

# Spatial variation of the native colon microbiota in healthy adults

Running title: Spatial variation of native colon microbiota

Kaitlin J. Flynn<sup>1</sup>, Mack T. Ruffin IV<sup>2</sup>, D. Kim Turgeon<sup>3†</sup>, and Patrick D. Schloss<sup>1†</sup>

† Corresponding authors: [kturgeon@umich.edu](mailto:kturgeon@umich.edu) and [pschloss@umich.edu](mailto:pschloss@umich.edu)

1. Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, Michigan 48109

2. Department of Family and Community Medicine, College of Medicine, Pennsylvania State University, Hershey, Pennsylvania 17033

3. Department of Internal Medicine, Division of Gastroenterology, University of Michigan Medical School, Ann Arbor, Michigan

Grant support: Rose and Lawrence C. Page Foundation

Disclosures: nothing to disclose

Word count: 4,205/5,000

Figures: 6

16 **Abstract**

17 The microbiome has been implicated in the development of colorectal cancer (CRC) and inflamma-  
18 tory bowel diseases (IBD). The specific traits of these diseases vary along the axis of the digestive  
19 tract. Further, variation in the structure of the gut microbiota has been associated with both  
20 diseases. Here we profiled the microbiota of the healthy proximal and distal mucosa and lumen to  
21 better understand how bacterial populations vary along the colon. We used a two-colonoscopy  
22 approach to sample proximal and distal mucosal and luminal contents from the colons of 20 healthy  
23 subjects that had not undergone any bowel preparation procedure. The biopsies and home-collected  
24 stool were subjected to 16S rRNA gene sequencing and Random Forest classification models were  
25 built using taxa abundance and location to identify microbiota specific to each site. The right  
26 mucosa and lumen had the most similar community structures of the five sites we considered  
27 from each subject. The distal mucosa had higher relative abundance of *Fingoldia*, *Murdochiella*,  
28 *Peptoniphilus*, *Porphyromonas* and *Anaerococcus*. The proximal mucosa had more of the genera  
29 *Enterobacteriaceae*, *Bacteroides* and *Pseudomonas*. The classification model performed well when  
30 classifying mucosal samples into proximal or distal sides (AUC = 0.808). Separating proximal  
31 and distal luminal samples proved more challenging (AUC = 0.599) and specific microbiota that  
32 differentiated the two were hard to identify. By sampling the unprepped colon, we identified  
33 distinct bacterial populations native to the proximal and distal sides. Further investigation of  
34 these bacteria may elucidate if and how these groups contribute to different disease processes on  
35 their respective sides of the colon.

36 Words: 233/250

37 Keywords: microbiome; colon cancer; proximal and distal colon

## 38 **Introduction**

39 The human colon is an ecosystem comprised of numerous microenvironments that select for  
40 different microbiota. Concentrations of oxygen, water, and anti-microbial peptides change along  
41 the gut axis and influence which microbiota reside in each location. Microenvironments differ not  
42 only longitudinally along the colon, but also radially from the epithelium to mucosa to intestinal  
43 lumen, offering several sites for different microbial communities to flourish. The identity of these  
44 specific microbiota and communities are important for understanding the etiology of complex  
45 diseases such as Colorectal Cancer (CRC) and Inflammatory Bowel Disease (IBD). CRC and IBD  
46 can be preceded or accelerated by perturbations of the structure of the gut microbiota (1–3). The  
47 manifestations of these diseases are known to vary based upon the location in which they occur.  
48 For instance, CRC that arises in the distal (left) colon are of hindgut origin and tend to have large  
49 chromosomal alterations indicative of chromosomal instability (1). In contrast, CRC arising in the  
50 proximal (right) colon are of midgut origin and tend to be sessile and microsatellite instable (MSI  
51 with BRAF and KRAS mutations) (1). In addition to the environmental gradients within the colon,  
52 the distal and proximal sides of the colon differ in the amount of inflammation present and the  
53 genomic instability of precancerous cells, respectively (1,4,5). In IBD patients, disease occurring  
54 in the distal colon extending proximally is usually indicative of ulcerative colitis (UC), whereas  
55 Crohn's disease (CD) can occur anywhere along the GI tract, most commonly in the ileum and the  
56 cecum (2). UC presents as continuous disease with only mucosal involvement, where as CD has  
57 skip lesions and full thickness involvement that may cause abscesses, strictures and fistulas (2).  
58 Thus, given the varied physiology of the proximal-distal axis of the colon and known differences  
59 in disease patterns at these sites, symbiotic microbiota and their metabolites likely vary as well,  
60 and may influence the heterogeneous disease prognoses of IBD and CRC. Because CRC can be a  
61 long-term complication of IBD, the distribution of microbiota is important to understanding the  
62 pathophysiology of both diseases.

63 Several recent findings have shown that development and progression of IBD or CRC can be

64 attributed to specific molecular events as a result of interactions between the gut microbiota  
65 and human host (1,3,6). For instance, comparison of the bacteria present on CRC tumors with  
66 those found on nearby healthy tissue has identified specific species that are tumor-associated  
67 (7). Specific bacteria have also been identified in fecal samples of patients with varying stages of  
68 colon tumorigenesis (8,9). These species include the oral pathogens *Fusobacterium nucleatum*  
69 and *Porphyromonas asacharolytica*. *F. nucleatum* has also been found to be elevated in the stool  
70 and biopsies of patients with IBD as compared to healthy controls (10,11). Furthermore, studies  
71 of *F. nucleatum* isolated from mucosal biopsies showed that more invasive *F. nucleatum* positively  
72 correlates with IBD disease level (10). Like many intestinal pathogens, the bacteria appear to  
73 have a high-impact despite being lowly-abundant in the community (2). The physiology of these  
74 rare taxa may contribute to the colonic disease state. These studies often examined only shed  
75 human stool or the small intestine, preventing fine-resolution analysis of paired samples from the  
76 proximal and distal sides of the colon. Similarly, comparisons of on- or off-tumor/lesion bacteria  
77 rarely have matched tissue from the other side of the colon from the same, disease-bearing patient,  
78 limiting what conclusions can be drawn about the colonic microbiome overall, let alone at that  
79 specific site (12). Due to these limitations, the contribution of the gut microbiota to CRC and  
80 IBD disease location in the colon is largely undefined. Characterizing these communities in healthy  
81 individuals could provide needed insight into disease etiology, including how the disruption of the  
82 healthy community could promote the initiation or proliferation of the distinct proximal and distal  
83 CRC tumors or IBD flares.

84 The few existing profiles of the microbial spatial variation of the colon have been limited by sample  
85 collection methods. The majority of human gut microbiome studies have been performed on whole  
86 shed feces or on samples collected during colonoscopy or surgery (5). While invasive methods  
87 allow investigators to acquire samples from inside the human colon, typically these procedures  
88 are preceded by the use of bowel preparation methods such as the consumption of laxatives to  
89 cleanse the bowel. Bowel preparation is essential for detecting cancerous or precancerous lesions  
90 in the colon, but complicates microbiome profiling as the chemicals strip the bowel of contents

91 and disrupt the mucosal layer (13,14). As such, what little information we do have about the  
92 spatial distribution of the microbiota in the proximal and distal colon is confounded by the bowel  
93 preparation procedure.

94 Here we address the limitations of previous studies and identify the microbes specific to the  
95 lumen and mucosa of the proximal and distal healthy human colon. We used an unprepared  
96 colonoscopy technique to sample the natural community of each location of the gut without prior  
97 disruption of the native bacteria in 20 healthy volunteers. To address the inherent inter-individual  
98 variation in microbiota, we used a machine-learning classification algorithm trained on curated  
99 16S rRNA sequencing reads to identify the microbiota that were specific to each location. We  
100 found that our classification models were able to separate mucosal and luminal samples as well  
101 as differentiate between sides of the colon based on populations of particular microbiota. By  
102 identifying the distinguishing microbiota we are poised to ask if and how the presence or disruption  
103 of the microbiota at each site contribute to the development of the tumor subtypes of CRC in the  
104 proximal and distal human colon.

