive clinical stool NAAT result, new-onset diarrhea, and a decision to treat for CDI. The diagnostic clinical stool sample was captured as a discarded sample; a discarded serum sample collected within 1 day of that stool sample was also captured. Patients were excluded if the diagnostic stool specimen was more than 72 hours old, if they had received CDI treatment for more than 24 hours before stool collection, or if they had a colostomy. Carriers were admitted for at least 72 hours, had received at least 1 dose of an antibiotic within the past 7 days, and did not have diarrhea in the 48 hours before stool sample collection, but had positive NAAT results on stool testing and were not treated for CDI. Patients with 2 or more loose stools within a 24-hour period were excluded; patients with 1 loose stool were included only if providers had recently administered a laxative. Patients were excluded if they had a colostomy; received oral or intravenous metronidazole, oral vancomycin, oral rifaximin, and/or oral fidaxomicin for more than 24 hours within the prior 7 days; had been diagnosed with CDI in the past 6 months; or had tested negative for *C difficile* within the past 7 days. Stool samples were collected prospectively under verbal informed consent. A discarded serum sample from within 1 day of the stool sample was also captured. Control groups included individuals without diarrhea who had screened as eligible for the Asymptomatic Carrier group but were NAAT negative on stool testing. Discarded serum samples were captured within 1 day of the stool sample. Patients who had antifungal medication 7 days before sample collection were excluded.

Serum Immune Marker Measurement

The measurement of host serum cytokine concentrations of interleukin (IL)2, IL4, IL6, IL8, IL10, IL13, IL15, IL1β, granulocyte colony-stimulating factor (GCSF), monocyte chemoattractant protein 1 (MCP1), vascular endothelial growth factor-A, and tumor necrosis factor (TNF)α was performed using a Milliplex magnetic bead kit and Luminex analyzer (MAGPIX) (Millipore Sigma, Inc., Burlington, MA) as per the manufacturer's instructions. Serum antibody levels against *C difficile* toxins A (anti-toxin A immunoglobulin [Ig] A, anti-toxin A IgG, and anti-toxin A IgM) and B (anti-toxin B IgA, anti-toxin B IgG, and anti-toxin B IgM) were measured by semi-quantitative enzyme-linked immunosorbent assay. All the experimental details have been reported previously.[2,7]

Fungal ITS2 Sequencing and Bioinformatics Analysis

ITS2 sequencing was conducted on an Illumina HiSeq platform (Illumina Hiseq 2500). Details of fecal DNA isolation, ITS2 sequencing, library preparation, data processing, and bioinformatics analysis are available online as supplementary methods.

Data Analysis

Alpha (ie, within-sample) diversity measures: Chao1 (estimated richness) and Shannon diversity of any 2 groups were compared using Wilcoxon rank sum test.[19] Permutational multivariate analysis of variance (PERMANOVA) was performed with the default 999 permutations based on the Adonis and the Bray-Curtis and unweighted UniFrac distance.[20] Note that in the PERMANOVA tests, we only included subjects with known information of age, sex, race, and ethnicity. ANCOM was conducted after removing spurious observations using default parameters with a Benjamini-Hochberg correction significance threshold of 0.05.[21] The P values of Ascomycota to Basidiomycota ratio was calculated based on Wilcoxon rank sum test.[22] Microbial correlation network was constructed using SparCC.[23] Correlated genus pairs were selected if the absolute value of sparse correlation $|r| > 0.1$ and $P < .05$. All statistical analysis was performed using R, except SparCC analysis (based on python).

Results

Study Population

Fecal samples from our clinical cohorts of 118 patients from Beth Israel Deaconess Medical Center were prospectively collected, including 58 stool samples of the CDI cohort, 28 of Carrier, and 32 stool samples of Control subjects. There was no difference in clinical characteristics of the participants including sex, age, gender, and ethnicity among the 3 groups ($P > .05$, Supplementary Table 1) as described in detail previously.[7] PERMANOVA showed that cohorts and clinical characteristics of the participants such as age, sex, race, and ethnicity had no significant effect on the mycobiome composition ($P > .05$, Supplementary Table 2).

Characteristics of the Sequence Datasets

With fungal ITS2 region sequencing, the total number of sequences was 7,418,956, with an average of 62,344 reads per sample, the average length of the reads was approximately 383 base pairs. Sequences were clustered into 712 operational taxonomic units (OTUs) based on their shared sequence similarity at a 97% threshold. Overall, a total of 6 phyla, 26 classes, 74 orders, 165 families, and 279 genera from fungi were identified, and 410 OTUs were identified to the species level.

Ecological Features of the Fecal Fungal Communities

A Shannon-Wiener curves analysis was performed to evaluate whether we obtained sufficient sequencing

sampling reads to perform a meaningful ITS2 analysis. The number of OTUs plateaued in all samples as the sample sequencing reads increased (Supplementary Figure 1*A*), suggesting that we acquired a sufficient number of sequencing sampling reads to reach plateau levels. All samples had a good depth of coverage as indicated by the Good's coverage estimates (>99.98%, data not shown).

The Venn diagram depicts those OTUs that were unique to 3 cohorts, or shared by them. Venn diagram showed that 128 of the total 712 OTUs were shared among the 3 groups, whereas 466 of 712 OTUs were unique for 3 groups (Supplementary Figure 1*B*).

To assess the variations of fungal biodiversity, the Chao1 index (estimated richness) and Shannon diversity were used to compare the 3 groups at the OTU levels. Compared with the Carrier and Control groups, the fungal richness and diversity were significantly decreased in the CDI group ($P <$ .01; Figure 1*A* and *B*).

To display fungal community composition among cohorts, we performed principal coordinate analysis using Bray-Curtis and the unweighted UniFrac distance. These data indicated that the fungal compositions of patients with CDI vary more prominently than Carrier. As expected, significant differences of fungal compositions were observed between CDI and Carrier ($P <$ .05) when analyzed by pairwise tests (Figure 1*C* and *D*). Interestingly, no significant difference was observed between Carrier and Control ($P >$ .05; Figure 1*C* and *D*). Meanwhile, by directly comparing the beta diversity of each group, we found that the CDI group had the largest variability, whereas the Carrier group showed lower variability (Figure 1*E* and *F*), indicating that the fungal compositions of participants within the CDI group vary more prominently than Carrier.

Taxonomic Composition of the Gut Mycobiota

Fungal phyla of Ascomycota, Basidiomycota, and unclassified fungi, together accounting for up to 90% of sequences on average, were the 3 dominant taxa in all 3 groups (Supplementary Figure 2). Fungal genera of *Saccharomyces*, *Candida*, *Nakaseomyces*, and *Penicillium* were the dominant taxa among these groups (Figure 2). The ternary plot showed that Carrier shared higher proportions of fungal communities (at the genus level) with Control subjects than with patients with CDI (Supplementary Figure 3*A*). Further classification at the genus level, a hierarchical heat map of the relative abundance of top-30 most abundant fungal genera (Supplementary Figure 3*B*) indicated that fungal communities of those 3 groups were quite unique.

Ascomycota:Basidiomycota Ratio

A previous work revealed that gut mycobiota is dysbiotic in patients with inflammatory bowel disease with much lower Ascomycota:Basidiomycota ratio than that of healthy Controls,[22] which prompts us to study the Ascomycota:Basidiomycota ratio in our cohort. Interestingly, we found that the Ascomycota:Basidiomycota ratio was dramatically higher in CDI than in Carrier ($P <$ .05, Figure 3). These results suggested that the Ascomycota:Basidiomycota ratio could represent a fungal dysbiosis index to differentiate CDI from Carrier.

Fungal Differential Abundance Analysis

When conducting differential abundance analysis, ANCOM detected 2 differentially abundant fungal genera (Supplementary Table 3; Figure 4), including genera *Aspergillus* and *Cladosporium*. ANCOM also detected 2 differentially abundant fungal OTUs (OTU657: *Aspergillus proliferans*; OTU252: *unclassified_g_Cladosporium*, an unclassified OTU within genus *Cladosporidium*) between CDI and Carrier, also detected 3 differentially abundant fungal OTUs (OTU252; OTU584: *unclassified_g_Aspergillus*; OTU687: *Candida dubliniensis*) between CDI and Control (Supplementary Table 4; Figure 4). No differentially abundant fungal genera or OTUs were found between Carrier and Control. These results suggested that differentially abundant fungal genera or OTUs could be used as potential biomarkers to differentiate CDI from Carrier.

Fungal Correlation Networks

To compare the fungal communities of the 3 groups at the network level, we constructed the fungal correlation network for each group using SparCC[23] (sparse correlations for compositional data). We found that the fungal correlation network of the CDI group has quite different structure compared with the other 2 groups. The overall fungal correlations in the CDI group are much weaker than those in the Carrier group (Figure 5). We also observed the disappearance of some fungal correlations in CDI compared with Carrier and Control. Strong positive correlations were found among *Aspergillus*, *Cladosporium*, and *Saccharomyces*, whereas Ascomycota and Basidiomycota exhibited the strong negative correlation in all 3 cohorts ($P <$.05, data not shown).

