Linking Spatial Structure and Community-Level Biotic Interactions through Cooccurrence and Time Series Modeling of the Human Intestinal Microbiota

Eric J. de Muinck,[a] Knut E. A. Lundin,[b] Pål Trosvik[a]

Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Oslo, Norway[a];
Department of Gastroenterology, Oslo University Hospital–Rikshospitalet, Oslo, Norway[b]

ABSTRACT The gastrointestinal (GI) microbiome is a densely populated ecosystem where dynamics are determined by interactions between microbial community members, as well as host factors. The spatial organization of this system is thought to be important in human health, yet this aspect of our resident microbiome is still poorly understood. In this study, we report significant spatial structure of the GI microbiota, and we identify general categories of spatial patterning in the distribution of microbial taxa along a healthy human GI tract. We further estimate the biotic interaction structure in the GI microbiota, both through time series and cooccurrence modeling of microbial community data derived from a large number of sequentially collected fecal samples. Comparison of these two approaches showed that species pairs involved in significant negative interactions had strong positive contemporaneous correlations and vice versa, while for species pairs without significant interactions, contemporaneous correlations were distributed around zero. We observed similar patterns when comparing these models to the spatial correlations between taxa identified in the adherent microbiota. This suggests that colocalization of microbial taxon pairs, and thus the spatial organization of the GI microbiota, is driven, at least in part, by direct or indirect biotic interactions. Thus, our study can provide a basis for an ecological interpretation of the biogeography of the human gut.

IMPORTANCE The human gut microbiome is the subject of intense study due to its importance in health and disease. The majority of these studies have been based on the analysis of feces. However, little is known about how the microbial composition in fecal samples relates to the spatial distribution of microbial taxa along the gastrointestinal tract. By characterizing the microbial content both in intestinal tissue samples and in fecal samples obtained daily, we provide a conceptual framework for how the spatial structure relates to biotic interactions on the community level. We further describe general categories of spatial distribution patterns and identify taxa conforming to these categories. To our knowledge, this is the first study combining spatial and temporal analyses of the human gut microbiome. This type of analysis can be used for identifying candidate probiotics and designing strategies for clinical intervention.

KEYWORDS 16S rRNA gene, microbiome, biotic interactions, intestine, microbial ecology, spatial structure, time series

Address correspondence to Eric J. de Muinck, ericdemuinck@gmail.com.

As has been well documented, the gastrointestinal (GI) tracts of humans contain a complex ecology of microorganisms with intimate and often long-term interactions with the host (1). The many recent studies of the human GI microbiome have

increased our knowledge of this relationship and its impacts on our development, physiology, immune system, and nutrition (2). However, much work remains to be done in order to come to an even partial understanding of this complex and dynamic ecosystem and its spatial distribution along the intestinal tract. The challenges in this area are due to several factors, including the high number of potentially interacting species, heterogeneity of habitats along the GI tract, and inherent difficulties of sampling (3). Additionally, and perhaps most importantly, there are large individual and temporal variations in microbiome composition (2).

Most studies rely on one or a few easily obtained fecal samples in order to profile an individual's gut microbiome (3). Individual variation seems to be the most important factor in microbiome compositions, and this supports the use of fecal samples to characterize and compare individuals. However, there is much evidence that fecal sampling may not represent the adherent microbial communities (4). The GI tract has gradients of pH, oxygen availability, and host factors (2). Given this, there have been conflicting reports as to the biogeographical distribution of members of the GI microbiota, with some finding evidence to support spatial segregation (5, 6) while others did not see differences between the different anatomical regions (7, 8).

The personalized and dynamic ecosystem that comprises the human GI microbiome is influenced by the host environment and by the interspecies bacterial interactions themselves (9). Ecological interactions within these GI communities, while distinct and specific to the individual, display universality in that aspects of the interspecies interactions may be generalizable (10). These findings highlight the importance of studying the microbial interactions at the individual level and relating these to the GI biogeography.

In order to address some of these issues and investigate the relationship between microbial interactions in the human GI tract and how these interactions could relate to the spatial segregation of the bacterial taxa along the GI tract, we obtained 139 daily fecal samples from a single individual (main study individual), as well as triplicate biopsy samples from the same individual from seven sites along the GI tract at a single time point. To our knowledge, this is the first study linking spatial cooccurrence of microbial taxa along the GI tract with the biotic interactions estimated by time series analysis. By profiling the relative bacterial/archaeal abundances in these samples using 16S rRNA gene amplicon sequencing, followed by time series analysis as described in references 9 and 11, we were able to determine interacting taxa. From the biopsy sampling, we observed six different categories of spatial distributions along the GI tract. We then compared the taxa that were found to be significantly interacting from the time series analysis with the cooccurrence information from the biopsy samples and found that interacting species were significantly more likely to cooccur. These findings shed light on the ecological segregation along a gastrointestinal tract and the relationship between these communities and the more easily obtainable fecal samples. The duration and frequency of sampling allowed for relating cooccurrence modeling with time series analysis in that very similar information can be obtained, but for both approaches, it is necessary to have a relatively large number of samples. Both cooccurrence and time series modeling do provide very similar information about which taxa are involved in pairwise interactions; however, time series modeling can describe the asymmetry of these interactions much more fully. Overall, these findings can be used for future experimental designs that address both the complexities of biotic interactions in GI microbial communities and the spatial distribution of community members.