## 105 **Methods**

### 106 **Human subjects**

107 The procedures in this study and consent were approved by the Institutional Review Board at  
108 the University of Michigan Health System with protocol number HUM00082721. Subjects were  
109 recruited using the online recruitment platform and were pre-screened prior to enrollment in the  
110 study. Exclusion criteria included: use of aspirin or NSAIDs within 7 days, use of antibiotics within  
111 3 months, current use of anticoagulants, known allergies to Fentanyl, Versed and Benadryl, prior  
112 history of colon disease, diabetes, abdominal surgery, respiratory, liver, kidney or brain impairments,  
113 undergoing current chemotherapy or radiation treatment and subjects that were pregnant or trying  
114 to conceive. 20 subjects that met the criteria were selected and provided signed informed consent  
115 prior to the procedure. There were 13 female and 7 male subjects ranging in age from 25 to 64.

116 18 of the 20 subjects had not used antibiotics within a year prior to the collection date and 2  
117 had not used antibiotics within 6 months. None of the subjects had medical conditions requiring  
118 frequent or extended antibiotic use.

### 119 **Sample collection**

120 At a baseline visit, subjects gave consent and were given a home collection stool kit (Zymo). One  
121 to seven days prior to the scheduled colonoscopy, subjects collected whole stool at home and  
122 shipped the samples to a research coordinator on ice. Notably, subjects did not undergo any bowel  
123 preparation method prior to sampling. On the procedure day, subjects reported to the Michigan  
124 Clinical Research Unit at the University of Michigan Health System. Subjects were consciously  
125 sedated using Fentanyl, Versed and/or Benadryl as appropriate. A flexible sigmoidoscope was first  
126 inserted about 25cm into the colon and jumbo biopsy forceps used to collect the luminal contents.  
127 Two luminal samples were collected and the contents immediately deposited into RNAlater (Fisher)  
128 and flash-frozen in liquid nitrogen. The forceps were withdrawn and new biopsy forceps were used  
129 to collect mucosal biopsies on sections of the colon that were pink and free of stool matter. Three  
130 mucosal biopsies were collected and flash-frozen in RNAlater. These samples comprised the distal  
131 colon samples. The sigmoidoscope was then withdrawn and a pediatric colonoscope was inserted  
132 to reach the proximal colon. The proximal samples were taken from the ascending colon proximal  
133 to the hepatic flexure at 75-120cm depending on the subject. Samples were then collected in the  
134 same manner as was done in the distal colon and the colonoscope withdrawn. All samples were  
135 stored at -80°C.

### 136 **Sample processing, sequencing and analysis**

137 DNA extraction was performed using the PowerMicrobiome DNA/RNA Isolation Kit (MO BIO  
138 Laboratories). For tissue biopsies, Bond-Breaker TCEP solution (Fisher) and 2.8mm ceramic  
139 beads (MO BIO Laboratories) were added to the bead beating step to enhance DNA recovery from  
140 mucosal samples. The resulting DNA was normalized to equal concentrations across all samples

141 and used as template for amplification of the V4 region of the 16S rRNA gene and fragments  
142 were sequenced on an Illumina MiSeq as previously described (15). Sequences were curated using  
143 the mothur software as described previously (16). The sequences were assigned a taxonomic  
144 classification using a naive Bayesian classifier trained using a 16S rRNA gene training set from the  
145 Ribosomal Database Project (RDP) (17) and clustered into operational taxonomic units (OTUs)  
146 based on a 97% similarity cutoff. Sequencing and analysis of a mock community revealed the  
147 error rate to be 0.018%. Samples were rarefied to 4231 sequences per sample in order to reduce  
148 the effects of uneven sampling bias.

149 Diversity analysis was performed using the Simpson diversity calculator and  $\theta_{YC}$  calculator metrics  
150 in mothur version 1.39.5 (16).  $\theta_{YC}$  distances were calculated to determine the dissimilarity between  
151 two samples. Random Forest classification models were built using the AUCRF R package using  
152 a leave-one-subject out approach (18). The Random Forest models were built using the full,  
153 non-rarefied, dataset as input. For each model the data were split into a 19-subject training set  
154 and a 1-subject test set. The model was built and cross-validated using 10-fold k cross-validation  
155 (AUCRFcv) on the training set to estimate the prediction error of the model. The resultant model  
156 was then used to predict the outcome the left-out subject. This process was repeated iteratively  
157 for all subjects and results plotted as Receiver Operator Characteristic curves using the pROC  
158 R package (19). Resultant models were used to identify the OTUs that were most important  
159 for classifying each location. Species-level information for sequences of interest was obtained by  
160 aligning the sequences to the GenBank nucleotide database using blastn. The species name was  
161 only used if the identity score was  $\geq 99\%$  over the full-length of the contig and matched a single  
162 reference.

## 163 **Statistical analysis**

164 Differences in community membership at the phyla level were tested using the analysis of molecular  
165 variance (AMOVA) metric in mothur. Differences in  $\theta_{YC}$  distances by location were tested using  
166 the Wilcoxon rank-sum test adjusted for multiple comparisons using the Benjamini-Hochberg

167 procedure.

## 168 **Data availability**

169 16S rRNA gene sequence reads and experiment metadata are available on the NCBI Sequence Read  
170 Archive (SRA) with accession number SRP124947 and PRJNA418115. A reproducible data analysis  
171 pipeline can be found at [https://github.com/SchlossLab/Flynn\\_LRColon\\_CancPrevRes\\_2017](https://github.com/SchlossLab/Flynn_LRColon_CancPrevRes_2017).

## 172 **Results**

### 173 **Microbial membership and diversity of the proximal and distal colon**

174 Luminal and mucosal samples were collected from the proximal and distal colon of 20 healthy  
175 individuals who had not undergone bowel preparation (Fig. 1). Subjects also collected stool at  
176 home one week prior to the procedure. To characterize the bacterial communities present at  
177 these sites, 16S rRNA gene sequencing was performed on DNA extracted from each sample. As  
178 expected, each site was primarily dominated by *Firmicutes* and *Bacteroidetes* (Fig. 2A) (20).  
179 Samples had varying levels of diversity at each site, irrespective of the individual (Fig. 2B).  
180 For example, the proximal mucosa was more diverse than the distal for some individuals while  
181 the opposite was true for others. Therefore we could not identify a clear pattern of changes in  
182 microbial diversity along the gut axis.

183 To compare similarity between the proximal and distal sides and within the lumen and mucosa, we  
184 compared the community structure of these sites based on the relative abundances of individual  
185 Operational Taxonomic Units (OTUs). Across all subjects we observed wide variation when  
186 comparing sample locations (Fig. 3A). Those ranges did not follow a clear pattern on an individual  
187 basis. However, when comparing median dissimilarity between the communities found in the  
188 proximal lumen and mucosa, the proximal and distal lumen, the proximal and distal mucosa, and  
189 the distal lumen and mucosa, we found that the proximal lumen and mucosa were most similar to  
190 each other than to the other samples ( $P < 0.005$ , Wilcoxon, BH adjustment).



191 **Fecal samples resemble luminal samples from the distal colon**

192 Next, we compared the luminal and mucosal samples to the fecal sample of each subject. Amidst  
193 the large inter-subject variation, we did identify significantly less dissimilarity between the distal  
194 luminal sample and the feces (Fig. 3B,  $P < 0.05$ , Wilcoxon, BH adjustment). Furthermore, there  
195 was an even larger difference in the communities found in the distal mucosa compared to the fecal  
196 communities, indicating that the mucosa is as different from the stool as compared to lumen ( $P$   
197  $< 0.0005$ , Wilcoxon, BH adjustment). These results suggest that the contents of the distal lumen  
198 were most representative of the subjects' feces, and the mucosal microbiota are distinct from the  
199 fecal and luminal communities.

200 **Interpersonal community variation is greater than the variation between sites**

201 To determine what factors may have driven the differences seen among the samples, we compared  
202 the community dissimilarity between samples from all subjects (interpersonal) versus samples from  
203 within one subject (intrapersonal). We found that samples from one individual were far more  
204 similar to each other than to matched samples from the other subjects (Fig. 3C); this is consistent  
205 with previous human microbiome studies that have sampled multiple sites of the human colon  
206 (21–23). Thus interpersonal variation drove the differences between samples more than whether  
207 the sample came from the proximal or distal side of the colon or from the lumen or mucosa.

