ORIGINAL ARTICLE

Spatial variation of the colonic microbiota in patients with ulcerative colitis and control volunteers

A Lavelle,[1,2] G Lennon,[1,2] O O'Sullivan,[3] N Docherty,[4] A Balfe,[1] A Maguire,[2] H E Mulcahy,[2] G Doherty,[2] D O'Donoghue,[2] J Hyland,[2] R P Ross,[3,5] J C Coffey,[6] K Sheahan,[2] P D Cotter,[3,5] F Shanahan,[5] D C Winter,[1,2] P R O'Connell[1,2]

[1]University College Dublin, School of Medicine and Medical Science, Dublin, Ireland
[2]Centre for Colorectal Disease, Saint Vincent's University Hospital, Dublin, Ireland
[3]Teagasc, Food Research Centre, Moorepark, Fermoy, County Cork, Ireland
[4]Department of Physiology, Trinity College Dublin, Dublin, Ireland
[5]Alimentary Pharmabiotic Centre, University College Cork, Cork, Ireland
[6]4i Centre for Interventions in Infection, Inflammation and Immunity, Graduate Entry Medical School, University of Limerick, Limerick, Ireland

Correspondence to
Professor P Ronan O'Connell, School of Medicine and Medical Science, University College Dublin, St Vincent's University Hospital, Elm Park, Dublin 4, Ireland; Ronan.OConnell@ucd.ie

ABSTRACT
Objectives The relevance of spatial composition in the microbial changes associated with UC is unclear. We coupled luminal brush samples, mucosal biopsies and laser capture microdissection with deep sequencing of the gut microbiota to develop an integrated spatial assessment of the microbial community in controls and UC.
Design A total of 98 samples were sequenced to a mean depth of 31 642 reads from nine individuals, four control volunteers undergoing routine colonoscopy and five patients undergoing surgical colectomy for medically-refractory UC. Samples were retrieved at four colorectal locations, incorporating the luminal microbiota, mucus gel layer and whole mucosal biopsies.
Results Interpersonal variability accounted for approximately half of the total variance. Surprisingly, within individuals, asymmetric Eigenvector map analysis demonstrated differentiation between the luminal and mucus gel microbiota, in both controls and UC, with no differentiation between colorectal regions. At a taxonomic level, differentiation was evident between both cohorts, as well as between the luminal and mucosal compartments, with a small group of taxa uniquely discriminating the luminal and mucosal microbiota in colitis. There was no correlation between regional inflammation and a breakdown in this spatial differentiation or bacterial diversity.
Conclusions Our study demonstrates a conserved spatial structure to the colonic microbiota, differentiating the luminal and mucosal communities, within the context of marked interpersonal variability. While elements of this structure overlap between UC and control volunteers, there are differences between the two groups, both in terms of the overall taxonomic composition and how spatial structure is ascribable to distinct taxa.

Significance of this study

What is already known on this subject?
► UC is a condition with a spatial pattern affecting the distal colon predominantly but which can extend proximally to involve the whole colon.
► UC results in mucosal inflammation with changes in the mucus gel layer and, in severe cases, ulceration of the colonic epithelium.
► Changes in the microbiota have been described in both Crohn's disease and UC.
► UC has been associated with increases in both mucosa-associated bacteria and mucolytic species.

What are the new findings?
► By systematically sampling both the luminal and mucosa-associated microbiota across the colon, in UC and control volunteers, this study demonstrates spatial variation between the luminal and mucosal communities in both cohorts.
► While there is overlap between the two cohorts, distinct bacterial groups are associated with this divide in UC.

How might this impact on clinical practice in the foreseeable future?
► The ability to identify bacterial groups that are spatially disturbed in UC may influence the future use of microbial modulation in the treatment of this condition.

INTRODUCTION
While a large body of research has investigated the relationship of diet,[1 2] disease,[3 4] temporal change[5] and host factors[6 7] in relation to the gut microbiota, a detailed picture of the spatial ecology of the large intestine has been difficult to elucidate due to its inaccessibility to direct sampling. Recent studies have noted marked heterogeneity between mucosal sites within the colon with differences between stool samples and mucosal biopsy.[8–10] While relatively homogenous when compared with the entire longitudinal axis of the GI tract, physiological differences are present within the colon, notably luminal changes associated with proximal colonic function as a bioreactor and fermenter of dietary polysaccharides, with corresponding longitudinal gradients in pH, water content, stool consistency and short-chain fatty acids.[11 12] Physiologically, the luminal environment also differs from that of the mucosal surface, which is lined by the mucus gel layer, composed of a sterile

inner layer and a loose outer layer of mucus that is densely colonised with the microbiota.[13] [14] Microbiologically, stool samples provide an unselected representation of microbial ecology that is not suitable for spatial ecological profiling.[14] [15] Finally, spatial patterns are evident in common diseases of the colon, including the distribution of colorectal cancer, UC and Crohn's disease.

We have previously developed a technique for sampling the lumen-associated microbiota using protected specimen brushing in combination with mucosal biopsy and have demonstrated that it differs in composition when compared with the mucosal microbiota using the molecular fingerprinting technique T-RFLP.[16] In the present study, we have combined protected specimen brushing with mucosal biopsy and laser capture microdissection (LCM) in control volunteers undergoing routine colonoscopy and patients with UC undergoing surgical colectomy to sample and deeply sequence the colonic microbiota at four colorectal levels, incorporating both the mucosal and luminal microbiota at each level, using the 16S rRNA marker gene (figure 1). The aims of our study were to determine how colonic ecology varies across multiple loci at a single time point within an individual and whether this spatial structure is conserved between individuals. Second, we wished to determine if the spatial structure of the microbiota, the degree of interpersonal variability or patterns in spatial variation differed between the colons of UC and control volunteers.

MATERIALS AND METHODS

Patient recruitment
Ethical approval was obtained from St. Vincent's University Hospital Ethics and Medical Research Committee. Four control volunteers undergoing routine day case colonoscopy were recruited. Each gave informed, written consent prior to the procedure, was more than 18 years of age and had not taken antibiotics in the previous 3 months. Controls had no evidence of active pathology (table 1). Patients with a history of inflammatory bowel disease (IBD), colon cancer, colonic resection, or hospital admission in the previous 6 weeks, active bleeding or emergency endoscopy were excluded. Bowel preparations used were polyethylene glycol based. Individuals with UC were undergoing total colectomies for either medically-refractory UC or an acute severe flair of UC, unresponsive to intravenous hydrocortisone and biological therapy. All patients undergoing colectomy received a single dose of intravenous antibiotics at induction of anaesthesia but had not received antibiotics in the preceding 3 months.

