Journal of Crohn's and Colitis, 2017, 191–203
doi:10.1093/ecco-jcc/jjw136
Advance Access publication July 27, 2016
Original Article

Original Article

Microbial Factors Associated with Postoperative Crohn's Disease Recurrence

Emily K. Wright,[a] Michael A. Kamm,[a] Josef Wagner,[c] Shu-Mei Teo,[b] Peter De Cruz,[a] Amy L. Hamilton,[a] Kathryn J. Ritchie,[a] Michael Inouye,[b,*] Carl D. Kirkwood[c,d,*]

[a]Department of Gastroenterology, St Vincent's Hospital and University of Melbourne, Melbourne, VIC, Australia [b]Centre for Systems Genomics, and Department of Pathology, University of Melbourne, Melbourne, VIC, Australia [c]Enteric Virus Group, Murdoch Children's Research Institute, Parkville, VIC, Australia [d]Department of Microbiology, La Trobe University, Melbourne, VIC, Australia

*Co-senior authors.

Corresponding author: Professor Michael Kamm, MBBS, FRACP, MD, St Vincent's Hospital, Victoria Parade, Fitzroy 3065, Melbourne, VIC, Australia. Tel.: + 61 3 9417 5064; fax: + 61 3 9416 2485; email: mkamm@unimelb.edu.au

This work was presented at the ECCO Congress, Barcelona, in February 2015 and at Digestive Disease Week, Washington, DC, in May 2015.

Abstract

Background and Aims: The intestinal microbiota is a key antigenic driver in Crohn's disease [CD]. We aimed to identify changes in the gut microbiome associated with, and predictive of, disease recurrence and remission.

Methods: A total of 141 mucosal biopsy samples from 34 CD patients were obtained at surgical resection and at colonoscopy 6 and/or 18 months postoperatively; 28 control samples were obtained: 12 from healthy patients [healthy controls] and 16 from hemicolectomy patients [surgical controls]. Bacterial 16S ribosomal profiling was performed using the Illumina MiSeq platform.

Results: CD was associated with reduced alpha diversity when compared with healthy controls but not surgical controls [$p < 0.001$ and $p = 0.666$, respectively]. Beta diversity [composition] differed significantly between CD and both healthy [$p < 0.001$] and surgical [$p = 0.022$] controls, but did not differ significantly between those with and without endoscopic recurrence. There were significant taxonomic differences between recurrence and remission. Patients experiencing recurrence demonstrated elevated *Proteus* genera [$p = 0.008$] and reduced *Faecalibacterium* [$p< 0.001$]. Active smoking was associated with elevated levels of *Proteus* [$p = 0.013$] postoperatively. Low abundance of *Faecalibacterium* [< 0.1%] and detectable *Proteus* in the postoperative ileal mucosa was associated with a higher risk of recurrence (odds ratio [OR] 14 [1.7–110], $p = 0.013$ and 13 [1.1–150], $p = 0.039$, respectively) when corrected for smoking. A model of recurrence comprising the presence of *Proteus,* abundance of *Faecalibacterium,* and smoking status showed moderate accuracy (area under the curve [AUC] 0.740, 95% confidence interval [CI] [0.69–0.79]).

Conclusions: CD is associated with a microbial signature distinct from health. Microbial factors and smoking independently influence postoperative CD recurrence. The genus *Proteus* may play a role in the development of CD.

Key Words: Microbiota; microbiome; inflammatory bowel disease; Proteus

1. Introduction

The pathogenesis of Crohn's disease [CD] involves interactions between the host genome, gastrointestinal [GI] microbiota, and mucosal immune system. Genetic alterations associated with the development of CD appear to lead to perturbation of the delicate relationship between the microbiota and the host, with a consequent dysregulated immune response and intestinal inflammation.[1] The microbiome plays a critical role in this process. [2,3]

Patterns of change in the GI microbiota have begun to emerge in CD patients, including reduction in within-sample biodiversity [α diversity] and decreased representation of several taxa including *Faecalibacterium*, and an increase in the Gammaproteobacteria.[3] Other specific taxonomic shifts have been reported in CD[2]; however, interpretation of these changes across different studies is difficult because of heterogeneity of study populations and methodology and differences in sampling site.

Surgery is required in a majority of CD patients. Following the resection of all macroscopic disease, endoscopically identifiable lesions are apparent within weeks to months,[4], most commonly in the neo-terminal ileum and at the surgical anastomosis. Recurrence occurs more commonly in smokers.[5] The neo-terminal ileum is recolonised by the microbiota postoperatively.[6] CD -specific microbe profiles have identified the ileum as the primary inductive site for all forms of disease.[7] The postoperative setting, with interrogation of the GI microbial community at the neo-terminal ileum, is an ideal clinical environment in which to explore the changes in the GI microbiome with respect to the presence or absence of disease recurrence.

Limited studies have examined the microbial communities postoperatively in patients with CD.[6,8,9,10] Sokol *et al.*[8] found that a lower proportion of *F. prausnitzii* in resected ileal Crohn's mucosa was associated with increased risk of endoscopic recurrence at 6 months [$p = 0.03$], suggesting that there may be a microbial signature, detectable at the time of resection, that is associated with the risk of recurrence postoperatively. These findings were replicated in smaller studies from Dey *et al.*[9] and De Cruz *et al.*[10] Most recently Mondot *et al.*,[11] in a study of 20 patients with Crohn's disease, described changes in the gut mucosal microbiota after ileocaecal resection. 'Bacterial dysbiosis' was associated with postoperative recurrence.

We performed a study in a well-characterised cohort of CD patients, followed from the time of surgery to 18 months postoperatively, to identify 16S microbiota profiles associated with postoperative CD disease recurrence. Here we report our next-generation sequencing analysis on 141 mucosal biopsy samples from 34 CD patients. The combination of our large sample size, longitudinal observations, concurrent sampling from different sites [ileum and anastomosis], and the use of unbiased high-throughput sequencing, are key strengths of this study.

2. Methods

2.1. Subjects and ethics approval

The present study was undertaken in parallel with a clinical study examining the management of Crohn's disease postoperatively (the Postoperative Crohn's Endoscopic Recurrence ['POCER'] study; Clinical Trial Registration: NCT00989560).

The clinical POCER study was a prospective, randomised, controlled trial which aimed to assess the value of postoperative endoscopic assessment and treatment step-up for early mucosal recurrence.[12] Patients were stratified according to risk of recurrence. Smokers, patients with perforating disease, or patients with one or more previous resection were classified as 'high-risk'; all others were 'low-risk'. Patients underwent resection of all macroscopic disease and then received 3 months of metronidazole. High-risk patients also received daily azathioprine [2 mg/kg/day] or mercaptopurine [1.5 mg/kg/day] or adalimumab [160 mg initially, then 80 mg 2 weeks later, then 40 mg every 2 weeks until study conclusion] if thiopurine intolerant.

Patients were randomised to colonoscopy at 6 months [active care] or no colonoscopy [standard care]. For endoscopic recurrence [Rutgeerts score ≥ i2] at 6 months, patients stepped-up to thiopurine, fortnightly adalimumab with thiopurine, or weekly adalimumab. The primary endpoint was endoscopic recurrence at 18 months.

Patients with a family history of bowel cancer, with an intact colon and normal colonoscopy, were recruited as healthy controls. Surgical controls included patients undergoing surveillance colonoscopy who had previously undergone right hemi-colectomy for colonic cancer and were in remission.

Mucosal samples from the CD patients were collected at the time of resection and at colonoscopy 6 and/or 18 months postoperatively, and in controls at the time of screening colonoscopy. Faecal samples were taken at 6 and 18 months postoperatively.