RESULTS

Overall diversity and spatial structure along the intestinal tract. Biopsy samples were obtained, in triplicate, from a healthy adult at seven sites along the GI tract: terminal ileum (TI), ileocecal valve (IV), ascending colon (AC), transverse colon (TC), descending colon (DC), sigmoid colon (SC), and rectum. We further obtained 139 daily fecal samples from the same individual collected both before and after biopsy sam-
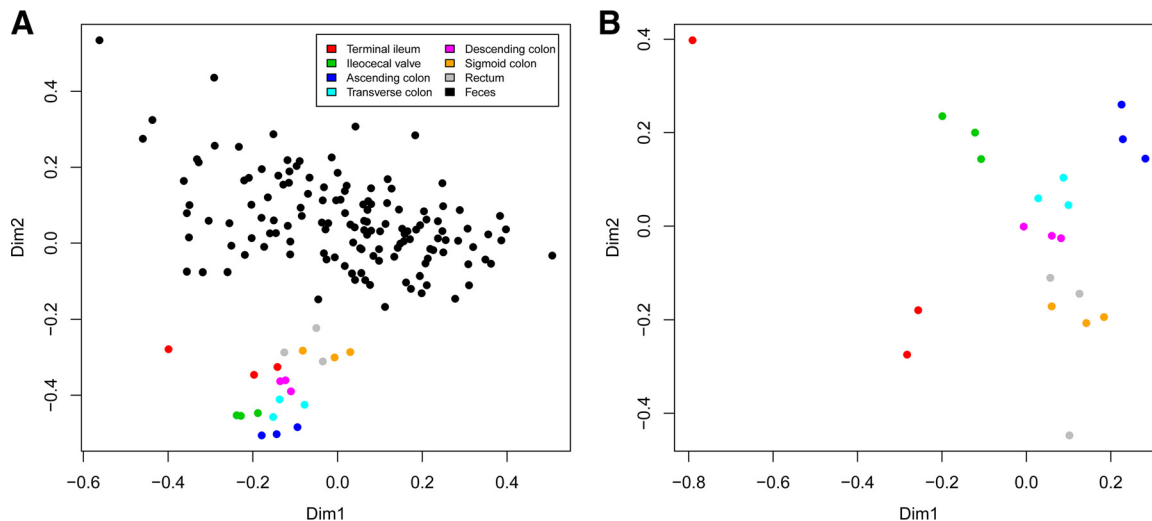
FIG 1 Nonmetric multidimensional scaling (NMDS) plots of fecal and biopsy sample-based Bray-Curtis distances computed from the relative abundances of the 258 most prevalent OTUs. (A) NMDS plot displaying fecal and biopsy samples from the main study individual. Colors represent the different sampling sites, and the distance between the circles represent similarity of the assemblage of OTUs comprising a sample. Clear separation was observed between the fecal samples and the biopsy samples ($P \ll 0.001$ by ANOSIM). Dim1, dimension 1. (B) NMDS plot of the biopsy samples obtained in triplicate from seven different locations along the GI tract. Significant structuring was observed between the anatomically distinct sampling sites ($P \ll 0.001$ by ANOSIM).

pling. 16S rRNA gene amplicon sequencing of the 160 samples generated a total of 22,302,434 reads. The mean read number per sample was 139,390 (standard deviation [SD], 71,119). A total of 2,238 operational taxonomic units (OTUs) were found using the criterion of 97% sequence identity for clustering to an OTU (see Table S1 in the supplemental material). The mean read depth of the biopsy samples was lower (59,831 ± 39,922 [SD]) than for the fecal samples. We attribute the lower number of reads to the relatively small amount of sample material in the biopsy samples.

For the following analysis, we applied common scaling to the size of the smallest biopsy sample library (10,662 reads). OTUs were filtered so that it had to be observed in at least 0.1% relative abundance (11 reads) in at least three samples (equivalent to the number of replicate biopsy samples taken at an anatomically distinct site), leaving the 258 most prevalent OTUs. Within the main study individual, the biopsy samples obtained were clearly distinct from the fecal samples (Fig. 1A, $P \ll 0.001$ by analysis of similarity [ANOSIM]). Highly significant structuring ($P \ll 0.001$ by ANOSIM) was observed for the biopsy samples depending on sampling location, with the main axis of variation separating the TI from the AC (Fig. 1B). We repeated the analysis using OTU tables collapsed to the genus level in order to test the robustness of the structuring. Both for the data set including both fecal and biopsy samples and also for biopsy samples alone, the genus-level analysis resulted in Bray-Curtis distance matrices that correlated with the OTU-level distance matrices with Pearson coefficients of >0.96. In order to put our data in a broader context, the biopsy and fecal samples were compared with 15 fecal samples obtained sequentially from three healthy adults. These 15 samples formed three distinct clusters clearly separated from both the fecal and biopsy samples of the main subject (Fig. S1, $P < 0.001$ by ANOSIM).

At the phylum level, we observed an inverse relationship between the two most dominant phyla, *Bacteriodetes* and *Firmicutes*, with sigmoid-shaped patterns of mean relative abundances from the TI to the rectum (Fig. S2). The mean relative abundances for *Actinobacteria* decreased from the TI to the DC, and the mean relative abundances for *Tenericutes* increased from the TI to the rectum. We observed the highest OTU diversity (Shannon entropy) in the TI. The diversity dropped precipitously toward the AC and then increased steadily toward the rectum (Fig. 2). Interestingly, this pattern was less related to overall OTU richness (Fig. S3A) than it was to community evenness (Fig. S3B).
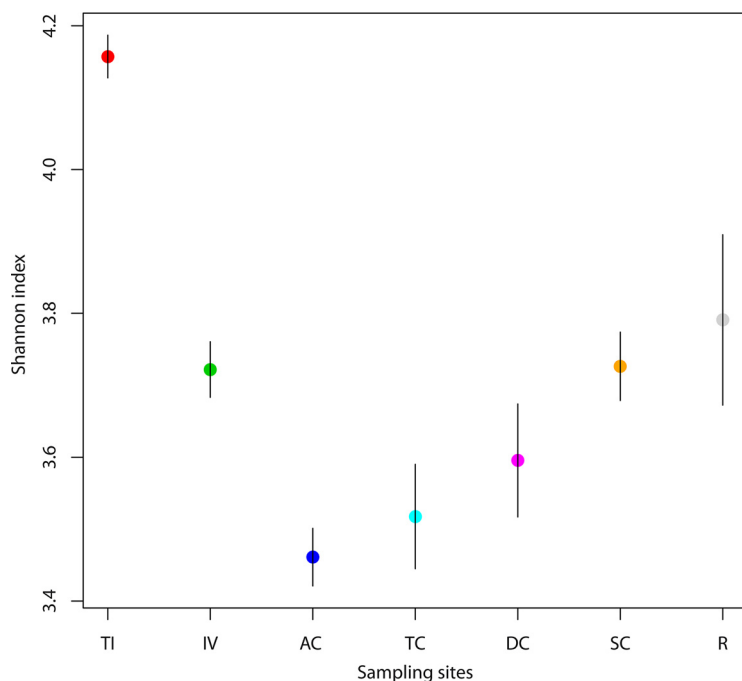
FIG 2 Bacterial diversity along the intestine as represented by the Shannon index. Bacterial diversity decreases from the terminal ileum to the ascending colon and then increases from the ascending colon until the rectum. Each circle represents the mean Shannon index for each sampling site, and gray bars represent the standard errors. The sampling locations are indicated on the x axis. Sampling site abbreviations: TI, terminal ileum; IV, ileocecal valve; AC, ascending colon; TC, transverse colon; DC, descending colon; SC, sigmoid colon; R, rectum.