Sample collection
Sampling levels were the caecum, transverse colon, descending colon and rectum (figure 1). Three samples of the microbiota were obtained at each of these levels: a luminal brush sample, a whole mucosal biopsy and a laser captured sample of the mucus gel layer. This gave a total of 12 samples for each individual.

In control volunteers, the luminal microbiota was sampled with a Microbiological Protected Specimen Brush (Hobbs Medical Inc, Stafford Springs, Connecticut, USA) as previously described.[16] Mucosal biopsies were taken using a Radial Jaw three biopsy forceps (Boston Scientific, Natick, Massachusetts, USA). One biopsy was placed into a sterile, nuclease-free micro-centrifuge tube and frozen on dry ice, while the second was suspended in Tissue-Tek O.C.T. compound within Tissue-Tek Cryomolds (Sakura Fintek USA, Inc, USA) and snap-frozen for LCM.
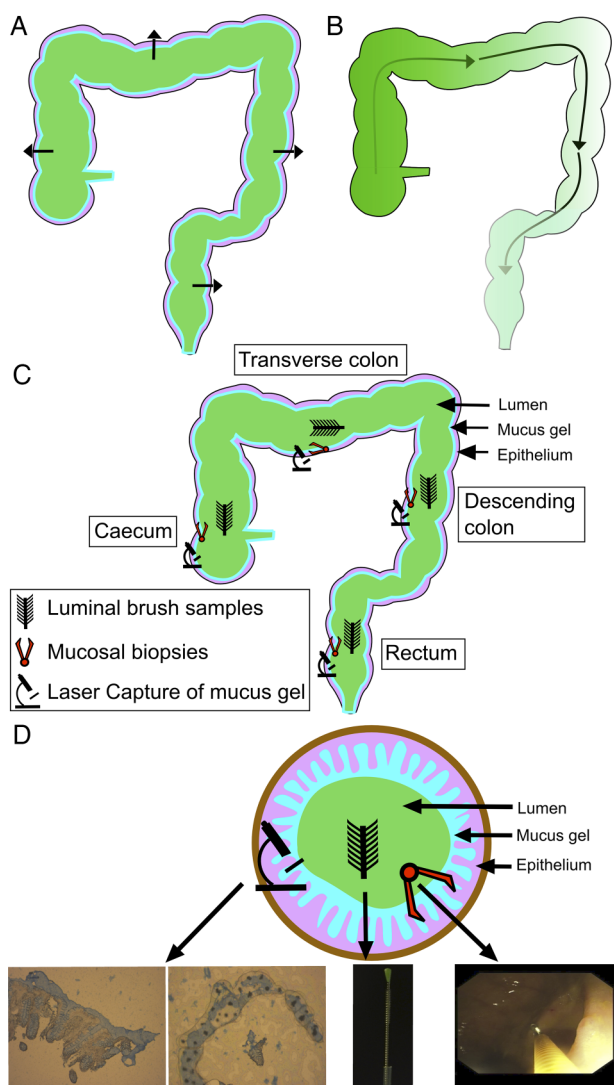

Figure 1 Schematic diagram of potential spatial gradients: (A) cross-sectional gradients and (B) longitudinal gradients. (C) Sampling occurred at four colorectal levels (caecum, transverse colon, descending colon and rectum). Three sampling layers were retrieved at each level (luminal, whole mucosal biopsy and mucus gel layer by laser capture microdissection). (D) A diagram of the cross-sectional organisation of samples with representative pictures from each sampling method.

Table 1 Characteristics of study volunteers

Clinical variable	Control volunteers	UC
Age (mean, range)	37.5 (21–55)	39.5 (23–58)
Gender (M/F)	2/2	3/2
Clinical indication	Microcytic anaemia (2/4) Change in bowel habit (1/4) Abdominal pain (1/4)	Medically-refractory UC (5/5)
Endoscopic findings	Normal	NA
Histological findings	No inflammation	(See table 2)
Medications (number/total)	L-thyroxine (1/4) Contraceptive pill (1/4)	Calcium/vitamin D (4/5) Pantoprazole (2/5) Fluoxetine (1/5)
Immunosuppression (number/total)	None	Corticosteroids (5/5) Aminosalicylates (3/5) Mercaptopurines (4/5) Infliximab (3/5) Adalimumab (1/5)

A third mucosal sample was formalin-fixed and paraffin-embedded for histopathological assessment of inflammation. Eight micro metres sections were mounted onto plain glass slides and stained with Mayer's Haematoxylin followed by Eosin counter staining (Sigma Aldrich, Dublin, Ireland) in a Leica Autostainer XL (Leica Microsystems, Ashbourne, Ireland). All sections were scored by a pathologist according to a system previously described.[17]

In patients undergoing a colectomy, the colon was opened along the antimesenteric border and sampled immediately post-removal. Sterile, nuclease-free cytology brushes were used to sample the luminal microbiota while whole mucosal biopsies were retrieved using sterile tweezers and surgical scissors. Three whole mucosal biopsies were collected at each level and all samples were processed identically to those taken from control volunteers. DNA was extracted using the Qiagen DNA mini kit (Qiagen, Germany) as previously described.[16] Briefly, for brush samples, the plug and the tip of the wire were dissociated using sterile wire cutters and placed in sterile, nuclease-free 1.5 mL micro-centrifuge tubes. DNA was extracted using a Qiagen DNA mini kit (Qiagen, Germany). An amount of 180 µl of buffer ATL was added along with 20 µL of proteinase K to each micro-centrifuge tube containing the sampling brush. This was vortexed vigorously for 1 min to dislodge adherent mucus, followed by pulse centrifugation and incubation at 56°C for 1 h.

Following brush removal, 200 µL of buffer AL was added, pulse vortexed and incubated at 70°C for 10 min. Finally, 100 µL of 100% molecular grade ethanol was added and the mixture loaded onto Qiagen columns (Qiagen, Germany) and processed as per manufacturer's instructions. The final eluate of DNA was in 200 µL of elution buffer.

Biopsy samples were processed in a similar manner using a Qiagen DNA mini kit. Samples were cut using a sterile blade and vigorously vortexed in 180 µL of buffer ATL and 20 µL of proteinase K and processed as described above. An extraction blank was used for each run, which was performed per patient. Extraction blanks were confirmed to be negative following PCR amplification of the V4 region of the 16S rRNA (see below), with all blank amplification values deemed negative if they were the same as non-template controls.

