

RESEARCH

Open Access



Depletion of core microbiome forms the shared background against diverging dysbiosis patterns in Crohn's disease and intestinal tuberculosis: insights from an integrated multi-cohort analysis

Aditya Bajaj^{1†}, Manasvini Markandey^{1†}, Amit Samal², Sourav Goswami², Sudheer K. Vuyyuru¹, Srikant Mohta¹, Bhaskar Kante¹, Peeyush Kumar¹, Govind Makharia¹, Saurabh Kedia¹, Tarini Shankar Ghosh^{2*} and Vineet Ahuja^{1*}

Abstract

Background/aims Crohn's disease (CD) and intestinal tuberculosis (ITB) are gastrointestinal (GI) inflammatory disorders with overlapping clinical presentations but diverging etiologies. The study aims to decipher CD and ITB-associated gut dysbiosis signatures and identify disease-associated co-occurring modules to evaluate whether this dysbiosis signature is a disease-specific trait or is a shared feature across diseases of diverging etiologies.

Methods Disease-associated gut microbial modules were identified using statistical machine learning and co-abundance network analysis in controls, CD and ITB patients recruited as part of this study. Module reproducibility was reinvestigated through meta-network analysis encompassing >5400 bacteriomes and ~900 mycobiomes. Subsequently, >1600 Indian gut microbiomes were analyzed to identify a central-core gut microbiome of 46 taxa, whose abundances aided in the formulation of an India-specific Core Gut Microbiome Score (CGMS) to measure the degree of core retention.

Results Both diseases witness similar patterns of alterations in [alpha]-diversity, characterized by a significant reduction in gut bacterial (i.e., bacterial/archaeal) diversity and a concomitant increase in the fungal [alpha]-diversity. Specific bacterial taxa, along with the diverging mycobiome enabled distinction between the diseases. Co-abundance network analysis of these taxa, validated by integrated meta-network analysis, revealed a 'disease-depleted' module, consistent across multiple cohorts, with >75% of this module constituting the central-core Indian gut microbiome. CGMS robustly assessed the core-microbiome loss across different stages of gut inflammatory disorders, in Indian and international cohorts.

Conclusions While the disease-specific gain of detrimental bacteria forms an important component of gut dysbiosis, loss of the core microbiome is a shared phenomenon contributing to various GI disorders.

[†]Aditya Bajaj and Manasvini Markandey have contributed equally to this work.

*Correspondence:
Tarini Shankar Ghosh
tarini.ghosh@iiitd.ac.in
Vineet Ahuja
vineet.aiims@gmail.com



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Keywords Gut inflammatory disorders, Crohn's disease, Intestinal tuberculosis, Gut microbiome, Mycobiome, Dysbiosis, Core gut microbiome, Machine learning, Microbiome meta-network analysis

Background

The gut microflora regulates a myriad of gut immunological processes, keeps a check on pathobiont expansion and forges dynamic interactions with the host physiological machinery, thereby acting as a fulcrum of gut homeostasis. Dysbiosis of the gut bacterial, archaeal and fungal communities has been strongly implicated in gut inflammatory disorders like inflammatory bowel diseases (IBD) [1–8].

IBD is a debilitating, chronic inflammatory disorder of the intestine, constituting Crohn's disease (CD) and ulcerative colitis (UC). It is a multifactorial disorder caused by a dysregulated immune response against the commensal microbiota, potentially triggered by environmental cues. Another such chronic inflammatory disease, but with an infectious etiology is intestinal tuberculosis (ITB)-an extra-pulmonary form of TB. With diverging etiologies, ITB and CD have overlapping clinical, pathological, radiological, and endoscopic manifestations [9, 10]. This poses a diagnostic dilemma in developing nations that are endemic to infectious diseases like ITB and are now witnessing a surge in lifestyle-associated non-infectious diseases like CD. Even though the gut bacterial and fungal communities have been extensively profiled in patients with CD, there are no reports of bacterial and fungal dysbiosis in intestinal tuberculosis. A recent report by He et.al., highlighted differential gut bacterial dysbiosis signatures associated with mucosa-associated microbiota in the two diseases, leaving a void for characterization of the dysbiosis in the less invasive luminal (faecal) stream [11, 12].

The present study aims to decipher ITB-associated alterations in the composition of gut bacteriome/archaeome and mycobiome and compare it with CD-associated dysbiosis. Through analysis of dysbiosis signature in the two disorders, we intended to elucidate the gut microbial modules specific to the disease type and the ones common between the disorders, despite their diverging etiologies. Besides identifying disease-specific microbiome alterations in bacteriome and especially the mycobiome, we observed a specific co-abundant microbiome-module depleted in both diseases, that was consistently reproduced across multiple cohorts. Interestingly, a panel of 46 members of this microbiome module (including 28 species-level taxa) were prevalent (detected in at least 70% of the samples) and occupied central positions in the gut microbial community of apparently healthy individuals collated from several Indian cohorts. This enabled the

identification of an Indian core gut microbiome, which led to the formulation of a Core Gut Microbiome Score (CGMS) that quantitatively assesses the extent of retention of this core in individuals.

Methods

Study cohort and sample collection

Patients with a confirmed diagnosis of ITB or CD were recruited prospectively at the IBD clinic, All India Institute of Medical Sciences, New Delhi. The patients were diagnosed with CD based on a combination of clinical, endoscopic and histological features as per the European Crohn's and Colitis Organization (ECCO) guidelines [13, 14]. Patients less than 18 years of age, with a history of anti-tubercular therapy (ATT), with co-existing infection (pulmonary/urinary), HIV seropositivity, and pregnant or lactating females, were excluded. The diagnosis of ITB was made in cases where the following criteria were present (1) caseating granulomas on biopsy, (2) presence of acid-fast bacilli by AFB staining or culture, (3) demonstration of active TB at an extra-pulmonary site(s). For indeterminate query cases, an ATT was given to the patients who did not fall into the above categories. A patient was categorized as having ITB if the patient had clinical and endoscopic/radiologic response to ATT (Paustian's criteria with Logan's modification) with a minimum follow-up of 12 months after completion of ATT and a diagnosis of CD was made if the patient showed no response, or worsened after initial improvement with ATT trial and subsequently showed a clinical and/or endoscopic response to oral steroids/CD specific therapy [15]. 10 g of faecal samples from healthy controls ($n = 17$) and from patients with CD ($n = 20$) and ITB ($n = 20$), were collected in sterile stool vials, homogenized, aliquoted and stored at -80°C , till further processing. The experimental protocols were approved by the institutional ethics review board of the All India Institute of Medical Sciences, New Delhi (Ref. No. IECPG/484/29.8.16).

DNA extraction and sequencing

For bacterial microbiome analysis, the total DNA was isolated from faecal samples using the protocol previously published by Bag et al. [16], with minor modifications. Briefly, frozen samples were thawed, and biopsy and faecal samples were weighed precisely to 2 mg and 200 mg respectively. Samples were homogenized using glass beads-2.3 mm (Biospec, USA). This was followed by

enzymatic cell lysis using lysozyme (10 mg/mL) (Sigma Aldrich), mutanolysin (25 KU/mL) (Sigma Aldrich) and lysostaphin (4 KU/mL) (Sigma Aldrich) at 37 °C for 1 h. Post-incubation samples were subjected to treatment with 4 M guanidine thiocyanate (Sigma Aldrich) and 10% N-lauryl sarcosine (Sigma Aldrich), before incubation at 37 °C for 10 min and 70 °C for 1 h. This was followed by mechanical lysis of cells by bead beating cycles before supernatants were transferred to fresh tubes and subjected to protein removal. Nucleic acids were pelleted using ice-cold ethanol (96%) and centrifugation at 14,000 g for 10 min at 4 °C. The final precipitation of DNA was achieved by adding 3 M sodium phosphate and 1 mL of 96% ethanol, subjecting the pellet to centrifugation at 14,000 g for 10 min at 4 °C.

Isolation of DNA for characterization of fungal members through ITS1 sequencing was performed by using an initial lyticase treatment (500U), followed by extraction using QIAamp Fast DNA Stool Mini Kit, using the manufacturer's instructions. Extracted DNA was sequenced using the Illumina MiSeq platform, following which high-quality reads were obtained by using Trimmomatic v0.38 to remove adapter sequences, ambiguous reads and low-quality reads (reads with more than 10% quality threshold (QV) < 20 Phred score).

Computation of bacteriome/archaeome (bacterial & archaeal abundance) and mycobiome (fungal abundance) profiles

Raw sequence reads were processed using QIIME2 version 2022 [17]. Forward and reverse paired-end reads were denoised, demultiplexed and subjected to chimera removal using the DADA2 plugin of QIIME2 [18]. DADA2 was utilized to generate amplicon sequence variants (ASVs) for the 16S (for Bacteriome and Archaeome) and ITS (Internal Transcribed Spacer; for Mycobiome datasets) sequences for each sample. The species and genus-level classification of each of the 16S-derived (Bacteriome/Archaeome) representative sequences were then obtained using the SPINGO classification tool [19]. For the ITS data, this classification was performed using the 'sintax' taxonomy classification tool implemented within the *vsearch* annotation pipeline with the UNITE database as the reference [20].

The methodology for the assessment of α and β -diversity variations across the three groups has been summarized in Supplementary Figure 1. Shannon and Pielou's evenness indices across the bacteriome and mycobiome datasets were determined using the 'diversity' and 'specnumber' functions of *vegan* (version 1.00.15) package, respectively. These matrices were compared

across three sample-types (controls, CD and ITB) using Kruskal–Wallis *H*-tests (using the *kruskal.test* function of the R programming interface), and *p* values indicative of significant variations in values obtained for pairwise comparisons across groups were computed using the Dunns' tests (computed using the *dunn.test* function in R and *p* values corrected using Benjamini–Hochberg by setting the "method" argument to "bh").

The beta diversity across the samples was investigated using three different distance measures, namely Kendall and Weighted-Jaccard. The abundance profiles were normalized using the Total Sum Scaling approach, that is counts for each taxa in a given sample divided by the total sum of all taxa in the sample.

For computing Kendall distances, the *cor.fk* function of the *pcaPP* package v2.0.2 was used to generate Kendall tau correlation matrix across the samples. The Kendall correlations were subsequently converted to Kendall distance matrix by using the following:

$$\begin{aligned} \text{Kendall Distance Matrix} \\ = \text{as.matrix}(1 - \text{cor.fk}(t(\text{abundances}))/2) \end{aligned}$$

Weighted Jaccard distances were on the other hand computed as:

$$\begin{aligned} \text{Jaccard Distance Matrix} = \text{as.matrix} \\ ((\text{vegdist}(\text{abundance}, \text{method} = \text{"jaccard"})) \end{aligned}$$

Each of the distance matrices captures different aspects of microbiome variations using different kinds of transformation (as detailed in Ghosh et al. [21]). For each distance measure, separate distance matrices were computed for microbiome and mycobiome profiles.

For each distance and profile (microbiome and mycobiome), the beta-diversity variations were then visualized, and the extent and significance of variations were investigated using Principal Coordinate Analysis (PCoA). Principal Coordinate Analysis was performed using the 'ade4' package v.1.7.17 of R. PCoA was performed separately for the four different distance measures, namely for the three distance measures (Kendall and Weighted Jaccard) applied for the two kinds of profiles (namely microbiome and mycobiome), resulting in $3 \times 2 = 6$ combinations. For each of the four plots, the separation between the three different groups were first visually investigated. Subsequently, the major PCoAs having significant variations across the three subject groups (that is the axes with the most significant splits), were compared using Kruskal–Wallis *H*-test with pairwise comparisons between the different groups performed using Dunns' tests (as described in the previous section).

