

## ORIGINAL ARTICLE

# Numerical ecology validates a biogeographical distribution and gender-based effect on mucosa-associated bacteria along the human colon

Daniel Aguirre de Cárcer<sup>1</sup>, Páraic Ó Cuív<sup>1</sup>, Tingting Wang<sup>1,2</sup>, Seungha Kang<sup>1</sup>, Daniel Worthley<sup>3</sup>, Vicki Whitehall<sup>3</sup>, Iain Gordon<sup>4</sup>, Chris McSweeney<sup>1</sup>, Barbara Leggett<sup>3</sup> and Mark Morrison<sup>1,5</sup>

<sup>1</sup>CSIRO Preventative Health National Research Flagship and Livestock Industries, Queensland Bioscience Precinct, Queensland, Australia; <sup>2</sup>School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China; <sup>3</sup>Conjoint Gastroenterology Laboratory, Royal Brisbane and Women's Hospital Research Foundation and Queensland Institute of Medical Research, Brisbane, Australia; <sup>4</sup>CSIRO Sustainable Ecosystems, Davies Laboratory, Australia and <sup>5</sup>The MAPLE Research Initiative and Department of Animal Sciences, The Ohio State University, Columbus, OH, USA

**We applied constrained ordination numerical ecology methods to data produced with a human intestinal tract-specific phylogenetic microarray (the Aus-HIT Chip) to examine the microbial diversity associated with matched biopsy tissue samples taken from the caecum, transverse colon, sigmoid colon and rectum of 10 healthy patients. Consistent with previous studies, the profiles revealed a marked intersubject variability; however, the numerical ecology methods of analysis allowed the subtraction of the subject effect from the data and revealed, for the first time, evidence of a longitudinal gradient for specific microbes along the colorectum. In particular, probes targeting *Streptococcus* and *Enterococcus* spp. produced strongest signals with caecal and transverse colon samples, with a gradual decline through to the rectum. Conversely, the analyses suggest that several members of the *Enterobacteriaceae* increase in relative abundance towards the rectum. These collective differences were substantiated by the multivariate analysis of quantitative PCR data. We were also able to identify differences in the microarray profiles, especially for the streptococci and *Faecalibacterium prausnitzii*, on the basis of gender. The results derived by these multivariate analyses are biologically intuitive and suggest that the biogeography of the colonic mucosa can be monitored for changes through cross-sectional and/or inception cohort studies.**

*The ISME Journal* (2011) 5, 801–809; doi:10.1038/ismej.2010.177; published online 2 December 2010

**Subject Category:** microbial population and community ecology

**Keywords:** biogeography; colorectum; microarray; numerical ecology

## Introduction

In addition to providing a variety of functional and metabolic capabilities relevant to host nutrition and well-being (Sonnenburg *et al.*, 2004; Backhed *et al.*, 2005; Mazmanian *et al.*, 2008), the microbiota resident within the human colorectum have also been observed to change in subjects suffering from inflammatory bowel diseases (IBD). Although these alterations are widely believed to be involved with disease (Campieri and Gionchetti, 2001; Gill and Rowland, 2002; Guarner and Malagelada, 2003; Sartor, 2008), the specific role(s) of individual species or groups of species in disease initiation

and/or progression are still largely unknown. Furthermore, IBD are clinically heterogeneous diseases; almost all cases of Crohn's disease involve a primary lesion in the vicinity of the ileocaecal junction, whereas ulcerative colitis usually affects the distal section of the colon, with a high incidence in the rectum (Maratka *et al.*, 1985; Gasche *et al.*, 1998; Radford-Smith *et al.*, 2002). Similarly, around 70% of the adenocarcinomas occur in the distal section of the colon and rectum (Office for National Statistics, UK, 2004), and although proximal cancers are more likely to be associated with changes in CpG island methylation (the CIMP pathway) (Weisenberger *et al.*, 2006), distal colorectal cancers are more commonly driven by chromosomal instability (Grady and Carethers, 2008). Such differences in IBD type and tumour biology might reflect, at least in part, differences in tissue microenvironments, including microbiological differences. To that end, it remains pertinent to

Correspondence: M Morrison, Queensland Bioscience Precinct, 306 Carmody Road, St Lucia, Queensland 4068, Australia.  
E-mail: mark.morrison@csiro.au

Received 21 June 2010; revised 20 October 2010; accepted 20 October 2010; published online 2 December 2010

elucidate how the microbiome might be structured within the human colon of both healthy and diseased subjects.

Ecological theory suggests that different environments give rise to different communities; therefore, the known anatomical and physiochemical differences along the large intestine should result in a 'biogeographical' distribution of microbes along and across the digestive tract. For instance, cultivation-independent studies have shown that the mucosa-associated microbiota from the distal small intestine and large intestine differ (Eckburg *et al.*, 2005; Wang *et al.*, 2005; Frank *et al.*, 2007). Earlier studies of sudden-death victims revealed that the abundance of methanogenic archaea, the activity of sulphate-reducing bacteria, and other parameters such as the pH and concentration of fermentation products, varied considerably along the human colorectum (Macfarlane *et al.*, 1992). Despite these observations, and the technological advances that have been applied to the analysis of gut microbial communities, evidence of a biogeographical pattern within the microbiome of the human large intestine has been elusive to obtain, principally because of the high degree of intersubject variability and constraints in sample number and size. A number of studies have shown a lack of variation in the mucosa-associated microbial communities along the human colon (Zoetendal *et al.*, 2002; Lepage *et al.*, 2005; Bibiloni *et al.*, 2006; Green *et al.*, 2006; Willing *et al.*, 2009) and between diseased and healthy mucosal tissue (Seksik *et al.*, 2005; Vasquez *et al.*, 2007). These studies have usually relied on alpha- and beta-diversity measures for the comparative analysis of microbial communities. Although such approaches were useful in revealing the main effect on variation—which in all cases was reported to be intersubject variability—it also hindered the identification of more subtle patterns.