Laser capture microdissection
Snap-frozen biopsy samples were cryosectioned at 12 µm thickness (Leica CM3050 S, Leica Microsystems, Germany) at −25°C. Samples were placed on sterile, nuclease-free ARCTURUS PEN Membrane Slides (Applied Biosystems, Paisley, UK). Fresh blades were used for each sample to prevent carryover of bacterial DNA between samples and every component of the Leica cryosectioner used was removed prior to each sample being sectioned, cleaned and sterilised with 100% molecular grade alcohol followed by 30 min under direct ultraviolet light in a sterile fume hood to prevent contamination with environmental DNA. Following sectioning, samples were immediately fixed in ice-cold 70% molecular grade ethanol solution and stored in airtight containers at −80°C. Samples were transferred directly on dry ice and were removed and stained individually prior to LCM. The staining protocol used Alcian blue (Sigma-Aldrich), to stain the preserved mucous gel layer,[18] which had been filter-sterilised prior to use and a counterstain with Mayer's Haematoxylin (Clin-Tech, UK). Laser capture isolated the closely adherent mucous gel layer and following approximately 80–100 targeted dissections (figure 1D), the caps were removed and the membrane from the cap placed in a sterile micro-

centrifuge tube. DNA was extracted using an identical protocol to that for whole biopsy and brush samples.

Amplicon sequencing and bioinformatics analysis
The V4 region of the 16S SSU rRNA gene was amplified by PCR with universal primers; 520F (5′-AYTGGGYDTAAAG NG-3′) and 926R (5′-CCGTCAATTYYTTTRAGTTT-3′). In all, 25 unique barcoded sequences were used to allow pooling of separate samples on each plate. An amount of 65 ng of template DNA was used in each PCR reaction, along with 200 nM of each primer and 25 µL of BioMix Red (Bioline, London, UK). The PCR reaction conditions were as follows: initialisation and denaturation at 94°C for 2 min, followed by 30 cycles of denaturation (94°C for 1 min), annealing (52°C for 1 min) and extension (72°C for 1 min) and then an elongation step (72°C for 2 min). Sequencing was performed on a Roche 454 GS-FLX using Titanium chemistry by the Teagasc 454Sequencing Platform.[19] Resulting reads were quality trimmed, clustered, aligned and checked for chimeras using the Qiime suite of tools.[20] α And β diversities were estimated using Qiime and a phylogenetic tree was generated using the FastTree package.[21] Principal coordinate analysis (PCoA) and hierarchical clustering based on unweighted Unifrac distances were also performed with Qiime. PCoA plots were visualised with KING.[22] For taxonomic classification, trimmed reads were blasted against the SILVA 16S specific database (V.110). Resulting Blast outputs were parsed using MEGAN[23] with a bit score cut-off of 86.

Statistical analysis
Significant differences in taxonomic abundance between groups were tested for using the Wilcoxon rank sum test. The unweighted UniFrac distance metric was used as an input to perform canonical analysis of principal coordinates (CAP).[24]

To model the distribution of the microbiota within individuals, an asymmetric Eigenvector map (AEM) analysis was performed.[25] This spatial analytic technique was chosen as it explicitly models directional ecological processes, accounting for the faecal stream in the colon.[26] A sampling map (see online supplementary figure S1) was created with a hypothetical upstream point from which a site-by-edges matrix is derived, upon which principal component analysis is performed (see online supplementary figure S1 for workflow and supplementary figure S2 for an illustrative example). The resulting AEM vectors are forward-selected into a redundancy analysis (RDA) to model species distribution. The adjusted R^2 value was used to estimate the percentage of variance in taxonomic distribution explained by the model. To model the entire cohort in the context of the AEM framework, a between-class analysis (BCA) was also performed using Hellinger-transformed family-level abundances.[27]

To determine if taxonomic differences in the microbiota could be used to classify samples into different cohorts, a machine learning algorithm, Random Forests, was applied to family-level abundances.[28] The Spearman rank correlation coefficient was used to assess the correlation between regional inflammation and diversity and regional inflammation and differences in the AEM analysis. A non-parametric Analysis of covariance (ANCOVA) was performed to assess if inflammation differently affected the luminal and mucosal microbiota. All statistical analyses were performed using R and associated packages.[29–33]

RESULTS
Overall, 98 samples from nine individuals were analysed using amplicon sequencing yielding 3 069 310 reads (mean
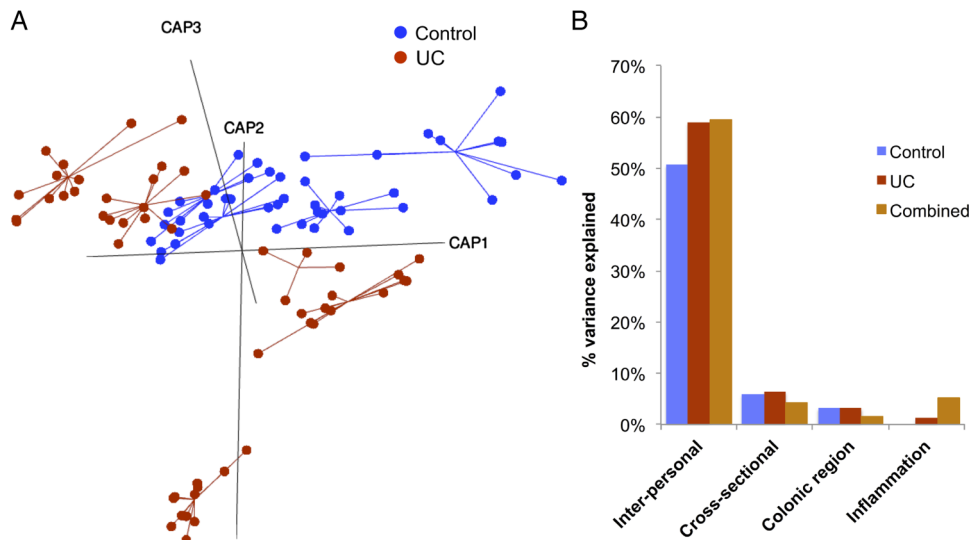
Figure 2 Constrained ordination by canonical analysis of principal coordinates (CAP) using unweighted UniFrac distance of all samples in the cohort (A) demonstrating that interpersonal variability explains the largest proportion of the variance. (B) Percentage of CAP variance described by the constraining variables interpersonal, cross-sectional (luminal vs mucosal), colonic region (caecum, transverse colon, descending colon and rectum) and inflammation.

sequencing depth per sample 31 642 (SD 16 216), online supplementary table S1).

