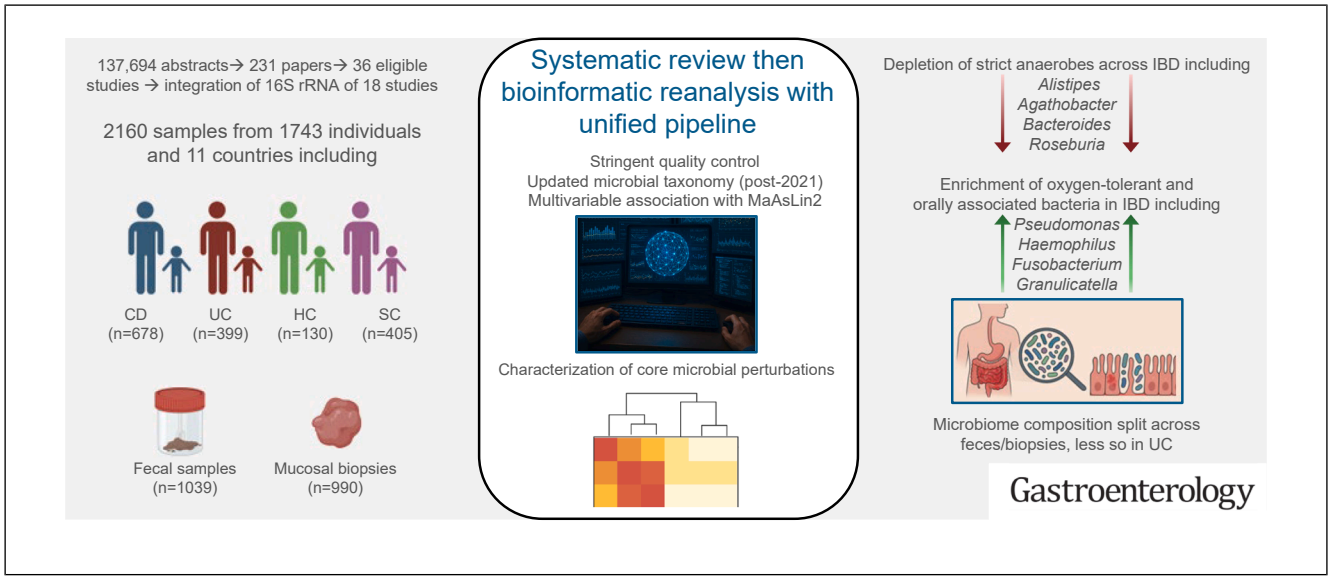
Gastroenterology 2026;170:539–556

The Gut Microbiome at the Onset of Inflammatory Bowel Disease: A Systematic Review and Unified Bioinformatic Synthesis

Peter Rimmer,[1,2,*] **Fan Zhang**,[3,*] Gregor Scott,[4] Microbiome Data Provision Group, **Georgina L. Hold**,[3,§] **Morris Gordon**,[5,§] **Tariq H. Iqbal**,[1,§] and **Richard Hansen**[6,§]

[1]Institute of Microbes, Infection and Microbiomes, University of Birmingham, Birmingham, United Kingdom; [2]Department of Gastroenterology, Queen Elizabeth Hospital Birmingham, Birmingham, United Kingdom; [3]Microbiome Research Centre, Faculty of Medicine and Health, University of New South Wales, Sydney, New South Wales, Australia; [4]National Health Service, Greater Glasgow and Clyde, United Kingdom; [5]School of Medicine, University of Central Lancashire, Lancashire, United Kingdom; and [6]School of Medicine, University of Dundee, Dundee, United Kingdom



BACKGROUND & AIMS: Few studies describe gut microbiome signatures in treatment-naïve new-onset inflammatory bowel disease (IBD). We present a novel secondary bioinformatic reanalysis of sequence outputs mapped to the latest microbial taxonomy. **METHODS:** MEDLINE and Embase searches were performed for microbiome studies in treatment-naïve IBD. Appraisal was completed with Risk Of Bias In Non-randomized Studies - of Exposures (ROBINS-E). Available 16S ribosomal RNA sequence data sets were downloaded and missing data sets requested. Integrated data were run through a unified QIIME2 bioinformatics pipeline. Multivariable models adjusting for methodologic differences were developed using MaAsLin2. **RESULTS:** There were 36 eligible studies; 18 contributed to bioinformatic reanalysis and 24 to supplementary meta-analysis. Samples from 1743 patients were included, comprising 678 from individuals with Crohn's disease (CD), 399 with ulcerative colitis (UC), 130 healthy controls (HCs), and 405 symptomatic controls (SCs); 990 of which were biopsy samples. Alpha diversity was reduced: feces-pediatric UC vs SCs, adult CD and UC vs HCs, and pediatric SCs vs HCs; pediatric biopsy samples-CD vs SCs, CD vs UC, and UC vs SCs. Beta diversity demonstrated clear distinctions between fecal and mucosal biopsy communities, least evident in UC, in addition to community separation by geography. Multivariate modeling revealed depletion of anaerobic and enrichment of aerobic and facultative anaerobic bacteria, alongside enrichment of oral genera across both CD and UC. **CONCLUSIONS:** Core microbial perturbations at onset of CD and UC are depletion of anaerobes and enrichment of oxygen-tolerant, orally associated bacteria. As we place greater emphasis on early diagnosis and prediction of IBD risk, this finding may support innovative diagnostic approaches. Microbiome-targeted intervention and alteration of luminal oxygen availability may offer novel therapeutic avenues for new-onset patients and identified high-risk groups.

Keywords: Microbiota; Crohn's Disease; Ulcerative Colitis; Treatment-Naïve.

Inflammatory bowel disease (IBD) comprises a group of heterogenous chronic immune-mediated inflammatory disorders of the gut. Crohn's disease (CD) and

***Authors share co-first authorship; §Authors share co-senior authorship.**

Abbreviations used in this paper: **adj, adjusted; CD, Crohn's disease; CI, confidence interval; HC, healthy control; IBD, inflammatory bowel disease; OF, observed feature; SC, symptomatic control; SCFA, short-chain fatty acid; SMD, standardized mean difference; UC, ulcerative colitis.**

Most current article

0016-5085
https://doi.org/10.1053/j.gastro.2025.09.014

ulcerative colitis (UC) are the most common subtypes. They often follow a relapsing-remitting course that can be significantly debilitating and require lifelong medical intervention, sometimes involving surgical resection. IBD incidence is rapidly increasing, particularly in developing counties and linked to "Westernization."[1–3] The global prevalence of IBD exceeds 0.3%, presenting an increasing health care burden. Hence the importance of ongoing research to understand pathogenetic mechanisms and develop novel therapies.[4]

The underlying etiology of IBD is not fully understood, but theories focus on the interplay between multiple contributors, including disruption of the gut microbiota, barrier dysfunction, genetic predisposition, and environmental factors.[5,6] Alterations in the gut microbiome have been widely reported in IBD, particularly reduced bacterial diversity.[7] The importance of bacteria in IBD pathogenesis has long been postulated, with early studies looking for specific causative pathogens.[8,9] The first study linking IBD with microbial community imbalance, using culture-based techniques, was published in 1968.[10] Although culture bias previously limited our understanding of complex microbial communities, the advent of molecular (DNA-based) methodologies has led to a wealth of compelling evidence in this regard.[11]

Microbiome perturbations in IBD are characterized by loss of microbial balance/harmony between commensal and pathogenic bacteria, resulting in a breakdown of homeostasis and dysregulated immune responses.[12] This paradigm describes a reduction in bacterial richness/evenness (assessed using alpha diversity indices), increased pathogenic bacterial numbers, and reduced beneficial bacterial species.[13,14] Although oxygen-dependent shifts in the gut microbial community are described in IBD, with depletion of obligate anaerobic bacteria and enrichment of facultative anaerobes and aerobes, establishing the *cause or consequence* significance of such changes is hampered by a paucity of studies of patients with newly diagnosed, treatment-naïve disease.[15] Patients who are posttreatment do not provide an adequate surrogate because IBD therapy and disease duration are important confounders.[16]

Published studies of treatment-naïve disease have shown reduced bacterial diversity, more so in patients with CD than UC.[13,14,17] However, reported patterns of disruption vary across studies, with inconsistent methodologies and quality control having an unquantified influence on results.[18] Studies in healthy individuals, and the largest study of patients with treatment-naïve CD, also noted significant differences between fecal and gut mucosal microbiome profiles.[14,19] Furthermore, most IBD studies using mucosal biopsy specimens are reliant on control populations with symptoms necessitating colonoscopy, rather than true "healthy" individuals. In the limited treatment-naïve data currently published, these factors prevent confident understanding of presenting microbial changes. *Faecalibacterium prausnitzii*, discussed extensively within the literature, illustrates this well. Eight studies reported its depletion in patients at IBD onset,[14,20–26] whereas 3 reported enrichment.[13,27,28]

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

Microbiome disruption has been reported at inflammatory bowel disease (IBD) onset, though studies show inconsistent patterns, and integrated analysis of the original sequence datasets has not been performed in treatment-naïve patients.

NEW FINDINGS

Methodological inconsistency rendered traditional meta-analysis unreliable, though through secondary bioinformatic reanalysis and modeling (2000 samples from 1700+ individuals) we demonstrated broad depletion of anaerobic bacteria across IBD subtypes. Conversely, a diverse pattern of enrichment is seen across aerobic, facultative anaerobic and microaerophilic bacteria, emphasizing genera associated with the oral cavity, in IBD.

LIMITATIONS

Some studies contributed disproportionately, with sequence data and high-quality metadata from others unobtainable. The scarcity of treatment-naïve metagenomic data in the literature meant analyses lacked the granularity to go beyond genus level or comment on microbial function.

CLINICAL RESEARCH RELEVANCE

While historically relevant, enrichment of oxygen-tolerant bacteria and depletion of anaerobes has not previously been demonstrated so starkly at disease onset and across multiple studies. Research targeting such perturbations at diagnosis might alter subsequent disease course and should be a priority. Further work is required to understand the processes driving the migration and apparent colonization of oral genera within the gut in IBD.

BASIC RESEARCH RELEVANCE

Our approach shifts the paradigm in assessing published microbiome datasets. We highlight the paucity of data from adult patients, and from parts of the world where IBD is increasing most rapidly. Future work must aim to establish an international repository of amalgamated and curated sequence to facilitate more rapid advances in our understanding of the role of gut microbiota in IBD.

The importance of the gut microbiota in patients who are newly diagnosed is underlined by its potential to aid in prediction of prognosis and therapeutic responses. For example, in patients with treatment-naïve UC, elevated levels of *Veillonella dispar* and *Haemophilus parainfluenzae* have been linked to colectomy within 12 months, whereas enrichment of *Roseburia* in CD has been linked to reduced disease severity and improved response to therapy.[28,29]

The erratic methodologic approaches and conflicting microbial signals within the literature highlight the potential limitations of conventional systematic review in the study of the microbiome. Previous attempts to undertake integrated analysis of published microbiome sequence data have been

made.[30] To date, comparisons between groups with heterogeneous treatment histories and the absence of multivariable adjustment in analyses prevent such works from truly identifying the core microbial perturbations at IBD onset. To bring us closer to this goal, we performed a novel secondary analysis of pooled treatment-naïve amplicon sequencing data using a unified bioinformatic pipeline.