208 **Random Forest classification models identify important OTUs on each side**

209 To identify OTUs that were distinct at each site, we constructed several Random Forest models  
210 trained using OTU relative abundances. We built the first model to classify the luminal versus  
211 mucosal samples for the proximal and distal sides, independently (Fig. 4A). The models performed  
212 well when classifying these samples (proximal AUC = 0.716, distal AUC = 0.862). The OTUs that  
213 were most predictive of each site were identified by their greatest mean decrease in accuracy when  
214 removed from the model. For distinguishing the proximal lumen and mucosa, OTUs affiliated with

215 the *Bacteriodes*, *Actinomyces*, *Psuedomonas* and *Enterobacteraceae* were included in the best  
216 model (Fig. 5A). The model to differentiate between the distal lumen and mucosa included OTUs  
217 affiliated with the *Turicibacter*, *Finegoldia*, *Peptoniphilus* and *Anaerococcus* (Fig. 5B). These  
218 results indicated that there were fine differences between the different sites of the colon, and that  
219 these could be traced to specific OTUs on each side.

220 Next, we built a Random Forest model to differentiate the proximal and distal luminal samples.  
221 The model performed best when distinguishing the proximal versus distal mucosa (Fig. 4B, AUC  
222 = 0.808) whereas the model to differentiate between the proximal versus distal lumen performed  
223 poorly (AUC = 0.599). The OTUs included in the model differentiating the distal and proximal  
224 mucosa included members of the *Porphyromonas*, *Murdochiella*, *Finegoldia*, *Anaerococcus* and  
225 *Peptoniphilus* (Fig. 6A). The model that attempted to separate the the proximal and distal  
226 lumen included OTUs affiliated with the *Bacteroides*, *Clostridium IV* and *Oscillibacter* (Fig. 6B).  
227 Interestingly, *Anaerococcus* and *Finegoldia* were distinct between the mucosa and lumen and also  
228 helped to differentiate between the proximal and distal sides.

## 229 **Bacterial OTUs associated with CRC and IBD are found in healthy individuals**

230 Given that specific bacterial species have been associated with colorectal cancer and IBD, we  
231 probed our sample set for these OTUs. Among our 100 samples, the most frequent sequence  
232 associated with the *Fusobacterium* genus was OTU179, which aligned via blastn to *Fusobacterium*  
233 *nucleatum subsp animalis* (100% over full length). This is the only species of *Fusobacterium*  
234 known to have oncogenic properties and be found on the surfaces of colorectal cancer tumors (24).  
235 There were 14 samples from 8 subjects with the *F. nucleatum subsp. animalis* sequences. Of the  
236 samples with the highest relative abundance of *F. nucleatum subsp. animalis*, four of the samples  
237 were from the proximal mucosa and three from the distal mucosa (Supplementary Fig. S1A). The  
238 second most frequent *Fusobacterium* sequence was OTU472, which aligned with 99% identity to *F.*  
239 *varium*. In addition to *F. nucleatum*, *F. varium* has been associated with IBD (25). Four subjects  
240 harbored *F. varium* and the samples were split evenly between the proximal and distal mucosa

241 (Supplementary Fig. S1B). OTU152 was similar to the members of the *Porphyromonas* genus  
242 and the most frequent sequence in that OTU aligned to *Porphyromonas asacharolytica* (99% over  
243 full length), another bacterium commonly detected and isolated from colorectal tumors. OTU152  
244 was only detected on the distal mucosa, and in fact was one of the OTUs the classification model  
245 identified as separating distal and proximal sides (Supplementary Fig. S1C). Among the 11 distal  
246 mucosa samples that were positive for *P. asacharolytica*, the relative abundances for this OTU  
247 ranged from 0.01% to 16%. Thus, disease-associated OTUs could be found in our sample set of  
248 20 healthy individuals.

## 249 Discussion

250 We identified bacterial taxa that were specific to the lumen and mucosa of the proximal and distal  
251 sides of the human colon using samples collected during an unprepared colonoscopy of healthy  
252 subjects. We found that all locations contained a range of phylum relative abundances and a  
253 range of diversity, but that there was a wide variability between subjects. Pairwise comparisons  
254 of each of the sites revealed that the proximal mucosa and lumen were most similar to each  
255 other. Further, comparison of colonoscopy-collected samples with fecal samples demonstrated  
256 that the distal lumen was most similar to feces. Random Forest models built using OTU relative  
257 abundances from each sample identified microbiota that were particular to each location of the  
258 colon. Finally, we were able to detect some bacterial OTUs associated with colonic disease in our  
259 healthy cohort. Using unprepped colonoscopies and machine learning, we have identified bacterial  
260 taxa specific to the healthy proximal and distal human colon.

261 When examining the relative abundance of the dominant phyla at each site (i.e. *Bacteriodes* and  
262 *Firmicutes*), there was a wide amount of variation. This likely reflects not only the variability  
263 between human subjects, caused by differences in age, sex, and diet, but may also reflect spatial  
264 patchiness in the gut microbiome within a subject. Patchiness refers to inconsistent distribution of  
265 microbial populations due to fluctuations in local resources (26). One study noted that the bacteria

266 recoverable from the same mucosal sample location can be vastly different when the samples  
267 are taken just 1 cm away from each other (27). Similar patchiness was also observed in luminal  
268 contents and fecal samples themselves; there was separation of different interacting microbes  
269 along the length of a stool sample, for instance (28). A third study that sampled six mucosal  
270 sites along the colon observed such patchiness in two of the three study subjects (21). While  
271 our subjects were not sampled frequently enough to draw specific conclusions about patchiness  
272 along the unprepped colon, we did observe some specific differences in mucosal versus luminal  
273 samples at the phylum level. The mucosal samples harbored more *Proteobacteria*, consistent  
274 with previous studies comparing mucosal swabs to luminal content in humans (4). However, we  
275 must still consider that the results from phyla analysis may have been impacted by inter-subject  
276 patchiness.

277 To get around the noisiness from a diverse set of samples, we built Random Forest classification  
278 models to identify the microbiota that were specific to each side and in the lumen and mucosa. For  
279 each comparison we identified the top five OTUs that were strongly predictive of one site or another.  
280 Generally, OTUs identified in each location were consistent with known physiological gradients along  
281 the gut axis (5). For instance, the proximal mucosa contains the highest oxygen concentrations  
282 of the colon and harbored mucosa-associated facultative anaerobes such as *Actinomyces* and  
283 *Enterobacteraceae* and aerobic *Pseudomonas*. The distal mucosa was far more likely to host  
284 strictly anaerobic species such as *Porphyromonas*, *Anaerococcus*, *Finnegoldia* and *Peptoniphilus*.  
285 Thus the gut microenvironment of each location likely enriches for these specific microbiota.

286 In addition to identifying features that are specific to each side of the gut, the ability of the  
287 Random Forest to classify samples can serve as a proxy for similarity. That is, a higher AUC value  
288 indicates the samples are more efficiently classified (and thus more different) than a model with a  
289 lower AUC value. For instance, the model separating the proximal and distal mucosa had an AUC  
290 of 0.850 whereas the model for classifying the proximal and distal lumen had a much lower AUC  
291 of 0.580. Further, the latter model required 44 OTUs to best separate the samples whereas the

292 models separating the mucosa only needed 10 OTUs. The much lower AUC and need for a high  
293 number of features compared to other models suggest these locations are the most similar of the  
294 comparisons tested. We speculate that the model was less effective at classifying the proximal  
295 and distal luminal contents because the mucosal microenvironments have more variable selective  
296 pressure along the colon than the luminal microenvironments.