Identification of diagnostic markers and differentially abundant taxa

Identification of taxonomic markers diagnostic of the different groups was identified using two parallel and complementary approaches, namely Random Forest-based identification and identifying taxa associated with the taxa-abundance-derived Principal Coordinates having the most significant variations across the three groups of individuals (controls, CD and ITB). In the first approach, we utilized Random Forest to identify the top diagnostic predictors discriminating between each of the three pairs of groups (Controls vs. ITB; Controls vs. CD; CD vs. ITB). The details of the approach are provided below. First, we created separate matrices containing the abundances of all the species and genus levels separately for the gut microbiome and mycobiome profiles. Then for a given profile and a given pair of groups, we created a Random Forest (RF) classifier to classify the samples belonging to the two groups based on the microbiome and mycobiome level features (separate Random Forest classifiers for microbiome and mycobiome). This was performed using the 'randomForest' function of the *randomForest* package v4.6.14. For a given level (microbiome or mycobiome) and a given pair of groups (Controls vs. CD; Controls vs. ITB; CD vs. ITB), after the generation of these RF models, the features were then ranked based on their feature importance scores (i.e. the mean decrease in GINI across 500 iterations). Subsequently, based on these ranks, we generated multiple RF models considering a varying number of top features (e.g. top 10, 20, 30, 40, till 250). The number of features for which the corresponding RF models generated the highest AUC values was taken as the set of the most discriminatory (or diagnostic) features for the classification of the two groups. After identifying the most diagnostic features for each pairwise group classification, we performed an additional variation to ascertain that these markers were the most diagnostic and reproducible and were not affected by biases in certain samples. For this, we performed 50 iterations, where in each iteration we selected 50% of the samples for training and the remaining 50% for testing. The objective was to ascertain that the AUC values observed in the overall models (taking all features) were similar even in the 50 iterations.

Understanding mutual relationships of biomarkers via network-based approach in the current dataset and validating the same across global meta-networks

For investigating the mutual co-abundance relationships between the above-identified set of markers specific for each subject group (Controls, ITB and CD), we utilized the permutation-renormalization bootstrap (ReBoot) approach [22], implemented within the R-based 'crepe'

module v1.34.0 (using Kendall-tau as the correlation score). The methodology adopted for this investigation with the detailed protocol for validating the reproducibility of marker relationships in global meta-networks using multiple publicly available microbiome and mycobiome datasets (listed in Table 1) has been described below [23–37].

Reboot-based co-abundance network computation

The co-abundance relationships (Kendall-tau > 0) with $FDR \leq 0.1$ were identified and these relationships were represented as marker vs. marker adjacency matrix. This adjacency matrix was subsequently represented as a network, such that any marker having co-abundance relationships with at least one another marker was represented as a node and any two nodes having a mutual co-abundant relationship were connected by an edge. To create this network, we utilized the 'igraph' package (v1.35.0) and specifically the 'graph_from_adjacency_matrix' function.

Validating reproducibility of marker relationships in global meta-networks using multiple publicly available microbiome and mycobiome datasets

We defined meta-network as a graph of co-abundance relationships amongst taxonomic features that are inferred based on performing meta-analysis across multiple studies. A meta-network contained the taxa as nodes. Two nodes were only connected by an edge if their abundances showed a positive association based on a Random Effects Model with an $FDR \leq 0.1$ and consistency of greater than 70%. The meta-analysis was performed using the random-effects model using the 'metafor' package (v2.0) and robumeta package (v3.8) of R.

Separate meta-analyses were performed to identify the intra-microbiome (focusing only on gut microbiome profiles), intra-mycobiome (focusing only on gut mycobiome profiles) and mycobiome-microbiome co-abundance associations (focusing on paired profiles with both gut microbiome and mycobiome data). For this purpose, we utilized 5437 gut microbiome profiles and 892 gut mycobiome profiles, encompassing 14 studies (Table 1). For microbiome profiles, we predominantly included datasets from studies in the Indian population [23–26]. These included a total of 1617 gut microbiome profiles from India. We also included 3559 gut microbiome profiles from healthy individuals from another population-level cohort from the neighboring country China [30]. One of these datasets also had gut microbiome profiles from diverse Asian populations from Japan and India [9]. This same dataset also contained matched gut mycobiome profiles ($n=97$). We included two more datasets from the EU focusing on

Table 1 Microbiome and mycobiome datasets utilized in the present study for the network meta-analysis, Indian core gut microbiome assignment and for validation of core gut microbiome score

Dataset	Reference	Datatype	Total Sample		Nationality	Purpose		
			Microbiome	Mycobiome		Network Meta-analysis	Indian Core Gut Microbiome	Core Gut Microbiome Score validation
AIIMS 2021	Kedia et al 2021 ²³ ; Das et al 2018 ²⁴	16S	162	0	India			
MicroDiab India	Pinna et al 2021 ²⁶ ; Alvarez et al 2021 ²⁵	16S	435	0	India			
LogMPic	Dubey et al 2018 ²⁷	16S	874	0	India			
Dhakan DB 2019	Dhakan et al 2019 ²⁸	Shotgun	88	0	India			
Gupta A 2019	Gupta et al 2019 ²⁹	Shotgun	60	0	India			
PRJDB7616	Pareek et al 2019 ³¹	16S/ITS	58	58	India			
		16S/ITS	39	39	Japan			
Liguori 2016	Liguori et al 2016 ³²	16S/ITS	47	47	Italy			
PRJEB423575	Jeffery et al 2020 ³⁴ ; Das et al 2021 ³³	16S/ITS	115	144	Ireland			
HMP2	Nash et al 201 ³⁵	Shotgun/ITS	0	390	US			
PRJNA439151	NA	ITS	0	70	China			
PRJNA647266	Hu et al 2022 ³⁷	ITS	0	75	China			
He et al	He et al 2019 ³⁰	16S	3559	0	China			
PRJNA662173	Jayasudha et al 2020 ³⁶	ITS	0	69	India			
Halfvarson 2017	Halfvarson et al 2017 ⁴⁵	16S	683	0	Sweden			
Lloyd-Price 2019	Lloyd-Price et al 2019 ⁸	Shotgun/16S	1627	0	USA			
Hall 2017	Hall et al 2017 ⁴⁴	Shotgun	259	0	USA			
Franzosa 2018	Franzosa et al 2018 ⁴¹	Shotgun	219	0	USA			

The respective purposes for which the study samples have been used, have been marked in blue

IBD and IBS (irritable bowel syndrome) that also contained both gut microbiome and mycobiome profiles [23, 24]. The paired gut microbiome and mycobiome profiles were utilized for meta-analyses investigating

microbiome-mycobiome associations. This set of mycobiome profiles was supplemented with three more datasets from the Human Microbiome Project v2, India and China for investigating the intra-mycobiome

$$\text{The rank scaled abundance of a given taxa 'j' in sample 'i'} = \frac{((\text{rank}(j \text{ in } i) - \min(\text{rank}(j \text{ across all } i)))}{((\text{max}(\text{rank}(j \text{ across all } i)) - \min(\text{rank}(j \text{ across all } i)))}$$

The core gut microbiome score for the various datasets was then calculated as:

$$\text{For a given microbiome } k, \text{ CGMS} = \sum_{\text{summation across all detected core taxa 'm'}} \text{Rank Scaled Abundance of taxa 'm' in 'k'}$$

* Percentile rank of 'm' in degree centrality in the meta
– network identified above

associations [35]. Network edges identified within our current study cohort (as described in the previous step) that were also observed to be present in the three meta-analyses performed in this step were identified as the core set of reproducible associations. All networks were visualized using the Cytoscape software [38].

Identification of an Indian core microbiome and computation of the core gut microbiome score (CGMS)

We collated 1617 gut bacteriome/archaeome from seven previous studies, especially focusing on Indian sub-populations [24–29, 31, 39]. The core gut taxa within the Indian population were identified using two properties—prevalence and centrality. The prevalent taxa were identified as those that were present in at least 70% of the samples in at least four of the seven datasets. To identify the central taxa, we built a meta-network based solely on the seven Indian studies. Within this network, taxa that were in the 70 percentiles in terms of their degree of centrality were identified as the central taxa. Taxa belonging to both the lists of the prevalent taxa as well as the central taxa were identified as the core gut taxa within the Indian population.

Once the core gut microbiome was identified, the core gut taxa were then sorted in a descending order based on their meta-network degree centrality. For any given gut microbiome dataset (irrespective of whether the gut microbiome belonged to the above list of 1617 datasets investigated), the list of all core gut taxa (as identified) that were detected in the dataset were identified, ranked across samples and then rank-scaled from 0 to 1 using the below formula:

Comparison of CGMS with other indicators of gut health in multiple cohorts utilized for this study

For each gut microbiome, we computed four additional measures of gut health, namely Shannon Diversity, Dysbiosis Score (introduced by Lloyd-Price et al. [8]), Kendall Uniqueness (introduced by Ghosh et al. [21]) and the Gut Microbiome Health Index (introduced by Gupta et al. [40]). The algorithm for the computation of each of the later three indices are described in the respective studies. For each cohort, microbiomes were either categorized as control or diseased (grouping multiple disease categories together). Comparison of each of the five indices (including CGMS) were performed using Mann–Whitney tests (as described previously). The consistency of the associations of these indices with health across cohorts was compared.

Investigation of inferred metabolic functionalities

Computation of inferred metabolite profiles

The inference of metabolic profiles for each gut microbiome was performed using methods as described in previous studies [41, 42]. We utilized taxa-metabolic function maps containing the experimentally validated metabolic-functional profiles (i.e. the production and consumption patterns) of ~300 metabolites from 992 species-level taxa as 0–1 information (1: detected and 0: Not detected). These were collated from multiple repositories/studies [43, 44]. Using this map, given the abundance of a gut microbiome, the metabolic functional profile can be inferred by performing an inner multiplication of the abundance profile of the

overlapping species with this taxa-function map. Computed in this manner, the abundance value of each metabolic function denotes the cumulated abundance of all taxa experimentally validated to have the given metabolic function.

Diversity analysis across different groups

Normalization was performed using Total-Sum-Scaling. Principal Co-ordinate analysis (PCoA) of the metabolic profiles of the 57 gut microbiomes was performed using the Bray–Curtis distance measure (using vegan version 2.6–8 package in R). We then performed PERMANOVA (using Euclidean distances between samples using the top 2 PCoAs), to judge the significance of the separability of the three groups. The distribution of the embeddings, PCo1 and PCo2 was visualized shown using a cowplot (using cowplot package, version 1.1.3 in R). The significance in difference between the PCoA1 values across each group have been investigated through a Mann–Whitney test using the stats (version 3.6.2) package in R.

Correlating the microbiome-specific inferred metabolite profiles with the PCo1 values

The inferred metabolite profiles were divided into two categories, namely production profiles and consumption/degradation profiles. Spearman correlation was then used to compute the association between the inferred metabolite profiles (both categories) and PCo1 was using the stats (version 3.6.2) package in R. Significantly ($q\text{-value} \leq 0.05$; $q\text{-value}$ obtained after adjustment using Benjamini–Hochb) positive (Spearman-Rho > 0) and negative (Spearman-Rho < 0) correlations were identified and were plotted and visualized as volcano plots, plotted using the ggplot package of R.

Replicative investigation of the inferred metabolite associations in six external previously investigated IBD cohorts

To validate the significantly enriched and significantly depleted metabolites in terms of production as well as consumption/degradation in Controls, a meta-analysis using Random Effect Model has been performed on the inferred metabolite profiles in six previously investigated global IBD cohorts [8, 32, 39, 45–47] of IBD patients and Controls.

Results

Patient demographics corresponding to the two disease cohorts

In our cohort of CD and ITB patients ($n=40$; CD=20 and ITB=20), the mean age of the patients with CD was

42.55 ± 10.75 years, while patients with ITB had a mean age of 33.9 ± 11.80 years. Most patients with CD had ileal involvement (L1 phenotype) and stricturing B2 phenotype (50%), while the ITB patients predominantly had ileocaecal (40%) and ileocolonic (30%) involvement. The summary of the demographic and clinical details of the patient and control subjects are shown in Table 2.

Both diseases witness differential patterns of α and β diversity variations of the gut microbiome (bacteriome/archaeome) and the gut mycobiome to controls

Subjects were categorized into three groups based on their clinical phenotypes—controls (non-CD and non-ITB individuals), and patients with CD or ITB. Both CD and ITB showed a significant reduction in gut microbiome (bacteriome/archaeome) α -diversity indices (Shannon Index and Pielou's evenness index) when compared with controls, with evenness undergoing a more drastic reduction than the Shannon diversity metric (Fig. 1; see "Methods" section). This significant disease-associated reduction evenness was constant across three taxonomic levels (Amplicon Sequence Variants (ASVs), species and genus). Interestingly, no marked differences were observed between the bacterial α -diversities between CD and ITB (Fig. 1). The comparison of the α -diversity in the gut fungal community revealed that the ITB subjects had significantly higher Shannon Indices and Pielou's Evenness Index when compared to the controls and CD patients across multiple taxonomic levels (Fig. 1). This indicates that both diseases are accompanied by shared alterations in the overall gut bacterial community structure, characterized by a decrease in the richness of the bacterial/archaeal members and the ITB-associated increase in the diversity and representation of the fungal members.

