



# Metagenomic Characterization of Microbial Communities In Situ Within the Deeper Layers of the Ileum in Crohn's Disease

## Citation

Pedamallu, C. S., A. S. Bhatt, S. Bullman, S. Fowler, S. S. Freeman, J. Durand, J. Jung, et al. 2016. "Metagenomic Characterization of Microbial Communities In Situ Within the Deeper Layers of the Ileum in Crohn's Disease." *Cellular and Molecular Gastroenterology and Hepatology* 2 (5): 563-566.e5. doi:10.1016/j.jcmgh.2016.05.011. <http://dx.doi.org/10.1016/j.jcmgh.2016.05.011>.

## Published Version

doi:10.1016/j.jcmgh.2016.05.011

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:31731821>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available. Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

## BRIEF REPORT

Metagenomic Characterization of Microbial Communities *In Situ* Within the Deeper Layers of the Ileum in Crohn's Disease

Chandra Sekhar Pedamallu,<sup>1,2,3,\*</sup> Ami S. Bhatt,<sup>1,2,\*</sup> Susan Bullman,<sup>1,2</sup> Sharyle Fowler,<sup>4</sup> Samuel S. Freeman,<sup>2</sup> Jacqueline Durand,<sup>4</sup> Joonil Jung,<sup>2</sup> Fujiko Duke,<sup>1,2</sup> Veronica Manzo,<sup>2</sup> Diana Cai,<sup>2</sup> Ashwin Ananthakrishnan,<sup>4</sup> Akinyemi I. Ojesina,<sup>1,2</sup> Aruna Ramachandran,<sup>1,2</sup> Dirk Gevers,<sup>2</sup> Ramnik J. Xavier,<sup>4</sup> Atul K. Bhan,<sup>4,5</sup> Matthew Meyerson,<sup>1,2,3</sup> and Vijay Jaynik<sup>4</sup>

<sup>1</sup>Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts; <sup>2</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts; <sup>3</sup>Harvard Medical School, Boston, Massachusetts; <sup>4</sup>Crohn's and Colitis Center, <sup>5</sup>Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts

## SUMMARY

Our study shows that the *in situ* microbial community of the diseased bowel in patients with Crohn's disease is distinct from inflammatory bowel disease-free individuals. In Crohn's disease patients, the microbial composition at the phyla level did not differ markedly between healthy and diseased areas, but at the species level an enrichment of potentially pathogenic organisms was observed in the diseased ileum.

extraction methods from paraffin sections and host nucleic acid depletion approaches to increase microbial read coverage. (*Cell Mol Gastroenterol Hepatol* 2016;2:563-566; <http://dx.doi.org/10.1016/j.jcmgh.2016.05.011>)

**Keywords:** Crohn's Disease; Deeper Mucosal Layers; Microbial Dysbiosis.

**BACKGROUND & AIMS:** Microbial dysbiosis and aberrant host-microbe interactions in the gut are believed to contribute to the development and progression of Crohn's disease (CD). Microbiome studies in CD typically have focused on microbiota in feces or superficial mucosal layers of the colon because accessing DNA from deeper layers of the bowel is challenging. In this study, we analyzed the deep tissue microbiome in patients who underwent surgical resection of the small intestine.

**METHODS:** Paraffin blocks were obtained from 12 CD patients undergoing ileocecal resection, and healthy ileum samples (inflammatory bowel disease-free controls) were obtained from 12 patients undergoing surgery for right-sided colon cancer. Diseased and healthy-appearing ileum was identified using microscopy, and paraffin blocks were macrodissected using a core needle to specifically isolate DNA. Illumina Whole Genome Sequencing was used for microbial sequence identification and subsequent taxonomic classification using the *PathSeq* tool.

**RESULTS:** We observed significant differences between the microbiome of CD samples vs inflammatory bowel disease-free controls, including depletion of *Bacteroidetes* and *Clostridia*. Notably, microbial composition at the phyla level did not differ markedly between healthy and diseased areas of CD patients. However, we observed enrichment of potentially pathogenic organisms at the species level.

**CONCLUSIONS:** Our study showed dysbiosis within deeper layers of the ileum of CD patients, specifically enrichment of enterotoxigenic *Staphylococcus aureus* and an environmental *Mycobacterium* species not described previously. Future studies with larger cohort sizes are warranted to confirm these findings. Studies would benefit from effective microbial DNA

Microbial communities in the gut lumen of patients with Crohn's disease (CD) are distinct from those observed in healthy controls and include depletion of commensal phyla such as *Firmicutes* and *Bacteroidetes*.<sup>1,2</sup> Although the etiology of this dysbiosis is unclear, it is thought to trigger intestinal inflammation.<sup>3</sup> Microbiome studies in CD typically have focused on the microbiota in feces or in the superficial mucosal layers of the colon because accessing tissue from the deeper layers of the bowel is challenging.

To identify potential pathogens we took an *in situ* approach to determine invasive tissue microbiota. We analyzed the deep tissue microbiome in patients who underwent surgical resection of the small intestine. Inflamed (diseased or involved) and uninfamed (healthy or uninvolved) areas of the resection specimens were selected from the same patient for analysis (Supplementary Table 1). Specifically, samples were obtained from 12 patients who underwent ileocecal resections at Massachusetts General Hospital (Boston, MA) for treatment of Crohn's disease and who were part of the Prospective Registry in Inflammatory Bowel Disease (IBD) study at Massachusetts General

\*Authors share co-first authorship.