Spatial patterns along the intestinal tract. The application of the above-described filtering procedure to the biopsy samples alone reduced the number of OTUs from 2,238 to the 136 most predominant OTUs (Table S1), accounting for an average of 93.5% (SD, 2.4%) of total per-sample reads. The distributions of these OTUs along the GI tract were relatively uniform (Fig. S4), but for several of the OTUs, we did observe significant spatial structure. We assigned these OTUs to six general categories (Fig. 3). For the monotonous gradient ascending (MGA) and monotonous gradient descending (MGD), we computed linear models under two basic assumptions. The GI tract can be considered a continuum or a series of physiologically distinct segments. In the former case, we regressed relative abundances of an OTU on the depth of sampling (centimeters from the anus), whereas in the latter, we assume equidistance between sampling sites by using the numbers 1 to 7 as the independent variable. Using both assumptions, we obtain similar results with 44 OTUs with significant ($P < 0.05$) trends under the first assumption and 59 OTUs with significant trends under the second assumption (Table S1). Similar numbers of OTUs were categorized as MGA (29 OTUs) and MGD (30 OTUs). Eighteen OTUs were classified as either gradient ascending with breakpoint (GAB) (5 OTUs) or gradient descending with breakpoint (GDB) (13 OTUs) with significant ($P < 0.05$) breakpoints in the AC (Table S1). We also observed lower numbers of OTUs with significant breakpoints in the TC (6 OTUs) and the DC (3 OTUs). We used exact tests for comparing one sampling site with all other sites in order to identify OTUs with significantly different occurrence ($P < 0.01$; false-discovery rate [FDR] < 0.05), and we categorized these as either habitat specialists (HS) or habitat avoiders (HA). We observed 36, 8, 36, and 2 OTUs with significantly different abundances in the TI, IV, AC, and rectum, respectively. No OTUs with significantly different abundances were observed in the TC, DC, or SC (Table S1).

Cooccurrence modeling of the biopsy and fecal samples and time series analysis of the fecal samples. We also analyzed 139 fecal samples obtained daily from the main study individual in order to identify interacting OTUs and relate these interactions
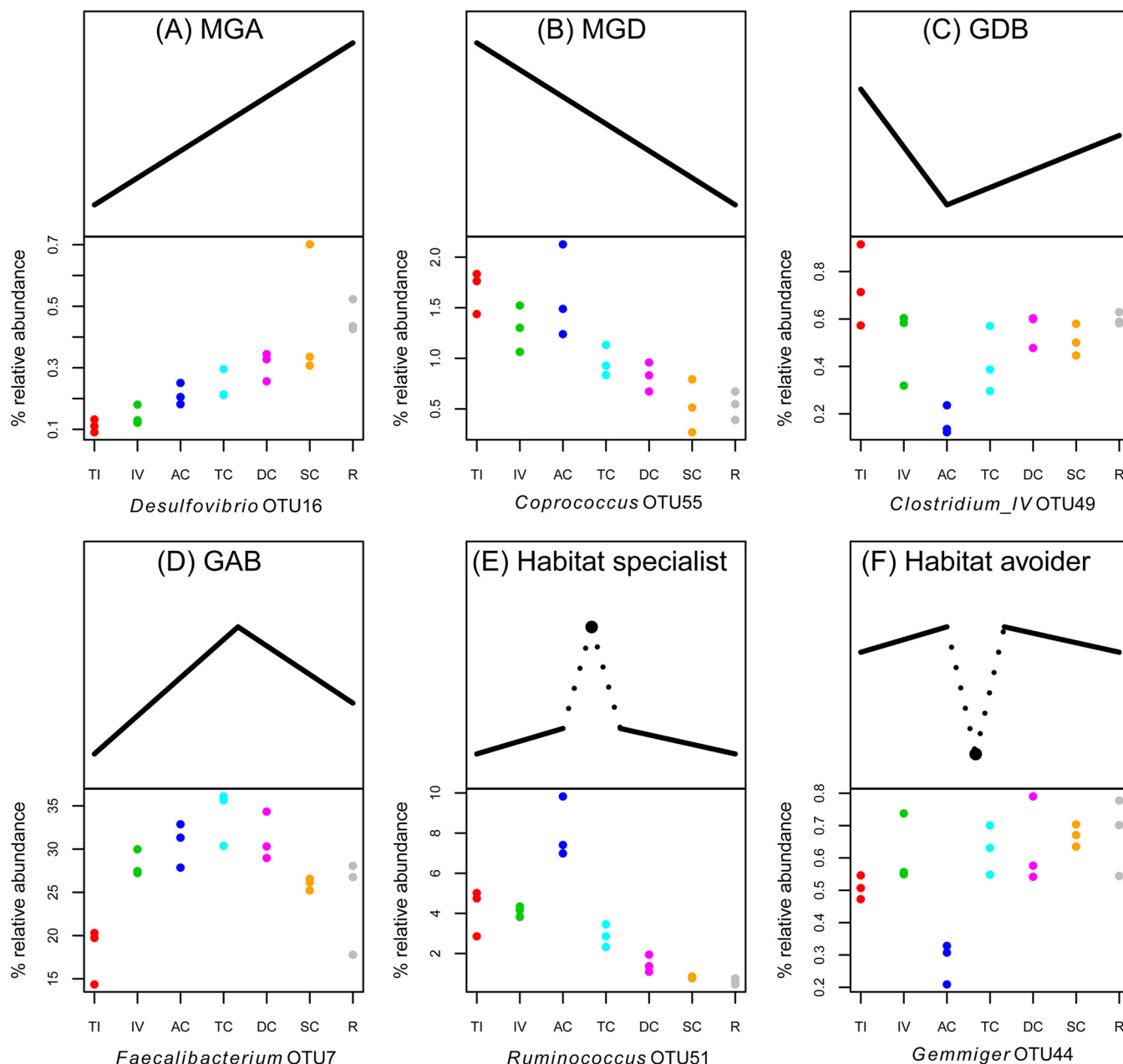
FIG 3 Examples of the six general categories of spatial structure observed along the GI tract. The top half of each panel illustrates the general patterns in relative abundance of a hypothetical OTU along the GI tract for each of the six different categories. (A) Monotonous gradient ascending (MGA). (B) Monotonous gradient descending (MGD). MGA and MGD refer to a significant positive or negative linear trend, respectively, in the relative abundances of an OTU along the intestinal tract, from the terminal ileum (TI) to the rectum (R). (C) Gradient descending with break point (GDB). (D) Gradient ascending with breakpoint (GAB). For GDB and GAB categorization, an OTU follows a "broken stick" pattern of relative abundances, where the linear regression coefficient changes sign in a segment along the GI tract. (E and F) Habitat specialists (HS) (E) and habitat avoiders (HA) (F). HS and HA refer to OTUs that have a significantly elevated or lowered relative abundance, respectively, in a particular segment of the intestine. The bottom half of each panel presents an example of an OTU identified in this study (along with a putative taxon assignment) that matches each of the observed categories of spatial patterning. The x-axis abbreviations are the same as those in Fig. 2. Note that using the above criteria, some OTUs can be placed into more than one category, e.g., OTU51 can be classified both as MGD and as HS (Fig. 3E).