Diagnostic Accuracy of CDI Classification Based on Host Immune Markers and Gut Mycobiota

To illustrate the diagnostic power of fecal mycobiota and immune factors, we constructed a random forest classifier to distinguish CDI from Carrier or Control. The classification performance was evaluated by the area under the curve (AUC) of the receiver operating characteristic. In classifying CDI and Carrier, we found that OTU486 (unclassified_o_*Pleosporales*) is the top feature with AUC ~0.664, and GCSF remains as the top immune feature as we previously reported[7] with AUC ~0.820 (Figure 6*A*). For the optimal marker sets of OTUs (or immune factors), we achieved AUC ~0.818 (or 0.8524), respectively. Notably, combining features of fungal OTUs with immune factors reached a superior classification with AUC ~0.924. The optimal set consisted of 4 fungal OTUs (OTU657: *Aspergillus_proliferans*, OTU35: unclassified fungi, OTU252: unclassified_g_*Cladosporium*, and OTU486) and 6 immune

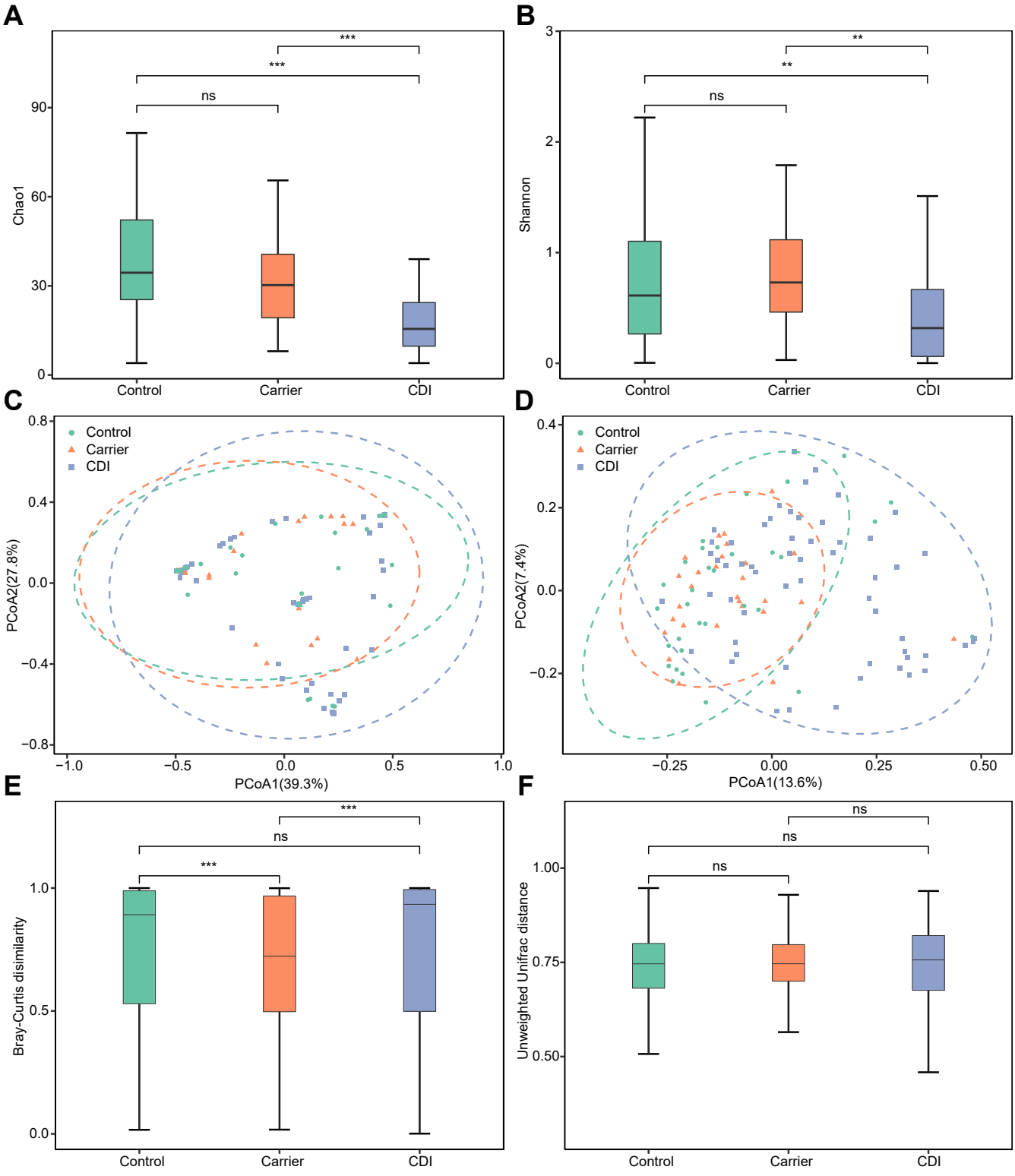


Figure 1. Alpha and beta diversities, and ordination analysis of the gut mycobiota with 3 distinct phenotypes: Control, Carrier and CDI. (*A*, *B*) The alpha diversity analysis was based on Chao1 index (*A*) and Shannon index (*B*). (*C*, *D*) Principal coordinate analysis (PCoA) of the fungal compositions at the OTU level based on the Bray-Curtis dissimilarity (*C*) and unweighted UniFrac distance (*D*). The ellipses represent the 95% of the samples belonging to each group. Dissimilarity was analyzed using Adonis statistical tests with 999 permutations based on Bray-Curtis dissimilarity: CDI vs Carrier ($R^2 = 0.0299$, $P = .032$), CDI vs Control ($R^2 = 0.0121$, $P = .337$), and Carrier vs Control ($R^2 = 0.0147$, $P = .491$). Similar analysis based on unweighted UniFrac distance yielded: CDI vs Carrier ($R^2 = 0.0363$, $P = .001$), CDI vs Control ($R^2 = 0.0361$, $P = .001$) and Carrier vs Control ($R^2 = 0.0118$, $P = .957$). (*E*, *F*) The beta diversity analysis was based on Bray-Curtis dissimilarity (*E*) and unweighted UniFrac distance (*F*). ns: $P > .05$, $^*P < .05$, $^{**}P < .01$, $^{***}P < .001$.

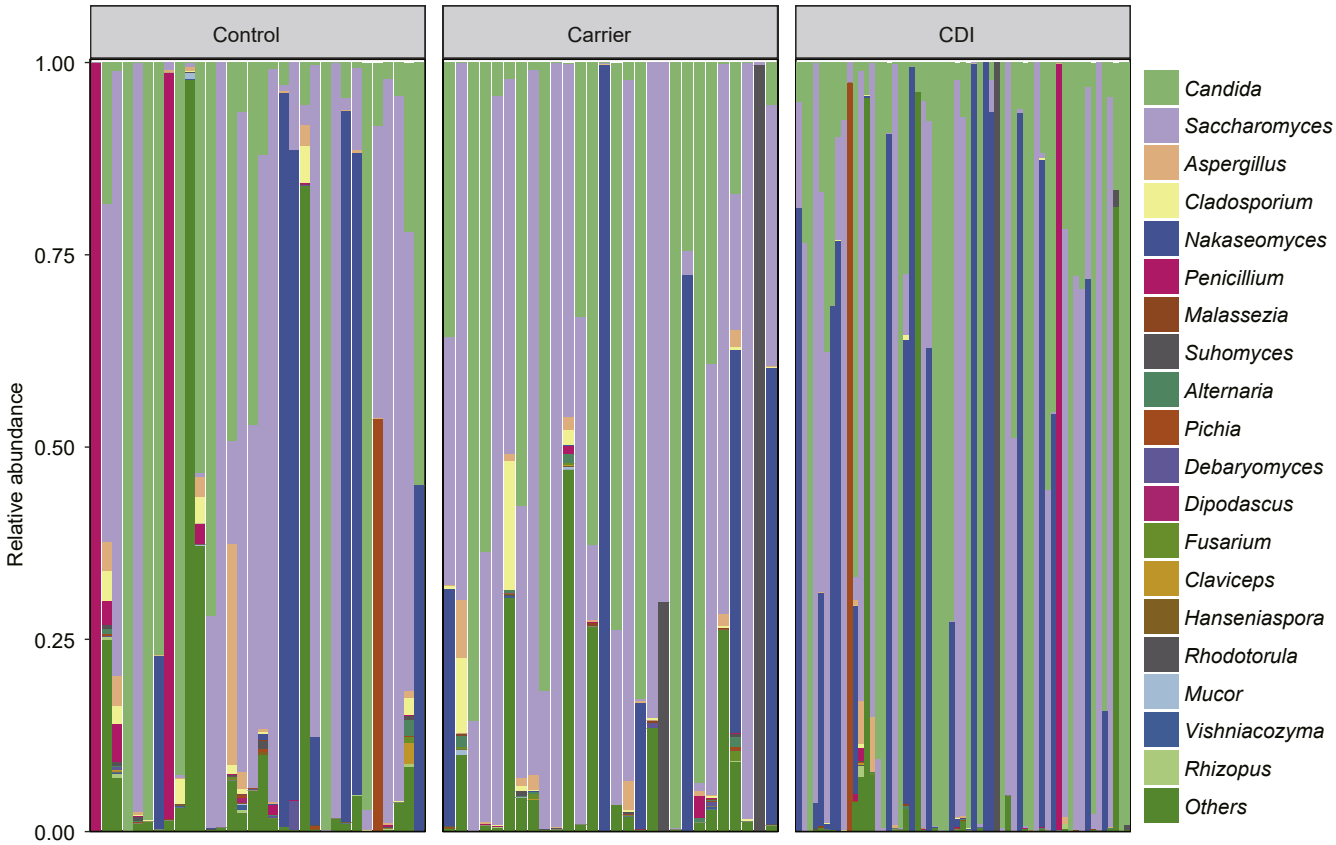


Figure 2. Genus-level taxonomic profiles of the gut mycobiota from 3 distinct phenotypes: Control, Carrier, and CDI. Only genera with ≥1% abundance in at least 1 sample were depicted. Otherwise, they were included in the category "others."