297 We detected *F. nucleatum* and *P. asacharolytica* in 8 and 5 of our subjects, respectively. These  
298 bacteria have been shown to be predictive of colorectal cancer in humans (9) and have oncogenic  
299 properties in cell culture and in mice (29). Though the bacteria are known to co-localize on CRC  
300 tumors, in our study *F. nucleatum* was found on both sides of the colon while *P. asacharolytica* was  
301 only detected in the distal mucosa. Not much is known about the distribution of *P. asacharolytica*  
302 along the healthy colon, but given its anaerobic lifestyle and asacharolytic metabolism, it is perhaps  
303 not surprising that our study detected the bacteria primarily in the less-oxygen-rich and protein-rich  
304 distal mucosa (4). In studies examining bacteria on colorectal cancer tumors, *F. nucleatum* was  
305 more commonly detected on proximal-sided tumors, and distribution of *F. nucleatum* decreased  
306 along the colon to rectum (30). Of the 8 (40%) individuals positive for *F. nucleatum* in our  
307 study, the bacterium was spread across the proximal mucosa, distal lumen and distal mucosa. The  
308 *Fusobacterium* species *nucleatum* and *varium* have been commonly isolated from mucosal biopsies  
309 of patients with IBD and UC (25,31). In our study, *F. varium* was only detected in three subjects  
310 and two of those samples were isolated from the proximal mucosa (Supplementary Fig. S1B). *F.*  
311 *varium* is most commonly isolated from UC patient biopsies from the ileum or cecum (adjacent to  
312 the proximal colon) (32), suggesting this species may exhibit preference for the higher oxygen  
313 content of these gastrointestinal sites.

314 Spatial organization of *Fusobacteria* and other bacterial species into polymicrobial biofilms that  
315 can invade the gut mucosa have been linked to CRC (33). The biofilms promote tumorigenesis by  
316 allowing bacteria to grow near the epithelium, inducing inflammation, genotoxicity and metabolic  
317 changes that favor tumor cell growth (33). In one study of CRC biofilms and tumors, all of the

318 proximal tumors examined contained a polymicrobial biofilm on the tumor mucosa (7). Further  
319 examination of these tumors identified *Fusobacteria* species as members of the biofilm. These  
320 results indicate that it is not only the presence of the bacterium but the tumor community as  
321 a whole that contributes to tumorigenesis (7). A 'driver-passenger' model has been proposed  
322 as a mechanism for biofilm assembly in the gut (34,35). In this model, 'driver' species such as  
323 *Fusobacterium spp* and *Porphyromonas spp* exert tumorigenic effects locally and create a niche  
324 for adherence of 'passenger' species that comprise the rest of the biofilm (34,35). Thus the  
325 the distribution of these disease-associated microbes in healthy patients is of interest as their  
326 presence can be predictive of disease prior to the onset of symptoms (9). A better understanding  
327 of the early microbial changes in the gut microbiome is essential for elucidating a mechanism for  
328 development of CRC or IBD subtypes in the proximal or distal colon.

329 Specific comparisons of our findings to previously published studies of spatial variation are  
330 confounded by the use of bowel preparation methods. A rare report of a matched-colonoscopy  
331 study sampled 18 patient's colonic mucosa and luminal contents prior to and after bowel cleansing  
332 (36). This study found that mucosal and luminal samples were distinguishable prior to bowel  
333 cleansing, but that bowel preparation resulted in an increase in shared OTUs between each site  
334 (36). After seven days, bowel cleansing not only made the samples more difficult to distinguish,  
335 but it also decreased the diversity observed across sites. Bowel preparation clearly biases the  
336 representation of microbiota recovered from sampling the lumen or mucosa.

337 By revealing specific differences in microbial populations at each location in the gut via sampling an  
338 unprepared bowel, we can begin to form hypotheses about how specific host-microbe interactions  
339 can affect disease progression of proximal and distal CRC and IBD subtypes. Future investigation of  
340 these samples using metagenomics and metatranscriptomics would illuminate the microbial activities  
341 in these gut microenvironments. Further, combining this approach with a more comprehensive  
342 sampling strategy along the unprepped colon could enhance microbiome-based screening and  
343 treatment modalities for colon disease.

344 **Acknowledgments**

345 We thank all the individuals who volunteered for the study. This work was supported by the Rose  
346 and Lawrence C. Page Foundation (DKT). We would also like to thank Brian Kleiner, Chelsea  
347 Crofoot, and Kirk Herman for their roles in study coordination, subject recruitment, sample  
348 collection and sample processing. Thanks also to GI Fellows Drs Amy Hosmer, Alexander Larson  
349 and Anna Tavakkoli who assisted Dr. Turgeon with the colonoscopies.

## 350 **References**

- 351 1. Yamauchi M, Lochhead P, Morikawa T, Huttenhower C, Chan AT, Giovannucci E, et al.  
352 Colorectal cancer: A tale of two sides or a continuum?: Figure 1. *Gut* [Internet]. *BMJ*;  
353 2012;61:794–7. Available from: <https://doi.org/10.1136/gutjnl-2012-302014>
- 354 2. Forbes JD, Domselaar GV, Bernstein CN. The gut microbiota in immune-mediated inflammatory  
355 diseases. *Frontiers in Microbiology* [Internet]. *Frontiers Media SA*; 2016;7. Available from:  
356 <https://doi.org/10.3389/fmicb.2016.01081>
- 357 3. Halfvarson J, Brislawn CJ, Lamendella R, Vazquez-Baeza Y, Walters WA, Bramer LM, et al.  
358 Dynamics of the human gut microbiome in inflammatory bowel disease. *Nature Microbiology*  
359 [Internet]. *Springer Nature*; 2017;2:17004. Available from: [https://doi.org/10.1038/nmicrobiol.](https://doi.org/10.1038/nmicrobiol.2017.4)  
360 2017.4
- 361 4. Albenberg L, Esipova TV, Judge CP, Bittinger K, Chen J, Laughlin A, et al. Correlation between  
362 intraluminal oxygen gradient and radial partitioning of intestinal microbiota. *Gastroenterology*  
363 [Internet]. *Elsevier BV*; 2014;147:1055–1063.e8. Available from: [https://doi.org/10.1053/j.gastro.](https://doi.org/10.1053/j.gastro.2014.07.020)  
364 2014.07.020
- 365 5. Donaldson GP, Lee SM, Mazmanian SK. Gut biogeography of the bacterial microbiota.  
366 *Nature Reviews Microbiology* [Internet]. *Springer Nature*; 2015;14:20–32. Available from: <https://doi.org/10.1038/nrmicro3552>
- 368 6. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. *Fusobac-*  
369 *terium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune mi-  
370 croenvironment. *Cell Host & Microbe* [Internet]. *Elsevier BV*; 2013;14:207–15. Available from:  
371 <https://doi.org/10.1016%2Fj.chom.2013.07.007>
- 372 7. Dejea CM, Wick EC, Hechenbleikner EM, White JR, Welch JLM, Rossetti BJ, et al. Microbiota  
373 organization is a distinct feature of proximal colorectal cancers. *Proceedings of the National*  
374 *Academy of Sciences* [Internet]. *Proceedings of the National Academy of Sciences*; 2014;111:18321–



- 375 6. Available from: <https://doi.org/10.1073/pnas.1406199111>
- 376 8. McCoy AN, Araújo-Pérez F, Azcárate-Peril A, Yeh JJ, Sandler RS, Keku TO. Fusobacterium  
377 is associated with colorectal adenomas. Goel A, editor. PLoS ONE [Internet]. Public Library  
378 of Science (PLoS); 2013;8:e53653. Available from: [https://doi.org/10.1371/journal.pone.](https://doi.org/10.1371/journal.pone.0053653)  
379 0053653
- 380 9. Baxter NT, Ruffin MT, Rogers MAM, Schloss PD. Microbiota-based model improves the  
381 sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Medicine [Internet].  
382 Springer Nature; 2016;8. Available from: <https://doi.org/10.1186/s13073-016-0290-3>
- 383 10. Strauss J, Kaplan GG, Beck PL, Rioux K, Panaccione R, DeVinney R, et al. Invasive  
384 potential of gut mucosa-derived fusobacterium nucleatum positively correlates with IBD status of  
385 the host. Inflammatory Bowel Diseases [Internet]. Ovid Technologies (Wolters Kluwer Health);  
386 2011;17:1971–8. Available from: <https://doi.org/10.1002/ibd.21606>
- 387 11. Brennan CA, Garrett WS. Gut microbiota, inflammation, and colorectal cancer. Annual  
388 Review of Microbiology [Internet]. Annual Reviews; 2016;70:395–411. Available from: <https://doi.org/10.1146/annurev-micro-102215-095513>
- 390 12. Flemer B, Lynch DB, Brown JMR, Jeffery IB, Ryan FJ, Claesson MJ, et al. Tumour-associated  
391 and non-tumour-associated microbiota in colorectal cancer. Gut [Internet]. BMJ Publishing Group;  
392 2017;66:633–43. Available from: <http://gut.bmj.com/content/66/4/633>
- 393 13. Jalanka J, Salonen A, Salojärvi J, Ritari J, Immonen O, Marciani L, et al. Effects of bowel  
394 cleansing on the intestinal microbiota. Gut [Internet]. BMJ; 2014;64:1562–8. Available from:  
395 <https://doi.org/10.1136/gutjnl-2014-307240>
- 396 14. Harrell L, Wang Y, Antonopoulos D, Young V, Lichtenstein L, Huang Y, et al. Standard  
397 colonic lavage alters the natural state of mucosal-associated microbiota in the human colon. Singh  
398 SR, editor. PLoS ONE [Internet]. Public Library of Science (PLoS); 2012;7:e32545. Available