**Abbreviations used in this paper:** CD, Crohn's disease; FDR, false-discovery rate; IBD, inflammatory bowel disease; LDA, linear discriminant analysis.

Most current article

© 2016 The Authors. Published by Elsevier Inc. on behalf of the AGA Institute. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

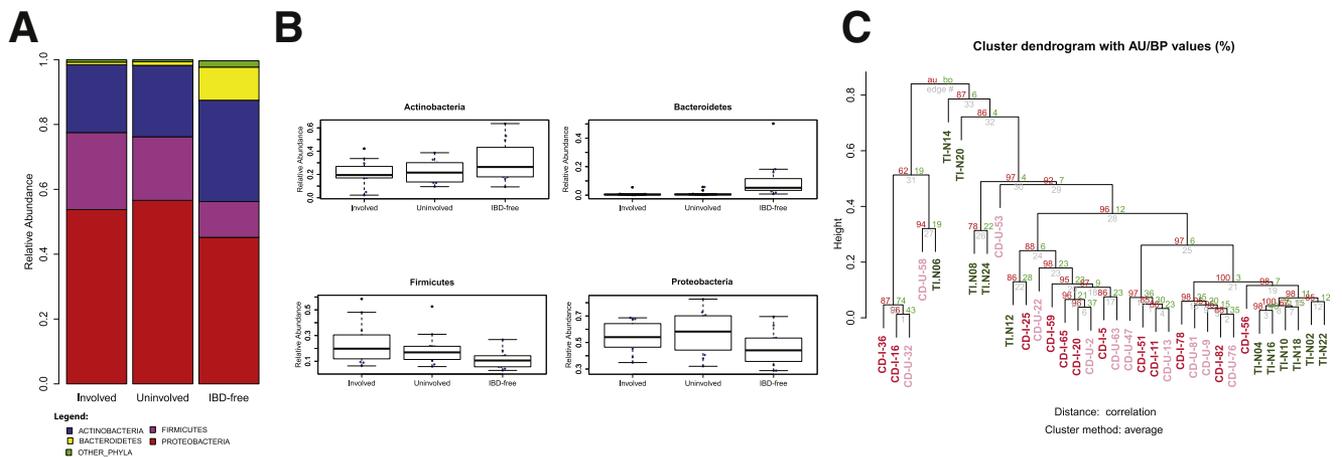
2352-345X

<http://dx.doi.org/10.1016/j.jcmgh.2016.05.011>

Hospital. We selected patients who had surgery early in the course of the disease (<5 years after diagnosis). Healthy ileum samples were obtained from patients undergoing a similar surgery for right-sided colon cancer and were designated as IBD-free controls. Deeper sections of the small bowel were macrodissected from these surgical specimens. The submucosal lymphoid areas, granulomas, and lymphoid reactions around fissures were identified by histologic examination of the tissue by a pathologist and co-investigator (A.K.B.). The diseased (involved) area was identified and cored on paraffin blocks (Supplementary Figure 1B and C) using a 1.5-mm dermal punch as described previously.<sup>4</sup> Punch biopsies were performed using sterilized, disposable needles and samples were placed directly into sterile Eppendorf tubes for further processing. Healthy-appearing (uninvolved) ileum from these CD patients and IBD-free ileum from patients undergoing surgery for right-sided colon cancer were cored similarly. DNA extraction and preparation of bar-coded DNA sequencing libraries were performed as previously described.<sup>4</sup> Microbial sequence identification and subsequent taxonomic classification was performed using the *PathSeq* tool.<sup>5</sup>

Microbial DNA was identified in the deeper bowel layers in all 3 experimental groups: involved (n = 12 samples) and uninvolved (n = 11 samples; 1 surgical resection did not contain an uninvolved segment) regions from ileal segments of 12 CD patients, and IBD-free control (n = 12 samples) ileal segments from 12 right-sided colon cancer patients. The microbiomes of all 3 groups were dominated by 4 phyla (*Actinobacteria*, *Firmicutes*, *Bacteroidetes*, and *Proteobacteria*). The relative abundance of these phyla, however, varied between the groups (Figure 1A). Reads corresponding to viral, archaeal and fungal sequences were virtually absent (Supplementary Tables 2 and 3). At the phylum level, the microbial community was similar in involved and uninvolved regions but differed markedly from the microbiomes of IBD-free patients (Figure 1A). In particular, we observed depletion of commensal bacterial