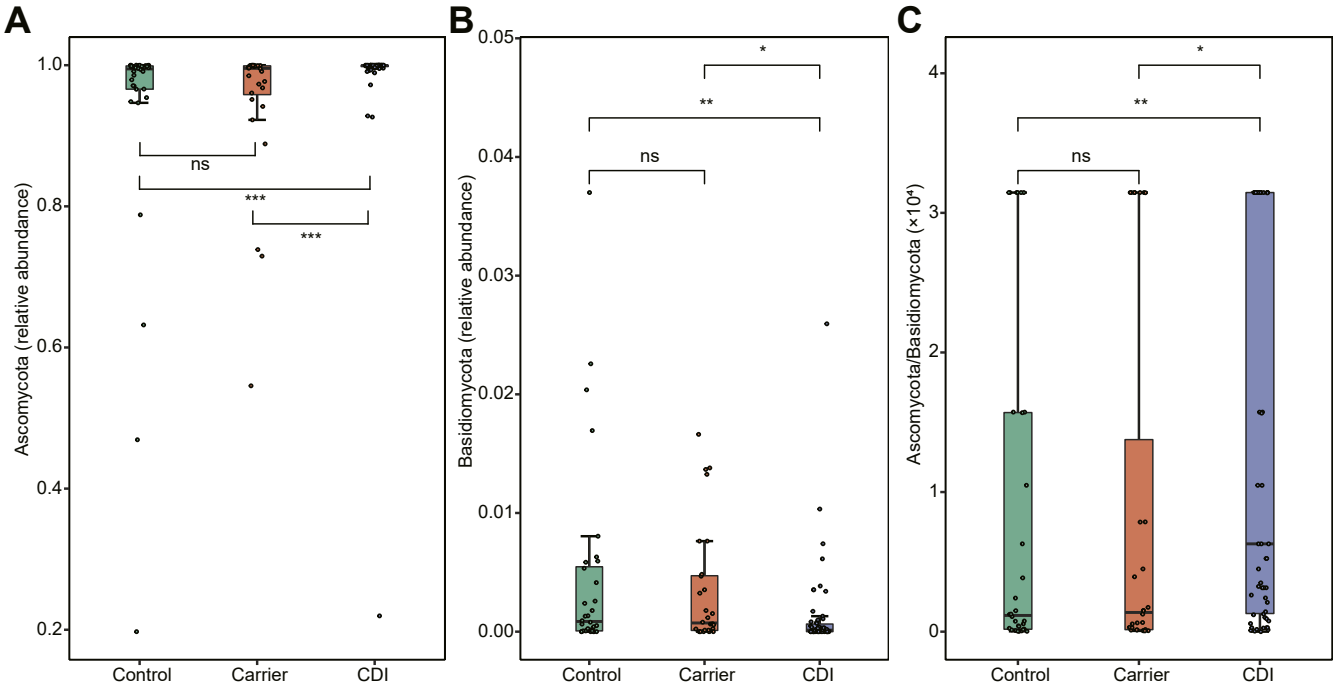


Figure 3. The relative abundance and ratio of Ascomycota to Basidiomycota of gut mycobiota from 3 distinct phenotypes: Control, Carrier, and CDI. (*A*) Ascomycota, (*B*) Basidiomycota, (*C*) Ascomycota to Basidiomycota ratio. Data are presented as median and 95% CI with *P* values based on Wilcoxon rank sum test. ns: $P > .05$, *$P < .05$, **$P < .01$, ***$P < .001$.

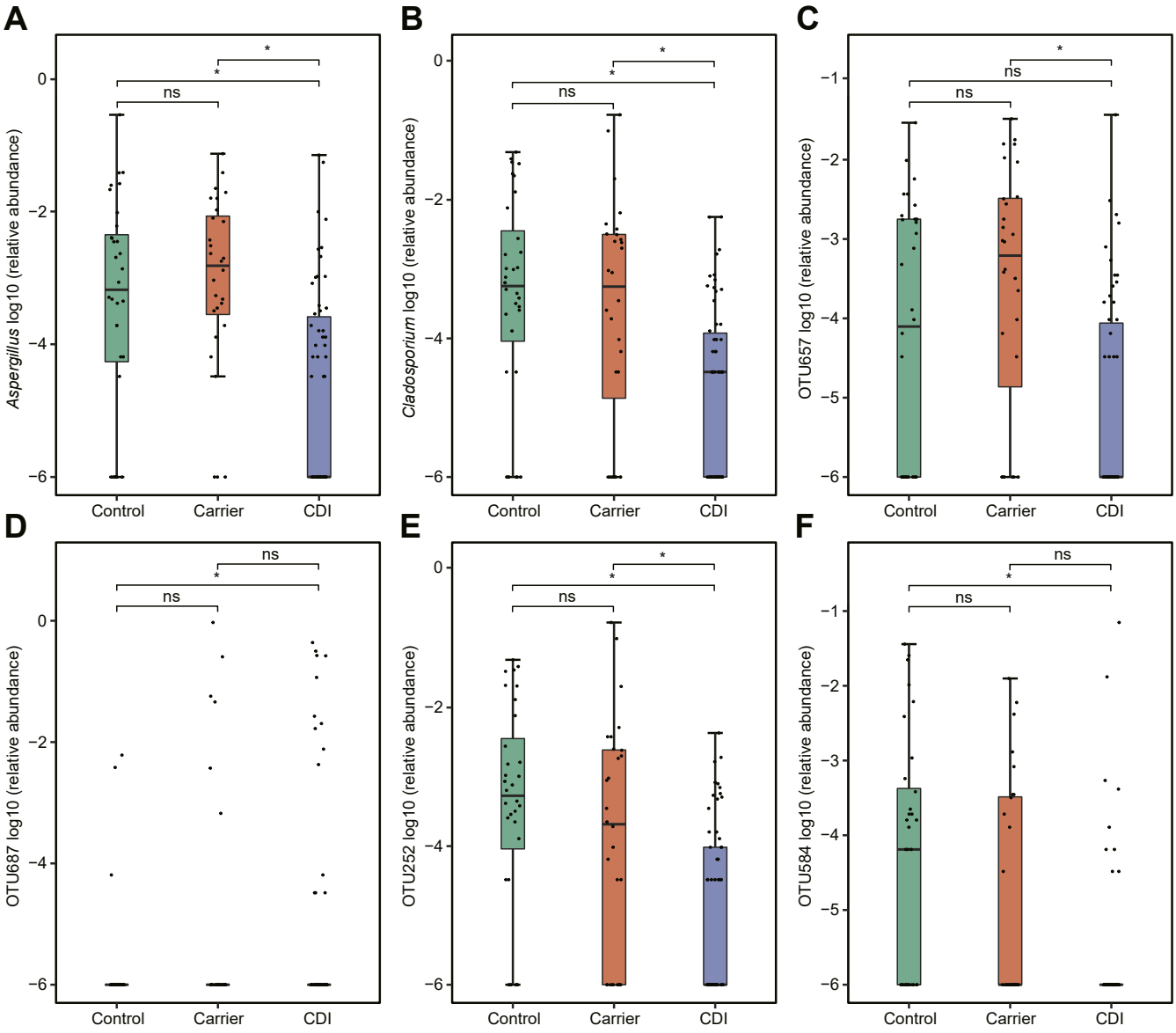


Figure 4. Differentially abundant fungal taxa among the 3 phenotypical groups: Control, Carrier, and CDI. Differentially abundant genera (*A*, *B*) and OTUs (*C*–*F*) were found using ANCOM. Note that for a taxon that is absent in most subjects, the interquartile range (difference between first quartile and third quartile) will be extremely small. ns: $P > .05$, $^*P < .05$.

markers (GCSF, IL6, IL8, IL10, TNFα, and IL4) (Figure 6*C*). In classifying CDI and Control, the mean AUC values were 0.751, 0.857, 0.746, 0.955, and 0.950 for the top OTU feature (OTU584: unclassified_g_*Aspergillus*), the top immune feature (GCSF), the optimal feature set of OTUs, the optimal feature set of immune factors, and the optimal combined feature set of fungal OTUs and immune factors, respectively (Figure 6*B*). The optimal combined feature set consisted of 1 fungal OTU (OTU584) and 5 immune markers (GCSF, TNFα, IL6, IL4, and MCP1) (Figure 6*D*). These results suggested that the random forest classifier based on a combined feature set of fungal OTUs and immune factors can achieve a powerful diagnostic performance in differentiating CDI from the Carrier (or Control) group.

Correlation Between Serum Biomarkers and Mycobiota Features

To reveal the interplay between the gut mycobiome and the host immune system, we calculated the correlations between fungal compositions (at the genus level) and the circulating levels of host immune markers. A total of 20 serum immune factors were measured for correlation with mycobiota features (Figure 7). Overall, the 3 groups have quite different correlations between gut fungal genera and host immune factors. More strongly positive associations between gut fungal genera and host immune factors were found in Carrier than in CDI. For example, in each group, we focused on the correlations between 2 main different genera (*Saccharomyces* and *Aspergillus*) with host immune factors.