No patient or control received antibiotics or probiotics in the month before the operation or colonoscopy. Patients had intestinal cleansing with polyethylene glycol on the day before surgery, and the same preparation was used in all patients and controls before colonoscopy. Controls were not matched for age, gender, or body mass index (BMI). All patients provided written informed consent.

2.2. Tissue collection and DNA extraction methods

In the CD patients, 5–10 mg tissue samples were obtained from the resection specimen at the affected ileum. At 6- and 18-month colonoscopy, tissue samples [each 5–10 mg] were collected from the anastomosis and neo-terminal ileum. In healthy controls, biopsies were taken from the ileum and in surgical controls, biopsies were taken from the anastomosis and neo-terminal ileum. Standard endoscopic forceps were used and tissue was placed in a sterile tube containing 1 ml RNA later [Ambion], held at 4°C overnight to allow full tissue penetration, then stored at -80°C. Individual tissue samples were subsequently thawed and homogenised using mechanical disruption in lysis buffer as described previously[13,14] and the DNA extracted from the homogenate using the QIAGEN AllPrep Mini Kit as per the manufacturer's instructions. Comparisons within CD patients and between CD and controls were made with respect to changes identified in the ileal mucosa. Unless otherwise stated, results reflect findings from the ileal mucosa.

2.3. Faecal collection

Faecal samples were collected for measurement of faecal calprotectin [FC]. Patients were instructed to collect stool samples no more than 3 days before the study visit, or if colonoscopy was to be performed, 3 days before colonoscopy. Samples were stored at -20 degrees Celsius in the patient's home freezer, transported on ice, and stored at -80 degrees Celsius at study centres until conclusion of the study. All samples were then analysed simultaneously in a central laboratory.

FC was measured by a quantitative enzyme immunoassay [fCAL™, Bühlmann, Schonenbuch, Switzerland] as per manufacturer's instructions. We have previously shown that an FC > 100 µg/g is sensitive for the diagnosis of CD endoscopic recurrence postoperatively[15] and selected this cut-off for analysis.

2.4. 16S rRNA gene sequencing

The bacterial 16S rRNA variable region 2 was amplified by polymerase chain reaction [PCR] with Illumina index/adaptors using the Expand High Fidelity PCR kit [Roche]. PCR products were purified

using the Qiagen DNA extraction kit [Qiagen] and quantified using a Nanodrop before Illumina MiSeq sequencing performed at the Australia Genome Research Facility [AGRF], using 250-cycle chemistry enabling 250bp sequencing from both ends.

2.5. Bioinformatics and statistical analyses

The MiSeq-generated overlapping paired-end sequence reads were first trimmed to 200 bp using FASTX-Toolkit version 0.0.14 [http://hannonlab.cshl.edu/fastx_toolkit/], and then merged using Flash version 1.2.7.[16]

The merged sequences were subsequently processed using QIIME v1.8.[17] Sequences were quality-filtered as follows: ≤ 3 low-quality bp [Phred quality score < 3] allowed before trimming, ≥ 189 consecutive high-quality bp with no uncalled bases [Ns].[18] Chimeras were removed using the UCHIME reference-based method.[19] A total of 2 464 848 sequences were filtered out [7%], leaving 35 034 316 for analysis. Quality-filtered sequences were assigned to operational taxonomic units [OTUs] using the subsampled open reference method[20] with the Greengenes 97% OTU reference set, version 13_5.[21] Briefly, the input sequences were pre-filtered against the reference set at a low percentage identity of 60% to remove sequences that are likely sequencing errors. Next, the closed reference OTU picking was applied on the filtered sequences [closed reference method uses UCLUST[22] to search each read against the database, and assigns the read to an OTU based on the best hit at ≥ 97% sequence identity]. All sequences that did not match the reference at the closed reference step were then de novo clustered at 97% similarity. OTUs with support from only a single read were discarded. Taxonomy was assigned using the RDP classifier[23] with the Greengenes 97% OTU reference set. Alpha diversity measures [number of OTUs, Shannon's Diversity Index, and Chao Diversity Index] were evaluated on 10 independent rarefaction runs per sample. Beta diversity was calculated using the unweighted UniFrac metric.[24]

All subsequent statistical analyses were performed using R, unless stated otherwise. Statistical analyses of alpha diversity were performed using Wilcoxon rank sum test where appropriate. For taxonomic analysis, the relative abundance of each taxon was calculated and summarised at phylum, family, and genus level. Statistical analyses of taxonomic differences were performed using Wilcoxon rank sum test. We acknowledge that performing OTU association analyses does incur a substantial multiple testing burden which we have addressed using a false discovery rate [FDR] adjustment. Statistical significance was defined as a p-value of < 0.05. When an FDR-adjusted p-value was < 0.05, results are referred to in the manuscript as being 'significant following FDR adjustment < 0.05'; otherwise, 'significance' implies a non-FDR adjusted $p < 0.05$. Regression analysis for endoscopic recurrence against smoking was performed using a generalised estimating equation [GEE] with unstructured correlation and robust standard errors to take into account multiple samples from the same patient. Principal coordinates analysis was performed on the unweighted UniFrac distance matrix. Statistical significance of differences in community composition were assessed using the Adonis method, a nonparametric multivariate analysis of variance [ANOVA] using QIIME v1.8[17]. [25] When Adonis was performed between specific groups, correction for multiple testing was performed using FDR adjustment.

2.6. Endoscopic visual assessment

At ileo-colonoscopy, mucosal recurrence at the anastomosis and neo-terminal ileum was assessed according to the Rutgeerts score[4] by the endoscopist. Endoscopic remission was defined as Rutgeerts score i0 [no lesions] or i1 [≤ 5 apthous lesions] and recurrence as i2 [> 5 aphthous lesions or larger lesions confined to the anastomosis], i3 [diffuse ileitis], or i4 [diffuse inflammation with large ulcers and/or narrowing].[4]

Photographs of the anastomosis and neo-terminal ileum were also scored by two investigators [PDC and MAK], blinded to both the endoscopist's score and the patient's identity and treatment. A final consensus score was determined by the two blinded assessors.

3. Results

3.1. The postoperative Crohn's disease cohort

In all, 34 CD patients (14 [41%] male, median age 28 [range 23–43] years) provided a total of 141 mucosal biopsy samples. These samples were taken from the surgical resection specimen [baseline] and from the ileum and anastomosis [biopsies] at colonoscopy 6 and/or 18 months postoperatively; 28 control samples were also obtained, including 12 terminal ileal samples from 12 healthy patients with a normal colon and terminal ileum [healthy controls] and 16 ileal and anastomosis samples from 8 surgical controls. Median time from right hemicolectomy to colonoscopy and tissue collection for the surgical controls was 18 months (interquartile range [IQR] 13–22); all patients were free of colonic cancer at the time of follow-up colonoscopy. Demographic details for CD patients and controls are shown in Table 1.

Of the 34 CD patients, 27 underwent colonoscopy at 6 months; of these, 17 were in endoscopic remission and 10 had disease recurrence. At 18 months, 27 patients underwent colonoscopy, of whom 13 were in endoscopic remission and 14 had disease recurrence [Figure 1].

An average of > 190 000 [range 75020–465400] taxonomy-assigned 16S sequences per sample were obtained using the Illumina MiSeq platform. Read counts were normalised by randomly subsampling each sample to 75 000 reads [rarefaction] before diversity calculations. Taxa were classified as being detected if any reads, regardless of number, were detected. Any taxa that was present in < 10% of all samples were excluded.

p-Value determination was performed both with and without false discovery rate [FDR] adjustment. Unless otherwise stated, p-values reported below are without FDR adjustment.