Spatial ecology

To determine if interpersonal variability was a significant component of the variance in our combined data set, CAP of unweighted unifrac distances was performed, demonstrating that the majority of explained variance in the combined cohort was due to interpersonal variability, with a much smaller proportion explained by the other measured spatial variables (figure 2). This was similar to the results of a BCA family-level abundance (see online supplementary figure S3).

To model spatial ecology within individuals and look for conserved patterns of spatial structure, we performed an AEM analysis. A representative graph of sampling locations was created, with luminal samples incorporating a unidirectional flow, from the proximal to distal colon (see online supplementary figure S1). The results of this modelling explained a high proportion of the intrapersonal variance (42%–81%). Fitted RDA values were plotted in the corresponding colonic regions to graphically represent the results (figure 3). In control subjects, segmentation between the luminal and mucus gel microbiota was evident (figure 3A, B).

In UC, four out of five individuals had the required number of samples to perform an AEM analysis (see online supplementary table S1). Inflammatory scores at each level and overall disease activity scores are presented in table 2. In all four of these individuals, the most significant variation also demonstrated segmentation between the luminal and mucus gel microbiota (figure 3C, D).

Taxonomic analysis

To broadly characterise bacterial gradients between the luminal and mucosal environments, the mean abundance of each bacterial family was assessed across the whole cohort in luminal, mucosal and mucus gel samples. Samples were categorised as demonstrating either a high to low gradient from the lumen to mucosa, a high to low gradient from mucosa to lumen or no apparent gradient (see online supplementary figure S4). Those

bacterial families that were significantly different across the three sampling layers by means of a Kruskal–Wallis test were additionally highlighted (see online supplementary table S2).

To determine if differential abundance in bacterial groups accounted for the observed separation between the two cohorts, relative abundances at the family level were compared between control volunteers and patients with colitis. Taxa were compared using the Wilcoxon rank sum test with a significance level of 5% after a Bonferroni correction for multiple comparisons. Samples were then ranked using the Z-score from the Wilcoxon test, from most abundant in controls to most abundant in UC (figure 4). To determine which families were most discriminatory between controls and colitis, we applied the machine-learning algorithm Random Forests to family level abundance data. Using a random sampling of a proportion of the whole cohort as a training set (approximately 40%) and testing the classifier on the remaining samples, the mean misclassification over 10 repeated runs was 1.8%, implying that the algorithm could correctly classify over 98% of samples based solely on family-level abundance as belonging to either controls or UC. Taxa that proved most discriminatory between the two cohorts were ordered according to the mean decrease in accuracy when removed from the analysis and compared, incorporating only those taxa with both discriminatory power and significance in terms of their Wilcoxon rank (figure 4). In controls, Bacteroidaceae and Akkermanseaceae were both significantly more abundant and highly discriminatory, while in UC, Clostridiaceae, Peptostreptococcaceae, Enterobacteriaceae, Ruminococcaceae, Bifidobacteriaceae, Actinomycetaceae and FJ440089, an uncultured member of the Prevotellaceae family, were both significantly more abundant and highly discriminatory (see online supplementary table S3 for family and table S4 for genus).

To determine the families that were differentially abundant between the lumen and mucus gel layer, a BCA was performed, which confirmed differentiation between luminal and mucosal samples in both controls and UC, with separation between the two cohorts (figure 5A). In both cohorts, Coriobacteriaceae, Bacteroideaceae, Ruminococcaceae and Family XIII Incertae
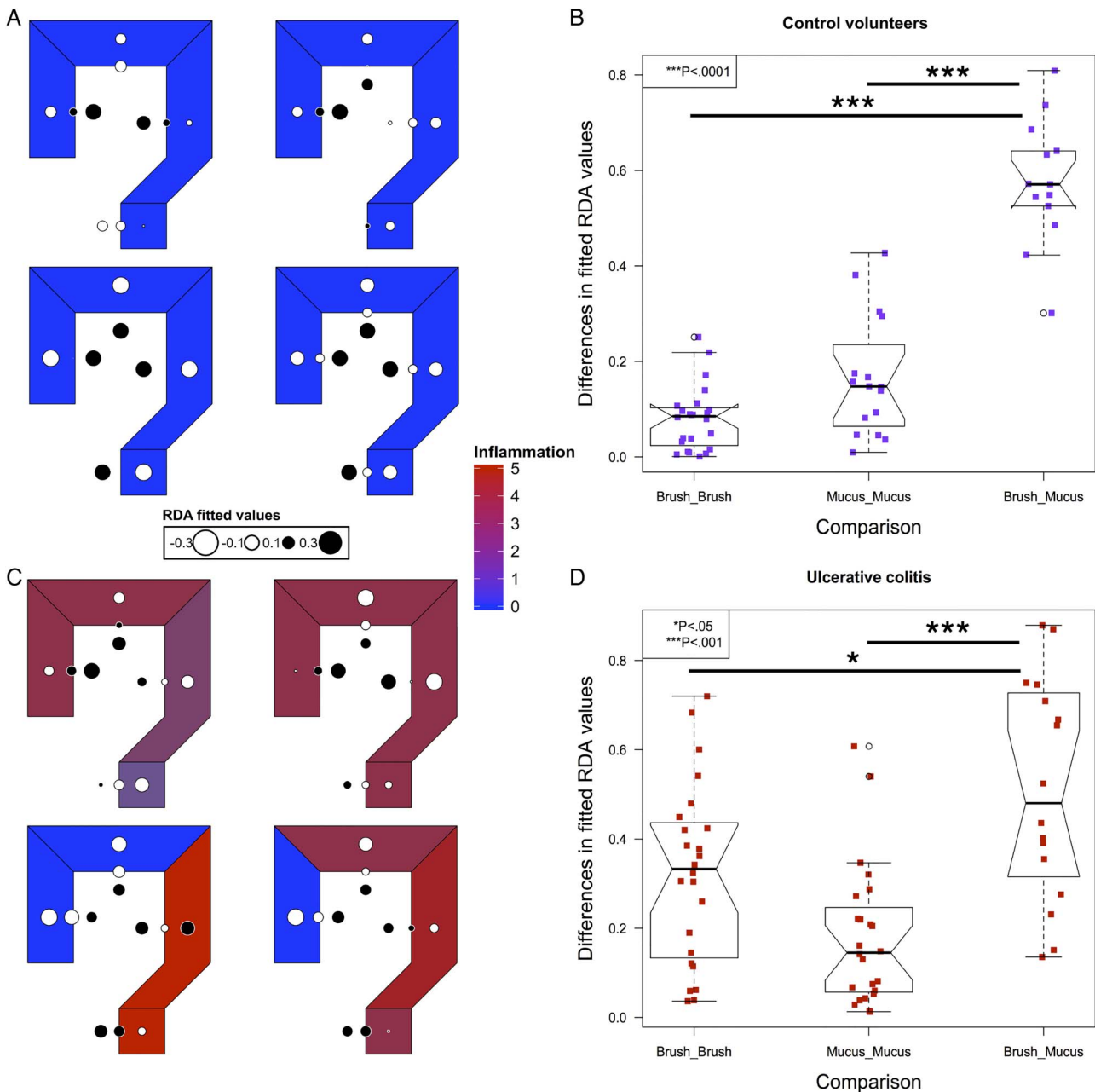
Lavelle A, et al. Gut 2015;**64**:1553–1561. doi:10.1136/gutjnl-2014-307873