Materials and Methods

Search Strategy

At conceptualization, this review was registered on PROSPERO (Registration ID: CRD42022371173, October 28, 2022). The study was performed in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[31] Searches (Supplementary Materials) were conducted on October 30, 2022, and updated December 23, 2024, across MEDLINE, Embase, and the Cochrane Gut register. Free-text and medical subject heading terms were both used. Searches were conducted from database inception to the date of the search. To reduce publication bias, ongoing relevant research studies were sought from controlled-trials.com, ClinicalTrials.gov, and online supplementary material. Abstract-only publications were eligible if sufficient information to judge inclusion was provided. Hand searching was undertaken of references in articles reviewed and relevant gray literature. Conference proceedings over the preceding 12 months were screened.

Study Selection and Eligibility

All experimental and nonexperimental study designs involving all ages were considered. Studies undertaking analysis of the gastrointestinal microbiome in patients with confirmed newly diagnosed, treatment-naïve IBD cohorts were eligible. The treatment-naïve state was defined by sampling before the initiation of conventional IBD therapy. Suitable sample types for microbial analysis included fecal samples, mucosal biopsy specimens, gastrointestinal washings, and oral samples. Both healthy asymptomatic control populations (HCs) and symptomatic "non-IBD" populations (SCs), whereby gastrointestinal symptoms were present but IBD had been excluded, were used as comparators. In cases where insufficient data were available to judge inclusion, primary authors were contacted for further information. Only those based on next-generation or high-throughput sequencing were included. Studies were screened independently in duplicate (by 2 of P.R., G.S., M.G., T.H.I., and R.H.), with disagreement resolved by discussion and consensus.

Outcome Assessment

Primary outcomes were to establish and quantify perturbations in the gastrointestinal microbiome at IBD diagnosis and across disease phenotypes. This included:

- Population-based changes in alpha and beta diversity, and
- Specific microbial taxa abundance differences between IBD subtypes and controls.

Secondary outcomes included:

- Quantification of methodologic variation across included studies and their impact on bioinformatic output.

Data Extraction

Data were extracted independently in duplicate by 2 researchers (P.R., G.S., T.H.I., and R.H.) (Supplementary Table 1). Given the uniform case-control nature of the included studies, appraisal for studies was not performed for the purposes of guiding inclusion decision-making. Instead, the methods data were extracted; especially surrounding bioinformatics, source of samples, and control type to guide analysis on methods basis.

Quality Assessment

The Risk Of Bias In Non-randomized Studies—of Exposure (ROBINS-E) tool examined the strength of evidence about presence of or nature of the potential effect of exposure (IBD) on outcomes.[32] A priori, key potential confounders were agreed, including age, sex, body mass index, baseline diet, and prior antibiotic use. Outcomes for signaling questions and overall risk of bias judgement were reported. Funnel plots reporting study effect size and standard error are presented as supplementary data.

Data Synthesis

Two forms of data synthesis were undertaken, specifically meta-analysis of published alpha diversity data and integrated bioinformatics analysis of available original amplicon sequencing output.

Meta-analysis of diversity data. Presented alpha diversity measures were extracted, and where sufficient, including measures of central tendency and spread, meta-analysis was undertaken. Different diversity measures were not combined in analysis. Where plots were presented without values, these were inferred using a web-plot digitizer.[33] We performed random-effects meta-analyses where $\geq$2 studies were available using the same diversity measure and patient groups and outcomes were sufficiently similar. We expressed the diversity measure as standardized mean difference (SMD) with 95% confidence intervals (CIs). Inconsistency was quantified and represented by the I^2 statistic according to Cochrane: <40% not important, 40% to 75% may represent heterogeneity, and >75% considerable heterogeneity.[34] Statistical analyses were performed using Cochrane Review Manager 5.4 (The Cochrane Collaboration).[35]

Integrated bioinformatics on pooled sequence data. Where available, amplicon sequencing data sets and metadata were downloaded. Unavailable data were requested from corresponding authors via email on 2 occasions a week apart. Once raw data and metadata were pooled, a bioinformatic pipeline was run on the combined data set. A quality control check was undertaken on amplicon data with dada2 embedded in qiime2.[36,37] Host contamination was removed using Bowtie 2 (version 2.4.2).[38] Only samples providing >10,000 clean reads were included. Taxonomy

annotation was performed using the qiime2 feature classifier plugin with the SILVA taxonomy release 138.2. R 4.2.2 (R Foundation for Statistical Computing) packages qiime2R and phyloseq were used for diversity analyses.[39,40] Mann-Whitney-Wilcoxon and Kruskal-Wallis were applied in the comparison of feature abundance and alpha diversity measures between groups. Shannon alpha diversity was calculated to align with the meta-analysis, whereas observed features (OFs) present a simpler measure of richness.

Adonis was used, based on Bray-Curtis distance, to investigate the effect of metadata factors on microbial composition.[41] MaAsLin2 analyses were performed to determine the multivariable association between microbial signatures and clinical data using the R package MaAsLin2 1.20.0.[42] Specifically, "fixed-effects" were sample type, pediatric vs adult, and diagnosis subtype (analysis repeated with HCs and SCs as reference). Age-group was chosen rather than actual age due to inconsistent patient-level data. "Random-effects" were sample geography (continent), target 16S domain, and subject. *P* values were adjusted for multiple testing where appropriate by Benjamini-Hochberg method, represented by P_{adj}.[43] The plots were constructed mainly in ggplot2 package.[44] Hierarchical clustering followed Ward's method (Ward D2).[45]

Results

Search Screening and Inclusion

As presented in the PRISMA flow diagram (Figure 1), the search strategy identified 15,256 studies, with 9 located from other sources.[46] After 1,471 duplicates were excluded, 13,794 abstracts were screened, with 0.4% (59 of 13,794) requiring consensus discussion. Thereafter, 233 studies (224 from the search and 9 other) were sought for full-text screening. Of these, 2 could not be retrieved, with another 195 excluded (Supplementary Table 2). Altogether, 36 studies were included. ROBINS-E judgements were completed (Supplementary Table 3). Studies were largely at low risk, although 4 had some concerns of bias, and 3 were at high risk due to poor control of confounders, for example, comparing adults and children directly. There was a paucity of metagenomic data, with only 3 studies containing data of this type. Consequently, presented data are based on amplicon sequencing only.

Meta-analysis of Published Shannon Alpha Diversity at Inflammatory Bowel Disease Onset

Bacterial diversity indices from IBD and controls were reported in 26 studies. IBD was stratified by subtype and controls by the presence of symptoms; 9 studies presented data from CD, 4 UC, and 6 from both CD and UC. All studies using mucosal biopsy samples compared against SCs. For studies of fecal samples, 14 studies (93%) compared with HCs, although 2 used healthy family members as controls (Supplementary Table 4). Shannon alpha diversity scores were available in 23 studies (88%), and this was chosen for the meta-analysis.[13,20,22,24,25,27,28,47–62] Analyses were split by sample type (Supplementary Figure 1).

Analysis of fecal samples demonstrated significant reductions in diversity in IBD (SMD, −0.59; 95% CI, −0.94 to −0.24), CD (SMD, −0.94; 95% CI −1.22 to −0.66), and UC (SMD, −0.70; 95% CI, −1.06 to −0.33) relative to controls. This gave a pooled SMD of −0.84 (95% CI, −1.12 to −0.56, $P < .00001$) with acceptable levels of heterogeneity observed ($I^2 = 23.9\%$). Regarding mucosal data from biopsy samples, neither CD (SMD, −0.16; 95% CI, −0.45 to 0.12; $P = .25$, $I^2 = 27\%$) nor UC (SMD, −0.00; 95% CI −0.30 to 0.30; $P = .99$, $I^2 = 0\%$) demonstrated a significant difference compared with SCs.

Considering the 26 included studies, methodologic variation was marked: 2 sample types (feces and biopsy specimens) from 4 continents, DNA extracted using 12 different kits, sequencing performed on 4 platforms targeting 8 16S ribosomal RNA hypervariable regions, and analyses reported using 4 reference databases. This alone generated 12,288 potential methodologic combinations before counting the 29 different analytic pipelines used. Inconsistent and incomplete reporting prevented adequate stratification of data, for example, by disease extent or severity (Supplementary Figure 1). Funnel plots highlighted greater asymmetry and significant outliers within fecal samples (Supplementary Figure 2).

Secondary Bioinformatic Analysis of Pooled Sequencing Data

Raw sequencing data and associated metadata were retrieved for 25 studies. After data processing and cleanup, 7 were excluded (in 3 all samples failed clean read cutoff, 2 failed denoising, 1 was metagenomic data, and 1 author shared an unanalyzable file format (Table 1).[13,14,20,23–25,27–29,48,49,51,53,56,57,63–72] The final presented analysis of 18 studies included 2160 samples (881 CD, 509 UC, 1 IBD unclassified, 122 IBD "not-specified," 130 HCs, 509 SCs, and 8 familial controls), originating from 1743 unique individuals (678 CD, 399 UC, 1 IBD-unclassified, 122 IBD "not-specified," 130 HCs, 405 SCs, 8 familial controls). There were 168 participants who provided both biopsy specimens and feces, and 249 provided biopsy samples from >1 site. For samples from patients with CD, 99 (11%) were from those with ileal disease, 47 (5%) colonic, 126 (14%) ileocolonic, 4 (<1%) upper gastrointestinal, and in 605 (69%), the subphenotype was not reported. For samples from patients with UC, 369 (73%) had extensive disease, 71 (14%) left-sided, 48 (9%) proctosigmoiditis, and in 21 (4%) it was not reported. Biopsy location was reported for 979 samples (99%), but inflammation status was not (Supplementary Table 5). All analyses were undertaken using the latest microbial taxonomy.[73]

Overall Comparison of Sample Types

Foremost, sample type was studied. There were no data from HCs from mucosal biopsy samples. In the remaining patients, mucosal biopsy samples were characterized by enrichment of Bacteroidota and Pseudomonadota, alongside