3.2. Alpha diversity

The alpha diversity analyses are summarised in Table 2 which details time, biopsy location, and time points used for each analysis. CD was associated with a decrease in alpha diversity compared with healthy controls but not with surgical controls when measured by number of OTUs [$p < 0.001$ and $p = 0.713$, respectively], Shannon Diversity Index [$p < 0.001$ and $p = 0.666$, respectively] and Chao Diversity Index [$p < 0.001$ and $p = 0.998$, respectively].

There were no statistically significant differences in alpha diversity in CD patients when ileal biopsies from baseline where compared with postoperative ileal samples at 6 and 18 months. A difference at baseline could not be detected between those who remained in endoscopic remission at 6 or 18 months versus those who went on to develop endoscopic recurrence. There were no statistically significant differences in alpha diversity between those in endoscopic remission versus recurrence [Rutgeerts ≥ i2] at 6 months or 18 months, nor between those with mucosal normality [Rutgeerts i0] versus severe recurrence [Rutgeerts i3 and i4]. Smokers did not differ significantly in alpha diversity compared with non-smokers when ileal samples from CD patients from all time points were compared. As a separate measure of recurrence and remission,

Table 1. Patient demographic data, disease phenotype, postoperative medical therapy, and 6- and 18-month postoperative endoscopic findings.

Demographics	Cases *n* = 34		Controls *n* = 20			
			Normal *n* = 12		Surgical *n* = 8	
	n	%	*n*	%	*n*	%
n [male]	14	41	4	33	5	63
Age, median	28	46	72			
Interquartile range [IQR]	2343	4060	4683			
Active smoker	11	32.35	3	25	1	12.5
Body mass index [BMI] mean [SD]	22 [3.7]					
Age at diagnosis						
A1 ≤ 16 years	4	12				
A2 17–40 Y=years	26	76				
A3 > 40 years	4	12				
Duration of Crohn's disease						
median [IQR]	6 [211]					
> = 10 years	12	35.29				
Disease location at surgery:						
L1 Ileum only	19	56				
L2 Colon only	1	3				
L3 Ileum and colon	14	41				
Disease phenotype at surgery:						
B1 Inflammatory	2	6				
B2 Stricture	8	24				
B3 Penetrating	24	71				
P Perianal Disease	1	3				
Indication for surgery:						
Failure of drug therapy	7	21				
Obstruction	4	12				
Perforation	23	68				
Type of surgical resection						
Ileocaecal resection	29	85				
Ileal resection	1	3				
Ileocaecal with sigmoid resection	4	12				
Postoperative complication	5	15				
Number of previous surgical resections						
0	29	85				
1	4	12				
2	0	0				
3 or more	1	3				
Immediate postoperative baseline drug therapy	*n* = 34					
Metronidazole alone	6	18				
Thiopurine	22	65				
Adalimumab	6	18				
6-month endoscopic outcomes	*n* = 27					
Remission	17	63				
Recurrence	10	37				
18-month endoscopic outcomes	*n* = 27					
Remission	13	48				
Recurrence	14	52				

SD, standard deviation.

FC > 100 μg/g was not associated with a significant change in alpha diversity when compared with FC ≤ 100 μg/g.

Within the CD cohort, there was no statistically significant difference in alpha diversity between samples taken from the neo-terminal ileum when compared with the anastomosis.

Figure 2 and Table 2 show the results of these analyses with the values shown for Shannon Diversity Index. Similar results were obtained using number of OTUs and Chao Diversity Index.

3.3. Beta diversity

3.3.1. Crohn's Disease and health

Microbial composition differed significantly between CD [at baseline] and healthy controls [$p < 0.001$].

In consideration of a postoperative ileal resection and ileocolonic anastomosis, CD patients, at both 6 and 18 months postoperatively, differed significantly from surgical controls [$p = 0.022$ and $p = 0.027$, respectively] Figure 3A.

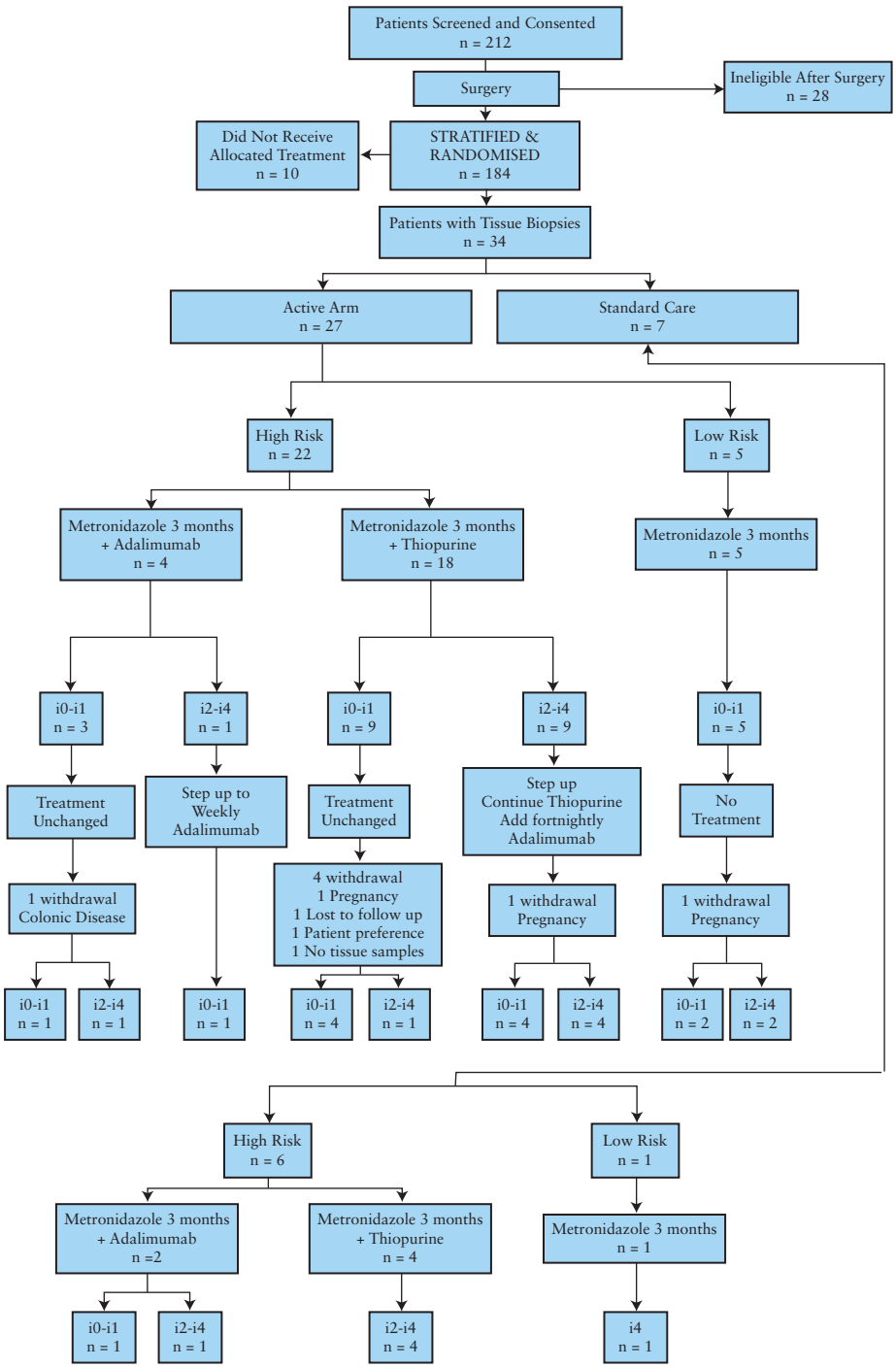

Figure 1. Consort diagram: The POCER study. Details the randomisation, drug treatment and endoscopic outcomes of the 34 Crohn's disease patients who supplied tissue for this study.