Figure 3 Modelling of intrapersonal variability by asymmetric Eigenvector maps analysis in the four control volunteers (A) and four out of five patients with UC (C). Values are centred at 0 with white corresponding to negative, black corresponding to positive and the radius of the circle corresponding to the magnitude of the value. The differences in the fitted redundancy analysis values for all luminal and mucus gel samples within an individual compared with brush versus mucus gel samples at matched locations within individuals, in controls (B) and UC (D), demonstrating that luminal samples are more similar to other luminal samples within individuals and mucus gel samples are more similar to mucus gel samples within individuals than luminal and mucus gel samples at each colorectal location. Significance values: *<0.05, ***<0.001.

Table 2 Inflammatory and disease activity scores in UC

| Individual | Inflammatory scores (0–5)[17] | | | | Mayo score |
	Caecum	Transverse	Descending	Rectum	
A	3	3	2	1	8
B	3	3	3	3	8
C	0	0	5	5	9
D	0	3	4	3	11
E*	3	3	3	5	10

*Not included in AEM analysis.
AEM, asymmetric Eigenvector map.

sedis were differentially abundant between the lumen and mucus gel layer (figure 5B, C). In controls, Lachnospiraceae and Plancomycetaceae were both additionally significant, while in UC, Porphyromonadaceae, Bifidobacteriaceae, Veillonellaceae, Peptostreptococcaceae and Enterobacteriaceae were also differentially abundant (see online supplementary table S5a,b and figure S5). The Random Forests algorithm had a misclassification error for discriminating between the lumen and mucus gel microbiota of 3.45% for controls and 0.88% for UC.