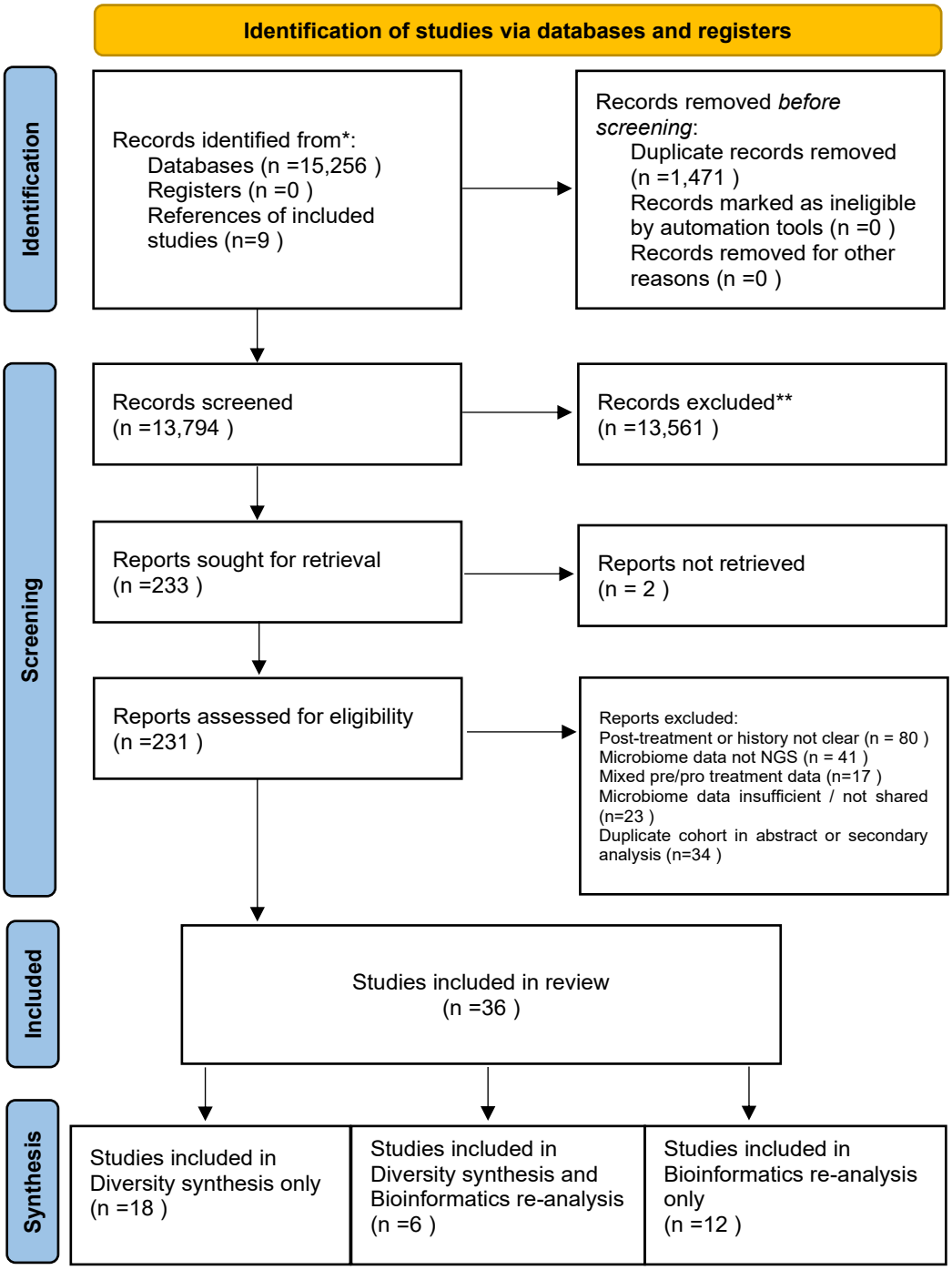


Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 flow diagram for new systematic reviews that included searches of databases and registers only. NGS, next-generation sequencing.

depletion of Actinomycetota and Bacillota compared with feces (Figure 2).[14,20,23,25,27–29,49,53,57,64,65–69,71] There was a significant reduction in alpha diversity in mucosal biopsy samples relative to feces, with a significant difference in Bray-Curtis community structure. Although this was observed across disease subtypes, community structure between feces and biopsy samples was most closely matched in UC. Because most published studies came from pediatric cohorts, which have distinct features compared with adult disease, we undertook an analysis of adult vs pediatric studies.[74,75] In analyses grouping all sample types and diagnoses, a significant reduction in Shannon alpha diversity was observed in children relative to adults (pediatric, mean 2.18; adult, mean 2.50; effect size, 0.21; $P_{adj} < .001$).

There was also significant separation in beta diversity ($R^2 = 0.0402$; $P_{adj} < .001$). Plots of beta diversity, split by chosen 16S sequence domain showed significant differences

Table 1. Studies Contributing Raw 16S rRNA Sequence Data to the Pooled Bioinformatic Pipeline, Including Proportion Passing Quality Control

Author	Sample type	Extraction kit	Platform	Domain	Depth (mean unless stated)	Country	Age group	Age (overall in study)		Number of patients with samples + metadata	Number of samples passing bioinformatic QC	Reason for loss of sample output	Overall risk of bias (assessed with ROBINS-E tool)[32]
								Mean (SD)	Median (IQR)		n (%)		
Kaakoush et al 2012[64]	Feces	Bioline ISOLATE fecal DNA kit	Roche 454	V1-3	2609	Australia	Pediatric	CD 11.6 (2.5) HC 9.5 (4.2)		CD 20 HC 22	0	All samples raw reads <10,000	Low
Hansen et al 2012[13]	Mucosal biopsy	Qiagen QIAamp mini kit	Roche 454	V3-4	21691	Scotland	Pediatric		CD 14.2 (ND) UC 13 (ND) SC 11.4 (ND)	CD 11 UC 11 SC 12	0	Failed in the denoise step	Low
Gevers et al[a] 2014[14]	Mucosal biopsy and feces	Qiagen AllPrep mini kit	Illumina MiSeq	V4	29914	USA	Pediatric	CD 12.4 (3.0) SC 12 (3.7)		Feces CD 166 SC 17 Biopsy CD 347 SC 217	851 (46.89)[a]	Metagenomic samples excluded[c]	Low
Perez-Brocal et al 2015[20]	Mucosal biopsy and feces	Qiagen QIAMP DNA stool mini kit	Roche 454	V1-3	10043	Spain	Adult	CD 45.4 (18.1) SC 33.5 (14.4)		Feces CD 14 HC 24 Biopsy CD 14	2 (3.85)	Majority <10,000 clean reads	Some concern
Assa et al 2016[27]	Mucosal biopsy	FastDNA Spin Kit	Illumina HiSeq	V6	292215	Canada	Pediatric		CD 14 (4.75) SC 14 (2)	CD 11 SC 15	26 (100)		Low
Grover et al[a,b] 2016[65]	Mucosal biopsy	ND	Illumina MiSeq	V6-8	ND	Australia	Pediatric	ND	ND	CD 23 SC 6	44 (51.76)[a]		Some concern
Mottawea et al 2016[48]	Mucosal luminal interface aspirate	FastDNA Spin Kit	Illumina HiSeq Roche 454	V6 V6	200000 13313	Canada	Pediatric		CD 14 (3.25) UC 15 (4.5) SC 14 (6)	Illumina CD 65 UC 23 SC 43 Roche CD 8 UC 8 SC 9	131 (83.44)	Mucosal-luminal aspirates excluded from analyses stratified by sample type as contribution only from one study	Low
Shah et al 2016[28]	Mucosal biopsy	Qiagen "buffers"	Illumina MiSeq	V4-6	2350 rarefied	USA	Pediatric	UC 12.9 (3.7) SC 13.9 (1.8)		UC 9 SC 13	5 (22.73)	Other samples <10,000 reads	Low
Shaw et al 2016[49]	Feces	Not stated	Illumina MiSeq	V4	66000 median	USA	Pediatric	ND	ND	CD 11 UC 3 HC 4 FC 6	23 (95.83)		High
Ashton et al 2017[66]	Feces	MP biomedical feces extraction kit	Illumina MiSeq	V4	ND	England	Pediatric	CD 13.6 (2.1) UC 10.1 IBDU 12.6 FC ND		CD 3 UC 1 IBD-U 1 FC 3	8 (100)		Some concern

Table 1. Continued

Author	Sample type	Extraction kit	Platform	Domain	Depth (mean unless stated)	Country	Age group	Age (overall in study) Mean (SD)	Age (overall in study) Median (IQR)	Number of patients with samples + metadata	Number of samples passing bioinformatic QC n (%)	Reason for loss of sample output	Overall risk of bias (assessed with ROBINS-E tool)[32]
Douglas et al[a] 2018[67]	Feces	ND	Illumina MiSeq	V6-8	13815	Scotland	Pediatric	CD 12.7 (2.4) SC 12.8 (2.4)		CD 20 SC 20	20 (17.39)	Some <10,000 reads Metagenomic data excluded[c]	Low
Schirmer et al[a] 2018[29]	Mucosal biopsy + feces	Bx Qiagen AllPrep mini kit Feces Chemagic DNA blood kit	Illumina MiSeq	V4	3000 rarefied	USA	Pediatric	UC 12.8 (3.3)		Biopsy UC 211 Feces UC 264	395 (83.16)		Low
Xu et al 2018[68]	Mucosal biopsy	Qiagen QIAamp stool mini kit	Illumina HiSeq	V3-4	ND	China	Adult	UC 48 (14)		UC 10	2 (10)[a]	Majority <10,000 clean reads	Low
Kansal et al 2019[51]	Mucosal biopsy	Qiagen AllPrep mini kit	Illumina MiSeq	V2	9188	Australia	Pediatric		CD 12 (ND) SC 12.3 (ND)	CD 88 SC 66	0	File format incorrect	Low
Levine et al 2019[69]	Feces	Mobio PowerFecal DNA kit	ND	V4-5	ND	Israel Canada	Pediatric	CD 14.1 (2.6)		CD 59	57 (96.61)		Low
Lloyd-Price et al[a] 2019[23]	Mucosal biopsy + feces	Qiagen AllPrep mini kit	Illumina MiSeq	V4	10000 rarefied	USA	Mixed	CD 20.2 (11.3) UC 24.7 (15.3) SC ND		Biopsy CD 56 UC 23 SC 22 Feces CD 14 UC 7 SC 3	130 (68.4)[a]		Low
Diederen et al 2020[63]	Feces	FastDNA spin kit	Illumina MiSeq	V1-2	4237	Netherlands	Pediatric		CD 14 (3) HC 13 (5)	CD 27 HC 15	22 (52.38)		Low
Hart et al 2020[70]	Feces	Mixed methods	Illumina MiSeq	V3	103341	Canada	Pediatric	CD 11.9 (3.2) UC 13.4 (2.0)		CD 19 UC 8	0	After filtering features with low abundance or observed in very few samples, zero remaining features	Low
Wang X et al 2021[25]	Feces	EZNA soil DNA kit	Illumina MiSeq	V3-4	ND	China	Pediatric	"IBD" 10 (5.3) HC 7.1 (3.8)		"IBD" 80 SC 48 HC 27	142 (91.61)	No metadata to identify IBD subtype; excluded from subgroup analyses	Low
Wang Y et al 2021[24]	Feces	Qiagen QIAamp stool mini kit	Illumina NovaSeq	V3-4	186189	China	Pediatric		CD 13 (3) HC 12 (2)	CD 23 HC 20	0	All samples have clean reads <10,000	Low
Galipeau et al 2021[53]	Feces	Qiagen QIAamp stool mini kit	Illumina MiSeq	V4	12510	Canada	Adult	UC 19.7 (7.2) FC 20.3 (7.3)		UC 7	7 (100)		Low

Table 1. Continued

Author	Sample type	Extraction kit	Platform	Domain	Depth (mean unless stated)	Country	Age group	Age (overall in study)		Number of patients with samples + metadata	Number of samples passing bioinformatic QC	Reason for loss of sample output	Overall risk of bias (assessed with ROBINS-E tool)[32]
								Mean (SD)	Median (IQR)		n (%)		
Paljetak et al 2022[56]	Mucosal biopsy	MasterPure DNA purification kit	Illumina MiSeq	V3-4	7916 median	Croatia	Adult		CD 46 (ND) UC 31 (ND) SC 31 (ND)	Biopsy CD 10 UC 13 SC 26 Feces CD 10 UC 12 SC 26	0	Failed in the denoise step	Low
Rimmer et al[b] 2022[57]	Feces	Qiagen QIAamp stool mini kit	Illumina MiSeq	V4	45079	England	Adult	CD 37.3 (16.3) UC 40.5 (14.8) SC 38 (12.4)		CD 53 UC 41 SC 52	145 (99.32)		Low
Rausch et al 2023[71]	Feces	Qiagen QIAamp stool mini kit	Illumina MiSeq	V1-2	11800	Malta	Adult	CD 37.8 (16.6) UC 47.4 (16.6) HC 44.7 (16)		"IBD" 56 HC 96	150 (98.68)	No metadata to identify IBD subtype; excluded from subgroup analyses	Low
Ning et al 2023[72]	Feces	HiPure stool DNA mini kit	Illumina NovaSeq	NA	ND	China	Adult	ND	ND	"IBD" 87 HC 45	NA	Metagenomic samples excluded[c]	Low
Total samples included in final dataset (% from pediatric patients)	All samples All IBD 1513 (87%)[c] CD 881 (91%) UC 509 (86%) IBD-U 1 (100%) All controls 647 (71%) HC 130 (27%) SC 509 (82%) FC 8 (100%)							Feces All IBD 770 (79%)[c] CD 367 (85%) UC 280 (83%) IBD-U 1 (100%) All controls 269 (45%) HC 130 (27%) SC 131 (60%) FC 8 (100%)			Mucosal biopsies All IBD 655 (93%) CD 449 (95%) UC 206 (89%) All controls 335 (88%) All control samples from symptomatic controls It was not possible to determine the age of patient for 39 SC		
Number of variations in methods	Location: 11 countries of origin Sample type: 3							Extraction kit: 11 Sequencing type: 1			Sequencing platform: 4 Sequence domain (if amplicon): 10		

CD, Crohn's disease; FC, familial control; HC, healthy control; IBD, inflammatory bowel disease patients where metadata did not allow stratification to disease subtype; IBD-U, inflammatory bowel disease unclassified; IQR, interquartile range; ND, data not presented; QC, quality control; SC, symptomatic control; SD, standard deviation; UC, ulcerative colitis.