3.3.2. Crohn's disease: time and sample site

Microbial composition changed within CD patients over time. Baseline resection samples were statistically significantly different from samples taken at colonoscopy at 6 [$p = 0.005$] and 18 months [$p = 0.001$]. These differences were also significant following FDR adjustment [$p = 0.023$ and 0.007, respectively]. The composition of samples at 6 months were also significantly different from those taken at 18 months [$p = 0.023$], although this was not significant following FDR adjustment [$p = 0.064$] Figure 3B.

Paired samples from the ileum and the anastomosis taken from the same patient at one time point were significantly more similar than samples taken from the same site in different patients at the same time point (mean unweighted Unifrac 0.39 versus 0.63, respectively, $p < 0.001$ [$p = 0.001$ following FDR adjustment]), Figure 3C and D.

3.3.3. Crohn's disease: endoscopic recurrence, faecal calprotectin, and smoking

There was no significant difference in beta diversity when comparing resection samples [baseline] from CD patients who went on to

Table 2. Microbiota diversity [alpha diversity] in relation to Crohn's disease, health, time, smoking status, biopsy location, and endoscopic recurrence.

Comparison	Group 1	*N* =	Group 2	*N*	Median SDI [IQR] Group 1	Median SDI [IQR] Group 2	*p*-Value [Wilcoxon-test]
Crohn's disease versus controls	CD [all samples]	141	Healthy controls [ileum]	12	4.454 [3.611–5.324]	6.382 [5.451–6.630]	< 0.001
	CD [all samples]	141	Surgical controls [all samples]	16	4.454 [3.611–5.324]	4.373 [4.159–5.242]	0.666
Crohn's disease over time	CD [baseline]	34	CD 6 months [ileum]	27	4.455 [3.672–5.220]	4.216 [3.599–5.028]	0.7062
	CD 6 months [ileum]	27	CD 18 months [ileum]	27	4.799 [3.675–5.514]	4.098 [3.624–5.318]	0.390
Smokers versus non-smokers	CD Smokers [all samples]	45	CD non-smokers [all samples]	96	4.098 [3.642–5.318]	4.514 [3.754–5.071]	0.837
Crohn's disease biopsy location	CD [ileum]	54	CD [anastomosis]	53	4.168 [3.746–5.535]	4.456 [3.647–5.059]	0.726
Recurrence and remission	CD baseline samples with recurrence [i2-i4] at 6 months [ileum]	10	CD baseline samples with remission [i0-i1] at 6 months [ileum]	17	4.733 [3.389–5.072]	4.730 [3.568–5.405]	0.604
	CD baseline samples with recurrence [i2-i4] at 18 months [ileum]	14	CD baseline samples with remission [i0-i1] at 18 months [ileum]	13	4.835 [4.422–5.398]	4.598 [3.468–5.749]	0.723
	CD recurrence [i2-i4] at 6 months [ileum]	10	CD remission [i0-i1] at 6 months [ileum]	17	4.458 [3.975–5.556]	3.910 [2.856–5.164]	0.155
	CD recurrence [i2-i4] at 18 months [ileum]	14	CD remission [i0-i1] at 18 months [ileum]	13	4.455 [3.553–4.978]	4.879 [4.100–5.083]	0.402
Severe recurrence and Mucosal normality	CD mucosal normality [i0] at 6 and 18 months [ileum]	13	CD severe recurrence [i3/i4] at 6 and 18 months [ileum]	12	4.476 [3.910–5.535]	4.623 [4.038–5.345]	0.936
FC ≤ 100 versus FC > 100	FC ≤ 100 at 6 and 18 months [ileum]	12	FC > 100 at 6 and 18 months [ileum]	19	3.886 [3.167–4.913]	4.454 [3.854–5.071]	0.326

CD, Crohn's disease; SDI,Shannon Diversity Index; FC, faecal calprotectin.

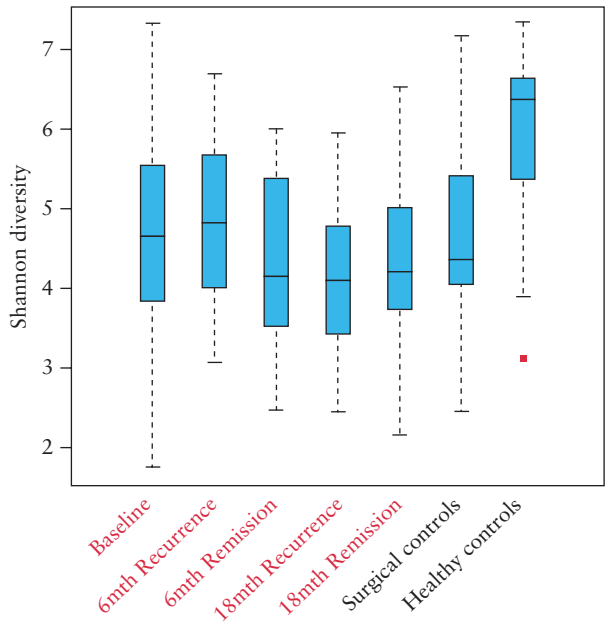

Figure 2. Alpha diversity measured by Shannon diversity index in patients with Crohn's disease (CD) compared to both healthy and surgical controls.

develop endoscopic recurrence and those who remained in endoscopic remission, at 6 or 18 months [p = 0.476 and p = 0.198, respectively].

Beta diversity did not differ significantly at 6 or 18 months [p = 0.926 and p = 0.074, respectively] between CD patients with endoscopic recurrence [i2, i3, and i4] and those in endoscopic remission [i0 and i1].

3.4. Bacteria associated with disease and recurrence

3.4.1. Crohn's disease and health

Compared with healthy controls, CD [baseline] was associated with an alteration in the abundance of several taxa which were significant following FDR adjustment [Table 3].

3.4.2. Crohn's disease: bacteria at resection and subsequent recurrence

At baseline, significant differences in specific taxa were observed between patients who, at 6 months postoperatively, remained in endoscopic remission compared with those who went on to develop endoscopic recurrence. These included significant increases in the Firmicutes phylum [p = 0.018], the Bacteroidaceae and Pasteurellaceae families [p = 0.046 and p = 0.020, respectively] and the *Bacteroides* genus [p = 0.046] in patients with subsequent remission when compared with those with subsequent recurrence. These baseline differences were not observed between patients who remained in remission or had recurrence at 18 months postoperatively. The abundance of *Faecalibacterium* or the presence or absence of *Proteus* at the time of resection were not predictive of later endoscopic remission or recurrence at 6 or 18 months.

3.4.3. Postoperative Crohn's disease over time and endoscopic recurrence

Bacterial composition in CD changed over time, with many taxonomic changes significant following FDR adjustment [Table 4].

At 6 months, endoscopic recurrence was significantly associated with an increased abundance of *Proteus* compared with remission [p = 0.008]. Using logistic regression to correct for smoking status, the detection of *Proteus* at 6 months was associated with a higher risk of recurrence compared with no detectable *Proteus* (odds ratio [OR] 13 [1.1–150], p = 0.039].

At 18 months, endoscopic recurrence was significantly associated with a reduced abundance of *Desulfovibrinaceae* [p = 0.004] and *Ruminococcaceae* [p = 0.014] at a family level and *Faecalibacterium* [p 0.001], *Desulfovibrio* [p = 0.011], and *Bilophila* [p = 0.022] at a genus level. When correcting for smoking status, only a low abundance [< 0.1%] of *Faecalibacterium* was a risk factor for endoscopic recurrence (OR 14 [1.7–110], p = 0.013].