399 from: <https://doi.org/10.1371/journal.pone.0032545>

400 15. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index  
401 sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq illu-  
402 mina sequencing platform. *Applied and Environmental Microbiology* [Internet]. American Society  
403 for Microbiology; 2013;79:5112–20. Available from: <https://doi.org/10.1128/aem.01043-13>

404 16. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing  
405 mothur: Open-source, platform-independent, community-supported software for describing and  
406 comparing microbial communities. *Applied and Environmental Microbiology* [Internet]. Ameri-  
407 can Society for Microbiology; 2009;75:7537–41. Available from: [https://doi.org/10.1128/aem.](https://doi.org/10.1128/aem.01541-09)  
408 01541-09

409 17. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive bayesian classifier for rapid assignment  
410 of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*  
411 [Internet]. American Society for Microbiology; 2007;73:5261–7. Available from: [https://doi.org/](https://doi.org/10.1128/aem.00062-07)  
412 10.1128/aem.00062-07

413 18. Calle ML, Urrea V, Boulesteix A-L, Malats N. AUC-RF: A new strategy for genomic profiling  
414 with random forest. *Human Heredity* [Internet]. S. Karger AG; 2011;72:121–32. Available from:  
415 <https://doi.org/10.1159/000330778>

416 19. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: An open-source  
417 package for r and s to analyze and compare ROC curves. *BMC Bioinformatics* [Internet]. Springer  
418 Nature; 2011;12:77. Available from: <https://doi.org/10.1186/1471-2105-12-77>

419 20. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. *Genome Medicine*  
420 [Internet]. Springer Nature; 2016;8. Available from: <https://doi.org/10.1186/s13073-016-0307-y>

421 21. Eckburg PB. Diversity of the human intestinal microbial flora. *Science* [Internet]. American  
422 Association for the Advancement of Science (AAAS); 2005;308:1635–8. Available from: <https://doi.org/10.1126/science.1110581>

423 [//doi.org/10.1126/science.1110591](https://doi.org/10.1126/science.1110591)

424 22. Cárcer DA de, Cuív PÓ, Wang T, Kang S, Worthley D, Whitehall V, et al. Numerical ecology  
425 validates a biogeographical distribution and gender-based effect on mucosa-associated bacteria  
426 along the human colon. *The ISME Journal* [Internet]. Springer Nature; 2010;5:801–9. Available  
427 from: <https://doi.org/10.1038/ismej.2010.177>

428 23. Zhang Z, Geng J, Tang X, Fan H, Xu J, Wen X, et al. Spatial heterogeneity and co-occurrence  
429 patterns of human mucosal-associated intestinal microbiota. *The ISME Journal* [Internet]. Springer  
430 Nature; 2013;8:881–93. Available from: <https://doi.org/10.1038/ismej.2013.185>

431 24. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, et al. *Fu-*  
432 *sobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Re-*  
433 *search* [Internet]. Cold Spring Harbor Laboratory Press; 2011;22:299–306. Available from:  
434 <https://doi.org/10.1101/gr.126516.111>

435 25. Lee Y, Eun CS, Lee AR, Park CH, Han DS. *Fusobacterium* isolates recovered from colonic  
436 biopsies of inflammatory bowel disease patients in Korea. *Annals of Laboratory Medicine* [Internet].  
437 Korean Society for Laboratory Medicine (KAMJE); 2016;36:387. Available from: [https://doi.org/](https://doi.org/10.3343/alm.2016.36.4.387)  
438 [10.3343/alm.2016.36.4.387](https://doi.org/10.3343/alm.2016.36.4.387)

439 26. Rougharden JD. Patchiness in the spatial distribution of a population caused by stochastic  
440 fluctuations in resources. *Oikos* [Internet]. JSTOR; 1977;29:52. Available from: [https://doi.org/](https://doi.org/10.2307/3543292)  
441 [10.2307/3543292](https://doi.org/10.2307/3543292)

442 27. Hong P-Y, Croix JA, Greenberg E, Gaskins HR, Mackie RI. Pyrosequencing-based analysis  
443 of the mucosal microbiota in healthy individuals reveals ubiquitous bacterial groups and micro-  
444 heterogeneity. Ahmed N, editor. *PLoS ONE* [Internet]. Public Library of Science (PLoS);  
445 2011;6:e25042. Available from: <https://doi.org/10.1371/journal.pone.0025042>

446 28. Stearns JC, Lynch MDJ, Senadheera DB, Tenenbaum HC, Goldberg MB, Cvitkovitch DG, et  
447 al. Bacterial biogeography of the human digestive tract. *Scientific Reports* [Internet]. Springer

- 448 Nature; 2011;1. Available from: <https://doi.org/10.1038/srep00170>
- 449 29. Sears CL, Garrett WS. Microbes, microbiota, and colon cancer. *Cell Host & Microbe* [Internet].  
450 Elsevier BV; 2014;15:317–28. Available from: <https://doi.org/10.1016/j.chom.2014.02.007>
- 451 30. Mima K, Cao Y, Chan AT, Qian ZR, Nowak JA, Masugi Y, et al. *Fusobacterium nucleatum* in  
452 colorectal carcinoma tissue according to tumor location. *Clinical and Translational Gastroenterology*  
453 [Internet]. Springer Nature; 2016;7:e200. Available from: <https://doi.org/10.1038/ctg.2016.53>
- 454 31. Ohkusa T. Induction of experimental ulcerative colitis by *Fusobacterium varium* isolated from  
455 colonic mucosa of patients with ulcerative colitis. *Gut* [Internet]. BMJ; 2003;52:79–83. Available  
456 from: <https://doi.org/10.1136/gut.52.1.79>
- 457 32. Ohkusa T, Sato N, Ogihara T, Morita K, Ogawa M, Okayasu I. *Fusobacterium varium* localized  
458 in the colonic mucosa of patients with ulcerative colitis stimulates species-specific antibody. *Journal*  
459 *of Gastroenterology and Hepatology* [Internet]. Wiley-Blackwell; 2002;17:849–53. Available from:  
460 <https://doi.org/10.1046/j.1440-1746.2002.02834.x>
- 461 33. Li S, Konstantinov SR, Smits R, Peppelenbosch MP. Bacterial biofilms in colorectal cancer  
462 initiation and progression. *Trends in Molecular Medicine* [Internet]. Elsevier BV; 2017;23:18–30.  
463 Available from: <https://doi.org/10.1016/j.molmed.2016.11.004>
- 464 34. Tjalsma H, Boleij A, Marchesi JR, Dutilh BE. A bacterial driverpassenger model for colorectal  
465 cancer: Beyond the usual suspects. *Nature Reviews Microbiology* [Internet]. Springer Nature;  
466 2012;10:575–82. Available from: <https://doi.org/10.1038/nrmicro2819>
- 467 35. Flynn KJ, Baxter NT, Schloss PD. Metabolic and community synergy of oral bacteria in  
468 colorectal cancer. McMahon K, editor. *mSphere* [Internet]. American Society for Microbiology;  
469 2016;1:e00102–16. Available from: <https://doi.org/10.1128/msphere.00102-16>
- 470 36. Shobar RM, Velineni S, Keshavarzian A, Swanson G, DeMeo MT, Melson JE, et al. The effects  
471 of bowel preparation on microbiota-related metrics differ in health and in inflammatory bowel  
472 disease and for the mucosal and luminal microbiota compartments. *Clinical and Translational*

473 Gastroenterology [Internet]. Springer Nature; 2016;7:e143. Available from: <https://doi.org/10.1038/ctg.2015.54>

## 475 **Figure legends**

### 476 **Figure 1**

477 Sampling strategy. A flexible sigmoidoscope was used to sample the distal colonic luminal contents  
478 and mucosa. The scope was inserted ~ 25cm into the subject and biopsy forceps were used to  
479 sample the luminal contents (D, inset). A separate set of biopsy forceps was used to sample the  
480 distal mucosa (D, inset). The sigmoidoscope was removed. A pediatric colonoscope was inserted  
481 and used to access the proximal colon (P, inset). Biopsies were taken of the proximal luminal  
482 contents and mucosa as described. One week prior to the procedure stool was collected at home  
483 and sent into the laboratory. Representative images from one individual are shown.