[a]Some individuals provided >1 sample (eg, fecal and biopsy).

[b]Data published in abstract only but expanded and shared by the authors.

[c]Analysis not undertaken for metagenomic data due to a paucity of raw sequences and/or metadata.

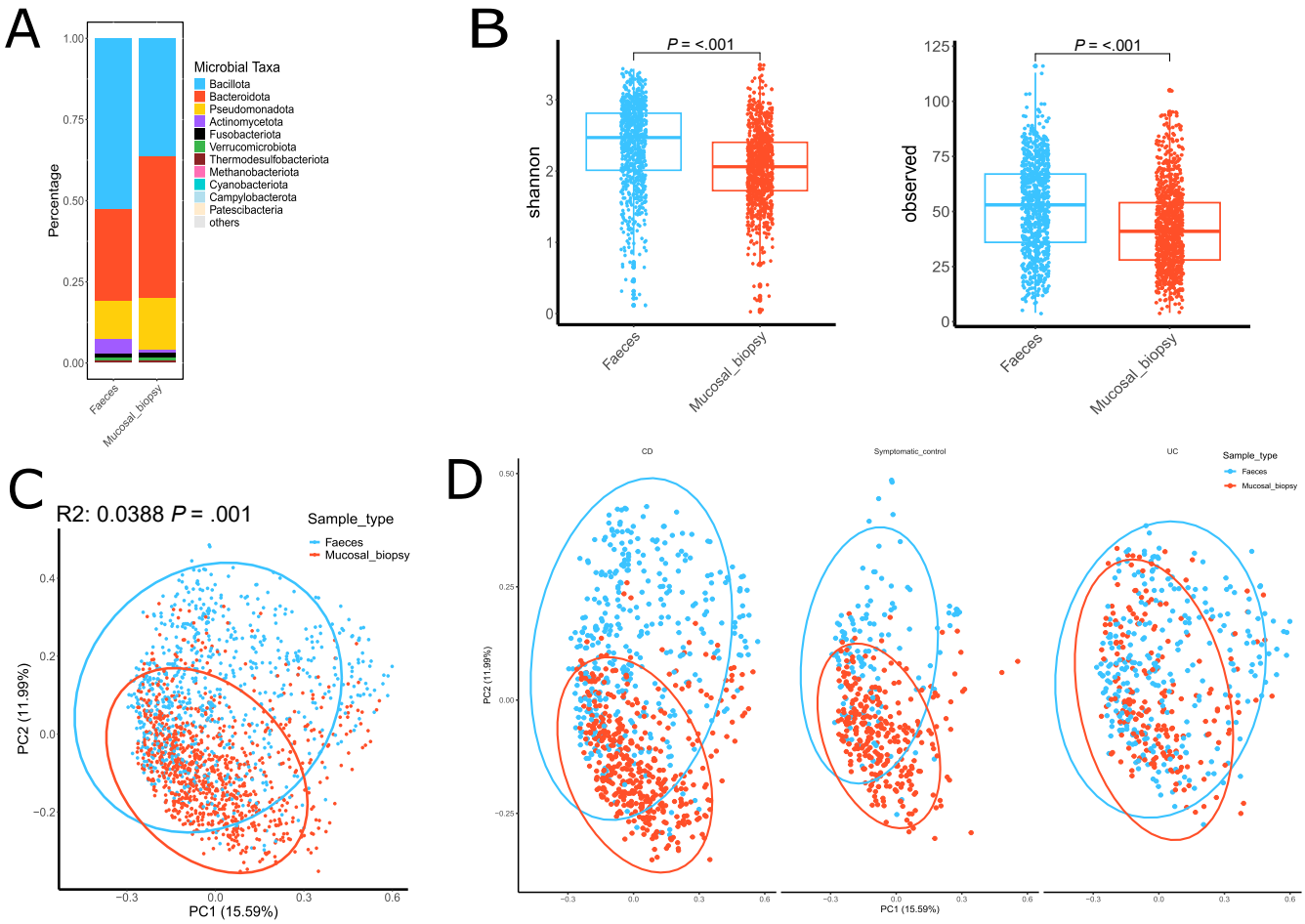


Figure 2. The difference between the microbial community identified in mucosal biopsy samples and feces.[14,20,23,25,27–29,49,53,57,63,65–69,71] All panels exclude HCs (no mucosal biopsy specimens). Panels *A–C* are derived from 901 feces samples (367 CD, 280 UC, 122 "IBD," and 131 SCs) and 990 mucosal biopsy specimens (CD, n = 449; UC, n = 206; SC, n = 335). Panel *D* includes only those where stratification is possible. The original pipeline data output are presented separately (Supplementary data set). (*A*) Microbial taxa bar plots at the phylum level. Mucosal biopsy specimens are characterized by a lower percentage of Actinomycetota and Bacillota, with a higher proportion of Bacteroidota (feces, 28%; biopsy specimen, 44%; effect size (EF), 0.33; $P_{adj} < .001$) and Pseudomonadota than in feces. (*B*) Increased alpha diversity is shown in fecal communities compared with biopsy samples. The *boxes* indicate the 25th percentile (*bottom border*), median (*center line*), and 75th percentile (*top border*), and the *vertical lines* show the maximum and minimum ranges excluding outliers. (*C*) An overall Bray-Curtis beta diversity principal coordinates analysis plot shows clear separation between feces and biopsy samples. (*D*) Bray-Curtis beta diversity PC analysis plots split by diagnostic subtype show a significant separation according to sample type across diagnoses, although this difference is smallest in UC. PC1, first principal component; PC2, second principal component.

(Supplementary Figure 3), although the relevance of this diminished given the different composition of cohorts across included studies.

Influence of Geography

Geographic origin has been shown to impact microbiome data.[76] Differences in community structure were observed across continents, particularly in fecal sample data. Comparisons were challenging due to an uneven spread of adult and pediatric patients among areas, with most samples originating from North America (Figure 3).[14,20,23,27–29,49,53,57,63,65–69]

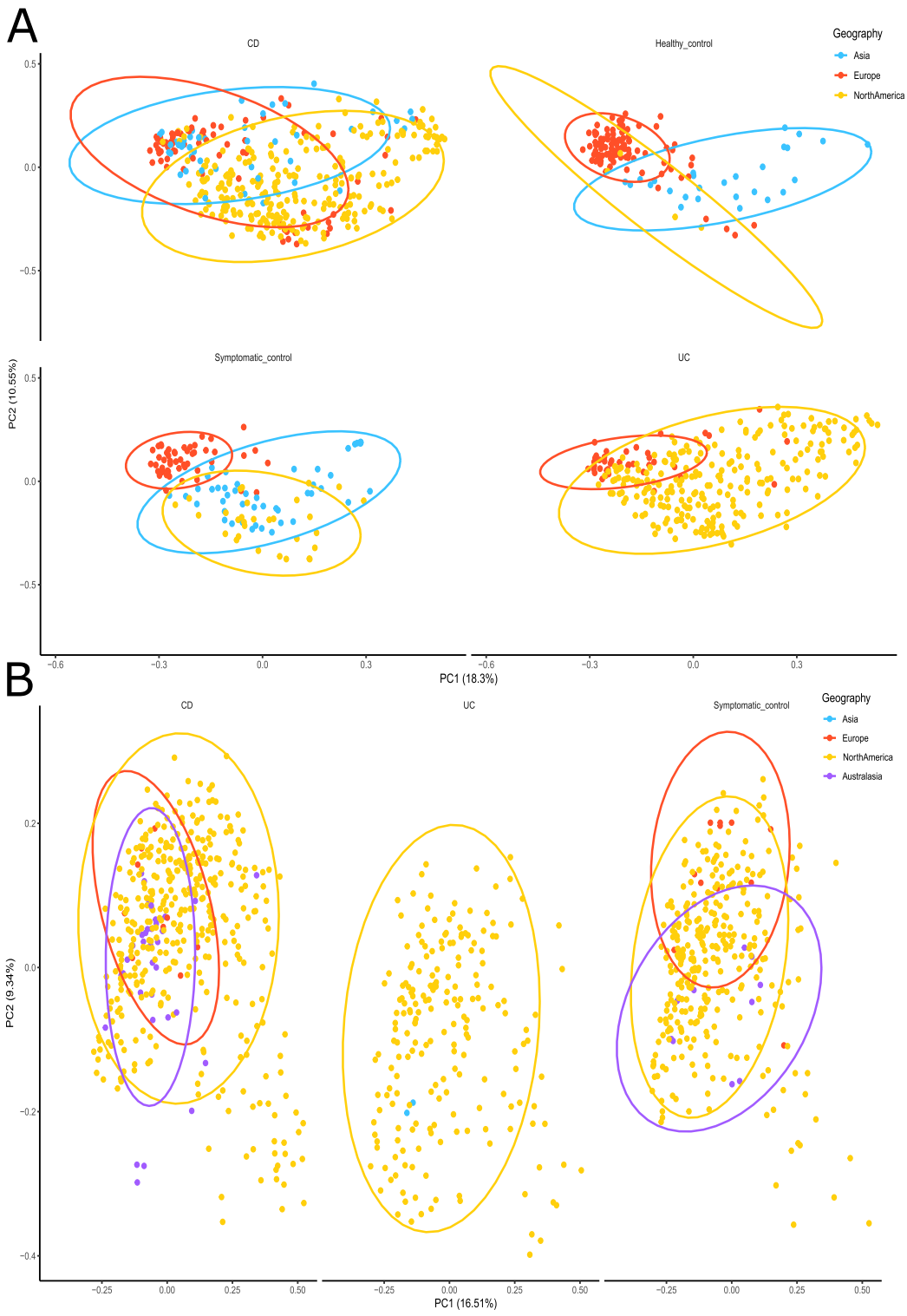
Alpha Diversity

In pediatric fecal samples, significantly lower Shannon diversity was observed in UC relative to SCs but not in CD (Figure 4).[14,23,29,49,53,57,63,66,67,69] For OF analysis, diversity was significantly lower for both UC and CD compared with SCs. In adults, significantly lower alpha diversity was seen CD, UC, and SCs relative to HCs (Figure 4). Differences between both CD and UC compared with SCs were not significant by Shannon but CD diversity by OF was significantly less than SCs. In pediatric patients, a distinct pattern was seen in controls, with significantly increased diversity by OF in SCs relative to HCs.