phyla such as *Bacteroidetes* in the involved and uninvolved regions compared with IBD-free ileum. The depletion in *Bacteroidetes* was especially significant ( $P = .0002$ ; involved vs IBD-free ileum), with nearly 10-fold greater relative abundance in IBD-free ileum (10.22% mean relative abundance, compared with 1.15% and 0.82% mean relative abundance in involved and uninvolved regions, respectively). The relative abundance of *Firmicutes* conversely was higher in the involved regions vs IBD-free ileum ( $P = .0204$ ), whereas abundance of the 2 other dominant phyla, *Actinobacteria* ( $P = .2189$ ) and *Proteobacteria* ( $P = .1005$ ), was not significantly different in involved vs IBD-free ileal specimens (Figure 1A). Here, it is important to note that the CD patients in this study were prescribed metronidazole and levofloxacin during the course of their treatment. Metronidazole specifically targets anaerobic and microaerophilic bacteria, including *Bacteroidetes* and members of the *Firmicutes* phylum containing the class *Clostridia*. We therefore cannot rule out the possibility that antibiotic treatment may have contributed to the depletion of members of these phyla in patients with Crohn's disease. Although all patients were treated with antibiotics, the duration of treatment varied. We decided to compare longer-term (>2 wk) to short-term ( $\leq 2$  wk) treatment in the involved (antibiotic treatment > 2 wk, n = 7 samples; antibiotic treatment  $\leq 2$  wk, n = 5 samples) and uninvolved (antibiotic treatment > 2 wk, n = 7 samples; antibiotic treatment  $\leq 2$  wk, n = 4 samples) regions. We observed neither a significant difference in the presence of *Bacteroidetes* or *Clostridiales* nor a clear enrichment of aerobic organisms and depletion of anaerobic or microaerophilic organisms in the longer-term treated group compared with short-term treatment (data not shown). Because we did not have an untreated cohort we could not determine reliably the consequences of antibiotic treatment on the microbiome in this study. However, Gevers et al<sup>6</sup> previously investigated the influence of antibiotic treatment for mucosal samples from CD



**Figure 1. (A)** Relative abundance of bacterial phyla in involved and uninvolved Crohn's and IBD-free ileum segments. **(B)** Box plots showing relative abundance of *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, and *Proteobacteria* phyla. **(C)** Unsupervised hierarchical clustering of bacterial genus level relative abundance from 3 sample groups (red, involved; pink, uninvolved; green, IBD-free). AU, approximately unbiased; BP, bootstrap probability.

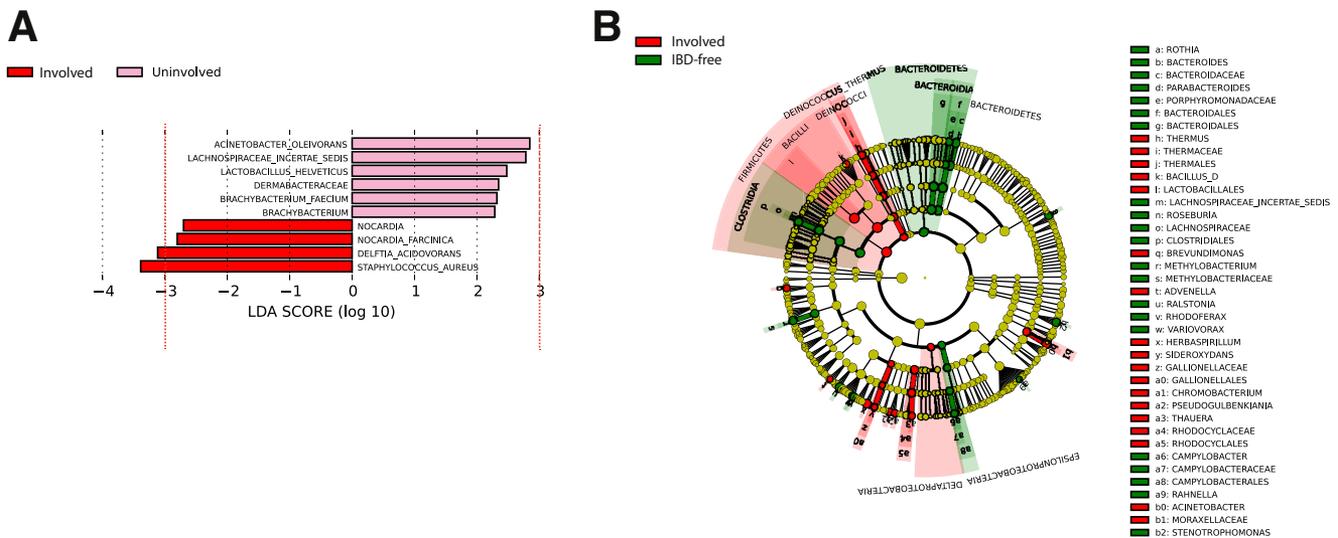
patients and noted a similar pattern of microbial dysbiosis between treated and untreated patients although the dysbiosis was more pronounced in the treated group.

We next performed unsupervised hierarchical clustering analysis based on genus-level bacterial abundance. Our results support the hypothesis that IBD-free ileum is significantly different than CD ileum, from both involved and uninvolved regions (Figure 1B). There are few reports on the microbiome of involved and uninvolved gut tissue in CD patients, although a study of colonic specimens has reported differences in the CD-associated microbiota between these regions.<sup>7</sup>

A metagenomic analysis of the 3 groups was performed using the linear discriminant analysis effect size (LEfSe) tool.<sup>8</sup> Analysis of the microbial community diversity showed that no phyla were enriched significantly in involved vs uninvolved regions, although the genera *Staphylococcus* and *Delftia* were increased modestly in inflamed regions using a linear discriminant analysis (LDA) cut-off score of 3.0 or greater (Figure 2A). At the species level resolution, we noted a significant enrichment (LDA score,  $\geq 3.0$ ) of *Staphylococcus aureus* in involved (average relative abundance, 0.45%) vs uninvolved tissue (average relative abundance, 0.05%). Because LEfSe analysis does not perform multiple hypotheses correction in identifying features significantly enriched in a group, we performed comparative marker selection using GENE-E (available: <http://www.broadinstitute.org/cancer/software/GENE-E>), to identify significantly enriched features in each group. Comparing involved (n = 12) vs uninvolved (n = 11) regions confirmed a marked but statistically nonsignificant enrichment of *S aureus* in the involved samples with a false-discovery rate (FDR) of 0.19 (P = .06) and a fold increase of 7.76. Analysis of the *S aureus* sequencing