### 484 **Figure 2**

485 Phylum-level relative abundance and diversity in the proximal and distal human colon. A) Relative  
486 abundance of the top five bacterial phyla in each sampling site. Each box represents the median  
487 and interquartile range. B) Simpson diversity of the microbial communities at each location. The  
488 horizontal lines represent the median values.

### 489 **Figure 3**

490 Comparison of microbial community structure between sites of the gut.  $\theta_{YC}$  distances are shown  
491 to indicate the interpersonal dissimilarities between two sites – each point represents one individual.  
492 In (A), comparisons of the proximal and distal mucosal and lumen are shown. In (B), comparisons  
493 of each site to the exit stool are shown. In (C), comparisons of samples from all subjects to each  
494 other (interpersonal) or within one subject (intrapersonal) are shown.

### 495 **Figure 4**

496 Random Forest classifies locations in the colon. A) Receiver Operator Characteristic curves are

497 shown for the Random Forest model classifying lumen and mucosal samples for the distal (red)  
498 and proximal (blue) sides of the colon. (B) Receiver Operator Characteristic curves are shown for  
499 the 10-fold cross validation of the Random Forest model classifying distal mucosa vs proximal  
500 mucosa (green) and distal lumen versus proximal lumen (purple).

501 **Figure 5**

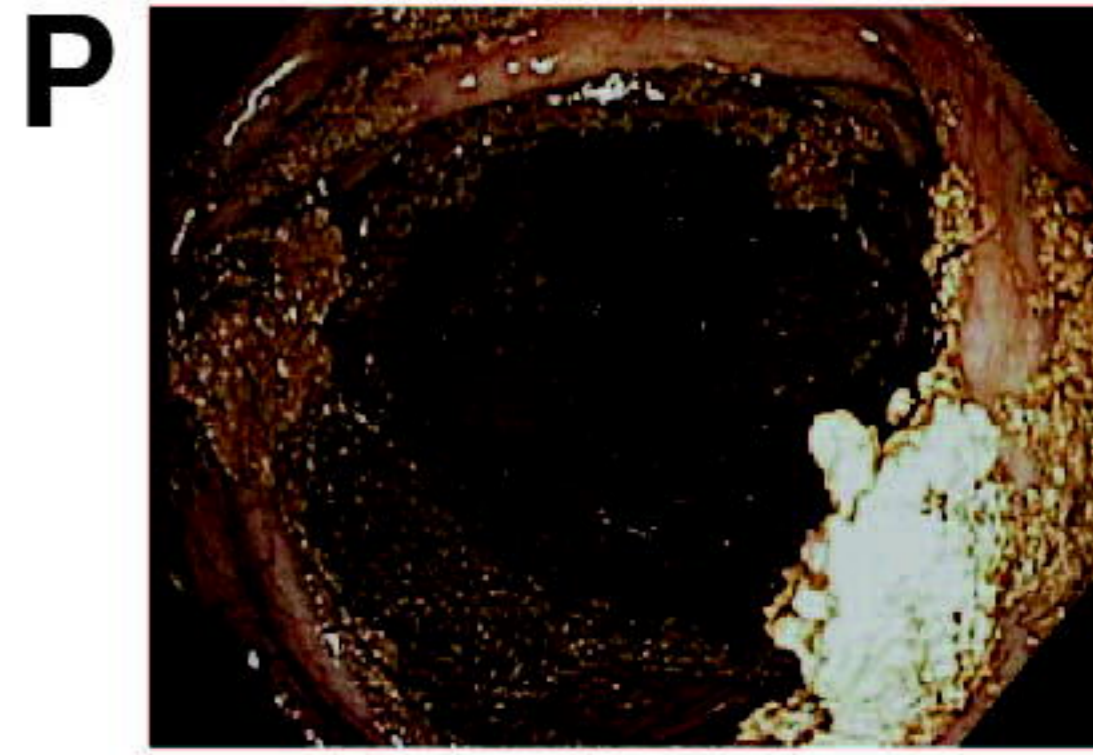
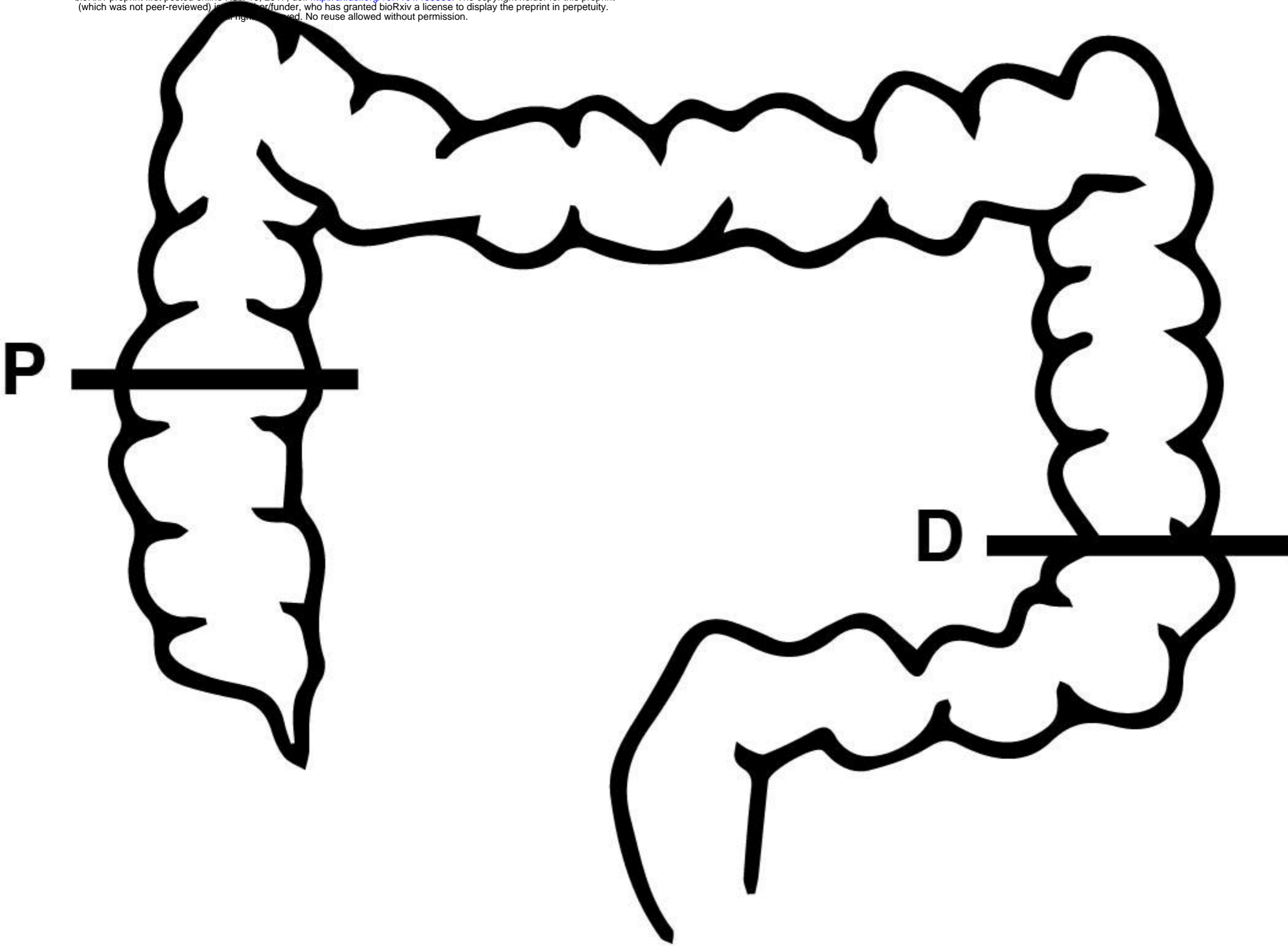
502 Taxa specific to the distal and proximal sides of the colon. Top five OTUs that are most important  
503 for the classification model for the distal mucosa and lumen (A) and the proximal mucosa and  
504 lumen (B). The vertical lines represent the median values for each OTU.