Proteus was detected at some time [baseline, and 6 and 18 months] in 14 CD patients [21 samples]. The median number of reads from each sample when present was low at 35([interquartile range [IQR] 7–82). *Proteus* was not detected in healthy or surgical controls. At 6 months, 5 of 6 patients who had *Proteus* detected had endoscopic recurrence. At 18 months, 1 patient had *Proteus* detected; this patient had endoscopic recurrence. Seven patients had *Proteus* detected at baseline and at either 6 or 18 months; of these, six had endoscopic recurrence including four with severe recurrence [i3 or i4].

At 18 months, severe endoscopic recurrence [Rutgeerts i3 and i4] when compared with complete mucosal normality [i0] was associated with an increase in the phylum Proteobacteria [p= 0.018]; reduced families *Ruminococcaceae* [p = 0.003], *Rikenellaceae* [p = 0.033] and *Turicibacteraceae* [p = 0.33], and reduced genera *Faecalibacterium* [p = 0.005], *Desulfovibrio* [p = 0.009], *Lachnobacterium* [p = 0.009], *Oscillospira* [p = 0.023], *Paraprevotella* [p = 0.033], *Atopobium,* [p = 0.033], *Odoribacter* [p = 0.033], and *Turicibacter* [p = 0.033]. These changes were not observed at 6 months.

3.4.4. Crohn's disease: effect of smoking

Active smoking in this cohort was associated with endoscopic recurrence (OR 3.3 [1–11]. p = 0.049], independent of the presence of *Faecalibacterium* and *Proteus*. Active smoking was associated with significant differences in several taxa compared with non-smokers at 6 and 18 months [Table 5]. At 6 months, *Proteus* was increased [p = 0.037] in smokers when compared with non-smokers.

3.4.5. Crohn's disease: faecal calprotectin

In these 34 CD patients, FC > 100 µg/g was associated with endoscopic recurrence (OR 6.4 [1.02–51.4], p= 0.032]. Median FC for those with recurrence was higher than those with remission at both 6 (241 µg/g [IQR 160–249] versus 92 µg/g [35–190]) and 18 months [243 µg/g [IQR 198–406] versus 72 µg/g [IQR 198–406]]. At 6 months, the abundances of *Pseudomonas* and *Parvimonas* were increased in patients with FC > 100 µg/g [p = 0.030 and p = 0.042. respectively].

3.4.6. *Proteus*, *Faecalibacterium*, smoking and endoscopic recurrence

Of the 24 patients with recurrence at either 6 and/or 18 months, 20 were associated with low *Faecalibacterium* abundance or the presence of *Proteus* [Figure 4].

A microbial profile at 6 or 18 months comprising a high abundance of *Faecalibacterium* and absent *Proteus* had a sensitivity for endoscopic remission of 84%, specificity of 69%, positive predictive value [PPV] of 70%, and negative predictive value [NPV] of 83%.

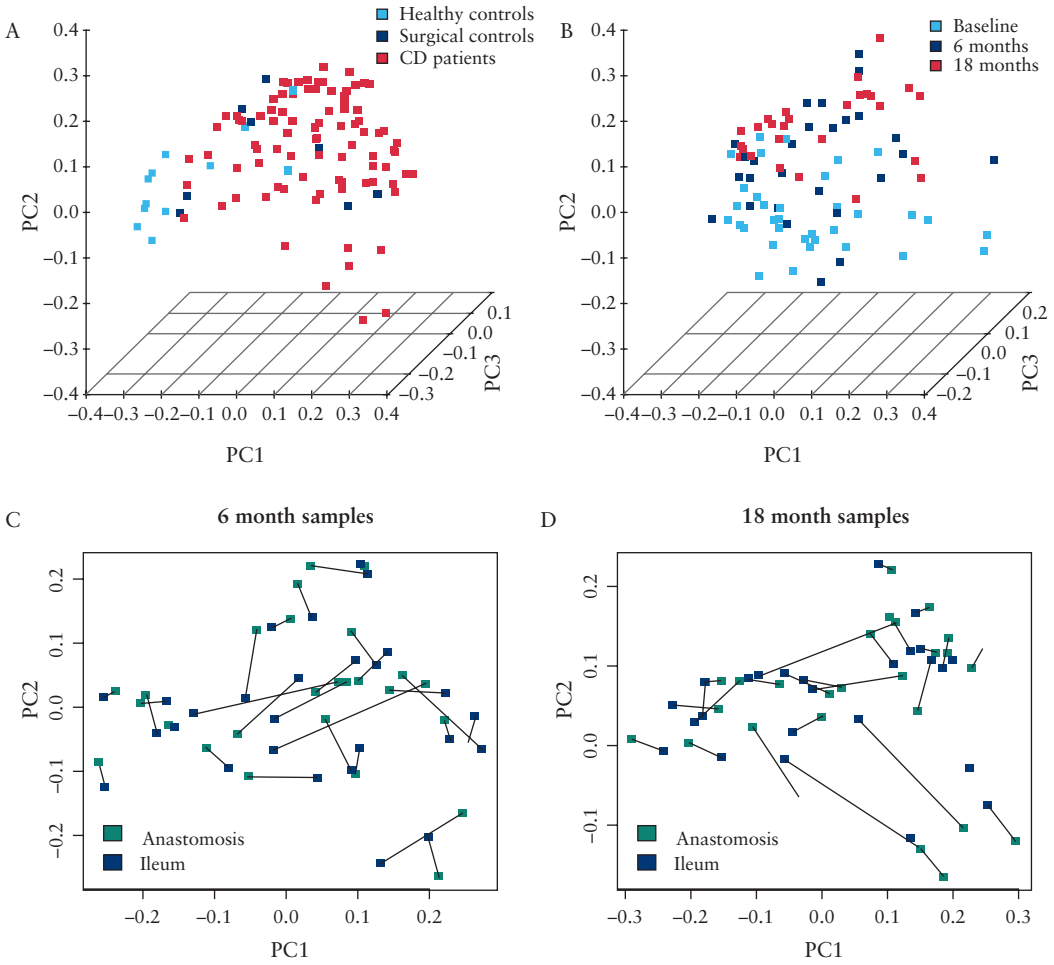

Figure 3. [A] and [B] A principal coordinate plot of the unweighted UniFrac distance with samples coloured according to patient group: Crohn's disease [from baseline, 6 and 18 months], healthy controls and surgical controls [A], and time point for Crohn's disease patients [B]. [C] and [D] A principal coordinate plot of the unweighted UniFrac distance with [C] 6-month samples and [D] 18-month samples, coloured based on the sample site [anastomosis and ileum]. A black line joins samples from the same patient at the same time point.

The accuracy of using microbial analysis of the ileal mucosa, with respect to the presence of *Proteus*, abundance of *Faecalibacterium* [> 0.1%], and smoking status, in the diagnosis of endoscopic recurrence, was modelled using receiver operating characteristic [ROC] curve analysis and yielded a moderate accuracy in predicting endoscopic recurrence (area under the curve [AUC] 0.740, 95% confidence interval [CI] 0.69–0.79).

3.5. Taxonomic identification at a species level

In this study cohort, 867 OTUs were identified as *Faecalibacterium*. Of these, 857 [99%] were identified as *F. prausnitzii* at species level.

Four distinct OTUs [Greengenes OTUs 4440497, 814112, 560629, and a de novo OTU] were identified as belonging to the *Proteus* genus, but these could not be identified at species level. Of these, OTU 814112 was most abundant [65% of all reads] and present, at least in part, in 19 of the 21 samples where *Proteus* was detected. When the representative full-length 16S sequence for this OTU was examined and compared with the National Center for Biotechnology Information [NCBI] BLAST reference database, it was most closely matched to the species *P. mirabilis* [query cover 100%, E value 0.0, identity 95%]. OTU 560629 was also found to be a best match for *P. mirabilis* [query cover 100%, E value 0.0, identity 98%].