reads showed the presence of an *S aureus* pathogenicity island in the involved tissue from 2 patients (CD-I-82 and CD-I-56). The *S aureus* pathogenicity islands are a large family of mobile, phage-related, genetic elements, most of which carry genes for 1 or more superantigens/enterotoxins.<sup>9</sup> A recent study reported infection with enterotoxigenic methicillin-resistant *S aureus* in a Crohn's patient with active colitis, which was resolved after targeted antibiotic treatment.<sup>10</sup> Superantigens secreted by these strains are potent T-cell mitogens that lead to a massive release of proinflammatory cytokines,<sup>11</sup> and several reports have highlighted the ability of enterotoxigenic strains to trigger intestinal inflammation.<sup>10,12-14</sup> These observations suggest that enterotoxigenic strains, coupled with compromised barrier function in intestinal epithelia, may contribute to localized lesions in CD.

We observed that the microbial community in the involved ileum was markedly different from the IBD-free ileum (Figure 2B and Supplementary Figure 2), including enrichment of some *Mycobacterium* species in the involved tissue. *Mycobacterium* species have been implicated previously in CD, specifically *Mycobacterium avium*.<sup>15</sup> Here, we observed an increase in *Mycobacterium abscessus* (average relative abundance, 0.09% and 0.002% in involved and normal tissue, respectively) (Supplementary Figures 2 and 3).<sup>16</sup> Comparative marker selection with GENE-E confirmed the enrichment of *M abscessus* in involved samples compared with IBD-free samples, with a FDR of 0.044 (P = .02) and a fold increase of 53. Our results show a bacterial imbalance in CD that is not restricted to mucosal surfaces, but also occurs in the deeper intestinal layers. We showed a marked depletion of *Bacteroidetes* in Crohn's patients, regardless of the presence of inflammation, which is consistent with studies suggesting that depletion of commensals such as



**Figure 2. Bacterial taxa in involved vs uninvolved regions of Crohn's patients, and involved regions of Crohn's patients vs IBD-free patients.** (A) LDA and effect size measurements identify a modest enrichment of *Staphylococcus* and *Delftia* genera in involved vs uninvolved ileal regions in Crohn's patients. Red, taxa enriched in involved regions; pink, taxa enriched in uninvolved tissues. Only taxa exceeding an LDA threshold of 3.0 (red dotted line) were considered for analysis. (B) A cladogram of the most enriched taxa in involved ileum of Crohn's patients vs IBD-free individuals. Red, taxa enriched in involved regions; green, taxa enriched in IBD-free tissues. LDA and effect size measurements for involved segments vs IBD-free are shown in Supplementary Figure 2.

*Bacteroidetes* (Supplementary Figure 4) can induce an aberrant immune response.<sup>2</sup> The *Firmicutes* phylum, including the *Bacilli* class, is enriched with a LDA score of 3.0 or greater in involved regions compared with IBD-free ileum, with enrichment of *Streptococcus* and *Bacillus* genera. We observed depletion of the *Clostridia* class, specifically the *Clostridiales* order, in involved ileum, which also is consistent with the literature.<sup>6</sup> An enrichment of potential bacterial pathogens emerged at the species level, including *S aureus* and *M abscessus*. Although this finding is intriguing and re-fuels the debate on the role of *Mycobacteria* in CD, further studies are required to confirm or negate their role in CD. Given that *Mycobacterium species* and *S aureus* are notoriously difficult-to-lyse organisms,<sup>17–19</sup> it is possible that we may in fact be underestimating their abundance in these samples. Moving forward, we intend to apply more effective lysis methods to investigate their prevalence in CD more rigorously. This approach may help reduce the intra-individual variation in isolation of bacterial nucleic acids from paraffin sections. Notably, our analysis of the deeper diseased ileum did not show an enrichment of the bacterial families *Enterobacteriaceae*, *Pasteurellaceae*, *Veillonellaceae*, and *Fusobacteriaceae* previously observed in the upper mucosal layers.<sup>6</sup> Although members of these families, such as *Fusobacteria*, have the potential to be invasive organisms, it is possible that they are not invasive in CD and remain at the mucosal interface. Interestingly, organisms belonging to the bacterial families that Gevers et al<sup>6</sup> noted as being enriched in the mucosal layers, including obligate anaerobes (all *Veillonellaceae* members and specific members of *Fusobacteriaceae*), were absent in our analysis of the deeper ileal layers. Rather, we observed aerobic (*M abscessus*) or aerotolerant (*S aureus*) species. A potential explanation for this is that inflammation within the deeper layers can result in increased oxygenated blood supply to these sites, promoting the growth of these species. It is possible that oxygen tension may determine which microbial species can colonize, and ultimately contribute to, pathology.