to the spatial patterns observed in the biopsy sample data. We focused on the most prevalent and consistently observed OTUs by filtering the data such that an OTU had to constitute at least 0.05% of the reads (27 reads) in at least 90% of samples, reducing the number of OTUs from 2,238 to 76. The resulting data set represented an average of 78.3% (SD, 4.9%) of total per-sample reads, showing no major lasting shifts in community structure, and thus indicating that the data would be suitable for time series analysis (Fig. S5).
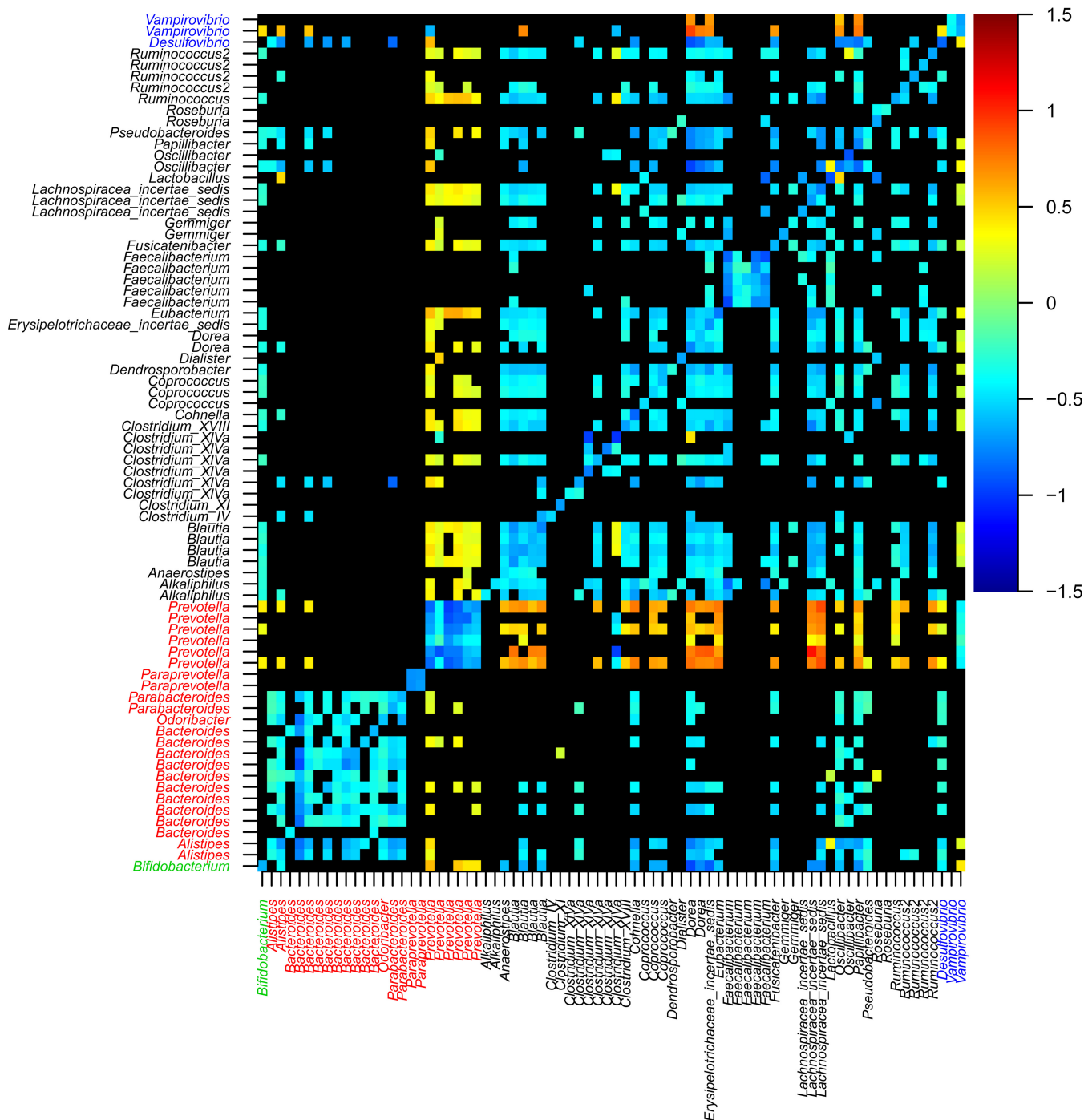
FIG 4 Bacterial interactions between 97% identified OTUs identified in the main study individual. The heat map shows the strength and direction of highly significant interactions after Benjamini-Hochberg correction for multiple testing ($P < 0.01$). Dependent variables are shown along the y axis, and independent variables are shown along the x axis, i.e., if you follow the column of a given OTU upward from the x axis until you reach a colored cell, that cell indicates the effect of the given OTU (independent) on the OTU indicated on the y axis (dependent). The color key on the right-hand side indicates the sign and magnitude of interactions that were significant. Cells representing nonsignificant relationships are black. Taxonomic assignments to the genus level are colored according to the phylum: green for *Actinobacteria*, red for *Bacteroidetes*, black for *Firmicutes*, and blue for *Proteobacteria*. Table S1 in the supplemental material lists the taxonomic assignments, in the order shown, with the matching OTU designations.