505 **Figure 6**

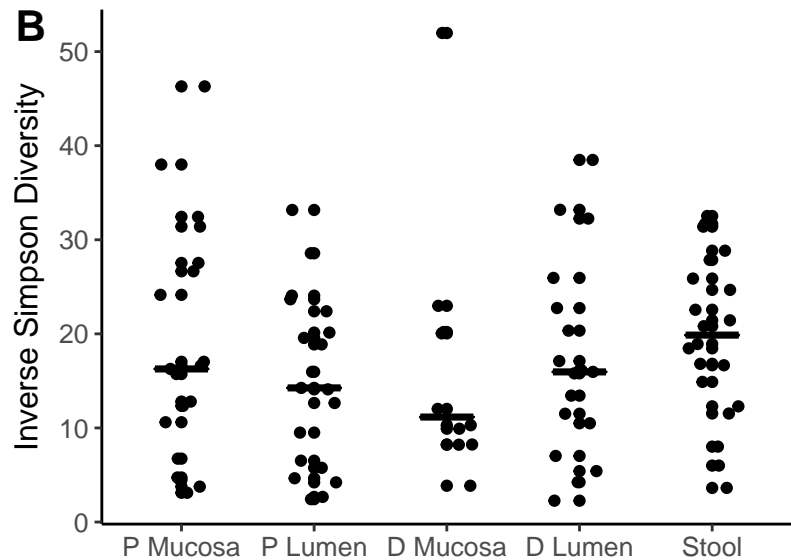
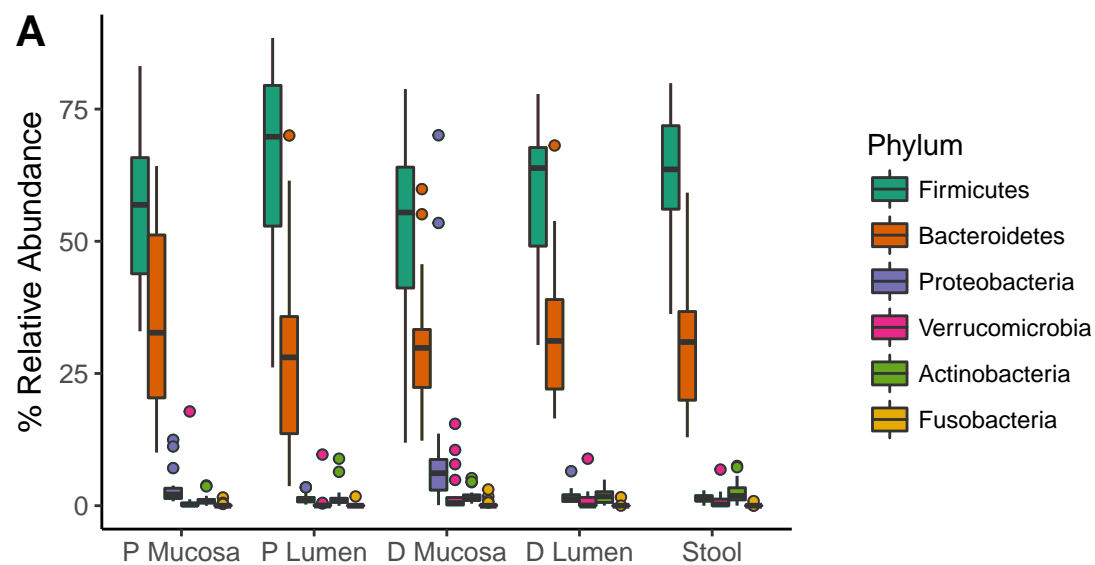
506 Taxa specific to the distal and proximal mucosa and lumen. The five OTUs that were most  
507 important differentiating the distal and proximal mucosa (A) and the distal and proximal lumen  
508 (B). The vertical lines represent the median values for each OTU.

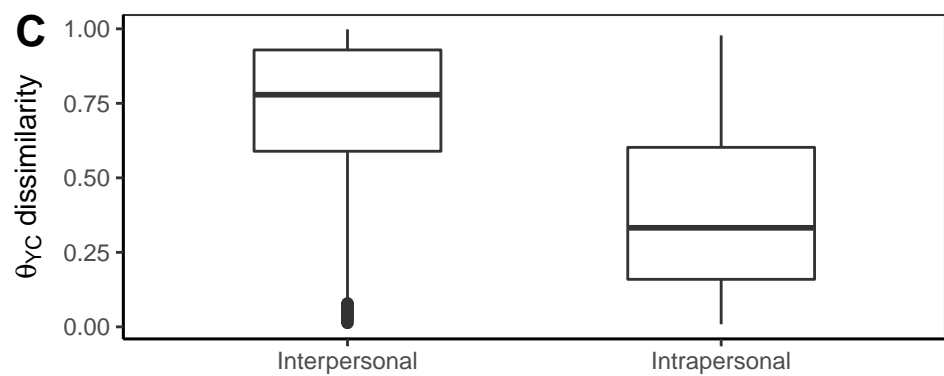
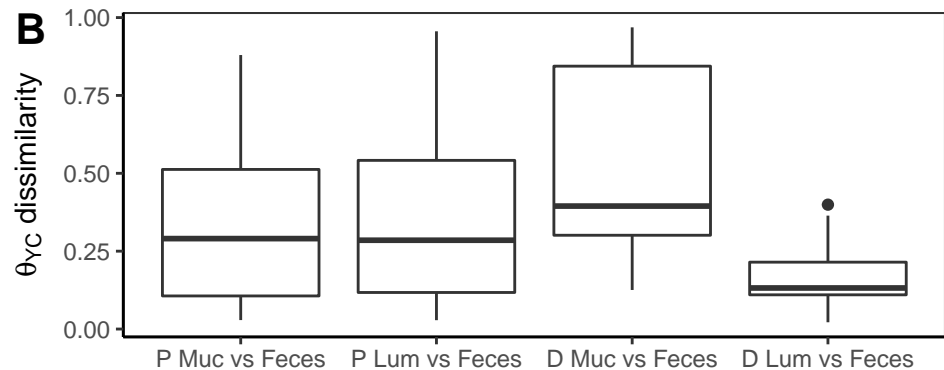
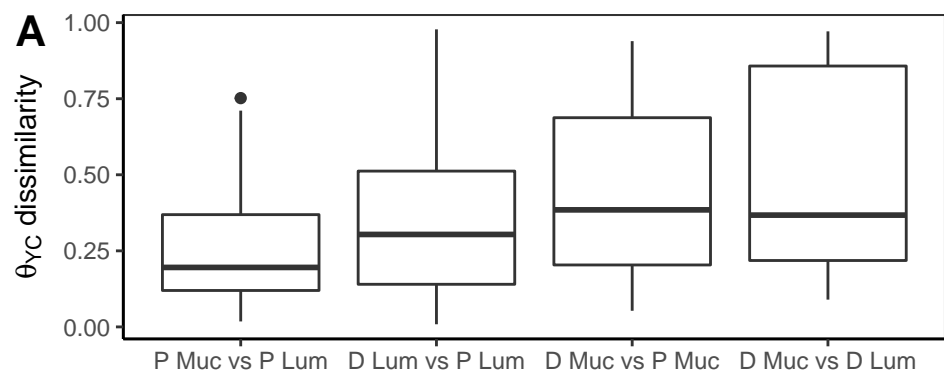
509 **Figure S1**