In summary, the *in situ* microbial community of patients with CD is distinct from IBD-free individuals. We show the enrichment of an enterotoxigenic strain of *S aureus* in diseased areas of the bowel as well as low levels of an environmental mycobacterial species in CD. These microbial communities may perpetuate stimulation of the host immune response, contributing to the ileal Crohn's phenotype and necessitating surgical intervention. Our study had limitations, including a smaller cohort size and the use of DNA extraction methods from paraffin sections that are designed for host, rather than microbial, cell lysis and nucleic acid extraction. To define the contribution of the pathogens identified in this study on the initiation and progression of

Crohn's disease, further investigation is warranted with a larger cohort size and optimized DNA extraction methods for intracellular pathogens in conjunction with sophisticated host nucleic acid-depletion approaches.

## References

1. Frank DN, et al. Proc Natl Acad Sci U S A 2007; 104:13780–13785.
2. Round JL, et al. Nat Rev Immunol 2009;9:313–323.
3. Chassaing B, et al. Gastroenterology 2011;140: 1720–1728.
4. Bhatt AS, et al. N Engl J Med 2013;369:517–528.
5. Kostic AD, et al. Nat Biotechnol 2011;29:393–396.
6. Gevers D, et al. Cell Host Microbe 2014;15:382–392.
7. Sepehri S, et al. Inflamm Bowel 2007;13:675–683.
8. Segata N, et al. Genome Biol 2011;12:R60.
9. Novick RP, et al. Plasmid 2003;49:93–105.
10. Bettenworth D, et al. World J Gastroenterol 2013; 19:4418–4421.
11. Proft T, et al. Clin Exp Immunol 2003;133:299–306.
12. Lu J, et al. Gastroenterology 2003;125:1785–1795.
13. Benjamin MA, et al. Infect Immun 1998;66:2193–2199.
14. Baca-Estrada ME, et al. Clin Exp Immunol 1995; 99:398–403.
15. Behr MA, et al. Inflamm Bowel Dis 2006;12:1000–1004.
16. Byrd TF, et al. Infect Immun 1999;67:4700–4707.
17. Willner D, et al. PLoS One 2012;7:e34605.
18. Vandeventer PE, et al. J Clin Microbiol 2011;49: 2533–2539.
19. Stuhlmeier R, et al. J Clin Pathol 2003;56:782–785.

---

Received January 21, 2016. Accepted May 15, 2016.

### Correspondence

Address correspondence to: Vijay Yajnik, MD, PhD, Crohn's and Colitis Center, Massachusetts General Hospital, Massachusetts 02114. e-mail: [vyajnik@mgh.harvard.edu](mailto:vyajnik@mgh.harvard.edu); or Matthew Meyerson, MD, PhD, Dana-Farber Cancer Institute, Boston, Massachusetts 02215. e-mail: [Matthew\\_Meyerson@dfci.harvard.edu](mailto:Matthew_Meyerson@dfci.harvard.edu); fax: 617-582-7880.

### Acknowledgments

The authors thank the Center for the Study of Inflammatory Bowel Disease, Massachusetts General Hospital (Boston, MA).

Current address of A.S.B. and V.M.: Department of Medicine, Department of Genetics, Stanford University, Stanford, California

Current address of A.I.O.: Department of Epidemiology, University of Alabama at Birmingham, Birmingham, Alabama

### Conflicts of interest

The authors disclose no conflicts.

### Funding

This work was funded by a Harvard Institute of Translational Immunology (HITI) pilot grant from the Helmsley Charitable Trust, a computational grant from Amazon Web Services and a grant from Center for the Studies on Inflammatory Bowel Disease (DK4331).

## Supplementary Materials and Methods

### Sample Selection and Pathologic Review

All patient samples were selected using a clinically annotated patient registry, (Prospective Registry in IBD Study at Massachusetts General Hospital), in which more than 1000 patients are enrolled with detailed treatment history, IBD genetics, and have consented to banking DNA, blood, and tissue. Twenty-seven formalin-fixed, paraffin-embedded ileal resection samples were identified for further investigation. The inclusion/exclusion criteria place an emphasis on selecting patients with preferably less than 6 months of exposure to antibiotics and immunosuppressive agents, which could alter the invasive gut microbiome. IBD-free ileum from right hemicolectomy specimens (obtained from colorectal carcinoma surgeries) were selected as negative controls. Institutional review board approval was obtained through the Massachusetts General Hospital and the Broad Institute. All patient samples were de-identified. All samples were reviewed histologically by a co-investigator who is a gastrointestinal pathologist with a specialization in IBD (A.K.B.).

In pilot experiments, we selected 2 submucosal CD samples, 2 deep fissure macrodissected CD samples, 2 submucosal biopsy specimens from patients with appendicitis, and 1 *Helicobacter pylori*-positive gastric submucosal sample (positive control sample) to validate our approach.

### DNA Extraction, Library Construction, and Sequencing

DNA was extracted from all formalin-fixed, paraffin-embedded samples using the Ambion Recoverall kit (ThermoFisher Scientific, Waltham, MA), as previously described.<sup>1</sup> DNA was quantified using both a Nanodrop apparatus and the Qubit quantification system (ThermoFisher Scientific). DNA was sheared using a Covaris apparatus (Covaris, Woburn, MA) to an average fragment length of 350–500 bp. Sequencing libraries were prepared using the Illumina TruSeq kit per the manufacturer's instructions (Illumina, San Diego, CA). Samples were pooled and multiplexed before sequencing on the Illumina V3 HiSeq platform (6 samples per lane, for an average depth of 3× human whole genome sequencing coverage).