We first computed a set of pairwise interaction models using the time series analysis approach (Fig. 4). Only 27% (1,345 of 5,700) of all the potential pairwise interactions were found to be significant at a 99% confidence level. Randomization of the order of the values in the independent variables in the time series models resulted in a set of
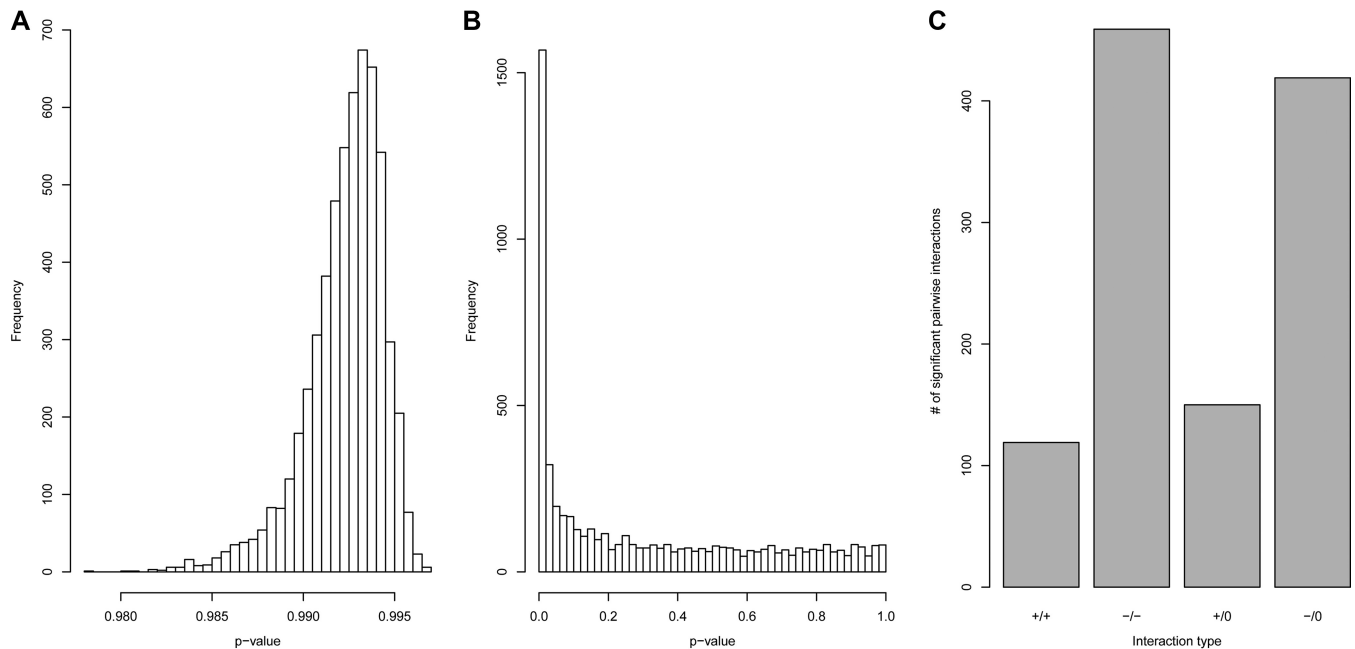
FIG 5 (A) Distribution of Benjamini-Hochberg-corrected *P* values resulting from recomputing the full set of time series models using randomly permutated dependent variables. (B) Distribution of Benjamini-Hochberg-corrected *P* values from the full set of original time series models. (C) Interaction categories from the time series analysis of the fecal sample data. The four categories are indicated on the *x* axis as follows: +/+ for cooperation, −/− for competition, +/0 for commensalism, and −/0 for amensalism. The *y* axis indicates the number of observed significant interactions in the specified categories.

P values normally distributed around 0.5. After applying the Benjamini-Hochberg correction for testing multiple hypotheses, not a single model resulting from variable randomization was found to be significant (Fig. 5A), as opposed to the original set of models with a distribution of *P* values that was highly skewed toward the low range (Fig. 5B). Of the interactions that were found to be significant, we observed four different categories of pairwise interaction (cooperation, competition, commensalism, and amensalism) in different proportions (Fig. 5C). The majority of the significant interactions were competitive, in line with previous observations (9, 14). Additionally, competition was more intense (i.e., regression coefficients were more strongly negative) for OTU pairs within a single phylum than between OTUs belonging to different phyla (mean coefficients of −0.2 versus 0.01; $P \ll 0.001$ by one-sided Wilcoxon's rank sum test). This observation was even more pronounced when comparing within-genus interaction to interactions between OTUs belonging to different genera (mean coefficients of −0.29 versus −0.08; $P \ll 0.001$ by one-sided Wilcoxon's rank sum test).