In addition, we applied the Random Forests classifier to both individuals and colorectal regions. While it performed extremely
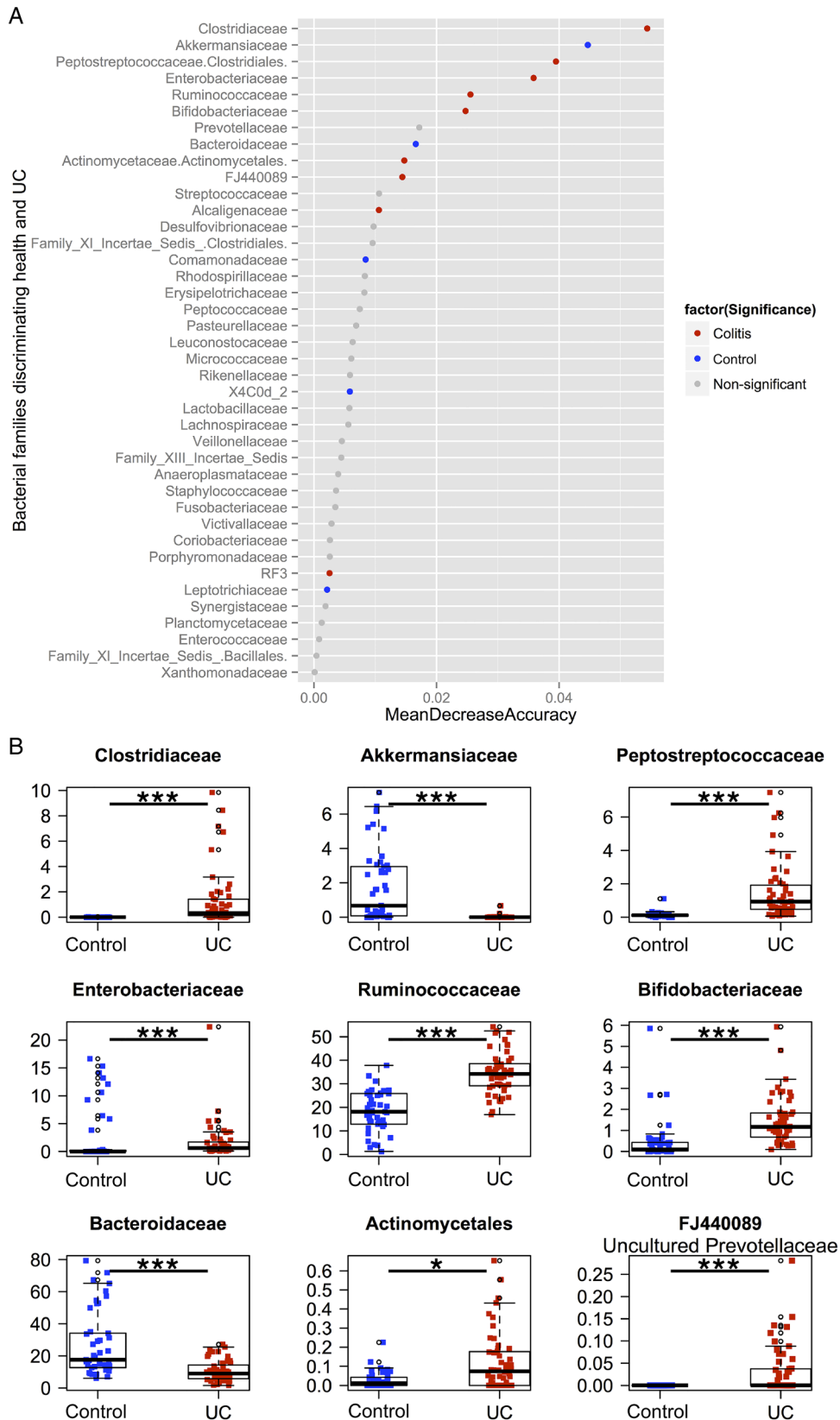
Figure 4 (A) Results from the Random Forests classifier, demonstrating bacterial families that are most discriminatory between the two cohorts in descending order. Samples are coloured by whether they are significantly increased in controls (blue), UC (red) or not (grey) using the Wilcoxon rank test with a Bonferroni correction for multiple comparisons. (B) Boxplots of the relative abundances of the nine bacterial families that were both discriminatory (with a mean decrease in accuracy of the classifier of greater than 0.01 when removed from the analysis) and significantly different in terms of abundance by the Wilcoxon rank test after a Bonferroni correction.
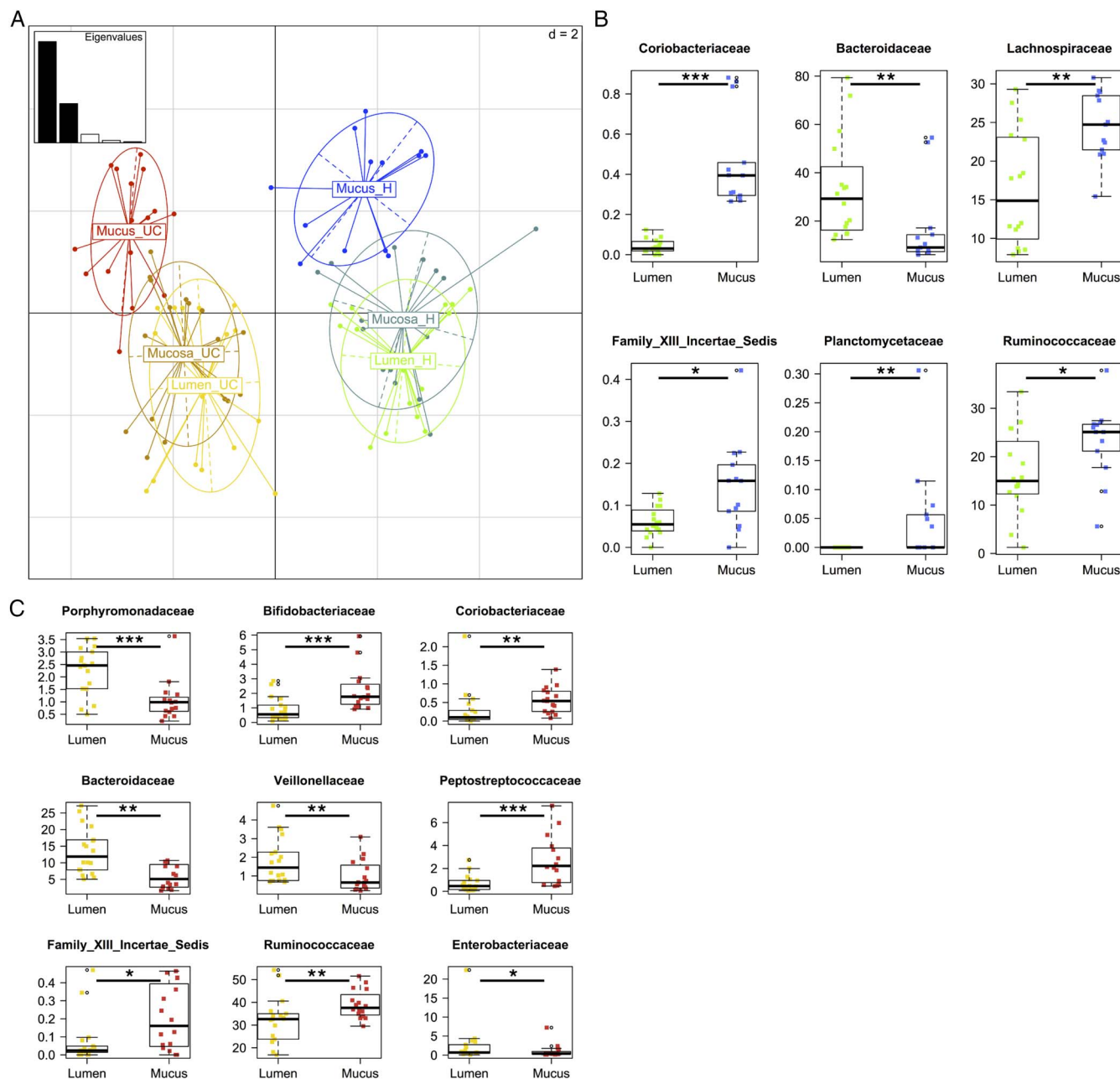
Figure 5 Luminal versus mucus gel samples in controls and UC. A between-class analysis (BCA) based on a principal component analysis of Hellinger-transformed family-level taxon abundance, comparing the luminal, mucosal and mucus gel microbiota of controls and UC (A). Bacterial families that were discriminatory between the lumen and mucus gel in controls (B) and UC (C) based on the Random Forests classifier and abundance using the Wilcoxon rank test. Lachnospiraceae and Planctomycetaceae were unique to controls, while Porphyromonadaceae, Bifidobacteriaceae, Veillonellaceae, Peptostreptococcaceae and Enterobacteriaceae were unique to UC.

well at classifying individuals correctly (a mean out-of-bag error rate of 0.103%), it was unable to distinguish samples from the caecum, transverse colon, descending colon or rectum (mean misclassification error of 96%).

α Diversity and inflammation
Using the Shannon Diversity Index (H), there was no significant difference in diversity between samples from controls and those from UC for any sampling technique (see online supplementary table S6 and figure S6). This was also true for the choa1 estimator, Simpson's index and Faith's phylogenetic diversity (PD) for luminal and mucosal samples. Using two of the four diversity indices, there was however a significantly increased diversity in samples from the mucus gel layer in UC when compared with controls (chao1 estimator: 565.8 (IQR 412.5) in controls vs 1355.4 (IQR 780.8) in UC, p<0.001; PD 27.5 (IQR 15.4) in controls vs 43 (IQR 15.5) in UC, p<0.05).