### PathSeq and Statistical Analysis

The *PathSeq*<sup>2</sup> algorithm was used to perform computational subtraction of human reads, followed by alignments of residual reads to human reference genomes and microbial reference genomes (which included bacterial, viral, archaeal, and fungal sequences; downloaded from National Center for Biotechnology Information in June, 2012). These alignments resulted in taxonomic classification of reads into bacterial, viral, archaeal, and fungal sequences in whole genome sequencing data.

Briefly, *PathSeq* is used to remove low-quality reads followed by subtraction of human reads by mapping reads to a database of human genomes (downloaded from NCBI in November 2011) using BWA<sup>3</sup> (release 0.6.1, default settings), MegaBLAST (release 2.2.25; cut-off E-value, 10<sup>-7</sup>;

word size, 16), and BLASTN<sup>4</sup> (release 2.2.25; cut-off E-value, 10<sup>-7</sup>; word size, 7; nucleotide match reward, 1; nucleotide mismatch score, -3; gap open cost, 5; and gap extension cost, 2). Only sequences with perfect or near-perfect matches to the human genome were removed in the subtraction process. In addition, low-complexity and highly repetitive reads were removed using *Repeat Masker*<sup>5</sup> (version open-3.3.0; libraries dated 2011-04-19).

Taxonomic classification is performed by residual read alignment using MegaBLAST (release 2.2.25; cut-off E-value, 10<sup>-7</sup>; word size, 16) to a database of bacterial sequences and followed by BLASTN (release 2.2.25; cut-off E-value, 10<sup>-7</sup>; word size, 7) to human reference genome and microbial reference genomes.

After the taxonomic classification of nonhuman DNA sequencing reads the relative abundance value for each bacterial organism was calculated as follows by using reads that maps with 90% or greater sequence identity and 90% or greater query coverage. Classifications were performed at the domain, then phylum, then genus, and then species level, requiring unique alignments (ie, reads with equivalent E-values to multiple taxa were removed from analysis).

At the species level, relative abundance for each organism was calculated as follows: relative abundance = (number of unique alignment positions in genome × 1,000,000)/(number of total aligned bacterial reads × genome size). The relative abundance values then were per-sample normalized such that the total relative abundance for each sample summed to 1. The resulting normalized relative abundance matrix was analyzed on LEfSe (online version available on August 2015)<sup>6</sup> between different sample group labels (uninvolved, involved, and IBD-free samples) with default parameters. However, we considered the bacteria that had a LDA score of 3.0 or greater for further analysis. In addition, the Wilcoxon rank-sum test was used to calculate the *P* value for phyla level relative abundance between groups. Unsupervised hierarchical clustering was performed using the *pvclust* R package (available: <https://cran.r-project.org/web/packages/pvclust/index.html>) with 1000 bootstrap iterations, Ward's clustering method, and otherwise default parameters on genus level bacterial relative abundance values.

Comparative marker selection was used to complement LEfSe analysis via GENE-E (<http://www.broadinstitute.org/cancer/software/GENE-E/>) with default settings and 10,000 permutations. This analysis was performed between involved (*n* = 12) and uninvolved (*n* = 11) samples, and between involved (*n* = 12) and IBD-free (*n* = 12) samples. This analysis returned a FDR, a fold change, and a *P* value based on a permutation test.

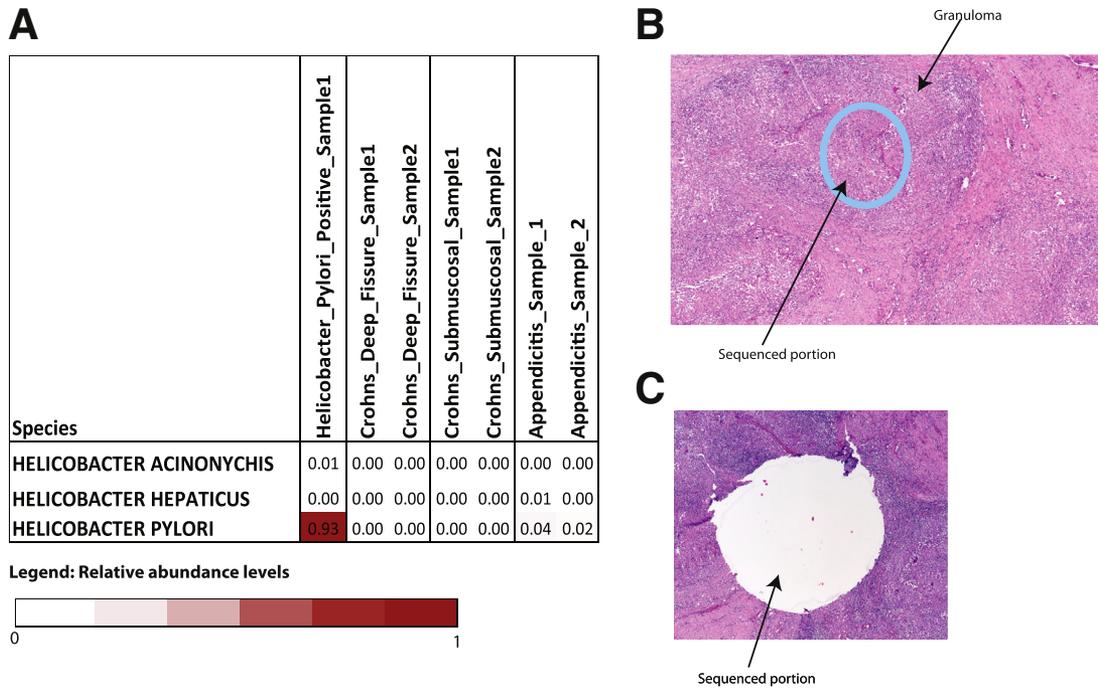
### Results of Sequencing and PathSeq Analysis