To further investigate the community variations, we then performed a series of Principal Coordinate Analyses (PCoAs) to find the variation patterns in the overall gut bacteriome/archaeome and mycobiome compositions across the three groups (see "Methods" section; Supplementary Figure 1 for a pictorial summary). The PCoA plots at the gut microbiome level (bacteriome/archaeome) depicted a marked overlap between the CD and ITB samples, while the controls clustered separately from both diseases with significant variations noted across the first PCoA axes (PCo1) (Fig. 2A, B). At the gut mycobiome level, however, the PCoA highlighted significant variations between the control and ITB samples, indicating a distinct make-up of the gut fungal community in the ITB patients, while the CD patients showed intermediate positioning between the controls and ITB patients (Fig. 2A, B). The pattern was consistent across all taxonomic levels and distance matrices (Supplementary

Table 2 Demographic and clinical details of healthy controls and patients with Crohn's disease and intestinal tuberculosis

Characteristics	Crohn's disease
Age, years, mean \pm SD	42.55 \pm 10.75
Sex, male, <i>n</i> (%)	13 (65.00%)
BMI, kg/m ² , mean \pm SD	21.26 \pm 5.08
Current alcohol intake, <i>n</i> (%)	3 (15%)
Current smoker, <i>n</i> (%)	2 (10%)
Diet (veg:non-veg)	11:9
Disease duration (years \pm SD)	4.5 \pm 3.9
<i>Phenotype (Montreal classification)</i>	
A1 (below 17 years)	1 (5%)
A2 (between 17 and 40 years)	10 (55.5%)
A3 (above 40 years)	9 (45%)
L1 (ileal involvement)	10 (50%)
L2 (colonic involvement)	4 (20%)
L3 (ileocolonic involvement)	5 (25%)
L4 (upper GI involvement)	5 (25%)
P (perianal involvement)	5 (25%)
B1 (nonstricturing, nonpenetrating disease)	6 (30%)
B2 (structuring disease)	19 (50%)
B3 (penetrating disease)	4 (20%)
Extra intestinal manifestations (EIMs)	4 (20%)
Intestinal tuberculosis	
Age, years, mean \pm SD	33.9 \pm 11.80
Sex, male, <i>n</i> (%)	12 (60%)
BMI, kg/m ² , mean \pm SD	18.90 \pm 3.299
Current alcohol intake, <i>n</i> (%)	2 (10%)
Current smoker, <i>n</i> (%)	3 (15%)
Diet (veg:non-veg)	1:1
Disease duration (years \pm SD)	2.602 \pm 4.34
Disease extent	
Ileal disease	4 (20%)
Ileocaecal disease	8 (40%)
Ileocolonic disease	6 (30%)
Colonic disease	1 (5%)
Upper GI disease	1 (5%)
Extra intestinal manifestations (EIMs)	3 (15%)
Caseating granuloma	2 (10%)
Controls	
Age, years, mean \pm SD	39.91 \pm 12.81
Gender, male, <i>n</i> (%)	7 (58.33%)
BMI, kg/m ² , mean \pm SD	20.72 \pm 2.22
Current alcohol intake, <i>n</i> (%)	0 (0%)
Current smoker, <i>n</i> (%)	3 (25%)
Diet (veg:non-veg)	1:2

CD Crohn's disease, ITB intestinal tuberculosis, HC healthy controls, BMI Body Mass Index, Veg vegetarian, non-veg non-vegetarian

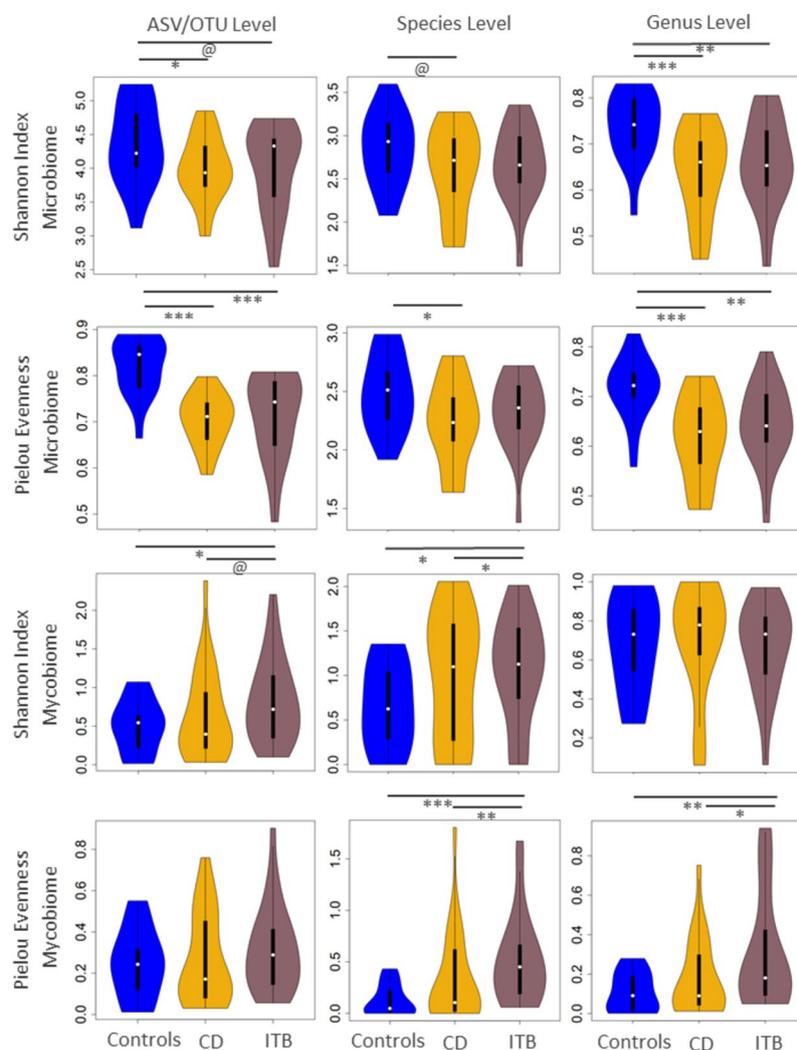


Fig. 1 Disease alterations in gut alpha diversity show clearly opposing patterns at the microbiome and mycobiome levels. Comparison of Shannon Index and Pielou Evenness Index for gut microbiomes and the gut mycobiomes of the three subject groups (namely Controls, CD and ITB) at the taxonomic levels of OTU/ASV, Species and Genus. Pairwise comparisons between the different groups are denoted by bars between corresponding boxes with the associated corrected p values (FDR) given in the notation: @ $0.05 < \text{FDR} \leq 0.1$; * $0.01 < \text{FDR} \leq 0.05$; ** $0.001 < \text{FDR} \leq 0.01$, *** $\text{FDR} \leq 0.001$

Figures 2–6; Supplementary Table 1). The α and β -diversity analysis indicates that both diseases undergo shared microbial dysbiosis but with distinct disease-specific patterns emerging for the bacteriome/archaeome and fungal components.

Specific gut microbial members diagnostically distinguish disease-associated dysbiosis and enable the distinction between CD and ITB

Next, we identified the taxa (or modules of specific taxa) that distinguished between CD and ITB and between the two diseases and controls, and the extent to which they contributed to the distinction. For this, we adopted the

machine learning-based Random Forest classification approach, which not only identified the most discriminatory markers for pairwise group classification using the entire dataset but also validated these markers using an iterative bootstrapped approach to preclude biases originating from outlier samples (see “Methods” section for a complete description of the methodology; Supplementary Figure 7 for a complete pictorial description). The diagnostic markers were assessed separately for the gut microbiome and mycobiome.

At the microbiome level, we identified the most discriminatory taxonomic features (at genus and species level) achieving the highest diagnostic accuracy for

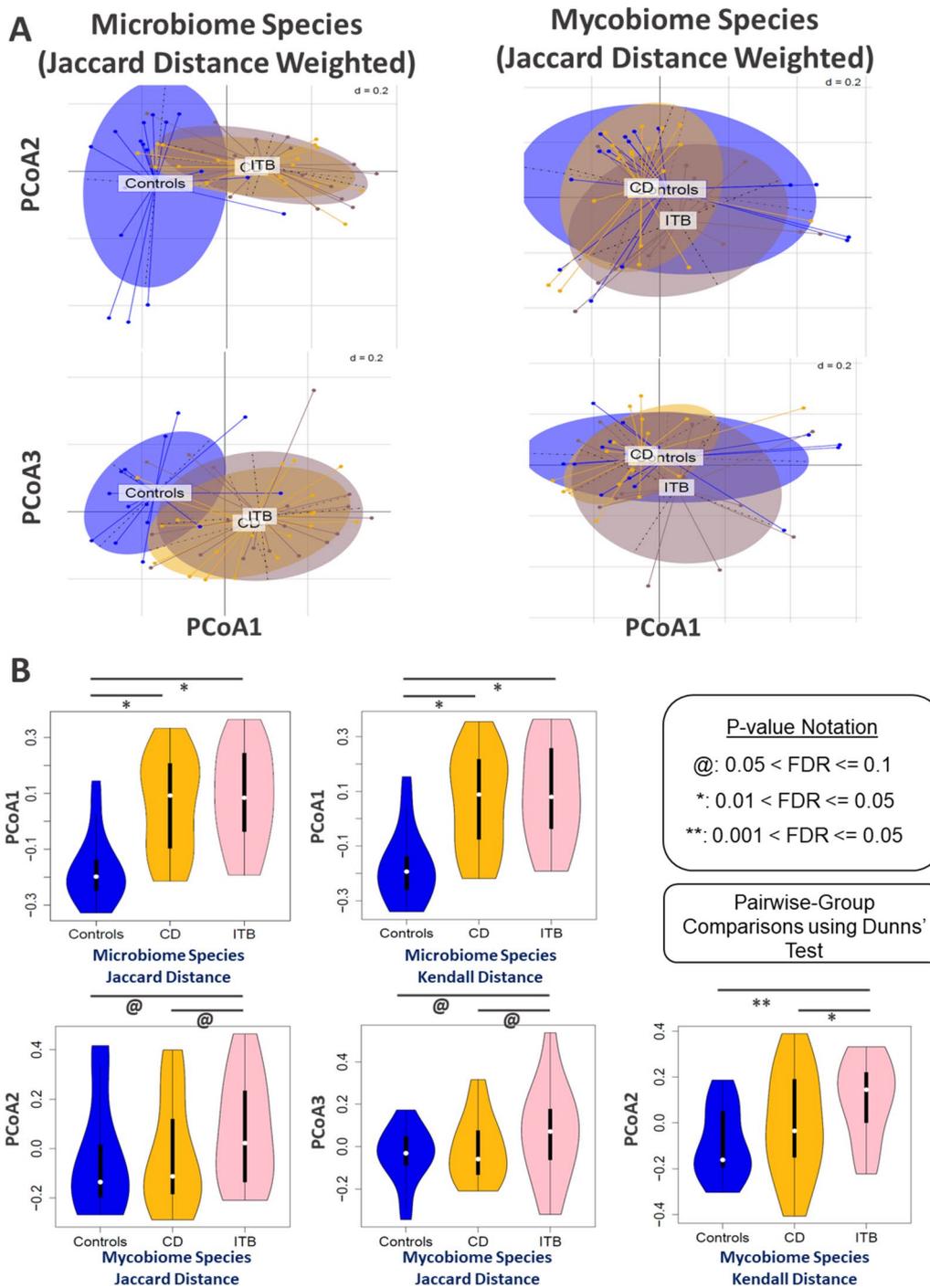


Fig. 2 ITB and CD indicate distinct patterns of microbiome- and mycobiome-level alterations with respect to the control group. **A** Principal Coordinate Analysis (PCoA) plots showing the variations of the Species-level Gut Microbiome profiles and Gut Mycobiome profiles of individuals belonging to the Control, ITB and CD groups. The PCoAs were performed using the weighted Jaccard distance measure. Species-level PCoAs obtained for the Kendall distance measure have been shown in Supplementary Figure 1. Similar plots for the other taxonomic levels using other distance measures are provided in Supplementary Figures 2, 3, 4, and 5. **B** Boxplot comparing the first species-level PCoA Axes obtained for the gut microbiome and mycobiome profiles obtained using the three different measures. Pairwise comparisons between the different groups are denoted by bars between corresponding boxes with the associated corrected *p* values (FDR) given in the notation: @: 0.05 < FDR <= 0.1; * 0.01 < FDR <= 0.05; ** 0.001 < FDR <= 0.01, *** FDR <= 0.001