In controls, there was no significant difference in diversity by any of the four metrics chosen between the different sampling techniques employed (luminal, whole mucosal biopsies or laser-capture dissected mucus gel) (see online supplementary table S6). In UC, there was similarly no difference by Simpson's index and the H, while there was again a significant difference using the chao1 estimator and PD.

To determine if regional inflammation altered the spatial structure of the microbiota, differences in RDA values from the AEM analysis between paired luminal and mucus gel samples were correlated with the degree of local inflammation. This analysis demonstrated no significant correlation between luminal-mucus segregation and the corresponding degree of inflammation (see online supplementary figure S7A). There was similarly no significant correlation between local inflammation and Shannon Diversity, in luminal, mucosal or mucus gel samples (see online supplementary figure S7B and table S7), a finding confirmed by CAP, with other variables factored out, illustrating that the direction of inflammation was orthogonal to that which separated the luminal and mucosal samples (see online supplementary figure S7C). Inflammation appeared to have a proportionally greater effect on the bacterial diversity of luminal samples than those from the mucus gel layer but this did not reach statistical significance when a non-parametric ANCOVA was performed.

DISCUSSION

We show that interpersonal variation is the most important variable in the microbiota composition of the human distal gut when sampled across different locations. As with temporal stability, individuals remain more like themselves than others when sampled at multiple spatial points, indicating that any consideration of spatial or regional variance must be subject to a person's signature microbiota. As such, stool sampling, which clearly demonstrates interpersonal variability, captures this most significant component of variance. This was noted in both our control volunteers and also in patients with severe, medically-refractory UC.

Sampling each individual at multiple locations allowed us to look for conserved patterns of spatial variation, which interpersonal variability would otherwise overshadow, by modelling spatial variation within each individual separately (AEM) and comparing the results across individuals. AEM is a relatively new technique developed to model spatial processes in ecology, such as river networks or ocean currents, where directional flow is a driving force in community assemblage. Eigenvector techniques such as this specifically allow a spatial representation of the sampling grid, so that the model 'knows', for instance, that the caecum and transverse colon are adjacent to one another. More precisely, it creates a set of patterns that could fit the data, such as large-scale longitudinal or cross-sectional patterns, as well as small-scale patterns that represent local differences (see online supplementary figure S2 for illustrative example). Each spatial pattern is treated as a unique variable, assessed individually and as in other multivariate techniques, only those patterns that fit the data are kept. The fitted values of this analysis are presented graphically by circles, with negative values represented as white, positive values as black and the magnitude of each value by the area of the circle. Visually, spatial gradients are thus represented from black to white. Additionally, this method was suitable to our sampling grid as it allowed us to model the direction of flow within the colon. As outlined above, there are physiological, anatomical and pathological influences in our sampling set and we wished to incorporate as much of the captured variance as possible (figure 3). Our findings demonstrate a degree of segregation between the luminal and mucus gel microbiota, with whole mucosal biopsies, the traditional method of sampling the mucosal microbiota, lying in between the two methods. Taxonomic analysis identified a group of bacterial families that accounted for this differentiation in controls, with an additional five families important only in UC (Porphyromonadaceae, Bifidobacteriaceae, Veillonellaceae, Peptostreptococcaceae and Enterobacteriaceae). Interestingly, three of these five families (Bifidobacteriaceae, Peptostreptococcaceae and Enterobacteriaceae) were also among the bacterial families that discriminated between controls and UC (figure 4). The spatial differentiation between lumen and mucus gel in our UC patients thus overlaps with controls while having additional unique components. Due to the large degree of inflammation and ulceration in UC, the LCM samples of adherent microbiota may represent something closer to a biofilm than a healthy mucus gel layer.[34] However, no direct correlation was observed between local inflammation and either the degree of segregation or microbial diversity, suggesting that the alterations observed in UC are more of a field effect and not exclusive to areas of severe inflammation (see online supplementary figure S7).

While our study was not powered to specifically distinguish taxonomic differences between controls and UC, numerous studies have previously looked at this question and our results are similar in terms of the reduction in Akkermansiaceae and the relative increase in members of Clostridiaceae and Enterococcaceae.[4 35–37] The profound decrease in Akkermansiaceae is noteworthy as the bacterium *Akkermansia muciniphila* has recently been shown to protect against obesity and reduce intestinal inflammation while enhancing the colonic barrier and mucus gel layer.[38 39] To attempt to correct for multiple sampling from the same individuals, we used the Random Forests classifier, a cutting-edge machine learning method that has previously been applied to microbiota research.[7] This allowed us to weigh those bacterial families that were most discriminatory between the two cohorts, in order of their predictive power instead of simply relying on statistical significance of raw abundance. The ability of Random Forests to correctly discriminate selected groups mirrored closely the results of our other analyses, with the classifier performing almost perfectly at classifying individuals (interpersonal variability), very well for classifying controls versus UC and lumen versus mucus gel in both controls and UC, while performing extremely poorly at discriminating colorectal regions. As well as internally validating our findings, this allowed us to isolate those bacterial taxa responsible for the patterns observed in our spatial models.

Our study has a number of limitations. First, to sample extensively in control volunteers, we had to recruit from individuals undergoing a colonoscopy who had all taken a bowel preparation preprocedure. Patients undergoing colectomy in our institution do not routinely have bowel preparation preoperatively as has become the standard practice internationally. This difference may account for the reduced diversity seen in our control volunteers, as most research to date has noted a reduced diversity in IBD compared with controls. While this means that diversity cannot account for the differences between the two cohorts, it puts constraints on the direct comparison between the two groups. We feel however that this limitation does not overly effect our conclusions, as many of the taxonomic differences previously described between health and UC are evident within our study group. Also, work on new-onset IBD in paediatric patients suggests that the reduced diversity noted is confined to Crohn's disease and is not a feature of UC, at least in treatment naive patients.[40] Finally, the numbers of patients in our study are small; however, to extensively sample and deeply sequence each sample, we wished to focus our efforts on a limited number. Reducing the number of sites would have reduced our ability to model the spatial variance completely, while reducing the sequencing depth would have reduced the taxonomic resolution and thus the robustness of our findings.

CONCLUSIONS

In our study, approximately half of the variance captured in the colonic microbiota was explained by interpersonal variability. Within individuals, differences were apparent between the luminal and mucosal compartments, while there was no evidence of longitudinal patterns in bacterial composition. At a taxonomic level, distinct bacterial taxa accounted for the differences between control volunteers and UC, as well as luminal and mucosal communities in both cohorts. While there was an overlap between the bacterial taxa that discriminated the luminal and mucosal communities in controls and UC, unique taxa had a spatial differentiation in UC.