In order to compare the methodologies, we also computed a set of pairwise interaction models using a cooccurrence approach based on contemporaneous correlations using both Spearman correlations between relative OTU abundances as well as the SparCC algorithm for computing Pearson correlations between transformed OTU counts (15) (Fig. S6). There was strong agreement between the time series modeling approach and both cooccurrence modeling techniques, with highly significant negative correlations between the regression coefficients from the time series models and the correlation coefficients from both the Spearman and SparCC approach (Spearman's rho = −0.85 and Pearson's *r* = −0.70, respectively; $P \ll 0.001$ for both comparisons). Interestingly, the correlation coefficients corresponding to significant time series models follow a strictly bimodal distribution in which the interactions discovered from the time series modeling that are negative are associated with positive contemporaneous correlations, while positive interactions are associated with negative correlations (Fig. 6A and C), while nonsignificant models follow a roughly normal distribution centered around zero (Fig. 6B and D).

We then compared the time series models with cooccurrence models based on spatial correlations between the same 76 OTUs in the biopsy samples (Fig. S7). In order
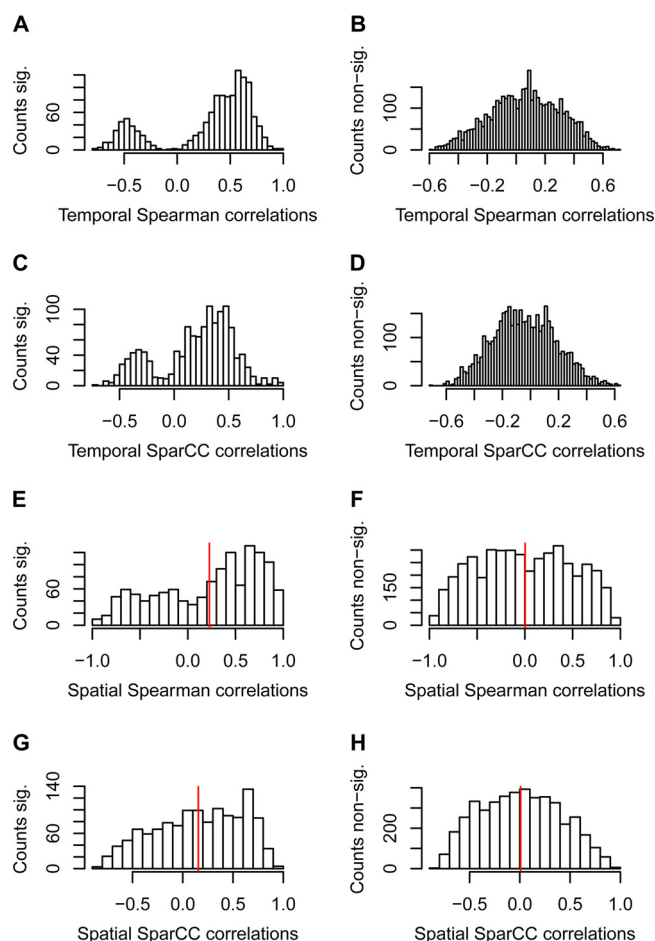
FIG 6 Distributions of pairwise correlation coefficients from cooccurrence models computed from temporal data (fecal) and spatial data (biopsy samples; means of three replicates). For each pair of panels, correlation coefficients were partitioned into significant (A, C, E, and G) or nonsignificant (B, D, F, and H) based on the significance values from the full set of time series models. (A and B) Temporal Spearman correlations; (C and D) temporal SparCC correlations; (E and F) spatial Spearman correlations; (G and H) spatial SparCC correlations. The vertical red lines in panels E to H represent the distribution means. Note that panels E and G have a significant positive skew (means of 0.22 and 0.16, respectively; $P \ll 0.001$ for both tests), indicating correspondence between the spatial and temporal correlation patterns.

to check for agreement between the modeling approaches, we again compared the distribution of spatial correlation coefficients corresponding to significant pairwise time series models to the distribution of spatial correlation coefficients corresponding to nonsignificant time series models. If there were agreement between the two approaches, we should, as above, observe a significant positive skew in the distribution of the correlation coefficients corresponding to significant time series models, while for the nonsignificant models, the distribution should be centered around zero. This is based on the rationale that interactions are predominantly negative due to resource competition, which would result in a predominance of positive cooccurrence patterns along the intestinal tract. We observed a highly significant positive skew for both the Spearman (means of 0.22 and 0.002 for correlation coefficients corresponding to significant and nonsignificant time series models, respectively; $P \ll 0.001$ by Wilcoxon's rank sum test; Fig. 6E and F) and SparCC models (means of 0.156 and 0.008, respectively; $P \ll 0.001$; Fig. 6G and H), indicating competing taxa that are actually colocalized and interacting. As a further comparison, we also checked for agreement between cooccurrence models as based either on spatial (biopsy samples) or contemporaneous (fecal samples) correlations as well as correspondence between spatial correlations and interaction coefficients from time series modeling. We observed noisy

but significant positive correlations between contemporaneous and spatial correlation coefficients both when using the Spearman approach ($r = 0.29$ and $P \ll 0.001$ by Pearson's correlation test; Fig. S8A) and the SparCC approach ($r = 0.31$ and $P \ll 0.001$; Fig. S8C). Comparison of spatial correlations with time series modeling coefficients gave similar results ($r = -0.28$ and -0.21 for Spearman and SparCC models, respectively, $P \ll 0.001$ for both tests, Fig. S8B and D), but with the expected negative relationship, signifying a preponderance of negative interactions.