pairwise discrimination between Controls vs. ITB (70 features; top AUC: 94.5%), Controls vs. CD (50 features; top AUC: 98.5%) and CD vs. ITB (10 features; top AUC: 76%), respectively (Fig. 3A, B). This indicated that while the microbiome alterations in the two gut inflammatory diseases were distinct from the controls, the disease groups also harbored subtle yet specific variations in their gut microbial community makeup. A similar pattern was also observed for the gut mycobiome community. We identified 10 fungal features enabling distinction between the controls and ITB group, and 30 fungal features discriminating the control vs. CD groups and the CD vs. ITB groups. The mycobiome alterations had the best capability to distinguish between Controls vs. ITB (94.4% AUC), followed by Controls vs. CD (83.3% AUC) and CD vs. ITB (75.7%), reflecting the trends observed for the gut microbiome. These observations highlight that the gut bacteriome/archaeome and mycobiome profiles couldn't just reliably distinguish between the patients (CD/ITB) and controls but also distinguished the diseases with appreciable accuracy. This entire investigation identified a total of 85 gut bacteria and 37 gut fungal taxa at the species and genus level that could facilitate diagnostically distinguishing between at least one pair of the three sample groups. These taxa have been summarized in Fig. 3C and D.

The microbiome taxa that could distinguish between controls and the two diseases and were enriched in the diseased group to controls included *Faecalibacterium* (*F. prausnitzii*), *Roseburia* (*R. inulinivorans*), *Coprococcus* (*C. catus*), *Eubacterium* (*E. ramulus*, *E. rectale*, *E. eligens*, *E. desmolans*), *Dorea* (*D. formicigenerans*, *D. longicatena*), *Catenibacterium*, *Fingoldia* (*F. magna*), *Blautia* (*B. faecis*, *B. luti*), *Prevotella*, *Megamonas* (*M. funiformis*) (fig. 3C). At the mycobiome level, the taxa with the same pattern of abundance alterations included *Blumeria*,

Aspergillus penicillioides and *Alternaria tenuissima*. On the other hand, the diagnostic taxa that were observed to be increased in both diseases with respect to (w.r.t) controls included *Fusobacterium*, *Hydrotaea* (*H. flava*), *Streptococcus* (*S. thermophilus*), *Collinsella*, *Gemella* (*G. haemolysans*), *Shewanella* (*S. amazonensis*), *Methylobacterium* (*M. aquaticum*), *Parabacteroides* (*P. distasonis*), *Veillonella* (*V. dispar*, *V. parvula*) at the microbiome level and; *Fusarium*, *Saccharomyces* (*S. cerevesiae*), *Parengyodontium* (*P. album*) and *Xenoacremonium* (*X. falcatum*) at the mycobiome level (fig. 3D).

At the bacteriome level, while *Bifidobacterium* and *Fingoldia* (*F. magna*) were specifically enriched in ITB w.r.t CD, *Bilophila* showed the opposite trend (enriched in CD w.r.t ITB). Additionally, the pathobiont *Ruminococcus gnavus*, along with *Lactobacillus* (*L. salivarius*), *Solobacterium moorei*, *Bacteroides stercoris*, *Collinsella* (*C. aerofaciens*) were significantly enriched in ITB w.r.t Controls (but unaltered in CD vs. controls analysis). Similarly, at the mycobiome level, *Candida tropicalis*, *Alternaria metachromatica* and *Phanerochaete* sp. showed significant enrichment in ITB w.r.t CD. Additionally, the overall abundances of *Malassezia*, *Candida*, and *Alternaria* showed a significant increase in ITB (w.r.t control) and a trend of enhancement in CD.

Differential taxa can be grouped into three distinct co-abundant modules in the gut microbiome

Next, we investigated if these diagnostic taxa could be arranged into specific co-abundant modules. For this purpose, we generated a co-abundance network of the above-identified 122 taxonomic features (85 gut bacterial and 37 gut fungal) (see “Methods” section). Separate co-abundance networks were constructed for the gut microbiome and mycobiome profiles.

The co-abundance network depicting the associations between the differentially abundant taxa consisted of

(See figure on next page.)

Fig. 3 Identification of differentially associating taxonomic features that are diagnostic of the three different groups (Controls, CD and ITB). **A** Identification of the most diagnostic features for each pairwise group classification, using multiple Random Forest models each considering only a varying number of top features (e.g. top 10, 20, 30, 40, till 250). For each pair of groups (Controls vs. CD, Controls vs. ITB, CD vs. ITB), the most diagnostic set of features were the ones corresponding to the model with the highest classification AUC. **B** Boxplots comparing the AUC ranges for the 50 iterative bootstrapped Random Forest (RF) model variants. As shown, the variants were generated for discriminating between each pair of groups (Controls vs. CD, Controls vs. ITB, CD vs. ITB), each considering only the top features identified (in **A**) for corresponding group-pair. For a given pair of groups, to create the RF variant in each iteration we randomly selected 50% of the samples (for generating the training RF model). This model was then tested on the rest 50% of the samples (corresponding to the concerned pair of groups). **C–D** Heatmaps showing the cross-group variation of the different taxonomic features identified in **A** to be amongst the top features discriminating across at least one pair of groups at the microbiome (i.e. Bacteriome/Archaeome) (**C**) and the mycobiome level (**D**). The upper heatmap groups these features based on their differential abundance/detection across each of the subject group pairs (Controls, ITB vs. Controls, ITB vs. CD; indicated by green color in the lower heatmap). In this scenario, for any pair in the notation ‘**A** vs. **B**’, the taxonomic features increased (in abundance or detection) in **A** with respect to **B** are highlighted in different shades of pink (FDR <=0.1) and those that are decreased are denoted in different shades of blue as denoted by the key. Markers that enable discrimination between controls vs. ITB/CD are highlighted in boxes with yellow boundaries, while those that facilitate distinguishing between ITB and CD are shown in boxes with blue lines. Abundance of features are denoted in blue font and detection are denoted in blue fonts

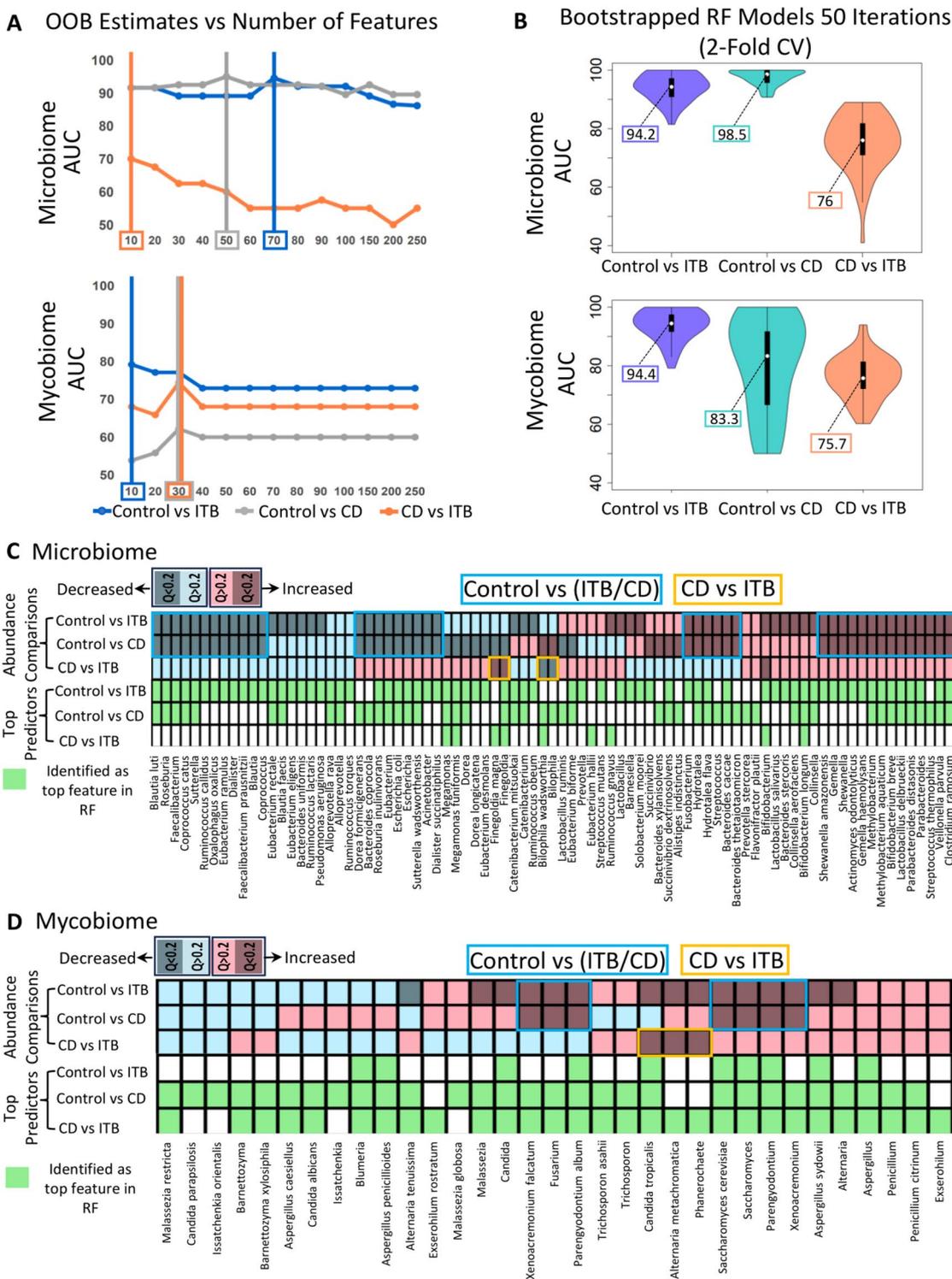


Fig. 3 (See legend on previous page.)