Eckburg *et al.* (2005) were the first to undertake large-scale *rrs* gene library construction and DNA sequencing to characterize the human gut microbiome, using matched tissue and faecal samples from three healthy subjects. Their pairwise comparisons also revealed that there were large differences between subjects, and between faecal and mucosal communities. However, they were unable to conclusively reveal evidence of biogeographical trends along the colonic mucosa, and suggested that higher resolution studies using microarray and/or next-generation sequencing technologies, as well as alternative methods of data analysis, were needed. Both phylogenetic microarrays (Paliy *et al.*, 2009; Rajilic-Stojanovic *et al.*, 2009; Kang *et al.*, 2010) and high-throughput sequencing methods are now used more frequently to examine gut microbiome diversity, but the development of robust quantitative methods to reveal the biogeographical features of the gut microbiota is considered a 'great challenge' and 'the diversity and biogeography of the gut microbiota needs to be defined at varying scales of

resolution' (Peterson *et al.*, 2008). Although some clustering and ordination methods of numerical ecology are commonly used to compare microbial community profiles through open source software (for example, QIIME (Caporaso *et al.*, 2010)), and increasingly used in gut microbiomics research (Fuentes *et al.*, 2008; Biagi *et al.*, 2010; Janczyk *et al.*, 2010), we contend that, in addition to those methods, there are additional useful numerical ecology methods that can be used. For instance, constrained ordination methods are being increasingly used for the analysis of mammalian gene expression (Culhane *et al.*, 2002; Baty *et al.*, 2006, 2008) and of the microbial communities present in terrestrial and aquatic samples (see Ramette (2007) for a recent review). To date, there are only a few instances in which these methods have been used with gut microbiome samples (Eckburg *et al.*, 2005; Mondot *et al.*, 2010; Van den Abbeele *et al.*, 2010 with *in vitro* cultures), even though such approaches may be useful to subtract the well-established intersubject variability that prevails in human-derived samples, thereby revealing more subtle patterns in gut microbiome structure and dynamics.

In this study, we present our findings obtained with a human intestinal tract-specific phylogenetic microarray (the Aus-HIT Chip) and, by applying constrained ordination numerical ecology methods to these data, reveal not only gender-specific differences in the human colonic mucosa, but for the first time a longitudinal gradient for specific microbes along the colorectum.

## Materials and methods

### Sampling

This study prospectively recruited subjects scheduled for outpatient colonoscopy at the Royal Brisbane and Women's Hospital in Brisbane and the Flinders Medical Centre in Adelaide, Australia. To be included in this study, subjects had to be between 45 and 70 years of age, non-smokers and should have a normal colonoscopy, including no diverticulosis and no past history of colorectal neoplasia. In addition, subjects were excluded if they were receiving anticoagulant therapy or had a history of IBD, familial adenomatous polyposis or hyperplastic polyposis. The first five women (age  $57.5 \pm 7.7$ ) and five men (age  $54.3 \pm 9.9$ ) who satisfied these criteria were selected for this study. All subjects gave written informed consent and the study was approved by the local institutional review boards. During colonoscopy, all patients had four mucosal biopsy samples taken, one each from the caecum, the transverse colon, the sigmoid colon and the rectum. The mucosal biopsy samples were snap frozen in liquid N<sub>2</sub> and stored at  $-80^{\circ}\text{C}$  until processing. DNA was extracted and purified from the frozen biopsy samples using the AllPrep DNA/RNA Mini Kit (Qiagen Pty Ltd, Hilden, Germany),

according to the manufacturer's instructions, and stored at 4 °C until processing.

#### Generation and fluorescent labelling of cRNA

The methods described by Kang *et al.* (2010) were used with several modifications. First, greater amounts of total DNA (100–500 ng) extracted from the biopsy samples were used for the PCR amplification of the *Archaea* and *Bacteria rrs* genes, using the primers described in Table 1. As an internal standard, the mitochondrial *rrs* gene was also PCR amplified using DNA extracted from blood as the template and the mitochondrial-specific primers listed in Table 1. All PCR reactions were performed in triplicate for each sample, then pooled and purified using the MiniElute PCR purification kit (Qiagen). Where necessary, the amplicons were concentrated using Pellet Paint (Novagen, Madison, WI, USA).

Single-stranded cRNA was then produced from 500 ng aliquots of the pooled PCR products, using the MEGAScript T7 *In Vitro* transcription kit (Ambion, Austin, TX, USA) as specified by the manufacturer. The resulting cRNA samples were purified using the MEGAclean Kit (Ambion), and the size and purity of the products obtained at each step were assessed by agarose gel electrophoresis and with an ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). Aliquots (500 ng) of the cRNA samples were spiked with 7 ng of the control mitochondrial cRNA, then labelled and fragmented using the Label IT  $\mu$ Array Cy5 labelling Kit (Mirus, Madison, WI, USA).

#### Microarray hybridization and image analysis

The cRNA samples described above were hybridized in quadruplicate to CustomArray 4 X 2K microarrays fabricated by CombiMatrix (Mukilteo, WA, USA). The relevant probe and other

methodological details have previously been described by Kang *et al.* (2010) and are accessible at the GEO database (<http://ncbi.nlm.nih.gov/GEO>, under GPL9543 and GSE18933). In brief, the microarrays include 593 probes specific for human intestinal bacteria and archaea derived from the published literature, another 163 newly designed probes using the GoArray algorithm (Rimour *et al.*, 2005), as well as 10 probes (6 positive and 4 negative mismatch controls) targeting the human mitochondrial *rrs* gene. The hybridized microarray slides were scanned using an Axon Genepix 4000A microarray scanner (Axon Instruments, Union City CA, USA). The images obtained were analysed using GenePix Pro 6.0 software (Axon Instruments), and probe signals were recorded as 635 nm foreground and 635 nm background intensities.