*PathSeq* was used to assign a taxonomic classification of each sequence based on the organism of origin. Results from pilot experiments (complete data not shown) used to validate our approach shows accurate identification of *H pylori* in *H pylori*-positive gastric submucosal samples (Supplementary Figure 1A). We obtained a median of 39.3

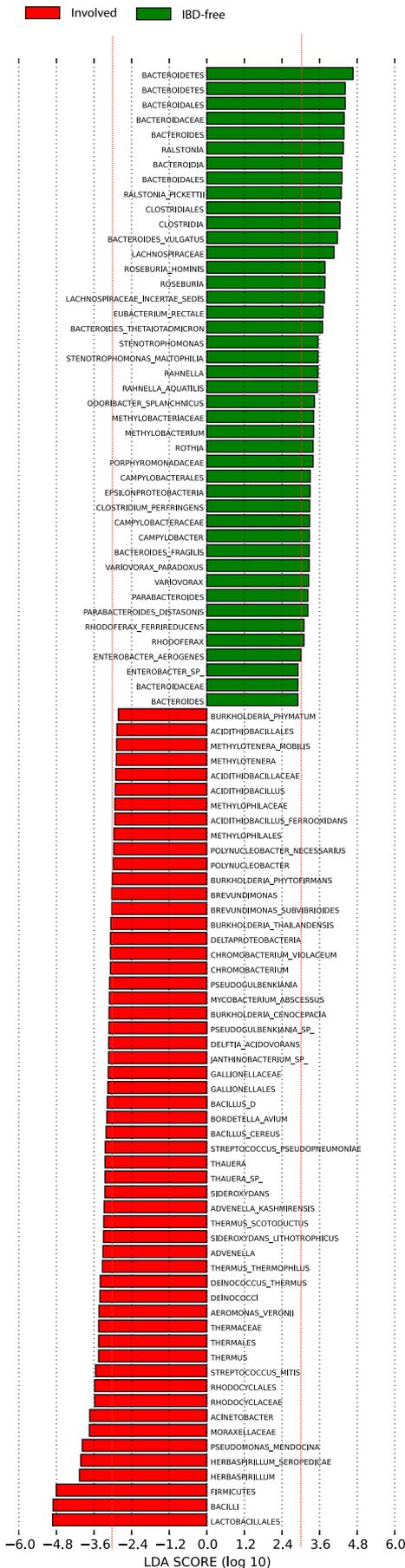
million human DNA sequencing reads (range, 14.5–188.8 million), 2330 bacterial sequencing reads (range, 244–19445 reads), and 8660 unmapped sequencing reads (range, 2110–78466 reads) per Crohn's disease or IBD-free sample. Microbial reads included those corresponding to known bacteria, archaea, fungi, and viruses. The number of human reads, microbial reads, and unmapped reads are presented in [Supplementary Table 2](#). The relative abundance for each bacterium at the species level is presented in [Supplementary Table 4](#).

## References

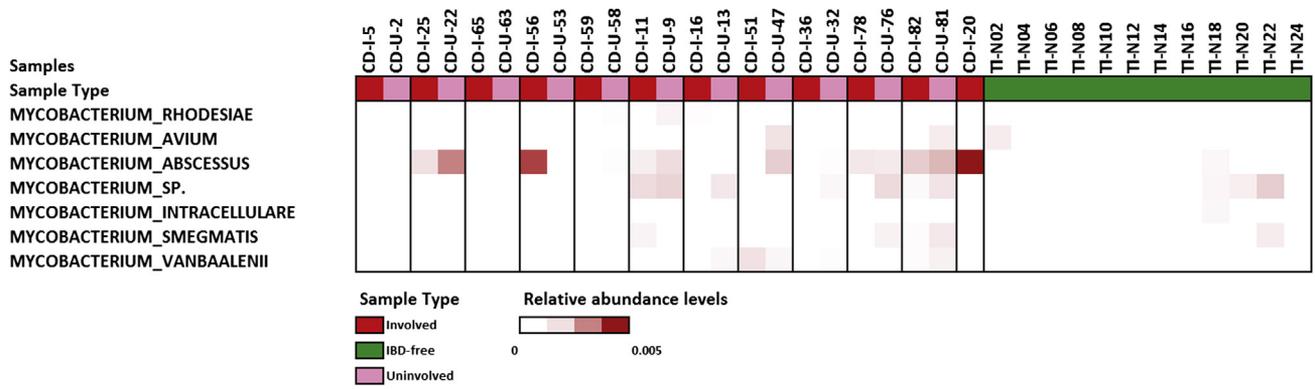
1. Bhatt AS, et al. *N Engl J Med* 2013;369:517–528.
2. Kostic AD, et al. *Nat Biotechnol* 2011;29:393–396.
3. Li H, et al. *Bioinformatics* 2009;25:1754–1760.
4. Altschul SF, et al. *Nucleic Acids Res* 1997;25:3389–3402.
5. Smit AFA, et al. Available from: <http://www.repeatmasker.org>. Accessed 2011.
6. Segata N, et al. *Genome Biol* 2011;12:R60.
7. Gevers D, et al. *Cell Host Microbe* 2014;15:382–392.