4. Discussion

Crohn's disease-specific host and microbe profiles have identified the ileum as the primary inductive site for all disease forms.[7] Postoperative studies of the microbial community at the neo-terminal ileum therefore have the potential to identify organisms of causative importance.

This study has identified significant differences in the ileal microbial profiles in patients with CD compared with healthy controls, in line with previous studies.[2,3] Additional taxa were identified as being associated with CD [Table 3]. This study has employed high-throughput Illumina sequencing and the potential identification of organisms with low abundance.

No significant differences in alpha diversity were observed between patients who went on to develop recurrence compared with those who stayed in remission or between those with postoperative recurrence compared with remission, but numbers were relatively small and analysis may have been underpowered.

Three key findings have emerged from this study. First, low abundance of *Faecalibacterium prausnitzii* and the presence of *Proteus* in the postoperative ileal mucosa are significantly associated with postoperative recurrence, independent of smoking. Second, the progression of recurrent disease, from low-grade mucosal inflammation [identified by an elevated FC] to severe

Table 3. Significant differences in abundance of microbiota taxa in Crohn's disease versus healthy and surgical controls

Taxonomic rank	Taxa	*p*-Value	*p*-Value [with FDR]	Abundance in CD
CD at the time of surgery versus healthy controls				
Phylum	Proteobacteria	< 0.001	0.002	Increased
	Bacteroidetes	0.003	0.007	Decreased
	Actinobacteria	< 0.001	< 0.001	Increased
	Fusobacteria	< 0.001	0.002	Increased
Family	*Enterobacteriaceae*	0.002	0.010	Increased
	Lachnospiraceae	0.051	0.087	Decreased
	Bacteriodaceae	0.015	0.041	Decreased
	Streptococcaceae	< 0.001	0.004	Increased
	Propionibacteriaceae	< 0.001	0.002	Increased
	Clostridiaceae	0.018	0.043	Increased
	Enterococcaceae	< 0.001	0.002	Increased
	Pasturellaceae	0.046	0.083	Increased
	Mirococcaceae	0.016	0.041	Increased
	Psuedomonadaeceae	0.044	0.083	Increased
	Fusobacteriaceae	< 0.001	0.005	Increased
	Moraxellaceae	0.001	0.009	Increased
	Comamonadaceae	< 0.001	0.001	Increased
	Actinomycetaceae	0.003	0.011	Increased
	Lactobacillaceae	< 0.001	0.001	Increased
	Corynebacteriaceae	0.002	0.010	Increased
	Sphingomonadaceae	0.003	0.012	Increased
	Christensenellaceae	< 0.001	0.005	Decreased
	Turicibacteraceae	0.046	0.083	Increased
Genus	*Bacteroides*	0.0154	0.057	Decreased
	Streptococcus	< 0.001	0.005	Increased
	Propionibacterium	< 0.001	0.003	Increased
	Trabulsiella	0.017	0.057	Increased
	Enterococcus	< 0.001	0.003	Increased
	Ruminococcus	0.023	0.079	Decreased
	Haemophilus	0.049	0.110	Increased
	Lachnospira	0.044	0.107	Decreased
	Veillonella	0.015	0.057	Increased
	Rothia	0.011	0.056	Increased
	Pseudomonas	0.044	0.107	Increased
	Fusobacterium	< 0.001	0.006	Increased
	Citrobacter	0.001	0.010	Increased
	Faecalibacterium	0.087	0.150	Decreased
	Butyricimonas	< 0.001	0.006	Decreased
	Odoribacter	< 0.001	0.006	Decreased
	Phascolarctobacterium	0.002	0.014	Decreased
	Bifidobacterium	0.016	0.057	Increased
	Lactobacillus	< 0.001	0.002	Increased
	Actinomyces	0.003	0.015	Increased
	Holdemania	< 0.001	0.003	Decreased
	Corynebacterium	0.002	0.013	Increased
	Lachnobacterium	< 0.001	0.003	Decreased
	Epulopiscium	< 0.001	0.006	Increased
	Turicibacter	0.046	0.109	Increased
	02d06	0.008	0.036	Increased
	cc_115	< 0.001	0.004	Decreased
CD at 6 months postoperatively versus surgical controls				
Family	Clostridiaceae	0.037	0.376	Decreased
	Enterococcaceae	0.042	0.376	Increased
	Prevotellaceae	0.035	0.376	Increased
	Christensellaceae	0.001	0.079	Decreased
Genus	*Trabulsiella*	0.003	0.232	Increased
	Enterococcus	0.043	0.669	Decreased
	Prevotella	0.035	0.669	Decreased

Table 3. Continued

Taxonomic rank	Taxa	*p*-Value	*p*-Value [with FDR]	Abundance in CD
CD at 18 months postoperatively versus surgical controls				
Phylum	Fusobacteria	0.012	0.112	Increased
Family	*Prevotellaceae*	0.022	0.193	Decreased
	Gemellaceae	0.036	0.241	Decreased
	Bifidobacteriaceae	0.042	0.241	Increased
	Christensellaceae	< 0.001	0.026	Decreased
Genus	*Trabulsiella*	0.004	0.198	Increased
	Prevotella	0.022	0.261	Decreased
	Fusobacterium	0.017	0.261	Increased
	Bifidobacterium	0.0418	0.414	Increased
	Holdemania	0.022	0.261	Decreased

CD, Crohn's disease; FDR, false discovery rate.

Table 4. Significant changes in abundance of microbiota taxa in Crohn's disease [CD] over time from the time of resection.

Taxonomic rank	Taxa	*p*-Value	p-Value [with FDR]	Abundance at baseline
CD at the time of resection versus 6 months postoperatively				
Phylum	Bacteroidetes	0.024	0.108	Decreased
	Actinobacteria	0.006	0.055	Increased
Family	*Lachnispiraceae*	0.027	0.134	Decreased
	Bacteroidaceae	0.037	0.142	Decreased
	Streptococcaceae	0.008	0.050	Increased
	Clostridiaecae	0.004	0.030	Increased
	Pseudomonadaceae	< 0.001	0.003	Increased
	Moraxellaceae	0.031	0.139	Increased
	Comamonadaeceae	< 0.001	0.008	Increased
	Bifidobacteriaeceae	0.007	0.089	Increased
	Lactobacillaceae	< 0.001	< 0.001	Increased
	Sphingomonadaceae	0.003	0.027	Increased
	Turicibacteriaceae	0.001	0.013	Increased
Genus	*Bacteroides*	0.037	0.170	Decreased
	Streptococcus	0.011	0.076	Increased
	Haemophilus	0.036	0.170	Increased
	Roseburia	0.037	0.170	Decreased
	Pseudomonas	< 0.001	0.004	Increased
	Clostridium	0.001	0.026	Increased
	Anaerostripes	0.004	0.038	Increased
	Eggerthella	0.045	0.197	Decreased
	Paraprevotella	0.018	0.107	Increased
	Bifidobacterium	0.007	0.060	Increased
	Lactrobacillus	< 0.001	< 0.001	Increased
	Epuloposcium	0.002	0.029	Increased
	Turicibacter	0.001	0.026	Increased
	02d06	0.002	0.029	Increased
Taxonomic rank	Taxa	p-Value	p-Value [with FDR]	Abundance at 6 months
CD postoperatively at 6 months versus 18 months				
Family	*Streptococcaceae*	0.048	0.315	Increased
	Enterococcaceae	0.011	0.113	Increased
Genus	*Enterococcus*	0.011	0.223	Increased
	Eggerthella	0.021	0.296	Increased
	02d06	0.044	0.343	Increased

FDR, false discovery rate.

endoscopic recurrence, is associated with changes in the gut microbial profile over time. Last, we have characterised a model of recurrence that comprises microbial and environmental factors. Although not yet useful clinically, this illustrates the future potential for assessing microbial factors in relation to disease diagnosis or course.