#### Data processing and analysis

Raw signal data were first processed using GeneSpring GX10 software (Agilent Technologies, Santa Clara, CA, USA). The probe intensity values were transformed to log<sub>2</sub>, and to avoid the generation of biased profiles the distribution of probe intensities for all the hybridizations were first checked by boxplot graphing before normalization using the quantile normalization method (based on the normalization of intensity distributions) (Bolstad *et al.*, 2003). Any probes that produced signal values lower than the highest negative control signals were also removed from further analysis. Of the 40 samples used in this study, one was consistently found as an outlier and was excluded from the analyses to avoid bias. Using these criteria, the remaining 39 samples produced a total of 115 profiles that were used for analysis, with no less than two hybridization profiles produced for each cRNA sample.

The averaged per-sample profiles were first subjected to hierarchical clustering, using the average linking method based on  $\chi^2$  dissimilarity measures

**Table 1** Primers used in this study

| Name                    | Sequence (5'-3')                                  | Specificity           | Reference                       |
|-------------------------|---|-----------------------|---------------------------------|
| 4F <sup>a</sup>         | TCCGGTTGATCCTGCCRG                                | <i>Archaea</i>        | Giovannoni <i>et al.</i> , 1988 |
| 27F <sup>a</sup>        | AGAGTTTGATCMTGGCTCAG                              | <i>Bacteria</i>       | Lane, 1990                      |
| T7-1492R <sup>a,b</sup> | <i>TCTAATACGACTCACTATAGGGGGYTACCTTGTACGACTT</i>   | Universal             |                                 |
| MitoF <sup>a</sup>      | TACTACCAGACAACCTTAGC                              | Mitochondrial 16S     | Kasashima <i>et al.</i> , 2008  |
| T7-MitoR <sup>a,b</sup> | <i>TCTAATACGACTCACTATAGGGGGTTTCGGGGTCTTAGCTTT</i> |                       |                                 |
| StrF                    | CGACGATACATAGCCGACCTGAG                           | streptococci          | Chalmers <i>et al.</i> , 2008   |
| StrR                    | TCCATTGCCGAAGATTCCCTACTG                          |                       |                                 |
| F-Ent                   | ATGGCTGTCGTGACGCTCGT                              | enterobacteria        | Leser <i>et al.</i> , 2002      |
| R-Ent                   | CCTACTTCTTTTGCAACCCACTC                           |                       | Sghir <i>et al.</i> , 2000      |
| EcocF                   | CCCTTATTGTTAGTTGCCATCATT                          | enterococci           | Rinttila <i>et al.</i> , 2004   |
| EcocR                   | ACTCGTTGTAAGTCCCATTTGT                            |                       |                                 |
| BacF                    | CGGCAACGAGCGCAACCC                                | <i>Bacteria</i>       | Denman and McSweeney, 2006      |
| BacR                    | CCATTGTAGCACGTGTGTAGCC                            |                       |                                 |
| Fprau420R               | CCGAAGACCTTCTTCCTCC                               | <i>F. prausnitzii</i> | Wang <i>et al.</i> , 1996       |
| Fprau223F               | GATGGCCTCGCGTCCGATTAG                             |                       | Ahmed <i>et al.</i> , 2007      |

<sup>a</sup>Primers used for preparation of target RNA for array hybridization.

<sup>b</sup>Modified by Kang *et al.* (2010) to include the T7 promoter (italicized) for subsequent cRNA synthesis.

using R package *ade4*. The profiles were then subjected to multivariate analysis methods derived from numerical ecology. For these analyses, both the averaged per-sample profiles ( $n = 39$ ) and all the microarray profiles ( $n = 115$ ) were used to examine the technical variation associated with the profiling approach used. Although both data sets produced similar results, only the former ( $n = 39$ ) was used in the assessment of biological effects. Between-group analysis (Culhane *et al.*, 2002) and analysis with respect to instrumental variables (Baty *et al.*, 2006) applied to correspondence analysis (Fellenberg *et al.*, 2001) were performed, using the R package *ade4* (Dray and Dufour, 2007). The correspondence analysis method, often referred to as reciprocal averaging, is considered to be specifically useful in ecological studies. Its underlying model assumes a unimodal distribution of species along environmental gradients (as stated by niche theory), and is based on  $\chi^2$  distance measures, thus avoiding the double-zero problem (Legendre and Legendre, 1998) (Ramette, 2007). These methods are expected to identify the probes that best explain the differences between *a priori* classifications, and permit the subtraction of particular effects (for example, inter-subject variability), which then allows the remaining effects to be effectively compared. Monte-Carlo tests (Baty *et al.*, 2006), in which sample labels are randomly permuted to assess the significance of the constraint being applied, and bootstrapping tests (Baty *et al.*, 2008) assessing the significance of probe contribution to the model were performed to assess the reliability and stability of the results obtained. Only probes showing a stable contribution ( $P < 0.05$ ) and a larger effect on the model than that of the control mitochondrial probes were considered for further analyses, to avoid spurious results. Pairwise differences between class (sampling sites) positions in the model were calculated by analysis of variance using the Statistical 6.0 software (StatSoft Inc., Tulsa, OK, USA).

#### Real-time PCR analysis

Some key bacterial groups, identified by the numerical ecology methods to vary with respect to gender and/or site were quantified by quantitative PCR. The primers used in these assays are listed in Table 1. The reaction mixtures were prepared using a Biomek2000 automated workstation (Beckman, Brea, CA, USA) and contained 2X iQ SYBR Green Supermix (Bio-Rad, Hercules, CA, USA), 400 nM each of the appropriate forward and reverse primer, 10 pg of template DNA (the pooled PCR amplicons obtained from the original sample) and sterile water. Quadruplicate 5  $\mu$ l aliquots of each reaction were dispensed into a 384-well plate, and a dilution series was also prepared from a pooled sample to determine the amplification efficiencies for each group-specific PCR reaction. Real-time PCR was performed using a 7900HT Sequence detection system (Applied Biosystems, Mountain View, CA, USA), and

after an initial denaturation step of 95 °C for 10 min, 40 cycles of 95 °C for 15 s and 58 °C for 30 s were performed. The resulting data for each specific group were normalized relative to the *Bacteria rrs* gene using the methods described by Pfaffl, 2001, and subjected to analysis of variance using Statistical 6.0 software. To visualize a possible site-related structure in these data, the datum produced for each individual sample was also divided by the subject average and then these values were plotted using the *triangle.plot* function from *ade4*.