For mucosal biopsy specimens, most analyses were based on pediatric samples, with no adult SC data available. Alpha diversity was significantly lower in pediatric CD compared with SCs and UC and significantly higher in UC vs SCs. In the smaller adult cohort, no significant difference was observed between CD and UC (Figure 5).[14,20,23,27–29,65,68]

Microbial Community Structure

Beta diversity assessment of fecal samples in pediatric CD and UC was significantly different from SCs and HCs. A similar pattern was observed in adults, although differences between CD and SCs did not reach significance. For mucosal samples, significant differences were observed



with SCs in both pediatric CD and UC (Supplementary Figure 4).

Differential Microbial Abundance

Differential abundance across sample types and diagnoses, adjusting for methodologic variations, were interrogated using MaAsLin2. All differentially abundant genera with a false discovery rate adjusted P value $<.05$ were considered in detail, and data regarding morphology, metabolism, and short-chain fatty acid (SCFA) production were documented (Supplementary Tables 6 and 7). Across IBD subgroups, depleted bacteria were obligate anaerobes, except for *Sutterella* in both CD and UC against HCs. Depletion of *Alistipes*, *Roseburia*, and *Phascolarctobacterium* were observed in both CD and UC compared with all controls (Figure 6).[14,20,23,27–29,48,49,53,57,63,65–69] Although these, and many other depleted genera, are known SCFA producers, many of the bacteria enriched in IBD also produce SCFAs. Enriched bacterial genera included aerobic *(Pseudomonas* and *Schaalia*), microaerophilic *(Campylobacter* and *Dialister*), and facultative anaerobic, including *Haemophilus*, *Enterococcus*, and *Rothia*. This pattern was seen across comparisons with both HCs and SCs. Enrichment of multiple genera found in the oral cavity was also seen, including *Fusobacterium*, *Peptostreptococcus*, *Haemophilus*, *Veillonella*, and *Granulicatella*.

Discussion

This study sought to improve understanding of the role the gut microbiota plays in newly diagnosed IBD by applying a unified bioinformatics analysis approach to existing published data sets. Leveraging the collective power of existing data sets is fundamental if we are to fully understand disease pathogenesis, identify new microbial therapeutic treatment avenues, and develop prognostic tools. This study is the first of its kind in the microbiome field to focus on treatment-naïve IBD. It brings together a vast sequence data set that has been fastidiously compiled, rigorously analyzed, and updated to the latest taxonomy to describe the core microbial perturbations at IBD onset.

Significant methodologic variation was identified across studies, whereas other important factors, such as sample acquisition/storage were poorly reported.[77] Some studies did not present standard diversity indices and used novel approaches without describing the underpinning parameters.[14,23] Contrasts in published data sets were also, in part, driven by using different control groups for different sample types. This rendered attempts to generate traditional cross-study comparisons, such as meta-analysis, largely futile; therefore, stepping back to the raw data stage was necessary to effectively combine data sets. By using a unified bioinformatics approach with consistent quality control standards, we were able to quantify, control, and account for impacts from methodologic factors.

Our approach to assessing gut microbiome study data has reinforced some established views but refuted others. There were clear differences between feces and mucosal biopsy samples, with reduced microbial diversity in biopsy samples compared with feces. These findings are compatible with previous work in healthy individuals.[19,78] However microbial community structure is more closely aligned between feces and biopsy samples in UC than in CD. This is likely partly attributable to the more distal inflammation in UC and perhaps greater mucosal shedding. Confirmation of this would require consistent availability of disease extent and severity reporting in metadata. Awareness of this similarity in UC may facilitate use of feces as a closer proxy to the mucosal microbiota in future studies.

Lower microbial diversity in children compared with adults has been previously described.[74] In our pediatric data, fecal Shannon diversity was reduced in UC compared with SCs but not HCs, with no significant reduction seen in CD. In adults, reductions in fecal alpha diversity were observed in comparisons of both UC and CD with HCs but in neither when compared with SCs. In mucosal biopsy specimens, alpha diversity indices were reduced in pediatric CD but increased in UC compared with SCs. These divergent patterns are likely multifactorial. As above, the microbial communities of feces compared with biopsy samples are more closely matched in UC. Furthermore, CD mucosal samples can be obtained directly from the site of inflammation rather than using a fecal sample as a distal surrogate. Also, given the closer association with distal colonic disease, in UC, mucosal samples may be less impacted by dietary and environmental factors, including transit time.[79]

Recent studies suggest the oral microbiome contributes to IBD.[80] Oral-gut transmission is considered to occur

Figure 3. The difference between microbial community structure, stratified by diagnosis and geography.[14,20,23,27–29,49,53,57,63,65–69] The original pipeline data output is presented separately (Supplementary data set). (*A*) A Bray-Curtis principal coordinates analysis plot splits fecal samples by diagnosis and continent of origin. This includes 367 patients with CD (Europe: 55 adult and 21 pediatric, Asia: 44 pediatric, North America: 247 pediatric), 280 with UC (Europe: 40 adult and 1 pediatric, North America: 232 pediatric and 7 adult), 131 SCs (Europe: 52 adult, Asia: 48 pediatric, North America: 31 pediatric), and 130 HCs (Europe: 95 adult and 4 pediatric, Asia: 27 pediatric, North America: 4 pediatric). Variations in age group compositions render meaningful comparisons challenging. Where comparable, separation sits just outside of significance when corrected for multiple testing. For example, patients with CD from Asia and North America (100% pediatric) (R^2 = 0.0398, P_{adj} = .055) and SCs from Asia and North America (100% pediatric) (R^2 = 0.0706; P_{adj} = .055). PC1, first principal component; PC2, second principal component. (*B*) An equivalent principal coordinates analysis plot for mucosal biopsy samples includes 449 patients with CD (Europe: 90 pediatric, Australasia: 34 pediatric, North America: 385 pediatric and 21 adult), 206 patients with UC (North America: 184 pediatric and 20 adult, Asia: 2 adult), and 335 SCs (Europe: 11 pediatric Australasia: 10 pediatric, North America: 275 pediatric, and 39 where age could not be determined). Comparable populations are seen in CD, where statistically significant separation is observed in CD from Australasia and both Europe and North America.

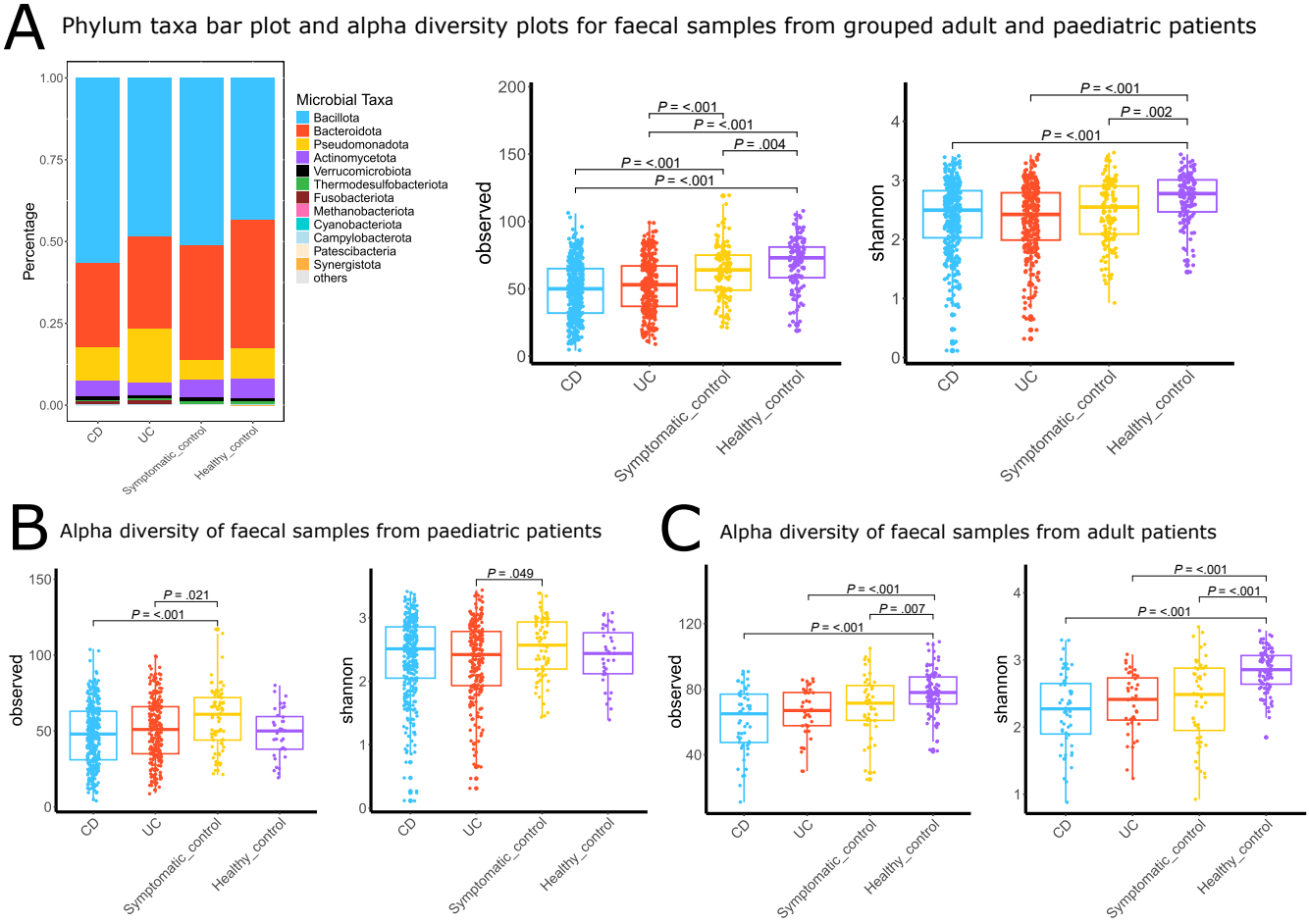


Figure 4. Fecal alpha diversity of pooled patients with treatment-naïve IBD patients and controls.[14,23,29,49,53,57,63,66,67,69] The original pipeline data output is presented separately (Supplementary data set). (*A*) A phylum-level taxa bar plot and alpha diversity plots are shown for the Shannon index and OF. All are derived from fecal samples from both adult and pediatric patients (CD, n = 367; UC, n = 280; SCs, n = 131; HCs, n = 130). The bar plot demonstrates prominent expansion Fusobacteriota and Pseudomonadota in UC over SCs and HCs. In CD, enrichment of Fusobacteriota is again observed compared with HCs and SCs. Bacteroidota are depleted in CD vs all comparators. For the Shannon index, significant reductions in alpha diversity are seen in CD and UC vs HCs but neither vs SCs. The *boxes* indicate the 25th percentile (*bottom border*), median (*center line*), and 75th percentile (*top border*), and the *vertical lines* show the maximum and minimum ranges excluding outliers. (*B*) Alpha diversity from pediatric patients (CD, n = 312; UC, n = 233; SCs, n = 79; HCs, n = 35) is presented. The Shannon index is significantly reduced in UC vs SCs. Significant reductions in OF are observed between CD and UC compared with SCs. HCs in pediatric patients and adults have differing patterns of diversity. (*C*) A nonsignificant reduction is observed in HC children relative to SCs, whereas in adults, HCs have significantly increased alpha diversity vs SC. Alpha diversity plots are shown for adult patients (CD, n = 55; UC, n = 47; SCs, n = 52; HCs, n = 95). SCs now have significantly reduced alpha diversity relative to HCs. For patients with CD and UC, Shannon alpha diversity is reduced relative to HCs. In the OF plot, reductions comparing CD with SCs do not stand after false discovery rate correction (P_{adj} = .10).

regularly, potentially markedly increasing during disease.[81,82] We confirmed consistent increases in bacteria associated with the oral cavity in the gut of patients with CD and UC. Additionally, we have shown this across samples of diverse geographic origin and age. Although the oral cavity may serve as a reservoir for pathobionts, further work is required to understand the processes driving the migration and apparent colonization of these genera within the gut.