Alterations in the gut microbiota may relate to the resection alone, removal of the ileocaecal valve, or removal of the appendix.

Table 5. Microbiota taxa in smokers and non-smokers with Crohn's disease.

Taxonomic rank	Taxa	*p*-Value	*p*-Value [with FDR]	Abundance in smokers
CD smokers versus non-smokers at 6 months postoperatively				
Phylum	Cyanobacteria	0.036	0.322	Reduced
Family	*Gemellaceae*	0.024	0.458	Reduced
	Moraxellaceae	0.041	0.458	Reduced
Genus	*Clostidium*	0.010	0.594	Reduced
	Proteus	0.037	0.594	Increased
CD smokers versus non-smokers at 18 months postoperatively				
Genus	*Phascolarctobacterium*	0.034	0.759	Reduced

CD, Crohn's disease; FDR, false discovery rate.

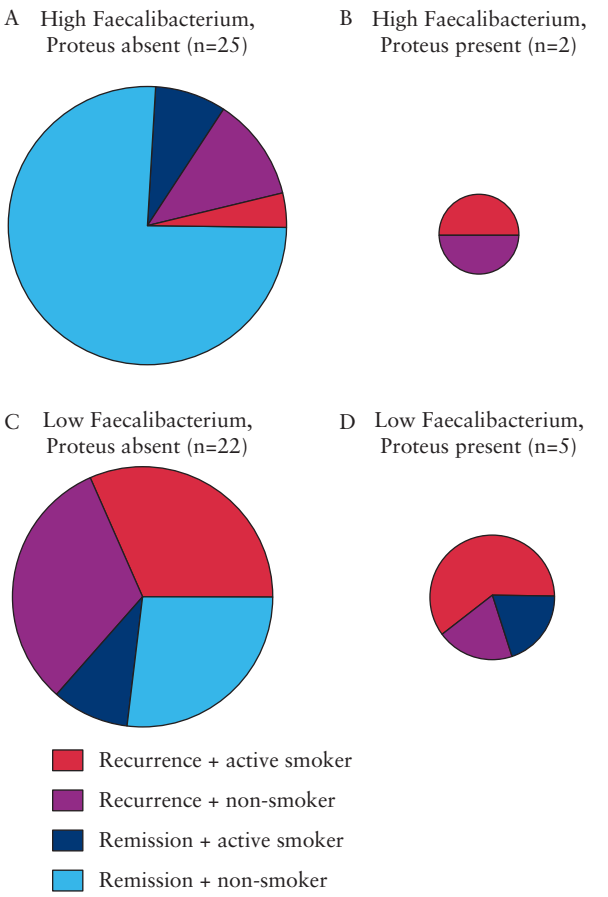

Figure 4. Proportion of patients with endoscopic remission and recurrence depending on the abundance of *Faecalibacterium,* presence or absence of *Proteus,* and smoking status. The size of each pie is proportional to the number of patients with that particular microbial profile. [A] Represents the most desirable microbial profile with high abundance of *Faecalibacterium* and absent *Proteus.* [D] Represents the least desirable microbial profile with low abundance of *Faecalibacterium* and detectable *Proteus.* [B] and [C] reflect microbial profiles with intermediate risk of recurrence.

The appendix may play a role in preserving and protecting beneficial or commensal microorganisms in the gut.[26] In this study, surgical controls were patients who had previously undergone right-hemicolectomy for colon cancer. A change in the microbiota of colon cancer patients has been described,[27,28] including an over-representation of Firmictutes, Fusobacteria and phylotypes closely related to Bacteroides, and a reduction in the abundance of the Proteobacteria, *Faecalibacterium* and *Roseburia*, compared with healthy controls.[28,29] This should be considered when interpreting our results.

Similar to previous studies, we have shown that the abundance of *F. prausnitzii* is associated with endoscopic remission postoperatively.[8,9] In our study, *Proteus spp.* were detected in 12 patients [41%] with CD. *Proteus* was not found in either healthy or surgical controls. When detected at 6 months postoperatively in the ileum, it was strongly associated with recurrence, independent of smoking, increasing the risk of recurrence 13-fold.

Faecalibacterium is a common gut bacterial genus, To test recurrence against low *Faecalibacterium* abundance, a cut-off value of zero reads could not therefore be used. Setting the cut-off at either 0.01% or 0.1% yielded the same results. For *Proteus* genus, however, abundance was very low and in many samples not detected at all. In samples with detectable *Proteus*, the median number of reads from each sample when present was low at 35 (interquartile range [IQR] 782). For this reason, a cut-off of any detectable *Proteus* was chosen.

Proteus spp. is a member of the *Enterobacteriaceae* family and can be found in the normal gastrointestinal flora. Although not traditionally thought to be pathogenic in the gastrointestinal tract, *Proteus spp.* are commonly associated with complicated urinary tract infections.[30]

The *Enterobacteriaceae* family have been found to be consistently increased in the intestinal tissue and faeces of patients with CD when compared with healthy controls,[7,31] although few studies have examined the presence or role of *Proteus* in inflammatory bowel disease. Garrett *et al.*[32] found that the presence of *Proteus mirabilis* in T-bet$^{-/-}$Reg$^{-/-}$ ulcerative colitis [TRUC] mice can elicit colitis in specific-pathogen free mice but not in wild-type mice. In a study of 25 paediatric patients with inflammatory bowel disease, *Proteus* was found in 16.7% of CD patients but was not found in patients with ulcerative colitis, indeterminate colitis, or healthy controls.[33]

Mondot *et al.*[11]have recently shown a relationship between the presence of *Proteus mirabilis*, when detected in the faeces of patients with Crohn's disease undergoing ileocaecal resection, and future endoscopic recurrence. These data concord with our findings of the association between *Proteus* and disease recurrence, in an examination of the mucosal microbiota in a large postoperative cohort of patients with Crohn's disease.

In this study, 88% [21/24] of patients with postoperative disease recurrence could potentially be accounted for by one or more of smoking, low abundance of *Faecalibacterium*, or the presence of *Proteus spp.* At least one of these factors was present in all 12 patients with severe recurrence [Rutgeerts i3 and i4].

Faecal calprotectin > 100 µg/g postoperatively identifies patients likely to have endoscopically identifiable recurrence.[15] *Pseudomonas*

was associated significantly with an FC > 100 μg/g in this study. *Pseudomonas spp*. have been found to be more prevalent in the ileal mucosa of children with CD compared with healthy controls and has been implicated in the pathogenesis of CD.[14] In the current study, although *Pseudomonas* was associated with an elevated FC, it was not more abundant in patients with endoscopic recurrence. In contrast, the detection of *Proteus* and a low abundance of *F. prausnitzii* were associated with endoscopic recurrence, but not increased FC. Only a low abundance of *F. prausnitzii* was associated with severe recurrence and this was observed, along with a significant reduction in beta diversity and other taxonomic differences, only at 18 months postoperatively. These findings suggest that there is an evolution of different stages of recurrent inflammation. These stages may be associated with different microbial profiles.

This study has explored the inter-relationship between different gut bacterial populations and CD recurrence. We have also explored the possible additive effect of the environmental factor, smoking, known to have the greatest effect on the risk of disease recurrence after surgery. Modelling based on data from these real observations may have an important role in identifying a patient's risk of recurrent disease. Further experimental work is also needed to explore the nature of the inter-bacterial relationship. At this stage, a causative relationship has not been established. The diagnostic model proposed in this study requires validation in an independent cohort.