## Results

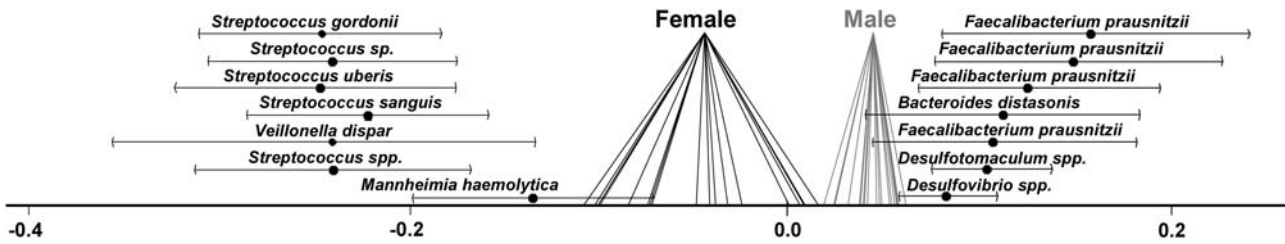
#### *Hierarchical clustering of the microarray profiles shows marked intrasubject grouping*

Similar to the findings of Eckburg *et al.* (2005), the initial clustering analysis of the microarray data showed that the microbial profiles produced from different tissues from the same subject clustered together (Supplementary Figure 1). The analysis did not reveal any clear evidence of a biogeography along the colorectum, other than an inconclusive tendency of the caecum-derived profiles to cluster with the transverse colon-derived profiles from the same patient (for 6 of the 10 subjects). However, the clustering analysis did indicate that there may be a possible gender effect, as the profiles derived from all but one of the male subjects clustered together (Supplementary Figure 1).

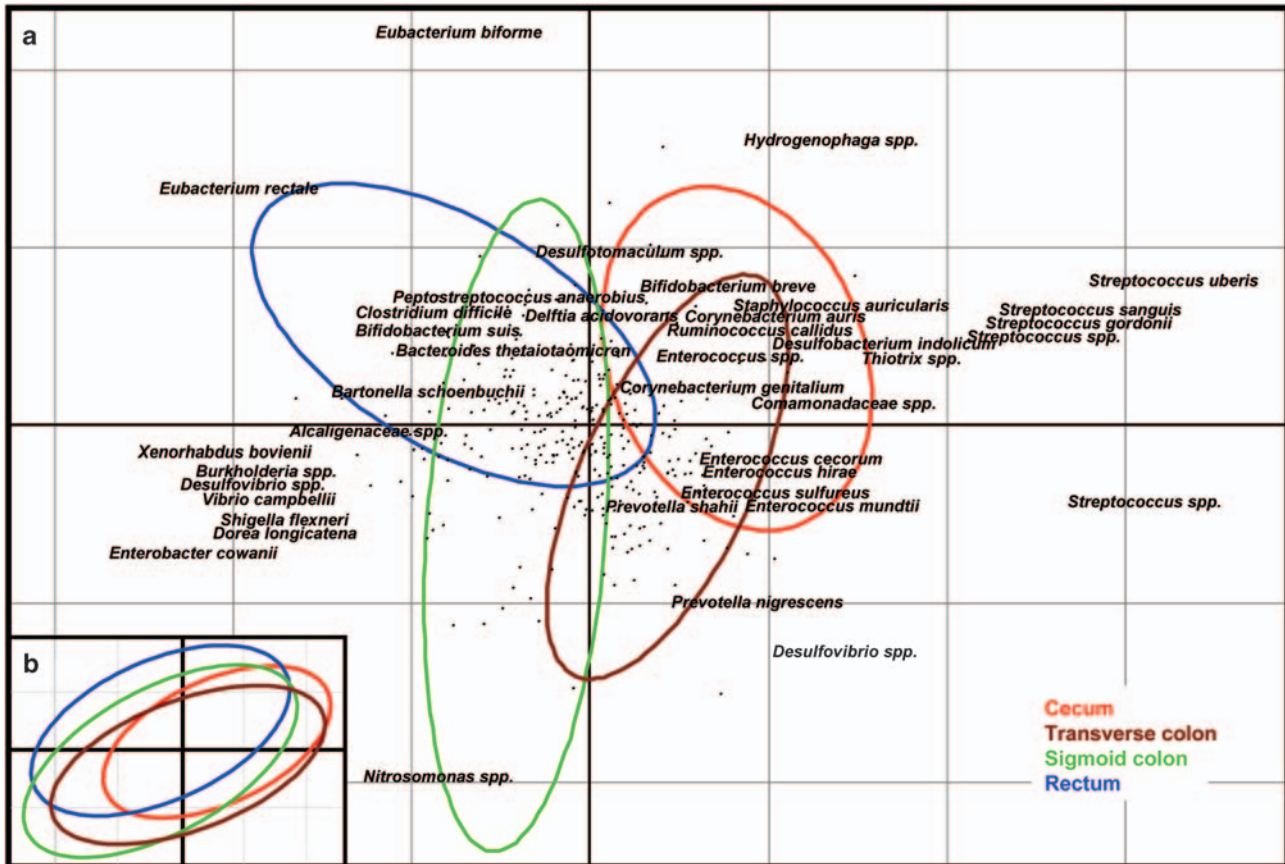
#### *Between-group analysis of the microarray data confirmed a gender-based bifurcation in microbiome profiles*

When the microarray data were subjected to analysis with constrained ordination (that is, between-group analysis using gender as the constraint), the results showed that gender had a statistically significant effect on the profiles and accounted for 15.5% of the variance in the data ( $P < 0.05$  compared with 13.3% and  $P < 0.05$  for the replicated ( $n = 115$ ) data set; Figure 1). The analysis also showed that the male-derived profiles were more similar to each other, whereas the female-derived profiles contained a greater amount of variation. The cRNA samples prepared from the female subjects produced relatively stronger signals from most of the *Streptococcus*, *Veillonella*, *Mannheimia* and *Ruminococcus* spp. probes, whereas samples from the male subjects produced relatively greater signals from several probes targeting *Faecalibacterium prausnitzii*, *Bacteroides*, *Clostridium*, *Prevotella*, *Enterococcus* and *Bifidobacterium* spp.