Within the multivariate model, the contrast between almost universal depletion of anaerobic bacteria with mixed enrichment, including aerobic, facultative anaerobic, and microaerophilic bacteria across IBD, was striking. The "oxygen hypothesis" has long been a point of discussion, but has arguably fallen from favor.[15] The relevance of this hypothesis has not previously been demonstrated so starkly in new-onset disease. Understanding the role of oxygen (and other altered luminal ecological factors) in the microbial etiopathogenesis of IBD is fundamental to our understanding of disease biology. Crucially, altering luminal oxygen availability may offer a novel therapeutic strategy of relevance to both new-onset patients and relatives considered at high risk of subsequent IBD development. As we enter an era of greater prediction of IBD risk, the development of novel non-immunosuppressant approaches aimed at avoiding progression to disease attains increasing importance.[83,84]

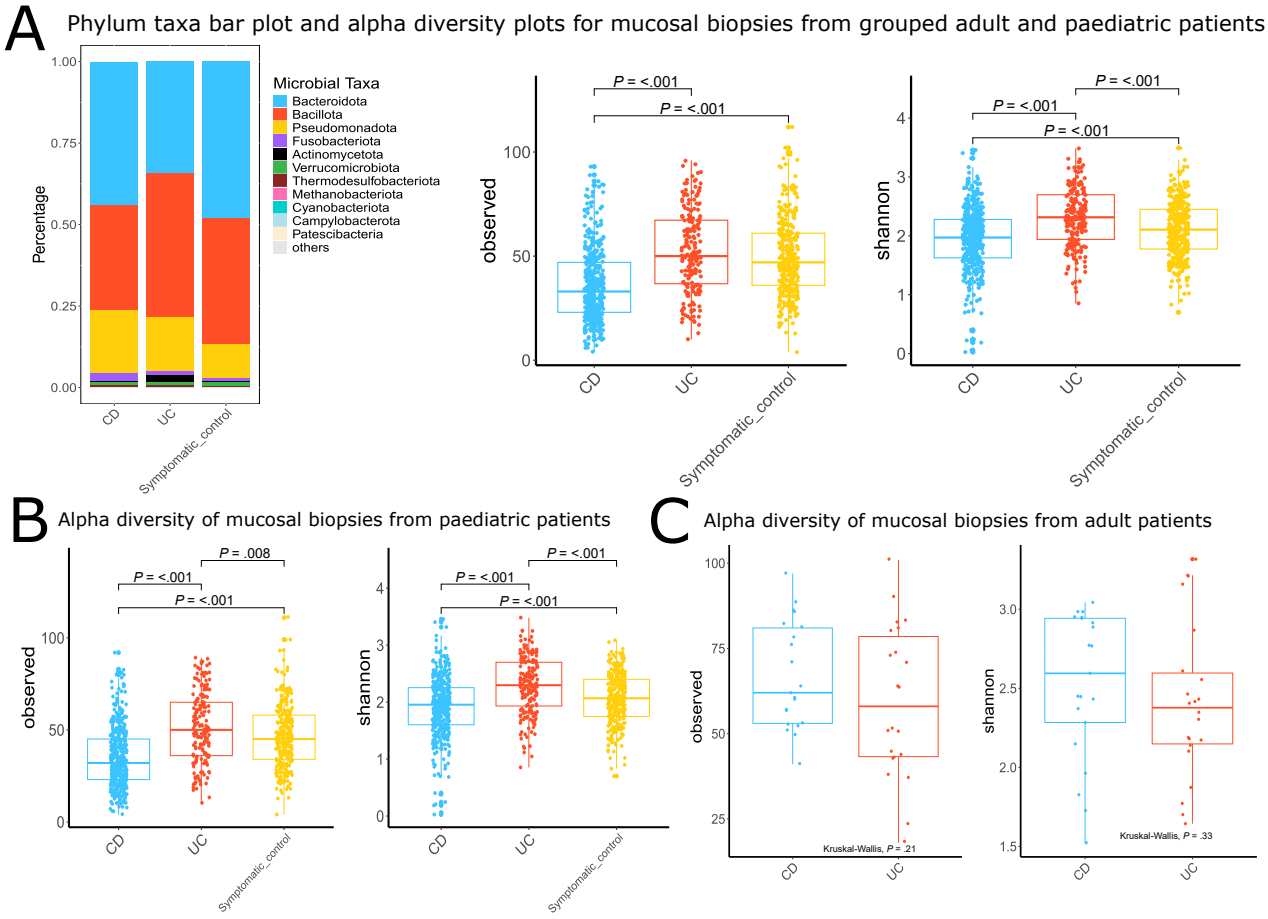
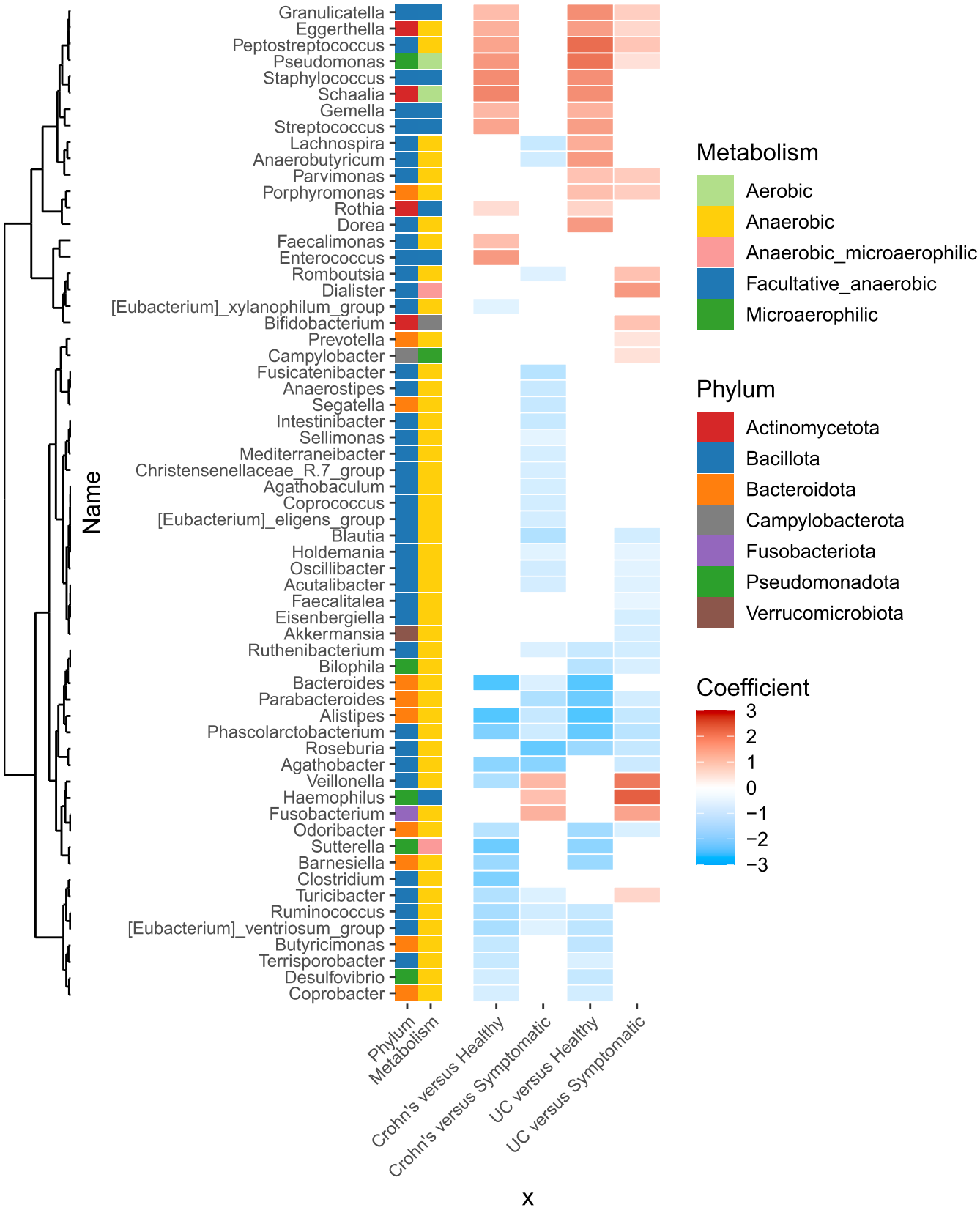


Figure 5. Mucosal alpha diversity of pooled patients with treatment-naïve IBD patients and controls.[14,20,23,27–29,65,68] The original pipeline data output is presented separately (Supplementary data set). (*A*) A taxa bar plot at phylum level and alpha diversity plots for the Shannon index and OF are shown. All are derived from gut mucosal biopsy specimens, with samples from both adult and pediatric patients (CD, n = 449; UC, n = 206; SCs, n = 335). The taxa bar plot demonstrates expansion of Fusobacteriota in CD vs SCs, but not UC. Enrichment of Pseudomonadota and depletion of Bacteroidota is observed in CD and UC vs SCs, but differential abundance is discordant for Bacillota and Actinomycetota (both depleted in CD and enriched in UC vs SCs). Alpha diversity is reduced in CD. For the Shannon index, this is the case vs both UC and SCs. Shannon index in UC is significantly higher than in SCs. (*B*) Data are shown for pediatric patients (CD, n = 428; UC, n = 184; SCs, n = 296). SCs are lost from Lloyd-Price et al[23] because the metadata did not allow the stratification of controls by age-group. CD had significantly lower Shannon diversity than UC or SC. Diversity in UC was significantly higher than SC. The *boxes* indicate the 25th percentile (*bottom border*), median (*center line*), and 75th percentile (*top border*), and the *vertical lines* show the maximum and minimum ranges excluding outliers. (*C*) Data are shown for adult patients. There is a paucity of mucosal biopsy specimen data with no controls. There is no significant difference in alpha diversity between UC and CD.

To address in methodologic inconsistency, generalized linear and mixed models were applied by treating batch effects, geographic origin (continent), and sequencing regions as covariates in models to statistically control for technical variability. The random effects in the model help account for heterogeneity across studies, therefore reducing bias and improving comparability. Disproportionately large contributions from some studies may influence conclusions. For example, 1 paper contributed 341 mucosal biopsy samples and 223 fecal samples from CD.[14] This represents 76% of biopsy samples and 61% of fecal samples for CD. A paucity of data from adults, particularly biopsy samples, limited targeted analyses, and enriching their availability should be another priority for the research community. Despite a move toward data availability and transparency, sequence data and high-quality metadata proved unobtainable in some cases. The scarcity of treatment-naïve metagenomic data in the literature meant our work could not reliably determine abundance beyond genus level, and data regarding microbial function or nonbacterial microbial domains were not available. The absence of metadata reporting inflammation status in mucosal specimens prevented reliable comparison of microbial composition between sites. Metadata for disease severity and treatment outcomes was inconsistently reported, precluding integrated analysis.