DISCUSSION

In this study, different DNA extraction methods were employed for the two sample types due to the limited amount of sample material in the biopsy samples (low DNA yield) and different requirements for purification (i.e., high levels of potential inhibitors to downstream PCR in the fecal samples). DNA extraction procedures can affect the observed distribution of OTUs, which is a known source of bias in microbiome studies (16, 17), although some reports have found this issue to be less serious (18, 19). Even though DNA extraction techniques differed between the biopsy samples and the fecal samples, we observed that, within an individual, the mucosa-associated microbiota clustered according to sampling site and seemed to cluster more closely with the feces as the distance from the anus decreased (Fig. 1A). Furthermore, we observed clustering by individual rather than by DNA extraction method (see Fig. S1 in the supplemental material). This is in line with what others have observed in that GI microbiomes are highly individual (4, 8).

There have been conflicting reports of differences in microbiota diversity along the GI tract (20, 21). However, we sampled a larger number of sites from the small intestine to the rectum. We report a very clear pattern with the highest diversity in the terminal ileum (TI), the lowest in the ascending colon (AC), and a gradual increase toward the rectum (Fig. 2). This is similar to the model postulated by Donaldson et al. (2), of higher diversity toward distal sections due to a decreasing gradient of antimicrobials toward the rectum. These structural patterns, as well as the inverse relationship between the relative abundances of *Bacteriodetes* and *Firmicutes* (Fig. S2), the two most dominant phyla, lend evidence to the notion that different segments of the intestine can promote distinct microbial community structures. We do not have a strong explanation for the reduced diversity observed in biopsy samples from the AC, but one possible hypothesis is that the AC represents a type of filter for promoting separation of the bacterial communities associated with the small intestine and the colon. This might provide a mechanism for the host to compartmentalize the functions provided by the GI micro-biota and may be related to the observation that OTU richness was relatively stable throughout the GI tract, while evenness increased from the AC to the rectum (Fig. S3).

Overall, in the biopsy samples, we found several OTUs that we could place into six different general categories of spatial occurrence. In addition to monotonous gradients in relative abundance (monotonous gradient ascending [MGA] and monotonous gradient descending [MGD]), which have been observed by others (2), we also observed several instances of "broken stick" patterns (gradient descending with breakpoint [GDB] and gradient ascending with breakpoint [GAB]), as well as OTUs with significantly higher or lower relative abundances in certain segments (habitat specialists [HS] or habitat avoiders [HA]). The OTUs that were found to have monotonous gradients in relative abundance along the intestine may be explained by the physiochemical gradients in pH, oxygen, nutrients, and antimicrobials, etc. (2), the categories GDB, GAB, HS, and HA suggest that other factors may be involved. However, there is little available information about the fine-scale environmental conditions along the GI tract. From a statistical standpoint, one would have more power to detect significant breakpoints in the transverse colon (TC), that being the midpoint of spatial sampling. However, we found that the majority of breakpoints were in the AC, suggesting that the environmental conditions in this compartment may strongly modulate the microbial community. This observation is also reflected in the diversity estimates of the AC.

Several of the most predominant OTUs displayed significant spatial structure along

the GI tract. The most represented OTU in the combined data set (OTU1, 100% BLAST identity to *Prevotella copri*) was significantly (linear model $P < 0.01$) more prevalent in the upper GI tract. The most dominant taxa in the biopsy samples (OTU7, 99% BLAST identity to *Faecalibacterium prausnitzii*), was a GAB with an AC breakpoint. Interestingly, *Sutterella* (OTU43, 100% BLAST identity to *Suterella* sp. strain Marseille-P2435 [accession no. LT223579]) and *Ruminococcus* (OTU51, 100% BLAST identity to *Ruminococcus faecis* strain Eg2 [accession no. NR_116747.1]) are both categorized as GAB with the breakpoint in the ascending colon. Both *Ruminococcus* and *Sutterella* GI tract colonization have been linked to autism spectrum disorder (22, 23). Another *Ruminococcus* (OTU27, 99% BLAST identity to *Ruminococcus bromii* strain ATCC 27255, NR_025930.1) was GDB with the breakpoint at the AC. This species has been described as a potential "keystone species" for resistant starch degradation in the human intestine (24).

Cooccurrence modeling is based on the rationale that negatively and positively correlated occurrence patterns arise from negative and positive interactions, respectively (25). In our analyses, we actually observed the opposite pattern, i.e., negatively interacting OTUs, as estimated by time series analysis, displayed positive contemporaneous correlations and vice versa. This result was robust to the algorithm used for computing correlations, and it was also reflected in the spatial correlation patterns observed for the biopsy samples. We do not have a definite explanation for these observations, but we offer the following hypothesis. If a pair of taxa have a preference for the same resource, they are likely to be found at locations where that resource is available, and thus have a positive cooccurrence pattern. On the other hand, they would have to compete for said resource, which would explain the preponderance of negative regression coefficients in the time series models. This hypothesis also favors habitat filtering rather than species assortment as the main community-level assembly rule, which is in agreement with previous research (26).

As observed in previous work (9), negative interactions were more intense between more closely related OTUs. This notion goes all the way back to Darwin (27) and forms the basis of the principle known as "reverse ecology" which uses metabolic overlap deduced from genome sequences to infer ecological interaction (26). This matches in particular the high levels of competition that we observed within the genera *Bacteroides* and *Parabacteroides*. Both the time series and cooccurrence modeling identified large numbers of positive interactions involving OTUs classified as *Prevotella* and OTUs from other genera, but the *Prevotella* had relatively few significant interactions with OTUs classified as *Bacteroides*. *Prevotella* spp. have been shown to often have a negative contemporaneous correlation in relative abundance with *Bacteroides* (28–30). We also made this observation. Taken together, these results suggest that the interaction between *Prevotella* and *Bacteroides* is not in fact directly antagonistic but is instead mostly mediated by other factors such as diet, an interpretation which is in concordance with work suggesting that certain dietary components enhance *Prevotella* abundance (12, 29, 31). An OTU had *Bdellovibrio* spp. (85% BLAST identity to *Vampirovibrio chlorellavorus* strain ICPB 3707), a predatory bacterium, as its closest known relatives. While negatively affecting *Prevotella*, this bacterium is positively affected by several other genera, supporting the classification as a gut predatory bacterium from an ecological basis.