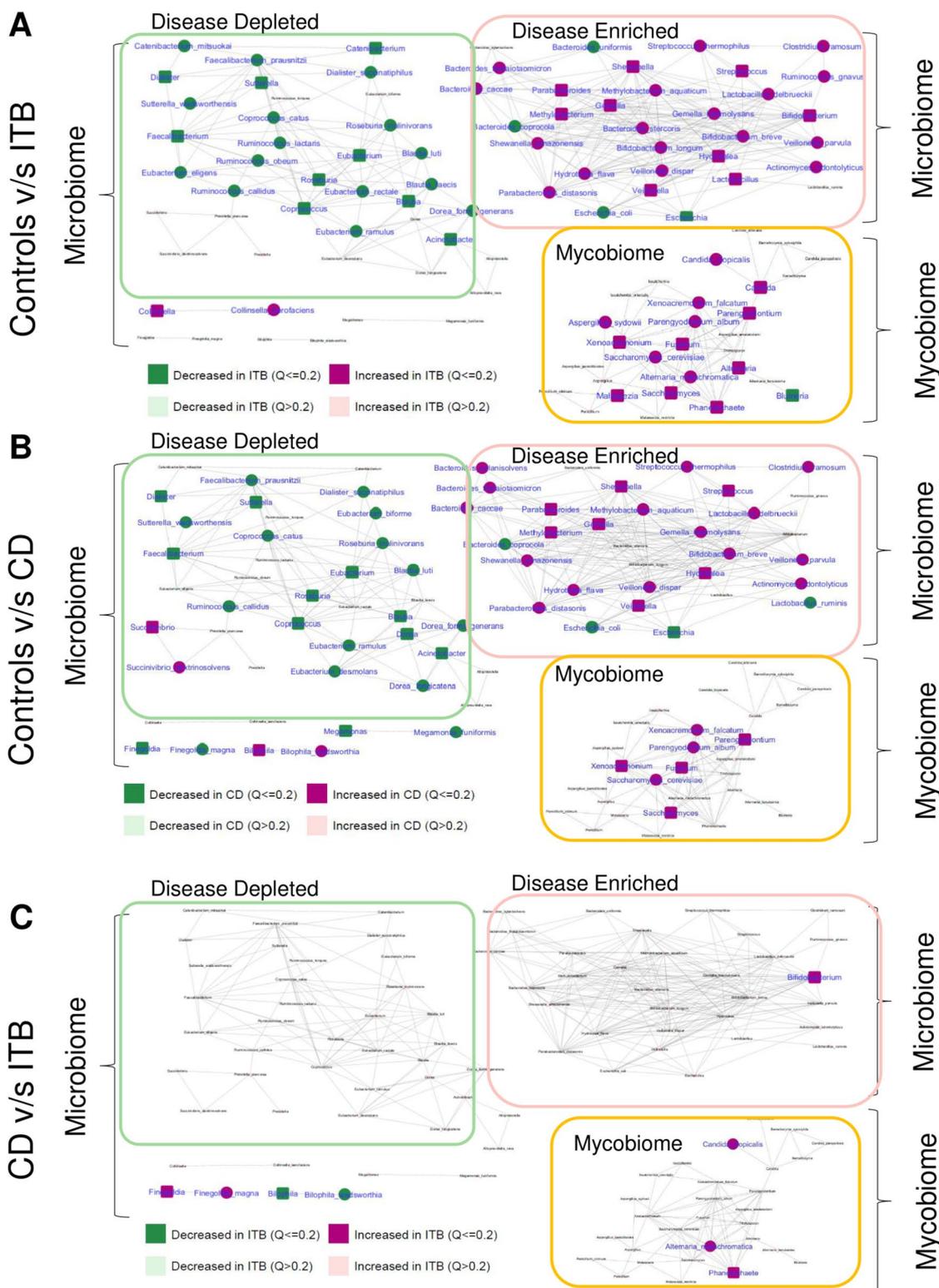


Fig. 4 Co-abundance network of diagnostic taxonomic features highlighting the three different modules. The three images in **A–C** show the same network, highlighting the only the differentially abundant taxa observed for **A** ITB vs. Controls, **B** CD vs. Controls, **C** ITB vs. CD. In this scenario, for any pair in the notation ‘**A** vs. **B**’, the taxonomic features increased (in abundance or detection) in **A** with respect to **B** are highlighted in pink ($FDR \leq 0.1$) and those that are decreased are denoted in blue as denoted by the key

three distinct modules. The “Disease Depleted” module consisted of putative beneficial bacterial taxa that were depleted across both diseases (highlighted in the previous section), such as *Faecalibacterium*, *Roseburia inulinivorans*, *Eubacterium eligens*, *Coprococcus catus*, *Coprococcus comes*, *Blautia luti*, *Dorea longicatena*, *Eubacterium ramulus*, etc. (Fig. 4A, B). Taxa that were enriched across both diseases, as identified in the previous section (Fig. 3C, D), comprised lineages like *Shewanella*, *Hydrotaea*, *Methylobacterium*, etc. (Fig. 4A, B). Amongst the ‘disease-enriched’ module, only *Bifidobacterium* was found to be enriched in ITB when compared to CD. Two sub-modules were also observed, one comprising *Fingoldia*, which increased in ITB compared to CD, and the other containing *Bilophila* (*B. wadsworthia*) enriched in CD compared to ITB (Fig. 4C).

The mycobiome co-abundance network was observed to consist of a central hub of multiple taxa linked through *Parengyodontium* to a *Candida*-specific sub-hub. While both the major hub and the sub-hub were observed to be increased in ITB vs. Controls, CD was enriched only for the major mycobiome hub to controls. There were also multiple disease-specific alterations within the mycobiome network. While *Candida tropicalis* was observed to be enriched in ITB vs. CD, *Alternaria metachromatica* and *Phanerochaete* showed the opposite trend.

To check the reproducibility of our findings on the co-abundance associations, we validated our module-specific co-abundance relationships using the global gut bacteriome/archaeome and mycobiome datasets. Identification of disease-associated modules reproduced in multiple global cohorts could indicate universal gut bacteriome/archaeome and mycobiome signatures which could be exploited as microbiome-based diagnostics and/or therapeutics. To evaluate this aspect, we developed intra-bacteriome and intra-mycobiome networks encompassing >5400 gut bacteriome profiles and ~900 gut mycobiome profiles collated from 14 studies across the globe (Table 1; Supplementary Figures 8, 9). We then investigated the reproducibility of the intra-modular edges observed within each module (Fig. 4), in the meta-networks identified in Supplementary Figures 8 and 9.

The intra-mycobiome associations showed distinctively low reproducibility across global cohorts. With threshold q value ≤ 0.1 , we could observe links between specific genera and their respective species (as expected) but no cross-genera associations. Relaxing the threshold to p value ≤ 0.05 increased the number of associations to 33 with only a few cross-clade links (*Malassezia restricta*—*Penicillium citrinum*; *Blumeria*—*Malassezia*; *Blumeria*—*Trichosporon asahii*; *Trichosporon*—*Aspergillus*). These results indicate the likelihood of intra-mycobiome associations being cohort-specific with low reproducibility

across cohorts (Supplementary Table 2; Supplementary Figure 9).

In contrast, the gut microbiome meta-networks revealed distinct patterns in the positioning of the taxa belonging to the ‘Disease-Depleted’ and ‘Disease-Enriched’ microbiome modules (Supplementary Figure 8A). Taxa belonging to the ‘Disease-Depleted’ module were observed to occupy central positions in the cross-cohort meta-network, showing significantly higher degree centrality measures (Mann–Whitney p value ≤ 0.002 ; Supplementary Figure 8A, B). We also observed stark differences in the reproducibility of these intra-modular edges across the different modules in the global meta-network with the intra-microbiome co-abundance network of taxa depicted in Fig. 4. Amongst the reproducible edges across the two co-abundance networks, more than 65% belonged to the ‘Disease-Depleted’ module.

This indicated that while the associations within the ‘Disease-Enriched’ and ‘Mycobiome’ modules are sporadic and observed only in the specific cohort of this study (as well as a likely consequence of indirect associations occurring due to other gut ecological changes), the taxa belonging ‘Disease-Depleted’ module not only constitute the central core of the global gut microbial community, but associations within this module are reasonably conserved across cohorts. This indicates a specific ecological role for these ‘disease-depleted’ taxa as ‘cornerstones’ within the microbiome whose depletion is linked with the onset of both diseases.

Loss of core gut microbiome is linked with the gut inflammation phenotype globally across multiple studies

Previous studies have indicated the loss of stability of the microbiome in IBD, with multiple studies including the current one, reporting the loss of diversity and diminishment of certain microbial taxa in gut inflammatory disorders [48–51]. Thus, we next checked if the two “Disease-Depleted” modules encompassed members of a core, health-associated gut microbiome, whose loss was driven by the inflammatory phenotype. Our hypothesis herein was that the “health-associated” microbiome module contained specific members that were not only prevalent across cohorts (as observed from their association with health and reproducible presence in global cohorts), but also influenced the stability of the microbiome (which could be a consequence of these taxa occupying the central nodes of the microbiome). To investigate this, we attempted to identify the core of the Indian gut microbiome. We collated 1617 gut microbiomes from seven studies. Subsequently, using a combination of prevalence and meta-network degree centrality (see “Methods” section), we identified a list of 46 taxa (28 species

and 18 genera) that were prevalent (Detection in at least four of the seven datasets; 70%) and central (degree centrality in the top 70 percentile) in the Indian gut meta-network (derived from the seven studies) (Fig. 5A, B; Supplementary Figure 10). More than 75% of the Indian core gut taxa belonged to the ‘Disease-Depleted’ Module-1 indicating that both the diseases were associated with the loss of this centrally connected core module in the gut microbial community.

Next, we probed if this loss of the core gut microbiome could be utilized to profile the extent of gut inflammation. We devised a simple quantitative score called the core gut microbiome score (CGMS). The CGMS for the microbiomes belonging to a given dataset was the summation of the rank-scaled abundances of the core Indian gut species (Fig. 5C) that were detected in the given dataset, with the abundances of each species being weighted by its percentile in the degree centrality in the meta-network (Fig. 5A).

We then investigated seven study cohorts (including the current study) from India, Sweden, the US and Italy and investigated the variation of the CGMS across different inflammation phenotypes. In the current study, the controls were observed to have significantly higher CGMS values as compared to CD and ITB (with no significant differences between the two disease types) (Mann–Whitney FDR ≤ 0.05). On the other hand, our previous study investigated two different inflammatory disorders (CD/UC and Acute Severe Ulcerative Colitis: ASUC) [23]. ASUC subjects have a higher degree of inflammation than CD/UC. This was also reflected in the CGMS scores, wherein the controls had CGMS values significantly higher than either CD/UC or ASUC (Mann–Whitney FDR ≤ 0.001 for both), and ASUC had significantly lower CGMS values compared to CD/UC (Mann–Whitney FDR ≤ 0.001), indicating a greater loss of core in ASUC.

We next investigated if the CGMS scores defined on the Indian population could predict inflammation phenotypes in geographically distinct datasets. For this purpose, we investigated the IBD cohort of the Human Microbiome Project [45]. Although defined on the gut

microbiome core of the Indian population, the CGMS scores could efficiently distinguish between Controls and IBD patients with the CGMS for patients being significantly lower (Mann–Whitney p value ≤ 0.001). We also investigated another Italian cohort [32], which had gut microbiome profiles sampled from Controls, and IBD patients with active disease and those in remission. Even in this case, we observed that the CGMS values of the controls were significantly higher than those of the patients. The values of individuals in remission were noticeably higher than the patients, indicating a partial core recovery. Two additional cohorts [Hall et al. ($n=259$) and Lloyd-Price et al. ($n=1627$)] [8, 45] consisting of gut microbiome profiles of healthy controls and of the patients with IBD, showed significantly low CGMS in the diseased samples when compared with controls. Swedish cohort [Halfvarson et al. ($n=683$)] too, showed a significant reduction in CGMS in CD [46]. Thus, these results indicate that the abundance profiles of the core members of the microbiome can be utilized to measure the inflammation status of individuals.

We then compared the efficacy of the CGMS (as an indicator of gut health) with four other indices previously proposed for this purpose. These included the Shannon Diversity and the GMHI (both known to be positively associated with controls or ‘healthy’ individuals), and Dysbiosis-Score and Kendall Uniqueness (known to be associated with diseased groups) This comparative evaluation showed CGMS as the only indicator showing significant reduction in the IBD-disease groups as compared to the controls across all seven cohorts. The GMHI and Dysbiosis-Score were second in performance showing significant decrease and increase in five out of the seven cohorts respectively, followed by Kendall Uniqueness (significantly increased in four cohorts). The worst performing indicator was Shannon Diversity (Fig. 5E). This indicates that the CGMS is able to perform equally or better than other indicators of gut health in the seven cohorts.

(See figure on next page.)