All CD patients in this study were exposed to metronidazole therapy for the first 3 months postoperatively, and some patients received a thiopurine or adalimumab. The detailed assessment of the postoperative ileal microbiome with respect to drug therapy was limited by the heterogeneous treatment regimens used in the clinical study and the relatively small numbers in each drug regimen cohort. No CD patient and no control patient received antibiotics in the 3 months preceding collection of tissue. Of the five patients with reported postoperative complications, three had postoperative ileus, one had an upper gastrointestinal bleed, and one had chest sepsis. The patient with chest sepsis was the only patient to require additional antibiotics. These were given immediately after surgery, but not at the time of later biopsies. Although CD patients received antibiotics and immune-suppressing drugs, the key analysis relates to the relationship between recurrent inflammation [despite previous antibiotics] and the microbiota.

Certain demographic aspects of the different studied patient populations could not be standardised. These included the difference in age, gender, and BMI between controls and CD patients. In addition, it was not possible to standardise diet or account for the wide range in dietary patterns of patients.

In summary, ileocaecal resection may be responsible for some of the bacterial differences observed in patients following intestinal resection in Crohn's disease. Microbial factors, such as the presence of *F. prausnitzii, Proteus spp*., and smoking may influence postoperative Crohn's disease recurrence through independent mechanisms. *Proteus spp*., when detected postoperatively in the neo-terminal ileum even in very low abundance, may play a role in the development of early postoperative recurrence. Quantitative PCR studies of tissue and faeces from our postoperative cohort are under way, which will focus on further defining the presence of *Proteus spp*. Functional studies, including 16s rRNA metagenomic surveys, are also required to elucidate the pathophysiological changes associated with these microbial changes.

Funding

This work was supported by AbbVie, Gutsy Group, Gandel Philanthropy, Angior Foundation, and Crohn's Colitis Australia. The National Health and Medical Research Council [NHMRC] supported EKW, ALH, SMT, PDC, MAK, MI and CDK. MI was supported by Fellowship #1061435, co-funded between the NHMRC and the Australian Heart Foundation.

Conflict of Interest

PDC has received travel grant support from AbbVie and Schering-Plough. MAK has acted as an adviser to AbbVie and Janssen, has received research support from AbbVie, and has acted as a speaker at symposiums sponsored by AbbVie and Janssen. ALH has received an educational grant from AbbVie.

Acknowledgments

We are grateful to colleagues at St Vincent's Hospital Melbourne and the Royal Melbourne Hospital for their participation and cooperation in obtaining surgical tissue and endoscopic biopsies.

Author Contributions

EKW, PDC, MAK: study concept and design; acquisition of data; data interpretation; drafting of the manuscript; critical revision of the manuscript; obtaining funding. JW, MI, SMT, and CDK: acquisition of data; statistical analysis and interpretation of data; critical revision of the manuscript. ALH and KJR: acquisition and monitoring of data.

Supplementary Data

Supplementary data are available at *ECCO-JCC* online.

References

1. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. *Nature* 2011;**474**:307–17.
2. Gevers D, Kugathasan S, Denson LA, *et al*. The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* 2014;**15**:382–92.
3. Morgan XC, Tickle TL, Sokol H, *et al*. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 2012;**13**:R79.
4. Rutgeerts P, Geboes K, Vantrappen G, *et al*. Natural history of recurrent Crohn's disease at the ileocolonic anastomosis after curative surgery. *Gut* 1984;**25**:665–72.
5. Sutherland LR, Ramcharan S, Bryant H, *et al*. Effect of cigarette smoking on recurrence of Crohn's disease. *Gastroenterology* 1990;**98**:1123–8.
6. Neut C, Bulois P, Desreumaux P, *et al*. Changes in the bacterial flora of the neoterminal ileum after ileocolonic resection for Crohn's disease. *Am J Gastroenterol* 2002;**97**:939–46.
7. Haberman Y, Tickle TL, Dexheimer PJ, *et al*. Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *J Clin Invest* 2014;**124**:3617–33.
8. Sokol H, Pigneur B, Watterlot L, *et al*. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* 2008;**105**:16731–6.
9. Dey N, Soergel DA, Repo S, *et al*. Association of gut microbiota with postoperative clinical course in Crohn's disease. *BMC Gastroenterol* 2013;**13**:131.
10. De Cruz P, Kang S, Wagner J, *et al*. Association between specific mucosa-associated microbiota in Crohn's disease at the time of resection and subsequent disease recurrence: A pilot study. *J Gastroenterol Hepatol* 2015;**30**:268–78.
11. Mondot S, Lepage P, Seksik P, *et al*. Structural robustness of the gut mucosal microbiota is associated with Crohn's disease remission after surgery. *Gut* 2016;**65**:954–6.
12. De Cruz P, Kamm MA, Hamilton AL, *et al*. Crohn's disease management after intestinal resection. *Lancet* 2015;**385**:1406–17.

13. Kang S, Denman SE, Morrison M, *et al*. An efficient RNA extraction method for estimating gut microbial diversity by polymerase chain reaction. *Curr Microbiol* 2009;**58**:464–71.
14. Wagner J, Short K, Catto-Smith AG, *et al*. Identification and characterisation of Pseudomonas 16S ribosomal DNA from ileal biopsies of children with Crohn's disease. *PLoS One* 2008;**3**:e3578.
15. Wright EK, Kamm MA, De Cruz P, *et al*. Measurement of fecal calprotectin improves monitoring and detection of recurrence of Crohn's disease after surgery. *Gastroenterology* 2015;**148**:938–97 e1.
16. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 2011;**27**:2957–63.
17. Caporaso JG, Kuczynski J, Stombaugh J, *et al*. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;**7**:335–6.
18. Bokulich NA, Subramanian S, Faith JJ, *et al*. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods* 2013;**10**:57–9.
19. Edgar RC, Hass BJ, Clemente JC, *et al*. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 2011;**27**:2194–200.
20. Rideout JR, He Y, Navas-Molina JA, *et al*. Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ* 2014;**2**:e545.
21. McDonald D, Price MN, Goodrich J, *et al*. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* 2012;**6**:610–8.
22. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;**26**:2460–1.
23. Wang Q, Garrity GM, Tiedje JM, *et al*. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 2007;**73**:5261–7.
24. Lozupone C, Lladser ME, Knights D, *et al*. UniFrac: an effective distance metric for microbial community comparison. *ISME J* 2011;**5**:169–72.
25. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austr Ecology* 2001;**26**:32–46.
26. Randal Bollinger R, Barbas AS, Bush EL, *et al*. Biofilms in the large bowel suggest an apparent function of the human vermiform appendix. *J Theor Biol* 2007;**249**:826–31.
27. Sobhani I, Tap J, Roudot-Thoraval F, *et al*. Microbial dysbiosis in colorectal cancer [CRC] patients. *PLoS One* 2011;**6**:e16393.
28. Wu N, Yang X, Zhang R, *et al*. Dysbiosis signature of fecal microbiota in colorectal cancer patients. *Microb Ecol* 2013;**66**:462–70.
29. Gao Z, Guo B, Gao R, *et al*. Microbiota disbiosis is associated with colorectal cancer. *Front Microbiol* 2015;**6**:20.
30. Coker C, Poore CA, Li X, *et al*. Pathogenesis of Proteus mirabilis urinary tract infection. *Microbes Infect* 2000;**2**:1497–505.
31. Willing BP, Dicksved J, Halfvarson J, *et al*. A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* 2010;**139**:1844–54 e1.
32. Garrett WS, Gallini CA, Yatsunenko T, *et al*. Enterobacteriaceae act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis. *Cell Host Microbe* 2010;**8**:292–300.
33. Conte MP, Schippa S, Zamboni I, *et al*. Gut-associated bacterial microbiota in paediatric patients with inflammatory bowel disease. *Gut* 2006;**55**:1760–7.