Subtraction of intersubject variability from the microarray data by application of numerical ecology methods revealed evidence of colonic mucosal biogeography. To assess whether there were differences between sites, an analysis with respect to instrumental variables (that is, a within-group analysis using subject as constraint, followed by a



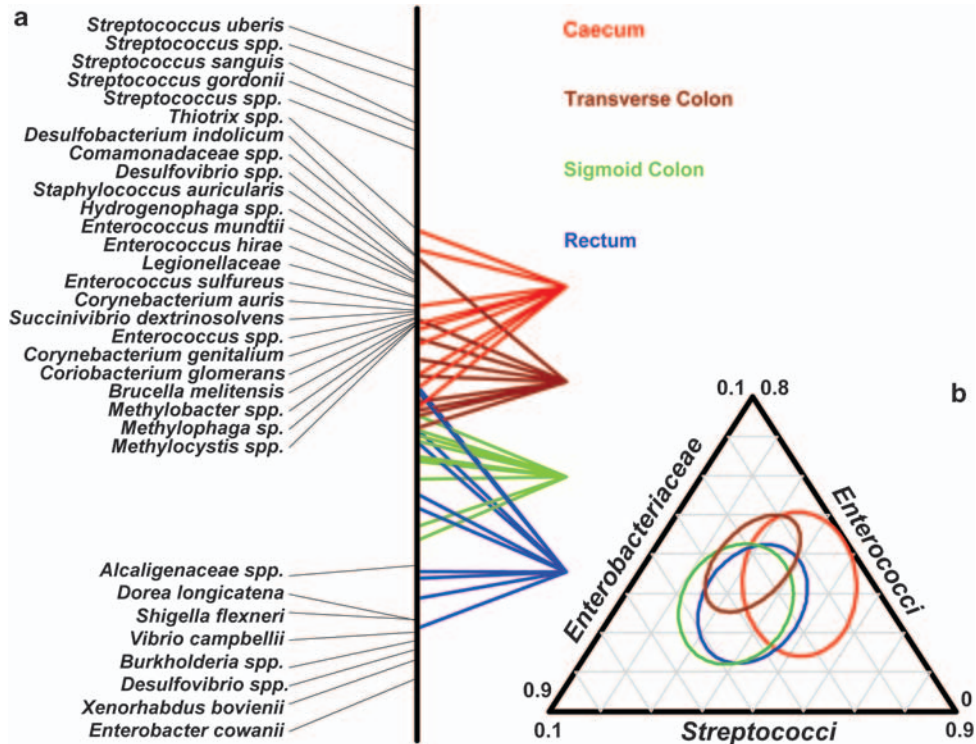
**Figure 1** Discrimination between female- and male-derived profiles. The diagram shows the single axis result of a between-group analysis applied to correspondence analysis on the data using gender as constraint. The seven most discriminant probes from either end of the axis are included (error bars are used to represent the spread of the probe coordinates on the model after bootstrapping).



**Figure 2** Decomposition of the data set variability according to site. The diagram shows the results of a between-group analysis applied to correspondence analysis using site as the constraint on the data, in which the ‘individual’ effect has been previously removed (a), or not removed (b). The coloured ellipsoids represent the collective variance for each site. In panel a, a maximum of 20 probes significantly contributing to each class are shown, and the axes comprise 53.13% (x) and 26.87% (y) of the variance explained by the constraint.

between-group analysis using sampling site as constraint) (Baty *et al.*, 2006) was performed with the data set. The analyses revealed that intersubject variability accounted for 94.2% of the variance in the data, and when this variability was not removed from the analysis no differences in the profiles with respect to site could be observed (Figure 2b,  $P < 1.0$ ). However, once the intersubject variability was removed, the sampling site was found to significantly contribute to the remaining variance (10.18% and  $P < 0.087$ , as compared with 4.8%,  $P < 0.01$  for the individual hybridizations). Indeed, the

between-site analysis applied to the residual variance once the interindividual effects had been removed shows the position of the sites along the colorectum to be the main (x) axis of variation (Figures 2a and 3a). Pairwise tests revealed that all classes (sampling sites) were significantly separated by one of the three axes of the model. Furthermore, there was also evidence of gradients in the intensity of some probe signals along the colorectum (Figure 3a). Probes targeting *Streptococcus*, *Comamonadaceae*, *Enterococcus* and *Corynebacterium* spp. showed strong associations with the caecum



**Figure 3** (a) Species contributing to the main (x) axis of variation shown in Figure 2a, plotted with respect to the microarray hybridization profiles generated from the individual samples. The positional location of the coloured lines represents their relative position along the axis of variation and their association with the listed probe identifiers. Only those probe identifiers that make a statistically significant contribution to the site differentiation are shown. (b) A triangle plot showing the quantitative PCR results produced from the individual samples using group-specific primers. The coloured ellipsoids represent the collective variance for each site, which concurs with the microarray results shown in Figure 2a.

and transverse colon classes, whereas several *Enterobacteriaceae* probes show association with the sigmoid colon and rectum classes. Taken together, the microarray data suggest that, in the subjects examined in this study, members of the *Streptococcus*, *Comamonadaceae*, *Enterococcus* and *Corynebacterium* spp. were present in greater abundance in tissue samples collected from the caecum and transverse colon, whereas members of the *Enterobacteriaceae* were present in relatively greater amounts in tissue samples collected from the sigmoid colon and rectum.