Fig. 5 Loss of the core gut microbiome is associated with gut inflammation and its severity across multiple studies. **A** Multi-study meta-analysis approach utilized for identification of the core gut microbiota in the Indian populations. **B** Co-abundance meta-network shown only for the core taxa in the India gut microbiomes derived from the seven studies. **C** Degree centrality of the identified core India gut taxa at the species level. Taxa are listed in descending order of their degree of centrality. **D** Comparison of the CGMS in different inflammation disease phenotypes in the current study, and the previous Kedia et al. study from India [13, 39]. The CGMS are also compared for five additional external studies from North America [8, 45–47], two studies from Europe (Liguori et al. from Italy [32]; Halfvarson et al. from Sweden [46]). **E**. Heatmap showing the association of four major indices of gut health along with CGMS with IBD-disease groups across the seven cohorts. Red boxes marked -1 denote a significant decrease (with Mann–Whitney test p value ≤ 0.05); blue boxes marked 1 denote a significant increase (with Mann–Whitney p value ≤ 0.05). Depicting the comparative evaluation for their association with non-diseased and IBD-patients across all investigated cohorts

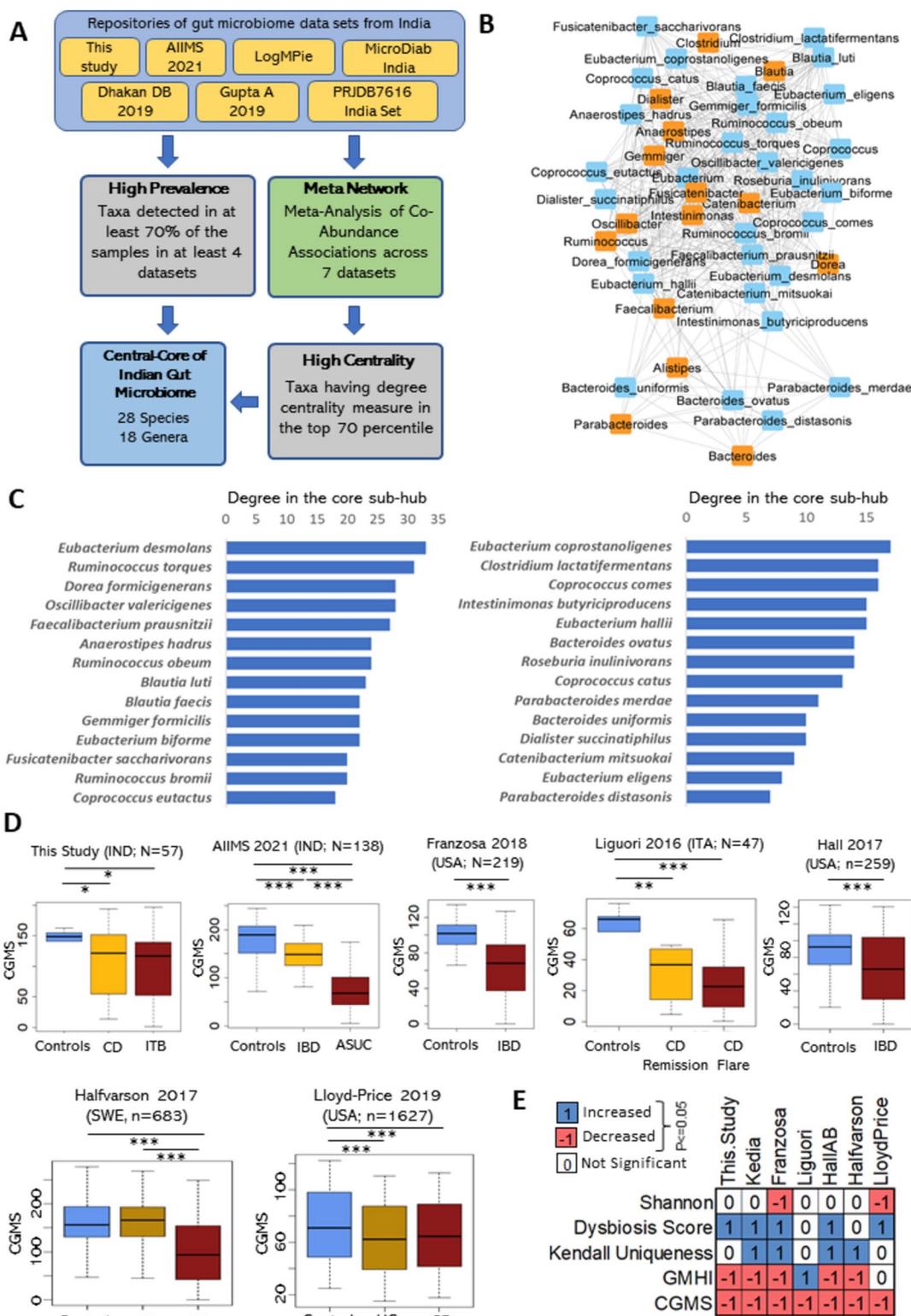


Fig. 5 (See legend on previous page.)

Investigation of taxa-inferred metabolic profiles identifies specific metabolites whose production or consumption/ degradation associated with loss of control-associated microbiome

In the last part of this investigation, we inferred the putative metabolic functionalities in the different gut microbiomes based on their taxonomic composition. For this purpose, we utilized taxa-to-metabolic functionality maps of experimentally validated metabolic functional profiles collated and used as part of multiple previous studies (See Methods). Given the species-level-taxonomic composition of a gut microbiome, this approach enabled inferring of metabolic profile given the abundances of its constituent taxa (see “Methods” section).

Comparing the inferred metabolic profiles (using Principal Coordinate Analysis or PCoA) across the three groups of subjects included in this study identified significant differences across the three groups (PERMANOVA using Bray–Curtis: R -squared=0.11 and p =0.009) (Fig. 6A), with functional variations across the three groups, most prominent along the first Principal Coordinate PCo1 (Fig. 6B). The most significant variation of the functional profiles was observed between the ITB and the Control subjects (Mann–Whitney p value=0.005), with the CD group at intermediate position.

We further identified the different metabolic functionalities whose abundances associated with Control-to-Disease variation along the PCo1, by correlating the abundance of each metabolic functionality (i.e. the cumulated abundance of the taxa having that metabolic functionality) with PCo1 values. We identified a total of 43 metabolite production profiles significantly associated with PCo1 values. While the production of all three short-chain fatty acids (Acetate, Propanoate and Butyrate), along with other metabolites like CO₂ and Methylamine, were putatively enriched in controls, the metabolite production profiles associated with

disease-states contained, besides other, the production of the two primary bile acids (BAs), cholic and chenodeoxycholic acids and the secondary hydrophobic BAs (deoxycholic and lithocholic acids) (Fig. 6C). We also identified 83 metabolic consumption profiles associated with PCo1 values. Notably, the list of most positively associated metabolite consumptions included the consumption of glycine- and taurine-conjugated BAs (Fig. 6D). This indicated that diseased microbiomes in our study were enriched for taxa consuming glycine- and taurine-conjugated BAs and producing primary and secondary BAs.

Since the analysis involved inferred metabolic functionalities, we further checked if the enrichment of these functionalities with disease were also replicated in global cohorts. We thus computed inferred metabolite profiles across six other global matched IBD-control cohorts (investigated in Fig. 5D, E) and performed a Random-Effect Model meta-analysis. Validating the associations, the positive association of taurine-conjugated BA consumption and an increased production potential of primary BA and secondary BA (de-oxycholic: DCA) with IBD phenotypes was replicated in this global meta-analysis (Fig. 6C, D; Supplementary Figure 11).

Thus, we next checked which specific microbiome members were associated with these specific functions in our study cohort. This revealed that the above conversion functionalities were primarily present in multiple *Bifidobacterium* and *Lactobacilli*, along with specific *Bacteroides* species (fragilis/vulgatus/thetaiotamicron) (Fig. 6E). Amongst them, six taxa, *Bifidobacterium longum/breve*, *Lactobacillus gasseri/johnsonii* and *Bacteroides fragilis/vulgatus* were observed to be significantly enriched in disease microbiomes. We further validated these associations across the global cohorts. Despite cohort-specific variations, despite being known probiotics, *B. longum*, *B. breve*, and *L. gasseri*, were significantly enriched in IBD patients across multiple cohorts, along

(See figure on next page.)

Fig. 6 Identification of metabolites associated with the progression from controls towards CD and ITB in terms of consumption, degradation and production through inferred metabolite profiling. **A** The Principal Co-ordinate Analysis (PCoA) plot describes the variation in microbiomes in three different groups (Controls, CD and ITB) which has been shown in different color legends. Bray–Curtis distance measure has been used to perform the PCoA analysis and the p value and R^2 value obtained from PERMANOVA using Bray–Curtis distance has been mentioned as well. **B** The top right plot (box plots) captures the values of Principle Co-ordinate 1 across three different groups (Controls, CD and ITB) and the significance in the difference between groups have been shown through the p value calculated through Wilcoxon Rank-Sum test. **C** Volcano plot showing the association of different metabolite production profiles with PCo1 (p values obtained from the Spearman correlation analysis were adjusted using Benjamini–Hochberg method for multiple corrections). The same volcano-plot for the consumption/degradation profiles is shown in **D**. The directionality of association of different functionalities with PCo1 as well as with Control or Diseased microbiome is also indicated. Functionalities that were also validated in the Random Effect Model meta-analysis in Supplementary Figure 11 are shown in red-boxes. **E** Heatmap on the left panel shows the different taxa associated with the different BA production/consumption profiles. The bar-plot in the middle shows the association of these taxa with PCo1 values of part **A** of this figure (positive values indicating disease association). The right most heatmap shows the enrichment or depletion in IBD patients across the global cohorts

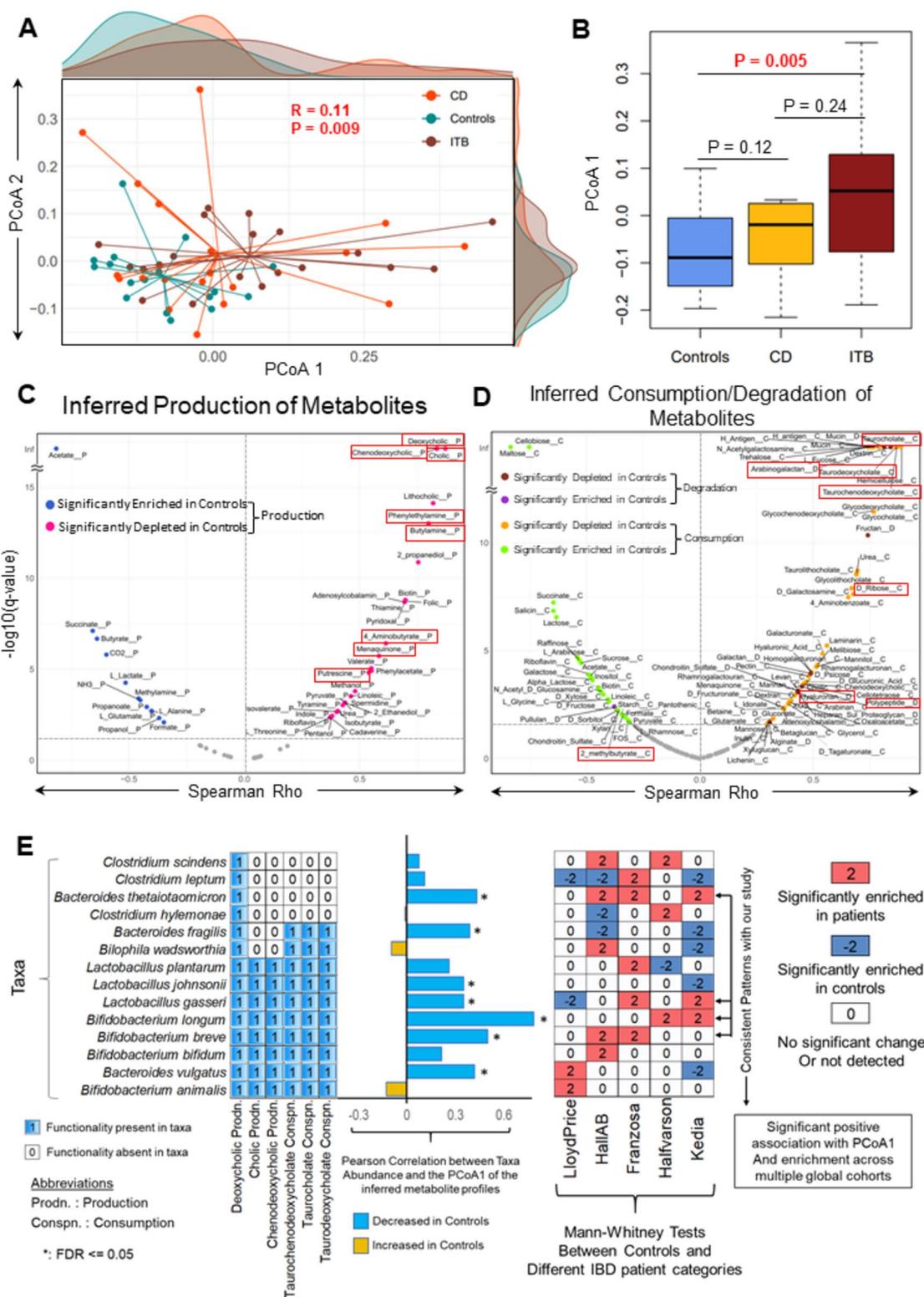


Fig. 6 (See legend on previous page.)

with *B. thetaiotamicron*, indicating that these alterations are replicable in other cohorts.