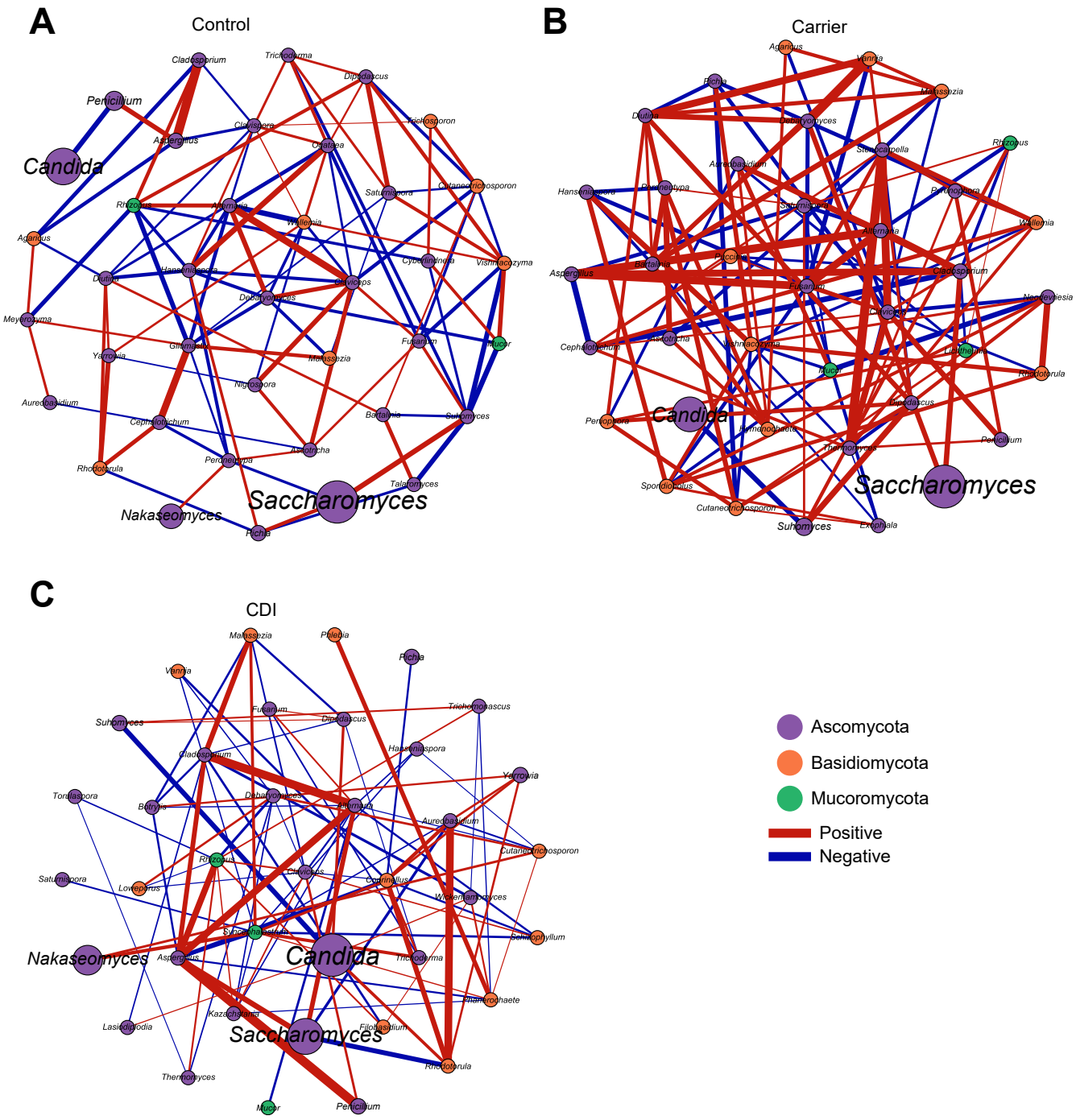


Figure 5. Fungal correlation networks of the 3 phenotypical groups: Control (*A*), Carrier (*B*), and CDI (*C*). Nodes represent genera and are colored based on their phylum. Edges represent fungal correlations: green/red means positive/negative correlations, respectively. Edge thickness indicates the absolute value of correlation coefficient, and only the high confidence interactions ($P < .05$) with high absolute correlation coefficients (> 0.1) were presented.

In the CDI group, we observed negative associations between *Saccharomyces* and IL6, GCSF. In the Carrier group, we observed positive associations between *Aspergillus* and IL1β, IL8, and TNFα, positive associations between *Saccharomyces* and MCP1, and negative associations between *Saccharomyces* and anti-toxin B IgA and anti-toxin A IgM. In the Control group, *Saccharomyces* was significantly positively associated with IL4. These results indicated that the correlations between gut fungal genera and host immune factors can be very sensitive to the colonization/infection status.

Discussion

Our study is the first to report a diagnostic model using fecal fungal OTUs and serum immune markers with a

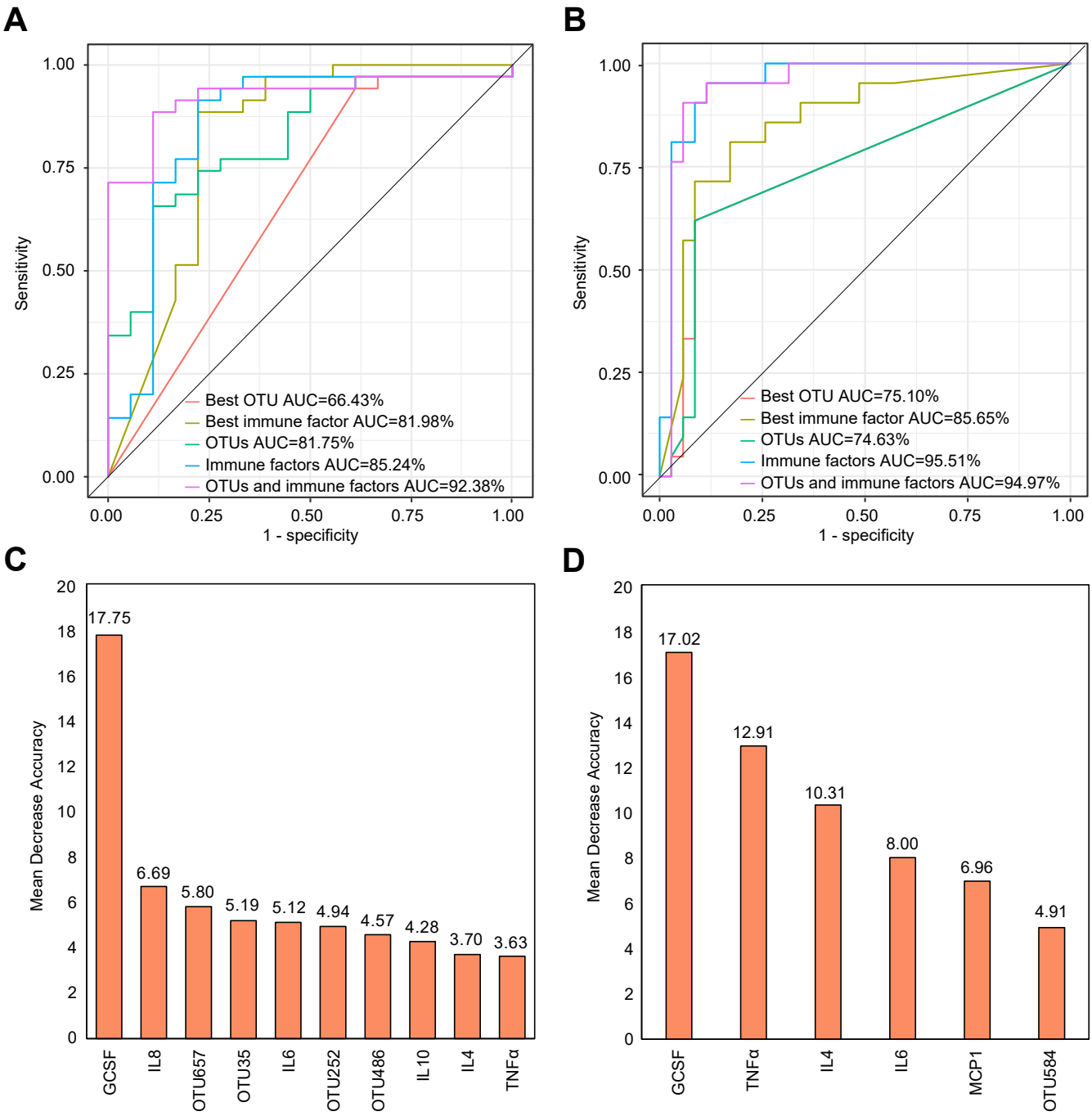


Figure 6. Classification analyses based on random forest models. (*A*, *C*) CDI vs Carrier. (*B*, *D*) CDI vs Control. For each classification analysis, we tried different types of features: best OTU, best immune factor, all OTUs, all immune factors, all OTUs, and immune factors. The receiver operating characteristic curves are shown in (*A*) and (*B*). The top features ranked based on their mean decrease accuracy are shown in (*C*) and (*D*). The lengths of the bars in the histogram represent the mean decrease accuracy, which indicates the importance of features (OTUs and immune factors) for classification. OTU657: *Aspergillus_proliferans*, OTU35: *unclassified fungi*, OTU252: *unclassified_g_Cladosporium*, OTU486: *unclassified_o_Pleosporales*, OTU584: *unclassified_g_Aspergillus*.

powerful diagnostic potential to differentiate patients with CDI from Carrier. We found that fungal alpha diversity (richness and diversity) and beta diversity were significantly lower in the CDI group compared with the Carrier group; and the abundance of several fungi at the phylum and genus levels between these 2 groups significantly differed. The Ascomycota:Basidiomycota ratio could represent a fungal dysbiosis index to differentiate CDI from Carrier and Control. The marked differences in the associations between mycobiome features and serum cytokines in the 3 different cohorts suggests interactions between the host systemic immune response and the gut mycobiome.