REFERENCES

1 Devkota S, Wang Y, Musch MW, *et al*. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice. *Nature* 2012;487:104–8.
2 Claesson MJ, Jeffery IB, Conde S, *et al*. Gut microbiota composition correlates with diet and health in the elderly. *Nature* 2012;488:178–84.
3 Greenblum S, Turnbaugh PJ, Borenstein E. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proc Natl Acad Sci USA* 2012;109:594–9.
4 Morgan XC, Tickle TL, Sokol H, *et al*. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 2012;13:R79.
5 Claesson MJ, Cusack Sn, O'Sullivan O, *et al*. Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc Natl Acad Sci USA* 2011;108(Suppl 1):4586–91.
6 Ley RE, Backhed F, Turnbaugh P, *et al*. Obesity alters gut microbial ecology. *Proc Natl Acad Sci USA* 2005;102:11070–5.
7 Yatsunenko T, Rey FE, Manary MJ, *et al*. Human gut microbiome viewed across age and geography. *Nature* 2012;486:222–7.
8 Hong PY, Croix JA, Greenberg E, *et al*. Pyrosequencing-based analysis of the mucosal microbiota in healthy individuals reveals ubiquitous bacterial groups and micro-heterogeneity. *PLoS ONE* 2011;6:e25042.
9 Zhang Z, Geng J, Tang X, *et al*. Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. *ISME J* 2014;8:881–93.
10 Stearns JC, Lynch MD, Senadheera DB, *et al*. Bacterial biogeography of the human digestive tract. *Sci Rep* 2011;1:170.
11 Macfarlane GT, Gibson GR, Cummings JH. Comparison of fermentation reactions in different regions of the human colon. *J Appl Bacteriol* 1992;72:57–64.
12 Bown RL, Gibson JA, Sladen GE, *et al*. Effects of lactulose and other laxatives on ileal and colonic pH as measured by a radiotelemetry device. *Gut* 1974;15:999–1004.
13 Johansson MEV, Larsson JMHn, Hansson GC. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc Natl Acad Sci USA* 2011;108:4659–65.
14 Zoetendal E, Von Wright A, Vilpponen-Salmela T, *et al*. Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Appl Environ Microbiol* 2002;68:3401.
15 Eckburg PB, Bik EM, Bernstein CN, *et al*. Diversity of the human intestinal microbial flora. *Science* 2005;308:1635.
16 Lavelle A, Lennon G, Docherty N, *et al*. Depth-dependent differences in community structure of the human colonic microbiota in health. *PLoS ONE* 2013;8: e78835.
17 Geboes K, Riddell R, Ost A, *et al*. A reproducible grading scale for histological assessment of inflammation in ulcerative colitis. *Gut* 2000;47:404–9.
18 Wang Y, Antonopoulos DA, Zhu X, *et al*. Laser capture microdissection and metagenomic analysis of intact mucosa-associated microbial communities of human colon. *Appl Microbiol Biotechnol* 2010;88:1333–42.
19 Claesson MJ, O'Sullivan O, Wang Q, *et al*. Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLoS ONE* 2009;4:e6669.
20 Caporaso JG, Kuczynski J, Stombaugh J, *et al*. QIIME allows analysis of high-throughput community sequencing data. *Nat Meth* 2010;7:335–6.
21 Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* 2010;5:e9490.
22 Chen VB, Davis IW, Richardson DC. KING (Kinemage, Next Generation): a versatile interactive molecular and scientific visualization program. *Protein Sci* 2009;18:2403–9.
23 Huson DH, Mitra S, Ruscheweyh HJ, *et al*. Integrative analysis of environmental sequences using MEGAN 4. *Genome Res* 2011;21:1552–60.
24 Anderson MJ, Willis TJ. Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology* 2003;84:511–25.
25 Blanchet FG, Legendre P, Borcard D. Modelling directional spatial processes in ecological data. *Ecol Model* 2008;215:325–36.
26 Blanchet FG, Legendre P, Maranger R, *et al*. Modelling the effect of directional spatial ecological processes at different scales. *Oecologia* 2011;166:357–68.
27 Arumugam M, Raes J, Pelletier E, *et al*. Enterotypes of the human gut microbiome. *Nature* 2011;473:174–80.
28 Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
29 R: A language and environment for statistical computing.: R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org/. R Development Core Team (2011).
30 Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer, 2009.
31 Oksanen J, Guillaume Blanchet F, Kindt R, *et al*. vegan: Community Ecology Package. R package version 20-2. 2011. http://CRANR-projectorg/package=vegan.
32 Blanchet FG. AEM: Tools to construct asymmetric eigenvector maps (AEM) spatial variables. R package version 02-5 2009. http://r-forge.r-project.org/R/?group_id=195
33 S D. spacemakeR: Spatial modelling. R package version 00-3/r49 2008. http://r-forge.r-project.org/R/?group_id=195
34 Swidsinski A, Loening-Baucke V, Theissig F, *et al*. Comparative study of the intestinal mucus barrier in normal and inflamed colon. *Gut* 2007;56:343–50.
35 Rajilic-Stojanovic M, Shanahan F, Guarner F, *et al*. Phylogenetic analysis of dysbiosis in ulcerative colitis during remission. *Inflamm Bowel Dis* 2013;19:481–8.
36 Vigsnæs LK, Brynskov J, Steenholdt C, *et al*. Gram-negative bacteria account for main differences between faecal microbiota from patients with ulcerative colitis and healthy controls. *Beneficial Microbes* 2012;3:287–97.
37 Negroni A, Costanzo M, Vitali R, *et al*. Characterization of adherent-invasive Escherichia coliisolated from pediatric patients with inflammatory bowel disease. *Inflamm Bowel Dis* 2012;18:913–24.
38 Everard A, Belzer C, Geurts L, *et al*. Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. *Proc Natl Acad Sci USA* 2013;110:9066–71.
39 Png CW, Linden SK, Gilshenan KS, *et al*. Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. *Am J Gastroenterol* 2010;105:2420–8.
40 Hansen R, Russell RK, Reiff C, *et al*. Microbiota of de-novo pediatric IBD: increased faecalibacterium prausnitzii and reduced bacterial diversity in Crohn's but not in ulcerative colitis. *Am J Gastroenterol* 2012;107:1913–22.