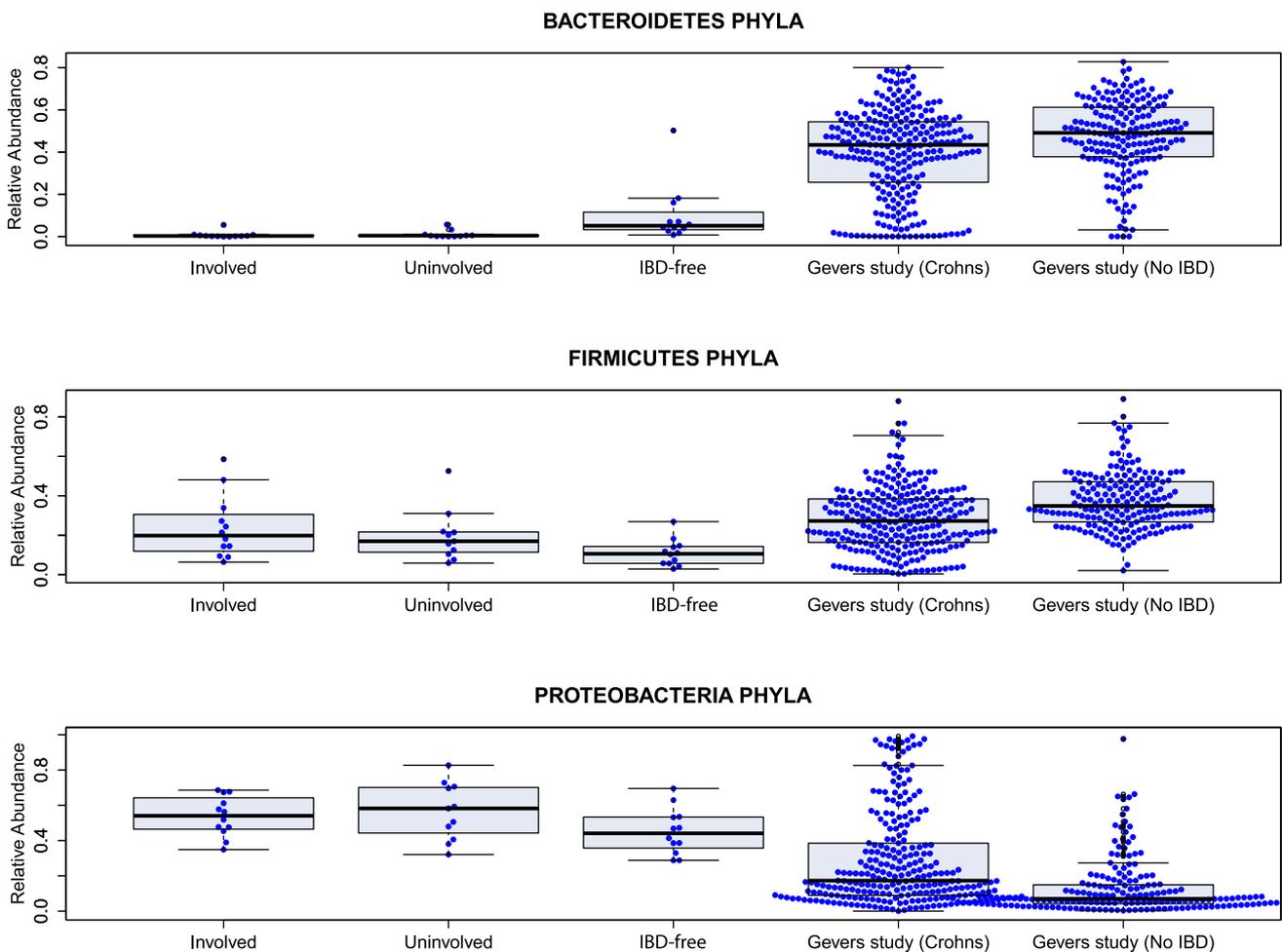
**Supplementary Figure 1. H&E sections of small intestine showing (A) relative abundance levels of various *Helicobacter* species in pilot set.** Results show accurate identification of *H pylori* in a *H pylori*-positive gastric submucosal sample. (B and C) Representative H&E section before and after macrodissection of diseased area of the small bowel is shown. (B) The sequenced portion of the bowel wall is shown as a circle and was removed from the paraffin section by the biopsy needle. (C) The rest of the bowel wall otherwise is intact.



**Supplementary Figure 2.** LDA coupled with effect size measurements identifies significant depletion of *Bacteroidetes* phyla and *Clostridia* class in *Firmicutes* phyla when compared with involved ileum segments with IBD-free ileum. In addition, there was significant enrichment of *Bacilli* class in *Firmicutes* phyla when comparing involved ileum segments with IBD-free. Involved enriched taxon have a negative score (red), and taxa enriched in IBD-free tissue have a positive score (green). Only taxa meeting an LDA threshold of 3.0 (red dotted line) were considered for analysis.



Supplementary Figure 3. Relative abundance of various *Mycobacterium* species in the cohort.



Supplementary Figure 4. Box plots show comparison of *Bacteroidetes*, *Firmicutes*, and *Proteobacteria* phyla between our study and the Gevers et al<sup>7</sup> study.