Greater unification of methodology and reporting approaches used in microbiome research is urgently needed. Reporting guidelines may exist, but future work should focus on defining best practice and aligning this with what is consistently deliverable in the microbiome field.[85] In the interim, establishing an international repository of


Granulicatella
Eggerthella
Peptostreptococcus
Pseudomonas
Staphylococcus
Schaalia
Gemella
Streptococcus
Lachnospira
Anaerobutyricum
Parvimonas
Porphyromonas
Rothia
Dorea
Faecalimonas
Enterococcus
Romboutsia
Dialister
[Eubacterium]_xylanophilum_group
Bifidobacterium
Prevotella
Campylobacter
Fusicatenibacter
Anaerostipes
Segatella
Intestinibacter
Sellimonas
Mediterraneibacter
Christensenellaceae_R.7_group
Agathobaculum
Coprococcus
[Eubacterium]_eligens_group
Blautia
Holdemania
Oscillibacter
Acutalibacter
Faecalitalea
Eisenbergiella
Akkermansia
Ruthenibacterium
Bilophila
Bacteroides
Parabacteroides
Alistipes
Phascolarctobacterium
Roseburia
Agathobacter
Veillonella
Haemophilus
Fusobacterium
Odoribacter
Sutterella
Barnesiella
Clostridium
Turicibacter
Ruminococcus
[Eubacterium]_ventriosum_group
Butyricimonas
Terrisporobacter
Desulfovibrio
Coprobacter
Name
Phylum
Metabolism
Crohn's versus Healthy
Crohn's versus Symptomatic
UC versus Healthy
UC versus Symptomatic
x
Metabolism
Aerobic
Anaerobic
Anaerobic_microaerophilic
Facultative_anaerobic
Microaerophilic
Phylum
Actinomycetota
Bacillota
Bacteroidota
Campylobacterota
Fusobacteriota
Pseudomonadota
Verrucomicrobiota
Coefficient
3
2
1
0
−1
−2
−3

amalgamated and curated sequence data sets with minimum requirements for metadata should be developed in a format that is usable and analyzable.

Conclusion

Through fastidious attention to detail and the deployment of multivariable modeling to correct for methodologic inconsistency, we have identified the core microbial perturbations at IBD onset. The depletion of anaerobes and enrichment of oxygen-tolerant bacteria, alongside enrichment of oral bacteria, may reveal novel diagnostic and therapeutic avenues for patients with new-onset disease or those in identified high-risk groups.

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.2025.09.014.

References

1. Henderson P, Dobson L. DOP50 PINPOINT: The epidemiology of pediatric-onset Inflammatory Bowel Disease in the United Kingdom—a prospective, national, cohort study. J Crohns Colitis 2024;18:i162.
2. Alatab S, Sepanlou SG, Ikuta K, et al. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol Hepatol 2020;5:17–30.
3. Ng SC, Shi HY, Hamidi N, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet 2017;390:2769–2778.
4. **Hracs L**, **Windsor JW**, Gorospe J, et al. Global evolution of inflammatory bowel disease across epidemiologic stages. Nature 2025;642:458–466.
5. Zhang YZ, Li YY. Inflammatory bowel disease: Pathogenesis. World J Gastroenterol 2014;20:91.
6. Kim DH, Cheon JH. Pathogenesis of inflammatory bowel disease and recent advances in biologic therapies. Immune Netw 2017;17:25–40.
7. Marchesi JR, Adams DH, Fava F, et al. The gut microbiota and host health: a new clinical frontier. Gut 2015;65:330–339.
8. Hawkins HP. An address on the natural history of ulcerative colitis and its bearing on treatment. BMJ 1909;1:765–770.
9. Seneca H, Henderson E. Normal intestinal bacteria in ulcerative colitis. Gastroenterology 1950;15:34–39.
10. Gorbach SL, Nahas L, Plaut AG, et al. Studies of intestinal microflora. V. Fecal microbial ecology in ulcerative colitis and regional enteritis: relationship to severity of disease and chemotherapy. Gastroenterology 1968;54:575–587.
11. Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005;437:376–380.
12. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. Nature 2011;474:307–317.
13. Hansen R, Russell RK, Reiff C, et al. Microbiota of De-Novo pediatric IBD: increased *Faecalibacterium prausnitzii* and reduced bacterial diversity in Crohn's but not in ulcerative colitis. Am J Gastroenterol 2012;107:1913–1922.
14. Gevers D, Kugathasan S, Denson LA, et al. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe 2014;15:382–392.
15. Rigottier-Gois L. Dysbiosis in inflammatory bowel diseases: the oxygen hypothesis. ISME J 2013;7:1256–1261.
16. Mah C, Jayawardana T, Leong G, et al. Assessing the relationship between the gut microbiota and inflammatory bowel disease therapeutics: a systematic review. Pathogens 2023;12:262.
17. Conte MP, Schippa S, Zamboni I, et al. Gut-associated bacterial microbiota in pediatric patients with inflammatory bowel disease. Gut 2006;55:1760–1767.
18. Knight R, Vrbanac A, Taylor BC, et al. Best practices for analyzing microbiomes. Nat Rev Microbiol 2018;16:410–422.
19. **Vaga S**, **Lee S**, Ji B, et al. Compositional and functional differences of the mucosal microbiota along the intestine of healthy individuals. Sci Rep 2020;10:14977.
20. Pérez-Brocal V, García-López R, Nos P, et al. Metagenomic analysis of Crohn's disease patients identifies changes in the virome and microbiome related to disease status and therapy, and detects potential

Figure 6. Differential microbial abundance modeled across IBD subtypes, sample types, and age-groups in newly diagnosed treatment-naïve IBD[14,20,23,27–29,48,49,53,57,63,65–69] Unified MaAsLin2 output from 2029 samples (CD, n = 881; UC, n = 509; HC, n = 130; SC, n = 509) considering sample type, age-group, and diagnosis subtype as fixed effects and sample geography, target 16S domain, and subject as random effects. Coefficients for all presented genera were differentially abundant at a false discovery rate $P_{adj} < .05$ (Supplementary Tables 6 and 7). Enriched bacteria in CD vs HCs included aerobic and facultative anaerobic bacteria. Depletion was observed across multiple anaerobes. Comparing CD with SCs, a facultative anaerobe, *Haemophilus*, was enriched alongside other oral bacteria *Fusobacterium* and *Veillonella*. Similar depletion of anaerobes was observed. Bacteria enriched in UC vs HCs were similar to CD, with the aerobic bacteria again significantly enriched alongside facultative anaerobes. Again, enrichment of genera typical of the mouth was noteworthy. Broad depletion of anaerobes was again noted. In comparisons between UC and SCs, aerobic and microaerophilic bacteria were again enriched, with the addition of *Campylobacter*. Increases in abundance of phenotypically oral bacteria and depletion across multiple anaerobes were repeated.

interactions and biomarkers. Inflamm Bowel Dis 2015;21:2515–2532.
21. Rojas-Feria M, Romero-García T, Caballero-Rico JÁF, et al. Modulation of fecal metagenome in Crohn's disease: Role of microRNAs as biomarkers. World J Gastroenterol 2018;24:5223–5233.
22. Kowalska-Duplaga K, Gosiewski T, Kapusta P, et al. Differences in the intestinal microbiome of healthy children and patients with newly diagnosed Crohn's disease. Sci Rep 2019;9:18880.
23. Lloyd-Price J, Arze C, Ananthakrishnan AN, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature 2019;569:655–662.
24. Wang Y, Gao X, Zhang X, et al. Microbial and metabolic features associated with outcome of infliximab therapy in pediatric Crohn's disease. Gut Microbes 2021;13:1–18.
25. **Wang X**, **Xiao Y**, Xu X, et al. Characteristics of fecal microbiota and machine learning strategy for fecal invasive biomarkers in pediatric inflammatory bowel disease. Front Cell Infect Microbiol 2021;11:711884.
26. Basha OM, Hafez RA, Salem SM, et al. Impact of gut microbiome alteration in ulcerative colitis patients on disease severity and outcome. Clin Exp Med 2022;23:1763–1772.
27. Assa A, Butcher J, Li J, et al. Mucosa-associated ileal microbiota in new-onset pediatric Crohn's disease. Inflamm Bowel Dis 2016;22:1533–1539.
28. Shah R, Cope JL, Nagy-Szakal D, et al. Composition and function of the pediatric colonic mucosal microbiome in untreated patients with ulcerative colitis. Gut Microbes 2016;7:384–396.
29. Schirmer M, Denson L, Vlamakis H, 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.
30. **Abbas-Egbariya H**, **Haberman Y**, Braun T, et al. Meta-analysis defines predominant shared microbial responses in various diseases and a specific inflammatory bowel disease signal. Genome Biol 2022;23:61.
31. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. PLoS Med 2021;18:e1003583.
32. Higgins JPT, Morgan RL, Rooney AA, et al. A tool to assess risk of bias in non-randomized follow-up studies of exposure effects (ROBINS-E). Environ Int 2024;186:108602.
33. Rohatgi A. Web Plot Digitizer, Version 4.7, 2021. Available at: https://automeris.io/. Accessed June 23, 2025.
34. Deeks JJ, Higgins JP, Altman DG. Analysing data and undertaking meta-analyses. Cochrane Handbook for Systematic Reviews of Interventions 2021;241–284.
35. The Cochrane Collaboration. Review Manager (RevMan). Version 5, 2020. Available at: https://revman.cochrane.org. Accessed June 23, 2025.
36. Callahan BJ, McMurdie PJ, Rosen MJ, et al. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 2016;13:581–583.
37. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 2019;37:852–857.
38. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357–359.
39. Bisanz JE. qiime2R: Importing QIIME2 artifacts and associated data into R sessions, 2018. Available at: https://github.com/jbisanz/qiime2R. Accessed June 23, 2025.
40. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 2013;8:e61217.
41. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecol 2001;26:32–46.
42. Mallick H, Rahnavard A, McIver LJ, et al. Multivariable association discovery in population-scale meta-omics studies. PLoS Comput Biol 2021;17:e1009442.
43. Benjamini Y, Drai D, Elmer G, et al. Controlling the false discovery rate in behavior genetics research. Behav Brain Res 2001;125:279–284.
44. Patil I. Visualizations with statistical details: The 'ggstatsplot' approach. J Open Source Softw 2021;6:3167.
45. Ward JH. Hierarchical grouping to optimize an objective function. J Am Stat Assoc 1963;58:236–244.
46. Haddaway NR, Page MJ, Pritchard CC, et al. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis. Campbell Syst Rev 2022;18:e1230.
47. Kellermayer R, Mir SAV, Nagy-Szakal D, et al. Microbiota separation and C-reactive protein elevation in treatment-naïve pediatric granulomatous Crohn disease. J Pediatr Gastroenterol Nutr 2012;55:243–250.
48. Mottawea W, Chiang CK, Mühlbauer M, et al. Altered intestinal microbiota-host mitochondria crosstalk in new onset Crohn's disease. Nat Commun 2016;23:13419.
49. Shaw KA, Bertha M, Hofmekler T, et al. Dysbiosis, inflammation, and response to treatment: a longitudinal study of pediatric subjects with newly diagnosed inflammatory bowel disease. Genome Med 2016;8:75.
50. Fossum Moen AE, Lindstrøm JC, Tannæs TM, et al. The prevalence and transcriptional activity of the mucosal microbiota of ulcerative colitis patients. Sci Rep 2018;8:17278.
51. Kansal S, Catto-Smith AG, Boniface K, et al. Variation of gut mucosal microbiome with anti-*Saccharomyces cerevisiae* antibody status in pediatric Crohn disease. J Pediatr Gastroenterol Nutr 2019;69:696–703.
52. Tang W, Huang Y, Shi P, et al. Effect of exclusive enteral nutrition on the disease process, nutrition status, and gastrointestinal microbiota for Chinese children with Crohn's Disease. J Parenter Enteral Nutr 2020;45:826–838.
53. **Galipeau HJ**, **Caminero A**, **Turpin W**, et al. Novel fecal biomarkers that precede clinical diagnosis of ulcerative colitis. Gastroenterology 2021;160:1532–1545.
54. Juyal G, Sood A, Midha V, et al. Correlation between fecal microbial taxa and ulcerative colitis in different phases of disease activity in a north Indian cohort. medRxiv 2021;12(12):21267614.