Discussion

The study highlights CD and ITB-associated gut bacterial and fungal dysbiosis signatures. It identifies disease-associated co-occurring modules to evaluate whether this dysbiosis signature is a disease-specific trait or a shared feature across GI diseases of diverging etiologies.

While CD is known to involve a reduction in gut microbial α -diversity, our results highlight that ITB undergoes an even sharper reduction in gut bacterial and fungal α -diversity (with the loss of fungal diversity is more pronounced in ITB rather than CD). β -diversity analysis also mirrored similar disease-specific patterns of dysbiosis. Based on bacteriome profiles, CD and ITB samples were found to be quite similar to each other and significantly divergent from controls. However, comparing mycobiome profiles between the three groups revealed remarkable differences between CD and ITB, with the latter converging with controls. This implicates the gut mycobiome to be a key distinguishing factor between the two diseases. ITB witnessed a significant loss of gut fungal diversity, while the gut mycobiome remained relatively unaltered in CD. *Malassezia*, a lipid-thriving fungus known to release pro-inflammatory free lipids from host tissues, was found to be enhanced in CD [52, 53].

A key finding of the current study is the multi-cohort-integrated identification of a panel of 46 gut microbial members that occupy central positions in the Indian gut microbial community. A combination of specific differentially abundant/diagnostic marker analysis revealed a disease-associated loss of these core microbiome members common in both CD and ITB. On the other hand, this loss of the core members is associated with a sporadic expansion of putatively detrimental disease-associated pathobionts. This phenomenon is yet another example of the Anna-Karenina principle applied to host microbiomes: *All happy microbiomes are alike; Each unhappy microbiome is unhappy in its way*. This has also been observed in multiple studies focusing on gut microbiome modulations in unhealthy aging, where in the healthy elderly have relatively similar core-enriched, youth-like microbiomes [21, 54]. The loss of this state is associated with an enrichment of multi-disease-associated taxa driving the host towards a physiological decline [21]. This loss of health-associated microbial signature in IBD had been previously highlighted by Halfvarson et al., where the healthy microbiomes were restricted to a small volume of ordination space, designated as a 'healthy plane', and the microbiome from the IBD subjects deviated significantly from this plane [46]. This observation was further

consolidated in another Human Microbiome Project study investigating the multiomics of the gut microbiome ecosystem in IBD. Microbiomes from IBD subjects with active disease showed higher dysbiotic scores and decreased gut microbiome stability when compared to the non-IBD controls and the subjects with inactive disease [8]. A study by our group highlighted similar attributes of gut dysbiosis in UC and acute severe colitis along with obliterated inter-microbe interactions. Co-abundance network analysis in this report showed significant negative associations between the "health-specific" and "disease-associated" genera in the healthy control, while such negative associations disappeared in the IBD microbiome [39]. Studies investigating the role of the Mediterranean diet on host health have also observed a diet-associated increase of specific core gut microflora, which are associated with improvements in host inflammation, cognition and physical well-being [21]. This significance of the core and central gut bacterial taxa to maintain gut homeostasis, propelled us in this study to investigate and formulate 28 species-level taxa panel-based core-gut microbiome score (CGMS). The CGMS performed rather well as an indicator of inflammation extent and phenotype, not only in Indian studies but also in distinguishing inflammation phenotypes of IBD patients (disease, non-diseased, remission) from other geographical locations and ethnicities (an analysis encompassing seven case-control microbiome datasets with >3000 gut microbiomes) [8, 34]. This is notable given the large-scale variabilities in the baseline gut microbiome composition across host characteristics like geography, demographics, ethnicity and lifestyle [55, 56]. Besides highlighting the role of scores like CGMS as clinically actionable targets, the results indicate that loss of specific core and central gut bacterial members (the panel of 28 species-level taxa) is a conserved inflammation-linked phenomenon across the global cohorts and suggests a pivotal role of these bacteria in health. These observations are a putative indication of how gut bacterial dysbiosis, rather than being a contributor to disease-specific etiology, could be an effect of the inflammatory milieu that prevails both in CD and ITB.

Investigation of the predicted metabolic profiles (both in the current and the global cohort), shows consumption of taurine-conjugated BAs and production of primary BAs (and one secondary BA: DCA), that are primarily derived from Bifidobacterium, Lactobacillus and specific Bacteroides, are enriched in inflammation phenotypes. While both an increase of Bifidobacteria and Lactobacilli [57], and an increased production of primary BAs [47], have been associated with IBD incidence, the increased DCA production is contrary to

previous observations. Nevertheless, a majority of taxa identified here have both functionalities. Furthermore, it is also not known whether enrichment of these specific pathways because of taxonomic alterations is a cause or a response to host physiological changes in gut inflammatory disorders (one of the limitations of the current study).

Even though the study gives important insights into the gut microbiome alterations in two GI diseases with overlapping clinical manifestations and is amongst the first reports delving into the ITB-associated gut microbial dysbiosis, the results must be interpreted keeping in mind a few limitations. The CD and ITB gut bacteriome and mycobiome have been derived from 16S and ITS-based amplicon sequencing analysis and must be validated by shot-gun metagenomics sequencing to gain deeper taxonomic assignment and functional significance of the predicted metabolic variations in the community. Even though the results describing the depletion of the core microbiome in inflammatory diseases have been backed by meta-networks constructed from samples from other cohorts, the analysis of ITB-associated bacterial and fungal dysbiosis must be validated by a larger sample size.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13099-024-00654-4>.

Additional file 1. Supplementary Table 1: The table presents the results of the beta-diversity investigations for the three different subject groups, by comparing the top three PCoA axes obtained using the three different distance measures for the gut microbiome composition at the level of A. Species B. Genus and C. ASVs; and the gut mycobiome composition at the level of D. Species, E. Genus and F. OTUs.

Additional file 2. Supplementary Table 2: Edge-list of the intra-mycobiome meta-network generated using A. q-value threshold of less than 0.1 B. p-value of less than 0.05.

Additional file 3. Supplementary Figure 1: The full Principal Coordinate Analysis (PcoA) based workflow adopted for investigating the microbiome-level and mycobiome-level beta diversity variations in the three different subject groups in the current study for the three different taxonomic levels using the three different distance measures, along with the identification of the Principal Coordinate Axes with the most significant variations across the groups.

Additional file 4. Supplementary Figure 2: Principal Coordinate Analysis (PCoA) plots showing the variations of the Specie-level Gut Microbiome profiles and the OTU-level Gut Mycobiome profiles of individuals belonging to the Control, ITB and CD groups using the Kendall distance measures.

Additional file 5. Supplementary Figure 3: A. Principal Coordinate Analysis (PCoA) plots showing the variations of the ASV and OTU-level Gut Microbiome profiles and Gut Mycobiome profiles of individuals belonging to the Control, ITB and CD groups. The PCoAs were performed using the weighted Jaccard distance measure. ASV and OTU-level PCoAs obtained for the Kendall distance measure have been shown in Supplementary Figure 4. B. Boxplot comparing the first ASV and OTU-level PCoA Axes obtained for the gut microbiome and mycobiome profiles obtained using the three different measures. Pairwise comparisons between the different

groups are denoted by bars between corresponding boxes with the associated corrected p-values (FDR) given in the notation: @: $0.05 < \text{FDR} < = 0.1$; *: $0.01 < \text{FDR} < = 0.05$; **: $0.001 < \text{FDR} < = 0.01$, ***: $\text{FDR} < = 0.001$.

Additional file 6. Supplementary Figure 4: Principal Coordinate Analysis (PCoA) plots showing the variations of the Specie-level Gut Microbiome profiles and the ASV and OTU-level Gut Mycobiome profiles of individuals belonging to the Control, ITB and CD groups using the Kendall distance measure.

Additional file 7. Supplementary Figure 5: A. Principal Coordinate Analysis (PCoA) plots showing the variations of the Genus-level Gut Microbiome profiles and Gut Mycobiome profiles of individuals belonging to the Control, ITB and CD groups. The PCoAs were performed using the weighted Jaccard distance measure. Genus-level PCoAs obtained for the Kendall distance measures have been shown in Supplementary Figure 4. B. Boxplot comparing the first Genus-level PCoA Axes obtained for the gut microbiome and mycobiome profiles obtained using the three different measures. Pairwise comparisons between the different groups are denoted by bars between corresponding boxes with the associated corrected p-values (FDR) given in the notation: @: $0.05 < \text{FDR} < = 0.1$; *: $0.01 < \text{FDR} < = 0.05$; **: $0.001 < \text{FDR} < = 0.01$, ***: $\text{FDR} < = 0.001$.

Additional file 8. Supplementary Figure 6: Principal Coordinate Analysis (PCoA) plots showing the variations of the Specie-level Gut Microbiome profiles and the ASV and OTU-level Gut Mycobiome profiles of individuals belonging to the Control, ITB and CD groups using the Kendall distance measure.

Additional file 9. Supplementary Figure 7: Overview of the bioinformatic workflow adopted for identification of the diagnostic markers for the three pairwise group classifications (Controls v/s CD, Controls v/s ITB and CD v/s ITB) using Random Forest models.

Additional file 10. Supplementary Figure 8: A. Layout of the intra-microbiome co-abundance meta-networks built using 5,347 gut microbiomes. Please refer to Table 1 for the detailed list of the datasets used. The taxa nodes belonging to the Disease Depleted and Disease Enriched Microbiome Groups are denoted in different colors. B. Comparison of the degree centrality measures of the Disease Depleted and Disease Enriched groups in the meta-network visualized as boxplots. The p-value of comparison of degree centrality measures is also indicated. C. Pi-chart showing the representation of the number of edges within of the Disease Enriched and Disease Depleted modules (as well as across the other taxa) identified in Figure 4 that were reproduced in the intra-microbiome co-abundance networks. Also indicated is the p-value of the comparison of the proportion of the number of reproducible edges between the Disease Depleted and Disease Enriched groups obtained using Fishers' exact test.

Additional file 11. Supplementary Figure 9. Layout of the intra-mycobiome co-abundance meta-networks built using ~900 gut mycobiomes. Please refer to Table 1 for the detailed list of the datasets used and Supplementary Table 2 for the detailed edge-list and thresholds used for generating the network.

Additional file 12. Supplementary Figure 10: Overlap of the taxa constituting the core of the Indian gut microbiome and the different taxonomic modules identified in the current study in Figure 5.

Additional file 13. Supplementary Figure 11: Enrichment and depletion of different metabolites (in terms of production, consumption and degradation) came from the analysis using Random Effect Model on six different cohorts. Beta values came from the analysis have been plotted against the negative logarithm of p-values and the enrichment and depletion has been shown using color legends. The common metabolites (in terms of production, consumption and degradation) between this plot and the plots of Figure 6C-D have been marked in red boxes for validation purpose.

Author contributions

Conceptualization; AB, MM, TSG, SKV, SM, SK, VA; Data curation; AB, MM, AS, SG, SK, TSG, VA; Formal Analysis and investigation, TSG, AS, SG, AM, MM, VA; Methodology; AB, MM, TSG, SK, VA; Investigation; AB, MM, AS, SG, SKB, SM, BK,

PK, GM, SK, TSG, VA; Writing-original draft; AB, MM, TSG, SG, VA; Writing-review and editing; TSG, VA, SK. All authors read and approved the final manuscript.

Funding

Indian Council of Medical Research: Center for Advanced Research and Excellence in Intestinal Diseases awarded to Professor Vineet Ahuja (grant number: 55/4/11/CARE-ID/2018-NCD-II).