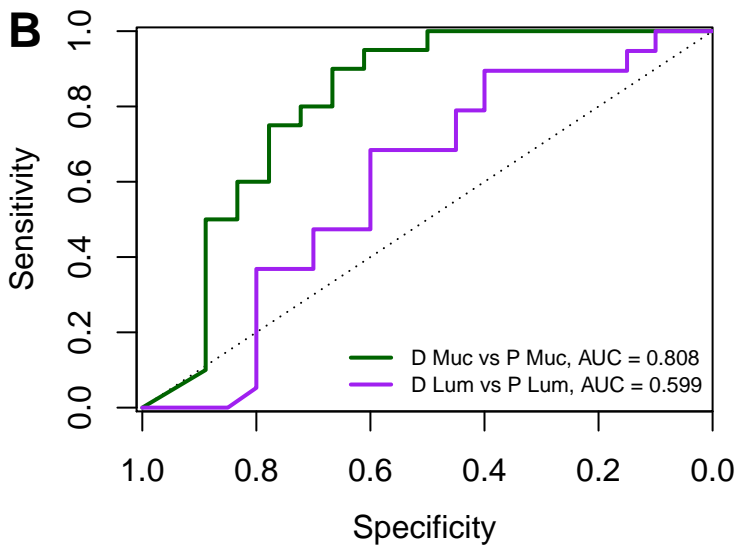
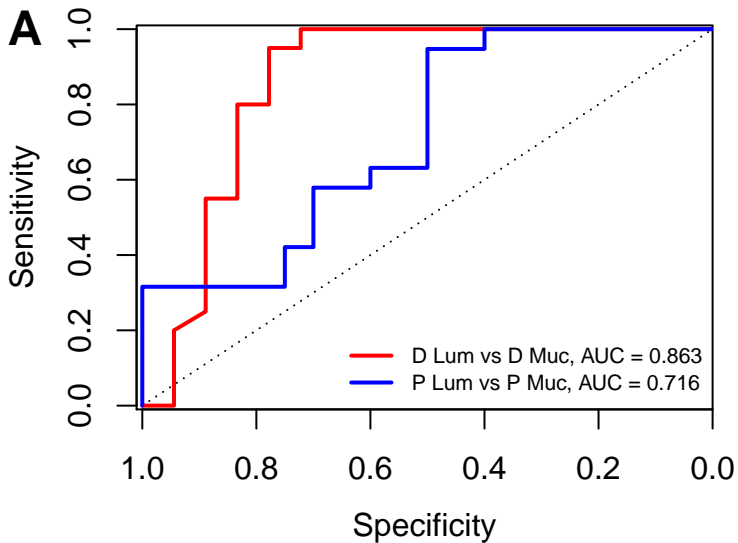
510 Location and relative abundance of cancer-associated OTUs. Relative abundance was calculated  
511 and plotted by sample site for each OTU of interest: (A) *Fusobacterium nucleatum subsp. animalis*  
512 (B) *Fusobacterium varium* and (C) *Porphyromonas asacharolytica*

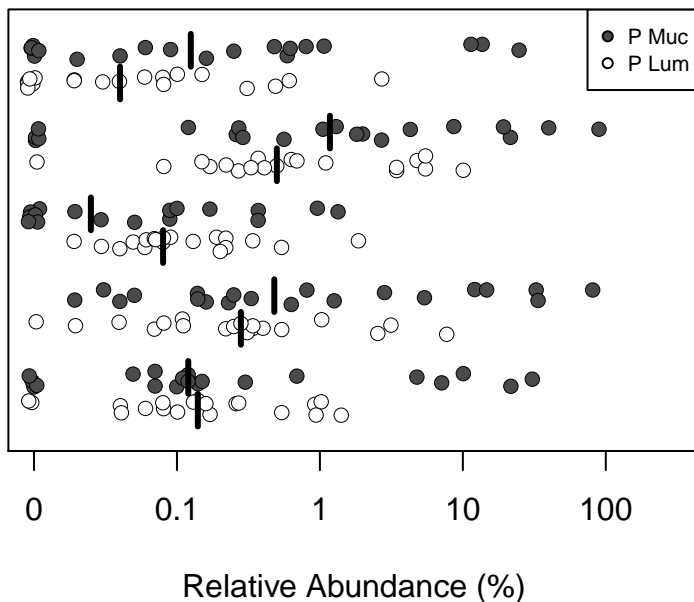
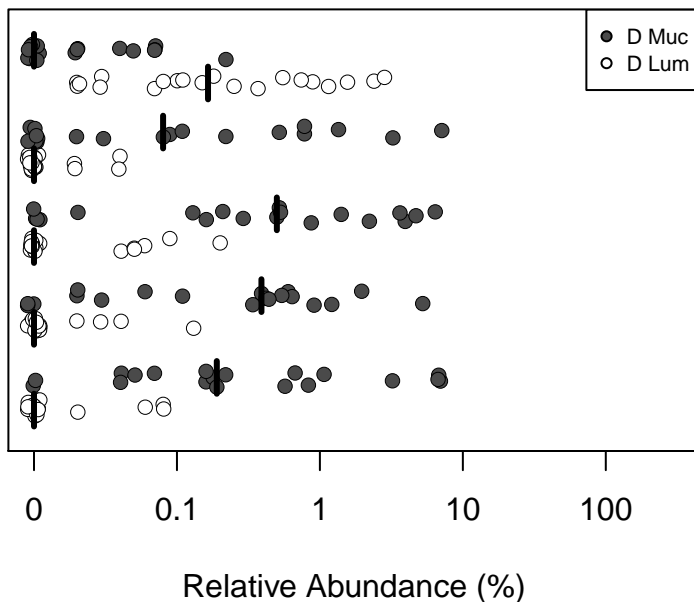


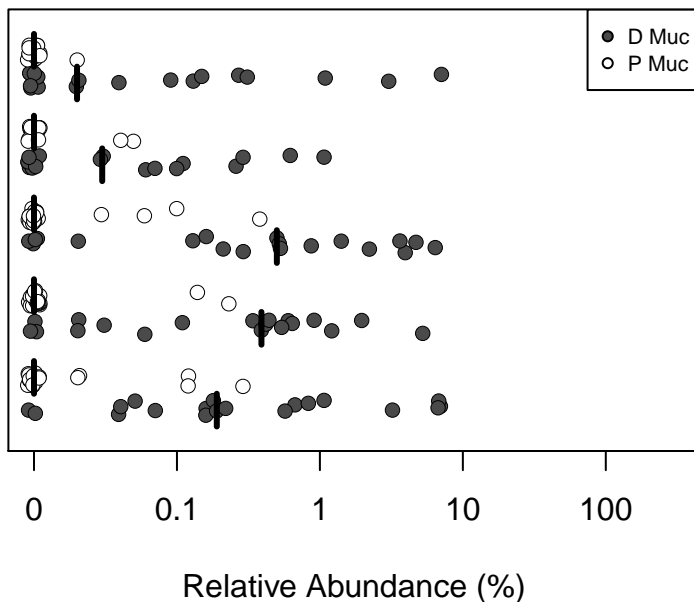








**A***Enterobacteriaceae* (OTU 87)*Bacteroides* (OTU 20)*Actinomyces* (OTU 99)*Enterobacteriaceae* (OTU 38)*Pseudomonas* (OTU 94)**B***Turicibacter* (OTU 138)*Anaerococcus* (OTU 143)*Fingoldia* (OTU 340)*Anaerococcus* (OTU 144)*Peptoniphilus* (OTU 129)

**A***Porphyromonas* (OTU 152)*Murdochiella* (OTU 680)*Finegoldia* (OTU 340)*Anaerococcus* (OTU 144)*Peptoniphilus* (OTU 129)**B***Bacteroides* (OTU 58)*Clostridium\_IV* (OTU 159)*Bacteroides* (OTU 20)*Bacteroides* (OTU 10)*Oscillibacter* (OTU 119)