#### Quantitative PCR analysis supports the findings derived from the microarray profiles

Overall, the results derived from the real-time data validate previous findings of the microarray analysis. The gender-based differences revealed by the microarray analyses were confirmed, with the female DNA samples found to possess fewer *F. prausnitzii* (female/male ratio: 0.41,  $P < 0.05$ ) and higher numbers of streptococci (female/male ratio: 1.29), although the latter was not statistically significant. The multivariate test applied to site-specific data revealed that there were statistically significant differences (Wilk's lambda,  $P < 0.01$ ), and that streptococci are present in lower abundance in

the sigmoid colon compared with other sites (ratio sigmoid/other sites  $\leq 0.63$ ,  $P < 0.05$ ). Furthermore, when the values obtained for each sample were plotted together (Figure 3b), the per-site distribution obtained was quite similar to results produced using the microarray data (Figure 2a). Taken together, these results are consistent with the findings derived from the microarray analyses and provide further evidence that select groups of bacteria may preferentially colonize different regions of the colorectum.

## Discussion

Gender-related differences in gut microbial communities have previously been observed in macaques (McKenna *et al.*, 2008) and mice (Schloss and Handelsman, 2006). In humans, gender-related differences have previously been reported in a cross-sectional study evaluating differences in a small set of members of the faecal microbiota (Mueller *et al.*, 2006). The study concluded that members of the *Bacteroides-Prevotella* group were in greater abundance in males than in females, which is in accordance with our microarray results. Gender-based differences in the gut microbiota could be indicative of underlying differences in

gut physiology or host effect on the microbiome (that is, diet and lifestyle). Regarding the biogeography of the colorectum, it is evident from the overlap observed between the classes (Figure 2a) that the model obtained could not perfectly resolve in all cases between different sampling sites. Nevertheless, the results obtained in this study clearly point to a consistent and common biogeography in the colorectum of the patients studied, as evidenced by the significant effect that sampling site had on the variance of the microarray profiles, and on the abundance of select bacterial taxa measured by quantitative PCR.

The main trend observed in the microarray analyses was a longitudinal gradient for specific microbes along the colorectum, with probes targeting members of the *Streptococcus*, *Comamonadaceae*, *Enterococcus*, *Corynebacterium* and *Lactobacillus* producing the strongest signals with caecal and transverse colon samples. These findings are not inconsistent with the observations reported by Frank *et al.* (2007), in which *Streptococcaceae* and *Corynebacteriaceae* were present in higher numbers from distal small intestinal samples compared with those from the colorectum. Conversely, probe signals for several members of the *Enterobacteriaceae* were found in our studies to increase towards the rectum, suggesting that these bacteria largely reside in the distal regions of the colorectum in healthy subjects. Interestingly, Willing *et al.* (2009) reported a large increase in *Enterobacteriaceae* in samples analysed from subjects with the ileocaecal form of Crohn's disease; our observations then further support the contention that this form of IBD is associated with a proliferation of specific bacteria 'outside' their normal biogeographical location. The environmental drivers of the differences for both axes of variation cannot be determined from this study, but the trend follows the main physicochemical changes known to occur through the colorectum: lumen content dehydration and pH increase towards the rectum (Bown *et al.*, 1974), as well as highest short-chain fatty acid, lactate and ethanol concentrations in the proximal colon, decreasing distally with a concomitant increase in products of protein fermentation (for example, ammonia, branched chain fatty acids and phenolic compounds) (Macfarlane *et al.*, 1992). There is also one report of there being differences in mucosal oxygen concentration along the human colon, especially between the caecum and distal regions ( $pO_2$  [mm Hg] being  $33.7 \pm 7.5$  in the caecum, compared with  $40.7 \pm 8.2$ ,  $41.2 \pm 8.7$  and  $39.7 \pm 7.3$  for the transverse colon, sigmoid colon and rectum, respectively (Sheridan *et al.*, 1987)). Interestingly, recent studies linking a higher abundance of aerotolerant bacteria with ileal CD (Baumgart *et al.*, 2007) and surgical intervention (Hartman *et al.*, 2009) suggest that perturbations in oxygen concentration may be associated with the observed dysbiosis. Our findings are not inconsistent with these reports,

and collectively, all three studies support the hypothesis that the dysbiosis seen with some types of IBD might initially be in response to the altered oxygenation state of the inflamed tissue.

In conclusion, it has long been recognized that there is a variation in the physicochemical conditions along the human colon, and it is biologically intuitive that such conditions would influence the biogeography of the colonic mucosal microbiome. However, the detection of such differences has been elusive, obscured behind intersubject variability. Our study shows that the application of numerical ecology methods to the analysis of data generated with a phylogenetic microarray is a powerful approach in the study of the mucosa-associated microbial community of the human colorectum. Not only did we find differences in the mucosa-associated microbiome with respect to gender but we also showed, for the first time, evidence of a distinctive biogeography along the length of the human colon. Given the region-specific nature of many types of colonic cancers and inflammatory diseases, our findings raise interesting questions on how the 'functional core' of the mucosa-associated gut microbiome might change along the length of the human colon, as well as the nature of host-microbe interactions at different regions.

## Acknowledgements

This research has been supported with funds provided by CSIRO's Preventative Health National Research Flagship, a CSIRO OCE Science Leader award (to MM) and funds from CSIRO's Transformational Biology Capability Platform. We thank Stuart Denman for technical advice, Antonio Reverter-Gomez and Maree O'Sullivan for helpful comments and advice with the microarray data normalization and statistical analyses, and Trevor Lockett and Carly Rosewarne for critical reading of the manuscript.