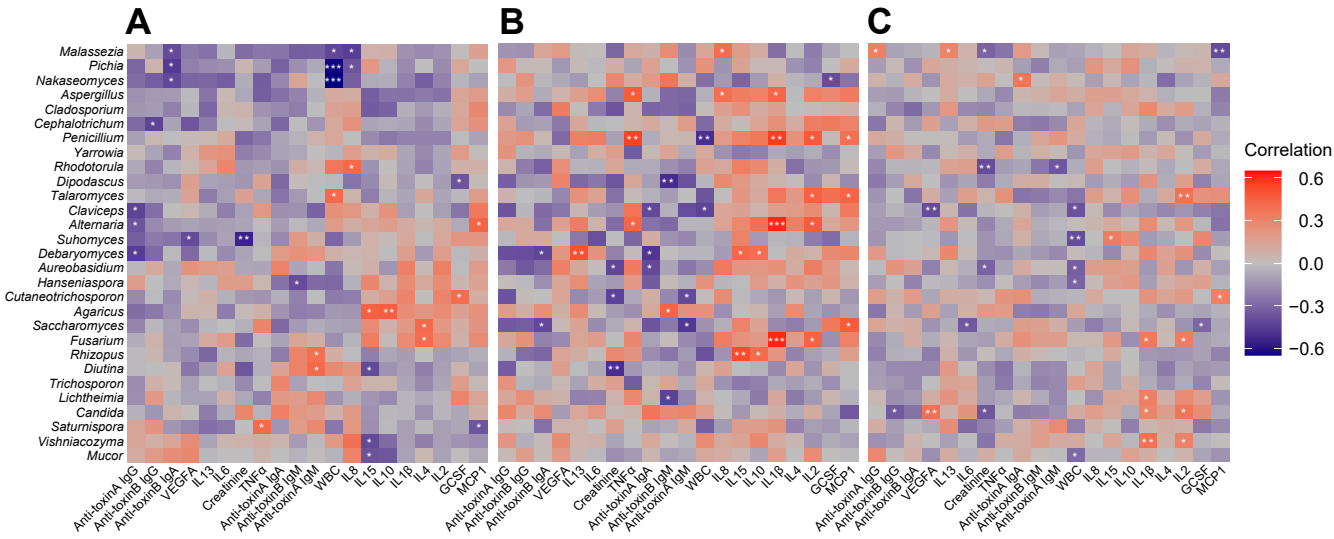


Figure 7. Spearman correlations between fungal abundances in stool samples and the circulating levels of host immune markers in serum samples from the 3 phenotypic groups: Control (*A*), Carrier (*B*), and CDI (*C*). For each heat map, rows correspond to fungal taxa at the genus level, columns correspond to immune factors. Red and blue represent the positive and negative correlations, respectively. The intensity of the colors denotes the degree of correlation between the genera abundances and the circulating levels of host serum immune factors. $^{*}P < .05$, $^{**}P < .01$, $^{***}P < .001$.

Comparisons between healthy and diseased cohorts have highlighted the importance of class discovery (detecting novel subtypes of a disease) and class prediction (forecasting the disease subtype of an individual or group).[14,24–26] Previous studies revealed that host inflammatory markers (including serum cytokines, calprotectin, and fecal lactoferrin) have diagnostic potential; however, they are not disease specific in CDI, and therefore are imperfect biomarkers.[27] Our classification analysis based on the optimal fungal OTU features achieved a powerful classification potential for distinguishing CDI from Asymptomatic Carrier (AUC ∼0.818). As we integrated gut fungal OTUs and host immune markers, we identified a specific immune-mycobiota signature for CDI that further enhances the classification performance in differentiating CDI from Asymptomatic Carrier (AUC ∼0.924). Our previous study found that serum GCSF concentration alone can achieve AUC ∼0.842 in discriminating CDI from Carrier.[7] Hence, the addition of gut fungal OTUs further enhances the discriminative power of GCSF.

We found that fecal fungal richness and diversity were significantly decreased in CDI compared with Carrier and Control groups. This is consistent with previous findings comparing CDI with healthy Controls.[14] Beta diversity was significantly different between CDI and Carriers, which is also consistent with previous reports in which principal coordinate analysis revealed significant clustering of samples between CDI and non-CDI.[24,25] These results indicate a significant global shift in gut mycobiota between Carrier and CDI, suggesting that an altered fungal community might play a role in CDI pathogenesis. Thus, greater diversity or richness in the fungal community is a sign of a relatively healthy gut mycobiota, even in patients receiving antibiotics.

Our current study showed that phyla of Ascomycota and Basidiomycota were the 2 dominant taxa in the 3 groups, which is consistent with a previous study.[28] The most commonly reported fungi found in the human gastrointestinal tract includes members of the genera *Candida*, *Saccharomyces*, *Penicillium*, *Aspergillus*, and *Cladosporium*.[28] Accordingly, our study found genera *Saccharomyces* and *Candida* to be the 2 dominant taxa in all 3 cohorts. Furthermore, we also observe that the abundance of phyla Ascomycota has a strongly negative correlation with that of Basidiomycota. Thus, the Ascomycota:Basidiomycota ratio was higher in patients with CDI than in Asymptomatic Carrier, suggesting this imbalance between Ascomycota and Basidiomycota may be correlated with CDI pathobiology, and could be used as a biomarker to differentiate CDI from Carrier.

Although *Candida* was among the most abundant genera in CDI, it is not identified to be differentiating CDI from Carrier in our study. This may be due to the high heterogeneity within the *Candida* genera and to the difficulty in identifying fungi at the species level using our sequencing approach. Several previous studies have evaluated the relation between CDI and *Candida* colonization and/or disease, and both positive and negative associations have been reported.[14,29–32] In this study, the fungal genus *Saccharomyces* was found to be depleted in CDI, suggesting a potential beneficial role of *Saccharomyces* abundance in the gut, consistent with our previous studies.[22,33–36] Interestingly, *Saccharomyces* abundance had a strong negative correlation with IL6 in the CDI cohort. Previous observations described the serum IL6 concentration correlating with CDI severity and mortality.[37,38] Although the directionality of these correlations is unclear, a previous study found that *Saccharomyces cerevisiae* inhibits the transcription and translation of IL6 in enterocytes.[39] The differential abundance analysis of fungal taxonomic composition, as conducted by ANCOM in this study, detected 2 differentially

abundant genera between CDI and Carrier (or Control): *Aspergillus* and *Cladosporium*. The decrease of *Aspergillus* and *Cladosporium* in CDI compared with Carrier (or Control) may suggest a beneficial role of these fungi in patients at risk for CDI. Another study suggested that *Aspergillus penicillioides* was more enriched in healthy individuals than in CDI; treating patients with CDI with fecal microbiota transplantation restored the abundance of this species.[14] The present study is the first to report increased abundance of *Cladosporium* in Carriers/Control compared with CDI. Thus, *Cladosporium* could perhaps play a protective role in patients at risk for CDI. Network analyses established strong fungal abundance correlations in the Carrier/Control groups, which were absent in the CDI group. Furthermore, the fungal correlations in the CDI group were weaker than those in the Carrier/Control group. This could be interpreted as CDI being a state in which physiological fungal correlations are disrupted. The absence of these correlations may reflect mycobiota-immune cross-talk that could mediate disease susceptibility, the directionality of these interactions remains to be further studied.

In conclusion, we describe previously unknown characteristics of the gut mycobiota in the *C difficile* colonization-infection continuum, pinpoint fungal taxonomic units that may play key roles in CDI pathogenesis, and identified specific fungal markers with promising diagnostic features. Gut mycobiota-targeted biomarkers together with immune factors could become potential diagnostic tools to discriminate CDI from Carrier. However, studies with larger cohorts need to be done to further validate the findings before this test can be used in the clinical diagnostic settings. Systematic investigation of the key fungal genera or OTUs by metagenomic sequencing may further improve the diagnostic value of these markers for CDI. Nonetheless, a method simpler and cheaper than sequencing of the mycobiome will need to be further developed.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at https://doi.org/10.1053/j.gastro.2021.02.069.

References

1. Lessa FC, Mu Y, Bamberg WM, et al. Burden of *Clostridium difficile* infection in the United States. N Engl J Med 2015;372:825–834.
2. Pollock NR, Banz A, Chen X, et al. Comparison of *Clostridioides difficile* stool toxin concentrations in adults with symptomatic infection and asymptomatic carriage using an ultrasensitive quantitative immunoassay. Clin Infect Dis 2019;68:78–86.
3. McDonald LC, Gerding DN, Johnson S, et al. Clinical practice guidelines for *Clostridium difficile* infection in adults and children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). Clin Infect Dis 2018;66:987–994.
4. Shim JK, Johnson S, Samore MH, et al. Primary symptomless colonisation by *Clostridium difficile* and decreased risk of subsequent diarrhoea. Lancet 1998; 351:633–636.
5. Gerding DN, Johnson S. Management of *Clostridium difficile* infection: thinking inside and outside the box. Clin Infect Dis 2010;51:1306–1313.
6. Wilcox MH, Gerding DN, Poxton IR, et al. Bezlotoxumab for prevention of recurrent *Clostridium difficile* infection. N Engl J Med 2017;376:305–317.
7. Kelly CP, Chen X, Williams D, et al. Host immune markers distinguish Clostridioides difficile infection from asymptomatic carriage and non-*C. difficile* diarrhea. Clin Infect Dis 2020;70:1083–1093.
8. Huang J, Kelly CP, Bakirtzi K, et al. *Clostridium difficile* toxins induce VEGF-A and vascular permeability to promote disease pathogenesis. Nat Microbiol 2019; 4:269–279.
9. Seekatz AM, Young VB. *Clostridium difficile* and the microbiota. J Clin Invest 2014;124:4182–4189.
10. Nycz BT, Dominguez SR, Friedman D, et al. Evaluation of bloodstream infections, *Clostridium difficile* infections, and gut microbiota in pediatric oncology patients. PLoS One 2018;13:e0191232.
11. Samarkos M, Mastrogianni E, Kampouropoulou O. The role of gut microbiota in *Clostridium difficile* infection. Eur J Intern Med 2018;50:28–32.
12. Schubert AM, Rogers MA, Ring C, et al. Microbiome data distinguish patients with *Clostridium difficile* infection and non-*C. difficile*-associated diarrhea from healthy controls. MBio 2014;5:e01021-14.
13. Zuo T, Wong SH, Lam K, et al. Bacteriophage transfer during faecal microbiota transplantation in *Clostridium difficile* infection is associated with treatment outcome. Gut 2018;67:634–643.
14. Zuo T, Wong SH, Cheung CP, et al. Gut fungal dysbiosis correlates with reduced efficacy of fecal microbiota transplantation in *Clostridium difficile* infection. Nat Commun 2018;9:3663–3674.
15. Kelly CP, Kyne L. The host immune response to *Clostridium difficile*. J Med Microbiol 2011;60:1070–1079.
16. Kyne L, Warny M, Qamar A, et al. Asymptomatic carriage of *Clostridium difficile* and serum levels of IgG antibody against toxin A. N Engl J Med 2000; 342:390–397.
17. Kyne L, Warny M, Qamar A, et al. Association between antibody response to toxin A and protection against recurrent *Clostridium difficile* diarrhoea. Lancet 2001; 357:189–193.
18. Chen X, Kokkotou EG, Mustafa N, et al. *Saccharomyces boulardii* inhibits ERK1/2 mitogen-activated protein kinase activation both in vitro and in vivo and protects against *Clostridium difficile* toxin A-induced enteritis. J Biol Chem 2006;281:24449–24454.
19. 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;75:7537–7541.

20. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 2005;71:8228–8235.
21. Mandal S, Van Treuren W, White RA, et al. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb Ecol Health Dis 2015; 26:27663.
22. Sokol H, Leducq V, Aschard H, et al. Fungal microbiota dysbiosis in IBD. Gut 2017;66:1039–1048.
23. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. PLoS Comput Biol 2012;8: e1002687.
24. Stewart DB Sr, Wright JR, Fowler M, et al. Integrated meta-omics reveals a fungus-associated bacteriome and distinct functional pathways in *Clostridioides difficile* infection. mSphere 2019;4:e00454-19.
25. Lamendella R, Wright JR, Hackman J, et al. Antibiotic treatments for *Clostridium difficile* infection are associated with distinct bacterial and fungal community structures. mSphere 2018;3:e00572-17.
26. Sangster W, Hegarty JP, Schieffer KM, et al. Bacterial and fungal microbiota changes distinguish C. difficile infection from other forms of diarrhea: results of a prospective inpatient study. Front Microbiol 2016; 7:789.
27. Martinez-Melendez A, Camacho-Ortiz A, Morfin-Otero R, et al. Current knowledge on the laboratory diagnosis of *Clostridium difficile* infection. World J Gastroenterol 2017;23:1552–1567.
28. Hallen-Adams HE, Suhr MJ. Fungi in the healthy human gastrointestinal tract. Virulence 2017;8:352–358.
29. Panpetch W, Somboonna N, Palasuk M, et al. Oral Candida administration in a *Clostridium difficile* mouse model worsens disease severity but is attenuated by Bifidobacterium. PLoS One 2019;14:e0210798.
30. van Leeuwen PT, van der Peet JM, Bikker FJ, et al. Interspecies interactions between *Clostridium difficile* and *Candida albicans*. mSphere 2016;1:e00187-16.
31. Markey L, Shaban L, Green ER, et al. Pre-colonization with the commensal fungus Candida albicans reduces murine susceptibility to *Clostridium difficile* infection. Gut Microbes 2018;9:497–509.
32. Raponi G, Visconti V, Brunetti G, et al. *Clostridium difficile* infection and Candida colonization of the gut: is there a correlation? Clin Infect Dis 2014;59:1648–1649.
33. Koon HW, Su B, Xu C, et al. Probiotic *Saccharomyces boulardii* CNCM I-745 prevents outbreak-associated *Clostridium difficile*-associated cecal inflammation in hamsters. Am J Physiol Gastrointest Liver Physiol 2016; 311:G610–G623.
34. Pothoulakis C, Kelly CP, Joshi MA, et al. *Saccharomyces boulardii* inhibits *Clostridium difficile* toxin A binding and enterotoxicity in rat ileum. Gastroenterology 1993; 104:1108–1115.
35. Castagliuolo I, LaMont JT, Nikulasson ST, et al. *Saccharomyces boulardii* protease inhibits *Clostridium difficile* toxin A effects in the rat ileum. Infect Immun 1996;64:5225–5232.
36. Czerucka D, Rampal P. Diversity of *Saccharomyces boulardii* CNCM I-745 mechanisms of action against intestinal infections. World J Gastroenterol 2019;25:2188–2203.
37. Rao K, Erb-Downward JR, Walk ST, et al. The systemic inflammatory response to *Clostridium difficile* infection. PLoS One 2014;9:e92578.
38. Saleh MM, Frisbee AL, Leslie JL, et al. Colitis-induced Th17 cells increase the risk for severe subsequent *Clostridium difficile* infection. Cell Host Microbe 2019; 25:756–765.
39. Zanello G, Berri M, Dupont J, et al. *Saccharomyces cerevisiae* modulates immune gene expressions and inhibits ETEC-mediated ERK1/2 and p38 signaling pathways in intestinal epithelial cells. PLoS One 2011;6: e18573.

Received December 11, 2020. Accepted February 26, 2021.

Correspondence
Address correspondence to: Yang-Yu Liu, PhD, Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, 181 Longwood Ave, Boston, Massachusetts 02115. e-mail: yyl@channing.harvard.edu; or Xinhua Chen, PhD, Division of Gastroenterology, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, 330 Brookline Avenue, Boston, Massachusetts 02215. e-mail: xchen1@bidmc.harvard.edu.

Acknowledgments
This study was approved by the institutional review board of Beth Israel Deaconess Medical Center (Institutional Review Board protocol number 2016P000026 and 2016P000054).

CRediT Authorship Contributions
Yangchun Cao, PhD (Data curation: Lead; Funding acquisition: Lead; Methodology: Lead; Writing – original draft: Lead). Lamei Wang, PhD (Data curation: Equal; Formal analysis: Equal; Methodology: Equal; Writing – original draft: Equal). Shanlin Ke, PhD (Data curation: Equal; Methodology: Equal; Software: Equal; Writing – original draft: Supporting). Javier A. Villafuerte Gàlvez, MD (Data curation: Supporting; Methodology: Supporting; Writing – original draft: Supporting). Nira R. Pollock, MD, PhD (Data curation: Supporting; Funding acquisition: Supporting; Investigation: Supporting). Caitlin Barret, Bachelor (Conceptualization: Supporting; Investigation: Supporting; Resources: Supporting). Rebecca Sprague, Bachelor (Conceptualization: Supporting; Data curation: Supporting; Resources: Supporting). Kaitlyn Daugherty, MD (Data curation: Supporting; Investigation: Supporting; Resources: Supporting). Hua Xu, Bachelor (Data curation: Supporting; Resources: Supporting). Qianyun Lin, MD, PhD (Data curation: Supporting). Junhu Yao, PhD (Funding acquisition: Supporting; Project administration: Supporting; Supervision: Supporting; Writing – original draft: Supporting). Yulin Chen, PhD (Conceptualization: Supporting; Project administration: Supporting; Supervision: Supporting). Ciaran P. Kelly, MD (Conceptualization: Equal; Funding acquisition: Supporting; Project administration: Supporting; Resources: Lead; Supervision: Equal). Yang-Yu Liu, PhD (Conceptualization: Supporting; Formal analysis: Equal; Funding acquisition: Supporting; Methodology: Equal; Software: Supporting; Supervision: Equal; Writing – original draft: Supporting; Writing – review & editing: Supporting). Xinhua Chen, PhD (Conceptualization: Lead; Funding acquisition: Lead; Resources: Supporting; Supervision: Lead; Writing – original draft: Supporting; Writing – review & editing: Equal).

Conflicts of interest
The authors disclose no conflicts.

Funding
This study was supported by the National Key Research and Development Projects of China (grant number 2017YFD0500500 and 2018YFD0501600 to Yangchun Cao and Junhu Yao), National Institutes of Health/National Institute of Allergy and Infectious Diseases (grant number 1R01AI116596–01 to Nira R. Pollock and Ciaran P. Kelly; grant number 1R01AI141529–01 to Yang-Yu Liu), National Institute of Diabetes and Digestive and Kidney Diseases (grant number T32 DK 07760 to Ciaran P. Kelly), and Irving W. and Charlotte F. Rabb Award (to Xinhua Chen)

Supplementary Methods

Fecal DNA Isolation and Quantification

Approximately 100 mg of fecal samples were thawed and resuspended in 50 mM Tris buffer (pH 7.5) containing 1 mM EDTA, 0.2% β-mercaptoethanol (Sigma, St Louis, MO) and 1000 U/mL of lyticase (Sigma). The mix was incubated at 37°C for 30 minutes and fungal genomic DNA was isolated by using QIAamp DNA Stool Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol.[1] The DNA was eluted with 1×TE, pH 8.0, and stored at −80oC until processing. The quality and quantity of the DNA was accessed using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE), 1% agarose gel electrophoresis (120 V, 40 minutes), and fluorometer (Qubit dsDNA BR Assay Kit, Life Technologies, Carlsbad, CA).

Fungal ITS2 Sequencing and Quality Control

The final fecal DNA for fungal sequencing was amplified based on ITS2 region using primers as below and PrimeSTAR HS DNA Polymerase kit (Takara Shuzo, Kyoto, Japan). For each sample, the ITS2 rRNA gene were amplified by polymerase chain reaction (PCR) using ITS2 amplicon PCR forward primer (ITS2-F: 5′-GCATCGATGAAGAACGCAGC-3′) and ITS2 amplicon PCR reverse primer (ITS2-R: 5′-TCCTCCGCTTATTGATATGC-3′).[1] The PCR reaction mixture (final volume, 50 mL) contained 5 mL of dNTPs (2 mM of each nucleotide), 5 mL of 10 × DNA polymerase buffer (Qiagen, Hilden, Germany), 1 mL of $MgCl_2$ (25 mM), 0.25 mL of HotStarTaq DNA polymerase (5 U; Qiagen), 1 mL of each primer (10 pmol/mL), and 5 mL of extracted DNA. Amplification was performed using the following cycling conditions: a preliminary step at 98 °C for 15 minutes; 38 cycles of 98 °C for 10 seconds, annealing at 59 °C for the primers used for 10 seconds, and 72 °C for 30 seconds; and a final extension step at 72 °C for 5 minutes, holding at 4°C until further processing. PCR products were detected by 1.2% agarose gel electrophoresis and purified with a Qiaquick gel extraction kit (Qiagen).