Availability of data and materials

All sequence data (corresponding to the Crohn's and Intestinal Tuberculosis patients along with the non-diseased controls) generated as part of this study have been deposited in the European Nucleotide Archive with the accession number PRJEB62109. The data shall be made publicly available upon acceptance of the manuscript. All codes and data files utilized for the current study are available at the GitHub repository: <https://github.com/tsg-microbiome/IIITD-AIIMS-ITB-Study>.

Declarations

Ethics approval and consent to participate

The experimental protocols were approved by the institutional ethics review board of the All India Institute of Medical Sciences, New Delhi (Ref. No. IECPG/484/29.8.16).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Gastroenterology, All India Institute of Medical Sciences, New Delhi 110029, India. ²Department of Computational Biology, Indraprastha Institute of Information Technology-Delhi, New Delhi, India.

Received: 8 August 2024 Accepted: 6 October 2024

Published online: 07 November 2024

References

- Hsia K, Zhao N, Chung M, Algarrahi K, Kouhsari LM, Fu M, et al. Alterations in the Fungal Microbiome in Ulcerative Colitis. *Inflamm Bowel Dis*. 2023;29(10):1613–21. <https://doi.org/10.1093/ibd/izad082>.
- Caruso R, Lo BC, Núñez G. Host-microbiota interactions in inflammatory bowel disease. *Nat Rev Immunol*. 2020;20:411–26.
- Matsuoka K, Kanai T. The gut microbiota and inflammatory bowel disease. *Semin Immunopathol*. 2015;37(1):47–55.
- Underhill DM, Braun J. Fungal microbiome in inflammatory bowel disease: a critical assessment. *J Clin Invest*. 2022;132(5): e155786. <https://doi.org/10.1172/JCI155786>.
- Palmela C, Chevarin C, Xu Z, Torres J, Sevrin G, Hirten R, et al. Adherent-invasive *Escherichia coli* in inflammatory bowel disease. *Gut*. 2018;67(3):574–87.
- Chehoud C, Albenberg LG, Judge C, Hoffmann C, Grunberg S, Bittering K, et al. Fungal signature in the gut microbiota of pediatric patients with inflammatory bowel disease. *Inflamm Bowel Dis*. 2015;21:1948–56.
- Schirmer M, Franzosa EA, Lloyd-Price J, McIver LJ, Schwager R, Poon TW, et al. Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat Microbiol*. 2017;3:337–46.
- Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*. 2019;569(7758):655–62.
- Makharia GK, Srivastava S, Das P, Goswami P, Singh U, Tripathi M, et al. Clinical, endoscopic, and histological differentiations between Crohn's disease and intestinal tuberculosis. *Am J Gastroenterol*. 2010;105(3):642–51.
- Kedia S, Das P, Madhusudhan KS, Dattagupta S, Sharma R, Sahni P, et al. Differentiating Crohn's disease from intestinal tuberculosis. *World J Gastroenterol*. 2019;25(4):418–32.
- He C, Wang H, Yu C, Peng C, Shu X, Liao W, et al. Alterations of gut microbiota in patients with intestinal tuberculosis that different from Crohn's disease. *Front Bioeng Biotechnol*. 2021;9: 673691.
- Khan IA, Nayak B, Markandey M, Bajaj A, Verma M, Kumar S, et al. Differential prevalence of pathobionts and host gene polymorphisms in chronic inflammatory intestinal diseases: Crohn's disease and intestinal tuberculosis. *PLoS ONE*. 2021;16(8): e0256098.
- Maaser C, Sturm A, Vavricka SR, Kucharzik T, Fiorino G, Annesse V, et al. ECCO-ESGAR guideline for diagnostic assessment in IBD part 1: initial diagnosis, monitoring of known IBD, detection of complications. *J Crohns Colitis*. 2019;13(2):144–64.
- Travis SPL, Stange EF, Lémann M, Öresland T, Chowers Y, Forbes A, et al. European evidence based consensus on the diagnosis and management of Crohn's disease: current management. *Gut*. 2006;55 Suppl 1(Suppl 1):i16–35.
- Haubrich WS. *Bockus gastroenterology 4. Irritable Bowel Syndrome*. 1985;4:2425–2444.
- Bag S, Saha B, Mehta O, Anbumani D, Kumar N, Dayal M, et al. An improved method for high quality metagenomics DNA extraction from human and environmental samples. *Sci Rep*. 2016;6:26775.
- Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019;37:852–7.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13:581–3.
- Allard G, Ryan FJ, Jeffery IB, Claesson MJ. SPINGO: a rapid species-classifier for microbial amplicon sequences. *BMC Bioinformatics*. 2015;16:324.
- Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ*. 2016;4: e2584.
- Ghosh TS, Shanahan F, O'Toole PW. Toward an improved definition of a healthy microbiome for healthy aging. *Nat Aging*. 2022;2:1054–69.
- Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, et al. Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol*. 2012;8: e1002606.
- Kedia S, Rampal R, Paul J, Ahuja V. Gut microbiome diversity in acute infective and chronic inflammatory gastrointestinal diseases in North India. *J Gastroenterol*. 2016;51(7):660–71.
- Das B, Ghosh TS, Kedia S, Rampal R, Saxena S, Bag S, et al. Analysis of the gut microbiome of rural and urban healthy Indians living in sea level and high altitude areas. *Sci Rep*. 2018;8(1):10104.
- Alvarez-Silva C, Kashani A, Hansen TH, Pinna NK, Anjana RM, Dutta A, et al. Trans-ethnic gut microbiota signatures of type 2 diabetes in Denmark and India. *Genome Med*. 2021;13(1):37.
- Pinna NK, Anjana RM, Saxena S, Dutta A, Gnanaprakash V, Rameshkumar G, et al. Trans-ethnic gut microbial signatures of prediabetic subjects from India and Denmark. *Genome Med*. 2021;13(1):36.
- Dubey AK, Uppadhya N, Nilawe P, Chauhan N, Kumar S, Gupta UA, et al. LogMPIE, pan-India profiling of the human gut microbiome using 16S rRNA sequencing. *Sci Data*. 2018;5: 180232.
- Dhakan DB, Maji A, Sharma AK, Saxena R, Pulikkan J, Grace T, et al. The unique composition of Indian gut microbiome, gene catalogue, and associated fecal metabolome deciphered using multi-omics approaches. *GigaScience*. 2019;8(3):giz004. <https://doi.org/10.1093/gigascience/giz004>.
- Gupta A, Dhakan DB, Maji A, Saxena R, Vishnu Prasoodanan PK, Mahajan S, et al. Association of *Flavonifractor plautii*, a flavonoid-degrading bacterium, with the gut microbiome of colorectal cancer patients in India. *mSystems*. 2019;4(6): e00438-19.
- He Y, Wu W, Zheng HM, Li P, McDonald D, Sheng HF, et al. Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nat Med*. 2018;24:1532–5.
- Pareek S, Kurakawa T, Das B, Motooka D, Nakaya S, Rongsen-Chandola T, et al. Comparison of Japanese and Indian intestinal microbiota shows diet-dependent interaction between bacteria and fungi. *npj Biofilms Microbiomes*. 2019;5(1):37.

32. Liguori G, Lamas B, Richard ML, Brandi G, da Costa G, Hoffmann TW, et al. Fungal dysbiosis in mucosa-associated microbiota of Crohn's disease patients. *J Crohns Colitis*. 2016;10:296–305.
33. Das A, O'Herlihy E, Shanahan F, O'Toole PW, Jeffery IB. The fecal mycobiome in patients with irritable bowel syndrome. *Sci Rep*. 2021;11(1):124.
34. Jeffery IB, Das A, O'Herlihy E, Coughlan S, Cisek K, Moore M, et al. Differences in fecal microbiomes and metabolomes of people with vs. without irritable bowel syndrome and bile acid malabsorption. *Gastroenterology*. 2020;158:1016–1028.e8.
35. Nash AK, Auchtung TA, Wong MC, Smith DP, Gesell JR, Ross MC, et al. The gut mycobiome of the Human Microbiome Project healthy cohort. *Microbiome*. 2017;5:153. <https://doi.org/10.1186/s40168-017-0373-4>.
36. Jayasudha R, Das T, Chakravarthy SK, Prashanthi GS, Bhargava A, Tyagi M, et al. Gut mycobiomes are altered in people with type 2 diabetes mellitus and diabetic retinopathy. *PLoS ONE*. 2020;15(12): e0243077.
37. Hu J, Wei S, Gu Y, Wang Y, Feng Y, Sheng J, et al. Gut mycobiome in patients with chronic kidney disease was altered and associated with immunological profiles. *Front Immunol*. 2022;13: 843695.
38. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498.
39. Kedia S, Ghosh TS, Jain S, Desigamani A, Kumar A, Gupta V, et al. Gut microbiome diversity in acute severe colitis is distinct from mild to moderate ulcerative colitis. *J Gastroenterol Hepatol*. 2021;36:731–9. <https://doi.org/10.1111/jgh.15232>.
40. Gupta VK, et al. A predictive index for health status using species-level gut microbiome profiling. *Nat Commun*. 2020;11:4635.
41. Ghosh TS, Das M, Jeffery IB, O'Toole PW. Adjusting for age improves identification of gut microbiome alterations in multiple diseases. *Elife*. 2020;9: e50240.
42. Ghosh TS, et al. Mediterranean diet intervention alters the gut microbiome in older people reducing frailty and improving health status: the NU-AGE 1-year dietary intervention across five European countries. *Gut*. 2020;69:1218–28.
43. Noronha A, et al. The virtual metabolic human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Res*. 2019;47:D614–24.
44. Sung J, et al. Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis. *Nat Commun*. 2017;8:15393.
45. Hall AB, Yassour M, Sauk J, Garner A, Jiang X, Arthur T, et al. A novel *Ruminococcus gnavus* clade enriched in inflammatory bowel disease patients. *Genome Med*. 2017;9(1):103. <https://doi.org/10.1186/s13073-017-0490-5>.
46. Halfvarson J, Brislawn CJ, Lamendella R, Vázquez-Baeza Y, Walters WA, Bramer LM, et al. Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat Microbiol*. 2017;2:17004.
47. Franzosa EA, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat Microbiol*. 2019;4:293–305.
48. Chu H, Khosravi A, Kusumawardhani IP, Kwon AHK, Vasconcelos AC, Cunha LD, et al. Gene-microbiota interactions contribute to the pathogenesis of inflammatory bowel disease. *Science*. 2016;352(6289):1116–20. <https://doi.org/10.1126/science.aad9948>.
49. Schirmer M, Denson L, Vlamakis H, Franzosa EA, Thomas S, Gotman NM, et al. Compositional and temporal changes in the gut microbiome of pediatric ulcerative colitis patients are linked to disease course. *Cell Host Microbe*. 2018;24:600–610.e4.
50. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, et al. The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe*. 2014;15:382–92.
51. Pascal V, Pozuelo M, Borrueal N, Casellas F, Campos D, Santiago A, et al. A microbial signature for Crohn's disease. *Gut*. 2017;66:813–22.
52. Wolf AJ, Limon JJ, Nguyen C, Prince A, Castro A, Underhill DM. *Malassezia* spp. induce inflammatory cytokines and activate NLRP3 inflammasomes in phagocytes. *J Leukoc Biol*. 2021;109(1):161–72. <https://doi.org/10.1002/JLB.2MA0820-259R>.
53. Limon JJ, Tang J, Li D, Wolf AJ, Michelsen KS, Funari V, et al. *Malassezia* is associated with Crohn's disease and exacerbates colitis in mouse models. *Cell Host Microbe*. 2019;25:377.
54. Pang S, Chen X, Lu Z, Meng L, Huang Y, Yu X, et al. Longevity of centenarians is reflected by the gut microbiome with youth-associated signatures. *Nat Aging*. 2023;3:436–49.
55. Shanahan F, Ghosh TS, O'Toole PW. The healthy microbiome—What is the definition of a healthy gut microbiome? *Gastroenterology*. 2021;160:483–94.
56. Shanahan F, Ghosh TS, O'Toole PW. Human microbiome variance is underestimated. *Curr Opin Microbiol*. 2023;73: 102288.
57. Wang W, et al. Increased proportions of bifidobacterium and the lactobacillus group and loss of butyrate-producing bacteria in inflammatory bowel disease. *J Clin Microbiol*. 2020;52:398–406.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.