## References

- Ahmed S, Macfarlane GT, Fite A, McBain AJ, Gilbert P, Macfarlane S. (2007). Mucosa-associated bacterial diversity in relation to human terminal ileum and colonic biopsy samples. *Appl Environ Microbiol* **73**: 7435–7442.
- Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JL. (2005). Host-bacterial mutualism in the human intestine. *Science* **307**: 1915–1920.
- Baty F, Facompre M, Wiegand J, Schwager J, Brutsche MH. (2006). Analysis with respect to instrumental variables for the exploration of microarray data structures. *BMC Bioinformatics* **7**: 422.
- Baty F, Jaeger D, Preiswerk F, Schumacher MM, Brutsche MH. (2008). Stability of gene contributions and identification of outliers in multivariate analysis of microarray data. *BMC Bioinformatics* **9**: 289.
- Baumgart M, Dogan B, Rishniw M, Weitzman G, Bosworth B, Yantiss R *et al.* (2007). Culture independent analysis of ileal mucosa reveals a selective increase

- in invasive *Escherichia coli* of novel phylogeny relative to depletion of *Clostridiales* in Crohn's disease involving the ileum. *ISME J* 1: 403–418.
- Biagi E, Nylund L, Candela M, Ostan R, Bucci L, Pini E *et al.* (2010). Through Ageing, and Beyond: Gut Microbiota and Inflammatory Status in Seniors and Centenarians. *PLoS ONE* 5: e10667.
- Bibiloni R, Mangold M, Madsen KL, Fedorak RN, Tannock GW. (2006). The bacteriology of biopsies differs between newly diagnosed, untreated, Crohn's disease and ulcerative colitis patients. *J Med Microbiol* 55: 1141–1149.
- Bolstad BM, Irizarry RA, Astrand M, Speed TP. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19: 185–193.
- Bown RL, Gibson JA, Sladen GE, Hicks B, Dawson AM. (1974). Effects of lactulose and other laxatives on ileal and colonic pH as measured by a radiotelemetry device. *Gut* 15: 999–1004.
- Campieri M, Gionchetti P. (2001). Bacteria as the cause of ulcerative colitis. *Gut* 48: 132–135.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Meth* 7: 335–336.
- Chalmers NI, Palmer RJ, Cisar JO, Kolenbrander PE. (2008). Characterization of a *Streptococcus* sp.-*Veillonella* sp community micromanipulated from dental plaque. *J Bacteriol* 190: 8145–8154.
- Culhane AC, Perriere G, Considine EC, Cotter TG, Higgins DG. (2002). Between-group analysis of microarray data. *Bioinformatics* 18: 1600–1608.
- Denman SE, McSweeney C. (2006). Development of a real-time PCR assay for monitoring anaerobic fungal and cellulolytic bacterial populations within the rumen. *FEMS Microbiol Ecol* 58: 572–582.
- Dray S, Dufour AB. (2007). The ade4 package: Implementing the duality diagram for ecologists. *J Stat Softw* 22: 1–20.
- Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M *et al.* (2005). Diversity of the human intestinal microbial flora. *Science* 308: 1635–1638.
- Fellenberg K, Hauser NC, Brors B, Neutzner A, Hoheisel JD, Vingron M. (2001). Correspondence analysis applied to microarray data. *Proc Natl Acad Sci USA* 98: 10781–10786.
- Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci USA* 104: 13780–13785.
- Fuentes S, Egert M, Jiménez-Valera M, Ramos-Cormenzana A, Ruiz-Bravo A, Smidt H *et al.* (2008). Administration of *Lactobacillus casei* and *Lactobacillus plantarum* affects the diversity of murine intestinal lactobacilli, but not the overall bacterial community structure. *Res Microbiol* 159: 237–243.
- Gasche C, Schober E, Turetschek K. (1998). Small bowel barium studies in Crohn's disease. *Gastroenterology* 114: 1349.
- Gill CI, Rowland IR. (2002). Diet and cancer: assessing the risk. *Br J Nutr* 88(Suppl 1): S73–S87.
- Giovannoni SJ, DeLong EF, Olsen GJ, Pace NR. (1988). Phylogenetic group-specific oligodeoxynucleotide probes for identification of single microbial-cells. *J Bacteriol* 170: 720–726.
- Grady WM, Carethers JM. (2008). Genomic and Epigenetic Instability in Colorectal Cancer Pathogenesis. *Gastroenterology* 135: 1079–1099.
- Green GL, Brostoff J, Hudspith B, Michael M, Mylonaki M, Rayment N *et al.* (2006). Molecular characterization of the bacteria adherent to human colorectal mucosa. *J Appl Microbiol* 100: 460–469.
- Guarner F, Malagelada JR. (2003). Gut flora in health and disease. *Lancet* 361: 512–519.
- Hartman AL, Lough DM, Barupal DK, Fiehn O, Fishbein T, Zasloff M *et al.* (2009). Human gut microbiome adopts an alternative state following small bowel transplantation. *Proc Natl Acad Sci USA* 106: 17187–17192.
- Janczyk P, Pieper R, Smidt H, Souffrant WB. (2010). Effect of alginate and inulin on intestinal microbial ecology of weanling pigs reared under different husbandry conditions. *FEMS Microbiology Ecology* 72: 132–142.
- Kang S, Denman SE, Morrison M, Yu Z, Dore J, Leclerc M *et al.* (2010). Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflamm Bowel Dis*, doi:10.1002/ibd.21319.
- Kasashima K, Sumitani M, Satoh M, Endo H. (2008). Human prohibitin 1 maintains the organization and stability of the mitochondrial nucleoids. *Experimental Cell Research* 314: 988–996.
- Lane D. (1990). 16S and 23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds). *Nucleic acid techniques in bacterial systematics*. John Wiley: New York, NY, pp 115–148.
- Legendre P, Legendre L. (1998). *Numerical Ecology*. Elsevier: Amsterdam.
- Lepage P, Seksik P, Sutren M, de la Cochetiere MF, Jian R, Marteau P *et al.* (2005). Biodiversity of the mucosa-associated microbiota is stable along the distal digestive tract in healthy individuals and patients with IBD. *Inflamm Bowel Dis* 11: 473–480.
- Leser TD, Amenuvor JZ, Jensen TK, Lindecrone RH, Boye M, Moller K. (2002). Culture-independent analysis of gut bacteria: the pig gastrointestinal tract microbiota revisited. *Appl Environ Microbiol* 68: 673–690.
- Macfarlane GT, Gibson GR, Cummings JH. (1992). Comparison of fermentation reactions in different regions of the human colon. *J Appl Bacteriol* 72: 57–64.
- Maratka Z, Nedbal J, Kocianova J, Havelka J, Kudrman J, Hendl J. (1985). Incidence of colorectal cancer in proctocolitis: a retrospective study of 959 cases over 40 years. *Gut* 26: 43–49.
- Mazmanian SK, Round JL, Kasper DL. (2008). A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* 453: 620–625.
- McKenna P, Hoffmann C, Minkah N, Aye PP, Lackner A, Liu Z *et al.* (2008). The macaque gut microbiome in health, lentiviral infection, and chronic enterocolitis. *PLoS Pathog* 4: e20.
- Mondot S, Kang S, Furet J, Aguirre de Cárcer D, McSweeney C, Morrison M *et al.* (2010). Highlighting new phylogenetic specificities of Crohn's disease microbiota. *Inflamm Bowel Dis*, doi:10.1002/ibd.21436.
- Mueller S, Saunier K, Hanisch C, Norin E, Alm L, Midtvedt T *et al.* (2006). Differences in fecal microbiota in different European study populations in relation to age, gender, and country: a cross-sectional study. *Appl Environ Microbiol* 72: 1027–1033.
- Paliy O, Kenche H, Abernathy F, Michail S. (2009). High-throughput quantitative analysis of the human