Library Preparation

Illumina paired-end adapters with unique indexes were ligated to 100 ng of ITS2 amplicons using a TruSeq DNA Sample Preparation (Illumina, San Diego, CA) where adapters and PCR primers were diluted 1:10 to accommodate lower input of amplicon mass for ITS2 preparations. Library enrichment was performed with 10 cycles of PCR and purified using Agencourt AmpureMagnetic Beads (Beckman Coulter, Fullerton, CA). After PCR purification, all libraries were subjected to quality control using quantitative PCR, Agilent 2100 Bioanalyzer to validate and quantitate library construction then pooled at equimolar concentrations. PCR samples were then sequenced on the Illumina HiSeq platform (Illumina Hiseq 2500).

ITS2 Data Processing

Raw reads were filtered at an expected error of less than 0.5% by SOAPnuke (v.1.5.3) developed by BGI (Shenzhen, Guangdong, China) as follows: (1) adapters removed, (2) read removed if N base is more than 3% of the read, (3) read removed if bases with quality low than 20 were more than 40% of read, and (4) all duplicates removed.[1] After quality filtering, reads were demultiplexed and quality filtered with the Quantitative Insights into Microbial Ecology (QIIME 1) software package,[2] using default parameters. The ITS2 read pairs were demultiplexed based on the unique barcodes. Paired reads were merged and chimeric sequences were removed using the UCHIME (v.7.2) reference dataset for the ITS2 library. The OTUs were picked and chimeras were removed using the UPARSE algorithm within USEARCH (v. 7.0) at 97% identity, and taxonomy assignment was performed against the fungi UNITE database using BLAST within QIIME 1.[2] The annotation was accepted when the bootstrap confidence estimation value was over 0.8, and the assignation stopped at the last well-identified phylogenetic level using Ribosomal Database Project classifier (v. 2.10) software.

Shannon-Wiener Curve and Venn Diagram Analysis

The Shannon-Wiener curve was drawn by using Mothur (v.1.30.1) and R (v.3. 6.3).[3] Venn diagram analysis of OTUs was carried out using R to detect the exclusive and shared OTUs between samples. The percent relative abundances of phylum- or genus-level taxa were conducted using R.

Clustering Analysis and Ternary Analysis

For clustering analysis of the genus-level fungal compositions, the average-linkage hierarchical clustering algorithm by R function "hclust" was applied to cluster stool samples and genera based on their similarities. A heatmap figure was generated based on the clustering result by the "gplots" package of R package. The ternary plot of genera relative abundance was generated with the "ggtern" extension package to R.

Wilcoxon Rank Sum Test and ANCOM Analysis

The *P* values of the Ascomycota:Basidiomycota ratio was calculated based on the Wilcoxon rank sum test. Differential abundance analysis was conducted at the genus and OTU levels using ANCOM (analysis of composition of microbiomes).[4] ANCOM was conducted after removing spurious observations using default parameters with a Benjamini-Hochberg correction significance threshold of 0.05. Note that ANCOM runs a bunch of pairwise tests. Each subhypothesis is structured as follows (for the 2-class case): $H_0(ij) : \text{mean}\{\log(x_i/x_j)\} = \text{mean}\{\log(y_i/y_j)\}$, where x_i is taxon-*i*'s abundance in sample-*x*, y_i is species-*i*'s abundance in sample-*y*, and so on. Here, the W-score of a taxon just counts how many times $H_0(ij)$ is rejected for taxon-*i*. The higher the W-score, the more significant differential is taxon-*i*.

Linear Discriminant Analysis Effect Size Algorithm

To identify specific taxa or KOs as biomarkers for each group, the linear discriminant analysis (LDA) effect size (LEfSe) algorithm was then performed on the Huttenhower lab Galaxy server (http://huttenhower.sph.harvard.edu/lefse/) by importing the microbial relative abundance values and associated sample metadata, with $P < .05$ considered significant and effect size calculate.[5] LEfSe first uses nonparametric factorial Kruskal-Wallis sum rank test to find significantly and differentially abundant features, then the Wilcoxon rank sum test to ensure the identified feature is biologically relevant. LDA is then performed on the identified features to determine the log10 effect size of each differentially significantly abundant feature. The threshold used to consider a discriminative feature for the logarithmic LDA score was set to >2.0.

Random Forest Classifier Construction

The random forest was also used to select differential OTUs and immune factors between 2 cohorts, and to verify the key discriminatory OTUs and immune factors that selected by random forest analysis.[6] For fungal sequencing and immune factors data, each OTU and immune factor was considered as a feature. For the selection of features, first, all the features were taken as training datasets with random forest algorithm using the rfcv function in an R package 'randomForest' and then each feature's importance score was calculated through permuting values of this feature and then calculating and normalizing the difference of out-of-bag errors before and after a permutation. Ten features were added one by one according to importance score of the feature (with descending order). Ten times cross validation for 500 times was performed to sift through the minimum OTU and/or immune factors combination with the lowest error rate and the lowest number that the mean prediction accuracy reaching the optimal value. Receiver operating characteristics (ROC) analysis was then performed to measure the quality of the classification models by the R software package pROC (v.1.16.2).[7] ROC curve results were plotted manually by the true positive rate against the false positive rate. ROC curves were constructed, and the AUC was used to designate the ROC effect. Meanwhile, mean decrease accuracy from the importance matrix was used to select features. A variable importance plot was produced according to the importance scores (mean decrease accuracy) of selected features (OTUs and immune factors) and their boxplots of selected features were drawn in R.

Fungal Correlation Network Analysis

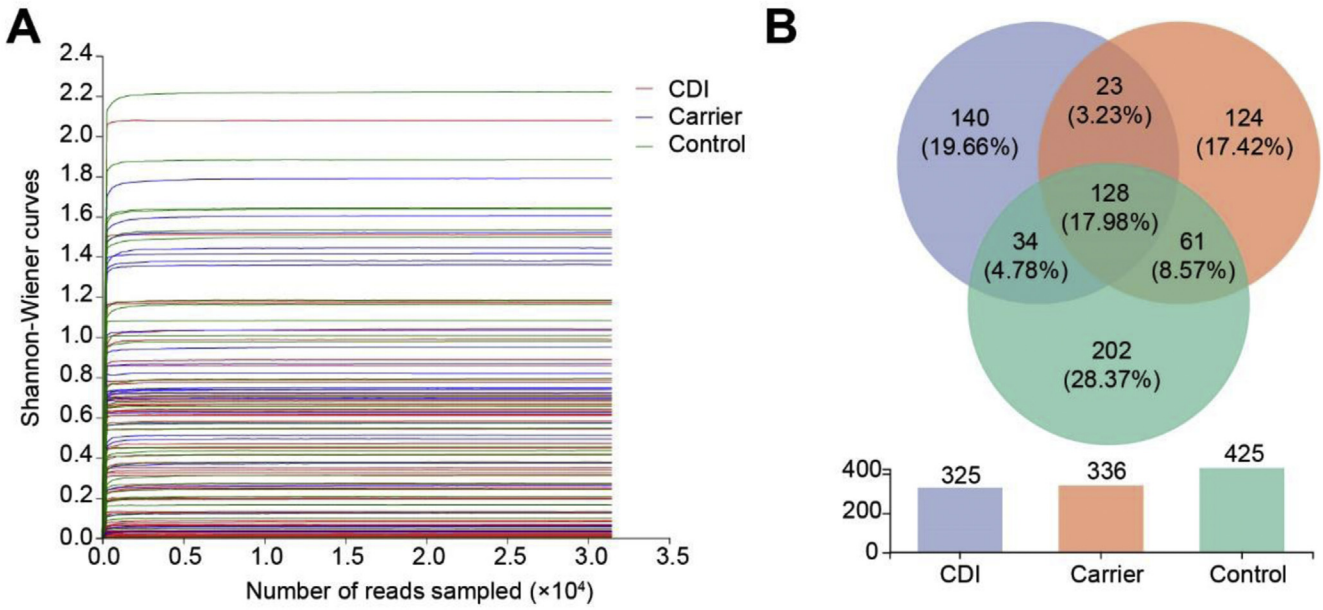
To further assess the potential interactions among fungal community members, network analysis was conducted. The top 40 dominant fungal genera were used for genera correlation network. The fungal correlation networks were constructed using SparCC (sparse correlations for compositional data, https://github.com/luispedro/sparcc). Significant interactions were determined by the bootstrapped results (n = 100) using the script PseudoPvals in SparCC. Significant correlations with absolute sparse correlations $|r| > 0.1$ and $P < .05$. The larger size of nodes represents the more relative abundance genera in the fungal community. Highly connected fungal taxa in each module can be considered as keystone taxa due to their central position in a mycobiota network.