55. El Mouzan MI, Winter HS, Assiri AA, et al. Microbiota profile in new-onset pediatric Crohn's disease: data from a non-Western population. Gut Pathog 2018;10:49.
56. Paljetak HČ, Barešić A, Panek M, et al. Gut microbiota in mucosa and feces of newly diagnosed, treatment-naïve adult inflammatory bowel disease and irritable bowel syndrome patients. Gut Microb 2022;14:2083419.
57. Rimmer P, Horniblow R, Cheesbrough J, et al. P59 The pre-treatment gut microbiome at IBD diagnosis: early insights into the Birmingham inception cohort study. Gut 2022;71:A67–A68.
58. Lv Y, Lou Y, Liu A, et al. The impact of exclusive enteral nutrition on the gut microbiome and bile acid metabolism in pediatric Crohn's disease. Clin Nutr 2023;42:116–128.
59. **Dovrolis N**, **Moschoviti A**, Fessatou S, et al. Identifying microbiome dynamics in pediatric IBD: more than a family matter. Biomed 2023;11:1979.
60. Mouzan ME, Mofarreh MA, Alsaleem B, et al. Bacterial dysbiosis in newly diagnosed treatment naïve pediatric ulcerative colitis in Saudi Arabia. Saudi J Gastroenterol 2025;31:14–21.
61. Orejudo M, Gomez M, de Francisco R, et al. P010 Faecal microbiota composition by shotgun metagenomic sequencing approach in a newly diagnosed cohort of inflammatory bowel disease patients: results from the IBDomics project. J Crohns Colitis 2024;18:i255–i256.
62. **Wang H**, **Wang Y**, Yang L, et al. Integrated 16S rRNA sequencing and metagenomics insights into microbial dysbiosis and distinct virulence factors in inflammatory bowel disease. Front Microbiol 2024;15:1375804.
63. Diederen K, Li JV, Donachie GE, et al. Exclusive enteral nutrition mediates gut microbial and metabolic changes that are associated with remission in children with Crohn's disease. Sci Rep 2020;10:18879.
64. Kaakoush NO, Day AS, Huinao KD, et al. Microbial dysbiosis in pediatric patients with Crohn's disease. J Clin Microbiol 2012;50:3258–3266.
65. Grover Z, Kang A, Morrison M, et al. 633 The relative abundances of *Dorea* and *Faecalibacterium* spp. in mucosa associated microbiome of newly diagnosed children with Crohn's disease are differentially affected by exclusive enteral nutrition. Gastroenterology 2016; 150:S132–S133.
66. Ashton JJ, Colquhoun CM, Cleary DW, et al. 16S sequencing and functional analysis of the fecal microbiome during treatment of newly diagnosed pediatric inflammatory bowel disease. Medicine 2017;96: e7347.
67. Douglas GM, Hansen R, Jones CMA, et al. Multi-omics differentially classify disease state and treatment outcome in pediatric Crohn's disease. Microbiome 2018;6:13.
68. Xu J, Ning C, Yang S, et al. Alteration of fungal microbiota after 5-ASA treatment in UC patients. Inflamm Bowel Dis 2019;26:380–390.
69. **Levine A**, **Wine E**, Assa A, et al. Crohn's disease exclusion diet plus partial enteral nutrition induces sustained remission in a randomized controlled trial. Gastroenterology 2019;157:440–450.e8.
70. Hart L, Farbod Y, Szamosi JC, et al. Effect of exclusive enteral nutrition and corticosteroid induction therapy on the gut microbiota of pediatric patients with inflammatory bowel disease. Nutrients 2020;12:1691.
71. **Rausch P**, **Ellul S**, Pisani A, et al. Microbial dynamics in newly diagnosed and treatment naïve IBD patients in the Mediterranean. Inflamm Bowel Dis 2023; 29:1118–1132.
72. **Ning L**, **Zhou YL**, **Sun H**, et al. Microbiome and metabolome features in inflammatory bowel disease via multi-omics integration analyses across cohorts. Nat Commun 2023;14:7135.
73. Oren A, Garrity GM. Valid publication of the names of forty-two phyla of prokaryotes. Int J Syst Evol Microbiol 2021;71:005056.
74. Radjabzadeh D, Boer CG, Beth SA, et al. Diversity, compositional and functional differences between gut microbiota of children and adults. Sci Rep 2020; 10:1040.
75. Van Limbergen J, Russell RK, Drummond HE, et al. Definition of phenotypic characteristics of childhood-onset inflammatory bowel disease. Gastroenterology 2008;135:1114–1122.
76. Gaulke CA, Sharpton TJ. The influence of ethnicity and geography on human gut microbiome composition. Nat Med 2018;24:1495–1496.
77. Jones J, Reinke SN, Ali A, et al. Fecal sample collection methods and time of day impact microbiome composition and short chain fatty acid concentrations. Sci Rep 2021;11:13964.
78. Eckburg PB, Bik EM, Bernstein CN, et al. Diversity of the human intestinal microbial flora. Science 2005; 308:1635–1638.
79. Procházková N, Falony G, Dragsted LO, et al. Advancing human gut microbiota research by considering gut transit time. Gut 2023;72:180–191.
80. **Wang A**, **Zhai Z**, Ding Y, et al. The oral-gut microbiome axis in inflammatory bowel disease: from inside to insight. Front Immunol 2024;15:1430001.
81. **Schmidt TS**, **Hayward MR**, Coelho LP, et al. Extensive transmission of microbes along the gastrointestinal tract. eLife 2019;8:e42693.12.
82. Kageyama S, Sakata S, Ma J, et al. High-resolution detection of translocation of oral bacteria to the gut. J Dent Res 2023;102:752–758.
83. Xue M, Leibovitzh H, Jingcheng S, et al. Environmental factors associated with risk of Crohn's disease development in the Crohn's and Colitis Canada - Genetic, Environmental, Microbial Project. Clin Gastroenterol Hepatol 2024;22:1889–1897.e12.
84. Lee SH, Turpin W, Espin-Garcia O, et al. Development and validation of an integrative risk score for future risk of Crohn's disease in healthy first-degree relatives: a multicenter prospective cohort study. Gastroenterology 2025;168:150–153.e4.
85. Mirzayi C, Renson A, Furlanello C, et al. Reporting guidelines for human microbiome research: the STORMS checklist. Nat Med 2021;27:1885–1892.

Author names in bold designate shared co-first authorship.

Received June 23, 2025. Accepted September 3, 2025.

Correspondence
Address correspondence to: Peter Rimmer, BMBS, MRCP, PhD, Institute of Microbes, Infection and Microbiomes, The University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom. e-mail: p.rimmer@bham.ac.uk.

Acknowledgments
The Microbiome Data Provision Group includes Nadeem Kaakoush,[1] Peter Lewindon,[2] James Ashton,[3] Shivani Kansal,[4] Johan Van Limbergen,[5] Rotem Sigall-Boneh,[6,7] Nikhil Pai,[8,9,10] Ken Croitoru,[11] Ting Zhang,[12] Hana Cipčić Paljetak,[13] and Mohammed Nabil Quraishi[14]; from the [1]School of Biomedical Science, University of Sydney, Sydney, New South Wales, Australia; [2]School of Medicine, University of Queensland, Brisbane, Queensland, Australia; [3]School of Medicine, University of Southampton, Southampton, United Kingdom; [4]Western Children's Health Centre, Laverton, Victoria, Australia; [5]Centrum Unverstieit van Amsterdam (AMC), Amsterdam, the Netherlands; [6]Paediatric Gastroenterology and Nutrition Unit, Pediatric Inflammatory Bowel Disease (PIBD) Research Centre, Edith Wolfson Medical Centre, Holon, Israel; [7]Tytgat Institute for Liver and Intestinal Research-Amsterdam Gastroenterology Endocrinology and Metabolism, University of Amsterdam, Amsterdam, the Netherlands; [8]School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; [9]Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; [10]McMaster University, Hamilton, Ontario, Canada; [11]Mount Sinai Hospital, Toronto, Ontario, Canada; [12]Shanghai Children's Hospital, Shanghai, China, [13]School of Medicine, University of Zagreb, Croatia; and [14]Sheikh Shakhbout Medical City, Abu Dhabi.

CRediT Authorship Contributions

Peter Rimmer, BMBS, MRCP, PhD (Conceptualization: Equal; Data curation: Lead; Formal analysis: Equal; Investigation: Lead; Methodology: Equal; Project administration: Lead; Visualization: Equal; Writing – original draft: Lead; Writing – review & editing: Lead)

Fan Zhang, PhD (Data curation: Equal; Formal analysis: Lead; Investigation: Equal; Methodology: Equal; Software: Lead; Visualization: Lead; Writing – review & editing: Supporting)

Gregor Scott, MB ChB (Data curation: Supporting; Investigation: Supporting; Writing – original draft: Supporting)

Georgina L. Hold, BSc, PhD (Formal analysis: Equal; Methodology: Equal; Resources: Equal; Supervision: Equal; Visualization: Supporting; Writing – original draft: Supporting; Writing – review & editing: Equal)

Morris Gordon, MB ChB, MRCPCH, PhD, MMed (Data curation: Equal; Formal analysis: Supporting; Methodology: Equal; Project administration: Equal; Supervision: Equal; Visualization: Supporting; Writing – original draft: Equal; Writing – review & editing: Equal)

Tariq H. Iqbal, MBBCHIR, FRCP, MD (Data curation: Supporting; Funding acquisition: Lead; Investigation: Supporting; Methodology: Supporting; Resources: Supporting; Supervision: Equal; Writing – original draft: Equal; Writing – review & editing: Supporting)

Richard Hansen, MB ChB, PhD, FRCPCH, FRCP Edin, AGAF (Conceptualization: Equal; Data curation: Equal; Formal analysis: Supporting; Investigation: Equal; Methodology: Equal; Project administration: Supporting; Supervision: Lead; Visualization: Supporting; Writing – original draft: Supporting; Writing – review & editing: Lead)

Conflicts of interest

These authors disclose the following: Peter Rimmer and Tariq H. Iqbal have received research grants from F. Hoffman La Roche. Peter Rimmer has received honoraria from AbbVie, Janssen Cilag, Ferring, and Takeda. Georgina L. Hold has received grant funding from Ferring and Pfizer and honoraria from AbbVie and Ferring. Tariq H. Iqbal has received honoraria from Pharmacosmos. The remaining authors disclose no conflicts.

Funding
Peter Rimmer and Tariq H. Iqbal are supported by funding from the National Institute for Health Research Birmingham Biomedical Research Centre (BRC) renewal 2023. The laboratory of Richard Hansen is supported by The Archie Foundation, and Richard Hansen has research time supported by Tayside Medical Science Centre.

Data Availability

Additional methodology and metadata are available upon reasonable request. References to repository locations and accession numbers for all publicly available sequence data sets are within the supplementary materials. Data references are included as supplementary data (Supplementary Table 8). Additionally shared data sets at present are available from their original publishing authors upon reasonable request, although establishing an accessible repository of these data sets is a future aim.