- intestinal microbiota with a phylogenetic microarray. *Appl Environ Microbiol* **75**: 3572–3579.
- Peterson DA, Frank DN, Pace NR, Gordon JI. (2008). Metagenomic Approaches for Defining the Pathogenesis of Inflammatory Bowel Diseases. *Cell Host Microbe* **3**: 417–427.
- Pfaffl MW. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**: e45.
- Radford-Smith GL, Edwards JE, Purdie DM, Pandeya N, Watson M, Martin NG *et al.* (2002). Protective role of appendectomy on onset and severity of ulcerative colitis and Crohn's disease. *Gut* **51**: 808–813.
- Rajilic-Stojanovic M, Heilig HG, Molenaar D, Kajander K, Surakka A, Smidt H *et al.* (2009). Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. *Environ Microbiol*, doi:10.1111/j.1462-2920.2009.01900.x.
- Ramette A. (2007). Multivariate analyses in microbial ecology. *FEMS Microbiol Ecol* **62**: 142–160.
- Rimour S, Hill D, Milton C, Peyret P. (2005). GoArrays: highly dynamic and efficient microarray probe design. *Bioinformatics* **21**: 1094–1103.
- Rinttila T, Kassinen A, Malinen E, Krogus L, Palva A. (2004). Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. *J Appl Microbiol* **97**: 1166–1177.
- Sartor RB. (2008). Microbial influences in inflammatory bowel diseases. *Gastroenterology* **134**: 577–594.
- Schloss PD, Handelsman J. (2006). Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures. *Appl Environ Microbiol* **72**: 6773–6779.
- Seksik P, Lepage P, de la Cochetiere MF, Bourreille A, Sutren M, Galmiche JP *et al.* (2005). Search for localized dysbiosis in Crohn's disease ulcerations by temporal temperature gradient gel electrophoresis of 16S rRNA. *J Clin Microbiol* **43**: 4654–4658.
- Sghir A, Gramet G, Suau A, Rochet V, Pochart P, Dore J. (2000). Quantification of bacterial groups within human fecal flora by oligonucleotide probe hybridization. *Appl Environ Microbiol* **66**: 2263–2266.
- Sheridan W, Lowndes R, Young H. (1987). Tissue oxygen tension as a predictor of colonic anastomotic healing. *Dis Colon Rectum* **30**: 867–871.
- Sonnenburg JL, Angenent LT, Gordon JI. (2004). Getting a grip on things: how do communities of bacterial symbionts become established in our intestine? *Nat Immunol* **5**: 569–573.
- Van den Abbeele P, Grootaert C, Marzorati M, Possemiers S, Verstraete W, Gerard P *et al.* (2010). Microbial Community Development in a Dynamic Gut Model Is Reproducible, Colon Region Specific, and Selective for *Bacteroidetes* and *Clostridium* Cluster IX. *Appl Environ Microbiol* **76**: 5237–5246.
- Vasquez N, Mangin I, Lepage P, Seksik P, Duong JP, Blum S *et al.* (2007). Patchy distribution of mucosal lesions in ileal Crohn's disease is not linked to differences in the dominant mucosa-associated bacteria: a study using fluorescence *in situ* hybridization and temporal temperature gradient gel electrophoresis. *Inflamm Bowel Dis* **13**: 684–692.
- Wang M, Ahrne S, Jeppsson B, Molin G. (2005). Comparison of bacterial diversity along the human intestinal tract by direct cloning and sequencing of 16S rRNA genes. *FEMS Microbiol Ecol* **54**: 219–231.
- Wang RF, Cao WW, Cerniglia CE. (1996). PCR detection and quantitation of predominant anaerobic bacteria in human and animal fecal samples. *Appl Environ Microbiol* **62**: 1242–1247.
- Weisenberger DJ, D Siegmund K, Campan M, Young J, Long TI, Faasse MA *et al.* (2006). CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* **38**: 787–793.
- Willing B, Halfvarson J, Dicksved J, Rosenquist M, Jarnerot G, Engstrand L *et al.* (2009). Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease. *Inflamm Bowel Dis* **15**: 653–660.
- Zoetendal EG, von Wright A, Vilpponen-Salmela T, Ben-Amor K, Akkermans AD, de Vos WM. (2002). 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* **68**: 3401–3407.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)