Fungal Abundances and Host Serum Immune Factors Correlation Analysis

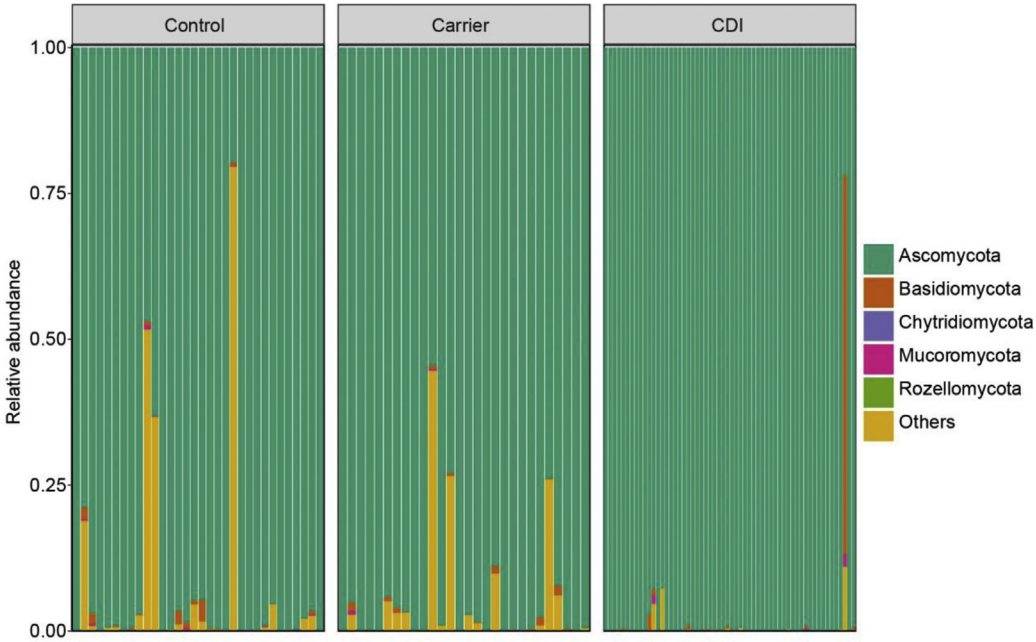
Correlations between the fungal abundances and host serum immune factors were determined by drafting a matrix of Spearman correlations. The level of significance was kept at the default of $P = .05$. All included genera were required to be detected in $\geq$ 15% of all samples in each group.

Supplementary References

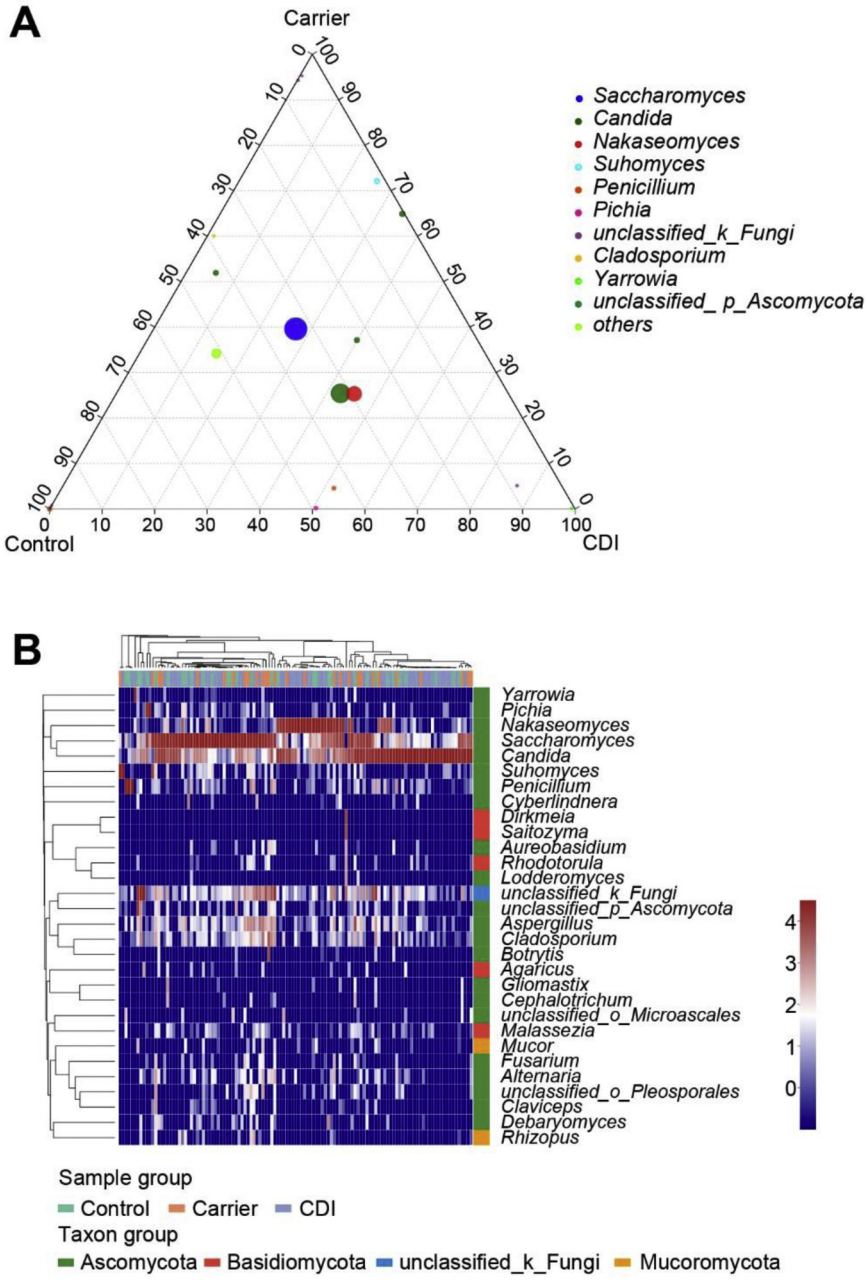
1. Zuo T, Wong SH, Cheung CP, et al. Gut fungal dysbiosis correlates with reduced efficacy of fecal microbiota transplantation in *Clostridium difficile* infection. Nat Commun 2018;9:3663.
2. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;7:335–336.
3. 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;75:7537–7541.
4. Mandal S, Van Treuren W, White RA, et al. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb Ecol Health Dis 2015; 26:27663.
5. Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation. Genome Biol 2011;12:R60.
6. Breiman L. Random forests. Machine learning 2001; 45:5–32.
7. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77.



Supplementary Figure 1. Shannon-Wiener curves and Venn diagram of fungal taxa in stool samples from 3 distinct phenotypes: Control, Carrier, and CDI. (*A*) Shannon-Wiener curves were all calculated at a 0.03 dissimilarity level. Shannon-Wiener curves were collated from each sample's Shannon diversity index. The graphic shows the estimated diversity plotted against the number of sequences per sample. Each line represents 1 sample. The plateau in each estimated diversity curve indicates the minimum number of sequences to capture diversity. (*B*) A Venn diagram showing shared and unique OTUs at a 0.03 dissimilarity level.



Supplementary Figure 2. Phylum-level taxonomic profiles of the gut mycobiota from 3 distinct phenotypes: Control, Carrier, and CDI. Only phyla with ≥1% abundances in at least 1 sample were depicted. Otherwise, they were included in the category "others."



Supplementary Figure 3. Taxonomic composition of the gut mycobiota from 3 distinct phenotypes: Control, Carrier, and CDI. (*A*) Ternary plot: A circle represents a genus, the circle size corresponds to the abundance (number of sequences) of that genus. Different colors correspond to different genera. The position of each circle is determined by the contribution of the indicated cohorts to the total abundance. The category "others" corresponds to all genus less than 1%. (*B*) Heat map: Each row represents an individual genus; relative values are color-coded and shown on the right side of the panel.

Supplementary Table 1. Demographic Characteristics of the Enrolled Subjects

Characteristics	Control (n = 32)	Carrier (n = 28)	CDI (n = 58)	*P* value
Sex, n (%)				.081[a]
Female	8 (25.00)	14 (50.00)	27 (46.55)	
Male	24 (75.00)	14 (50.00)	31 (53.45)	
Age, Avg ± SD	59.81 ± 14.31	60.61 ± 18.91	65.38 ± 16.51	.211[b]
Ethnicity, n (%)				.181[a]
Hispanic	0 (0.00)	0 (0.00)	2 (3.45)	
Non-Hispanic	28 (87.50)	21 (75.00)	51 (87.93)	
Unknown	4 (12.50)	7 (25.00)	5 (8.62)	
Race, n (%)				.693[a]
White	25 (78.12)	20 (71.43)	45 (77.59)	
Others/unknown	7 (21.88)	8 (28.57)	13 (22.41)	

[a]χ^2 test.
[b]One-way analysis of variance,

Supplementary Table 2. PERMANOVA in Mycobiota Compositions

Characteristics	*F*	R^2	*P* value
Cohorts	1.448	0.025	.150
Sex	0.674	0.006	.602
Age	1.873	0.016	.109
Race	1.935	0.016	.088
Ethnicity	1.066	0.018	.384

NOTE. Race: White and others. Ethnicity: Hispanic, Non-Hispanic, and Unknown. Here *F* represents the F-statistic: a larger *F* value indicates that the between-group variation is greater than within-group variation. *R*2 represents the variation explained by the model. *P* represents the *P* value calculated from permutation.

Supplementary Table 3. Detection of Differentially Abundant Genera and OTUs Between Control and CDI Using ANCOM

			Relative abundance ($\times 10^4$)	
Taxonomy level	Taxa	W-score	Control	CDI
Genus	*Cladosporium*	19	74 ± 137	4 ± 11
Genus	*Aspergillus*	17	149 ± 507	27 ± 117
OTU	OTU252 (*unclassified_g_Cladosporium*)	49	72 ± 135	2 ± 7
OTU	OTU584 (*unclassified_g_Aspergillus*)	41	34 ± 85	15 ± 94
OTU	OTU687 (*Candida dubliniensis*)	32	3 ± 13	255 ± 848

Supplementary Table 4. Detection of Differentially Abundant Genera and OTUs Between Carrier and CDI Using ANCOM

			Relative abundance ($\times 10^4$)	
Taxonomy level	Taxa	W-score	Carrier	CDI
Genus	*Aspergillus*	20	82 ± 158	27 ± 117
Genus	*Cladosporium*	16	115 ± 36	4 ± 11
OTU	OTU657 (*Aspergillus proliferans*)	50	43 ± 77	8 ± 48
OTU	OTU252 (*unclassified_g_Cladosporium*)	33	111 ± 361	2 ± 7