Microbiome studies of the GI tract are almost commonplace. 16S rRNA gene amplicon sequencing provides a limited view of complex bacterial communities. However, there remain many unanswered questions that these types of surveys are well equipped to address. Here we present data showing clear spatial structure in the microbiota associated with distinct regions of the GI tract. We also identified OTUs that conformed to six general categories of spatial occurrence patterns. Finally, we have demonstrated that the spatial structure of the microbiota in the GI tract relates to the interaction structure inferred from the time series analysis of fecal samples. In a previous study (4), one fecal sample from each of three subjects was obtained at a time point 3 months after the biopsy samples were taken in order to lessen the potential influence of the colonic cleansing on the GI microbiota. Interestingly, we performed the

biopsy sampling during an approximate midpoint in the sampling period and observed no marked effect of bowel cleansing on the overall structuring of the microbiota (Fig. S5). We also observed significant general agreement between the interaction models derived from analyzing either biopsy samples or longitudinal fecal data, lending some support to the use of fecal samples to gain an understanding of the adherent microbiota. The strong accordance between cooccurrence and time series modeling approaches to the fecal data suggests that both approaches are valid for identifying biotic interactions. The time series approach has the benefit of being able to describe directionality in nonsymmetric interactions. However, combining both techniques can facilitate the discovery of interactions that are not directly antagonistic or facilitative but that are mediated by environmental factors, as may be the case of *Bacteriodetes* and *Prevotella* in this study.

MATERIALS AND METHODS

Sampling procedures. The study obtained ethical approval from the regional ethical committee of Norway, and the participants gave signed informed consent. The subjects were not taking any medication during the course of the study. The main study individual had not used antibiotics within the past 2 years or during the study; this individual underwent standard bowel cleansing with Fleet Phospho-Soda the evening before the colonoscopy. Endoscopically, the colonic mucosa appeared grossly normal. Colonic tissue biopsy samples (approximately 1 by 2 mm each) were collected in triplicate from the terminal ileum (TI; about 155 cm from the anus), ileocecal valve (IV; about 150 cm from the anus), ascending colon (AC; about 142 cm from the anus), transverse colon (TC; about 109 cm from the anus), descending colon (DC; about 64 cm from the anus), sigmoid colon (SC; about 20 cm from the anus), and rectum (about 10 cm from the anus) during colonoscopy. Samples were then placed in cryovials and snap-frozen in liquid nitrogen. The cryovials were transferred to an −80°C freezer for storage until further processing. Fecal samples were frozen immediately upon collection at −20°C and then transferred to −80°C until further processing.

DNA extraction. The biopsy samples were transferred to FastPrep tubes with 210 μl of 20% SDS, 500 μl phenol-chloroform-isoamyl alcohol (25:24:1), 500 μl H$_2$O, and ~200 mg acid-washed glass beads (≤106 μm) (Sigma-Aldrich). The FastPrep tubes were then shaken at 4 m/s for 1 min (FastPrep-24; MP Biomedicals) and centrifuged at 13,200 rpm for 5 min, and ~600 μl of the aqueous phase was transferred to a clean 1.5-ml microtube. Sixty microliters of ice-cold 3 M sodium acetate and 450 μl ice-cold isopropanol were added, and the solution was mixed. The tubes were then incubated at −20°C for 30 min and centrifuged at 13,200 rpm at 4°C for 15 min. The supernatant was then removed, and the pellet was washed with 1 ml ice-cold ethanol (EtOH) before resuspension in 20 μl buffer AE (Qiagen). An additional cleaning step was then carried out using molecular biology-grade glycogen (Thermo Fisher, Waltham, MA, USA) according to the manufacturer's instructions. Approximately 240 mg of fecal material was used for DNA extraction with the PowerSoil 96-well DNA isolation kit (Mo Bio Laboratories, Inc., Carlsbad, CA, USA).

Illumina sequencing and data processing. Library preparation for Illumina sequencing was conducted by the procedure of de Muinck et al. (13). Sequencing was done on an Illumina HiSeq 2500 apparatus (Illumina, San Diego, CA, USA) using the 250-bp paired-end rapid-run mode. Low-quality reads were trimmed and Illumina adapters were removed using Trimmomatic v0.36 (32) with default settings. Reads mapping to the PhiX genome (NCBI identifier [ID] or accession no. NC_001422.1) were removed using BBMap v36.02 (33). Demultiplexing of data based on the dual index sequences was performed using custom scripts available at github (https://github.com/arvindsundaram/triple_index-demultiplexing). Internal barcodes and spacers were removed using cutadapt v1.4.1 (34), and paired reads were merged using FLASH v1.2.11 (35) with default settings. Sequence files were then converted from fastq to fasta, and primers were trimmed from merged read ends. Further processing of sequence data was conducted using a combination of vsearch v2.0.3 (36) and usearch v8.1.1861 (37). Specifically, dereplication was performed with the "derep_fulllength" function in vsearch with the minimum unique group size set at 2. Operational taxonomic unit (OTU) clustering, chimera removal, taxonomic assignment, and OTU table building were carried out using the uparse pipeline (38) in usearch. Taxonomic assignment to the genus level was done against the RDP-15 training set. Samples from the main study individual were clustered as a single data set so that a given OTU number from a biopsy sample corresponds to the same OTU number in a fecal sample. Clustering was redone when including additional adult samples.

Statistical analyses. Read depths were normalized by common scaling (39). This entails multiplying each OTU count in a given library with the ratio of the smallest library size to the size of the individual library. This procedure replaces rarefying (i.e., random subsampling to the lowest number of reads), as it produces the library scaling one would achieve by averaging over an infinite number of repeated subsamplings. After library scaling, the data were filtered according to the criteria stated below or in each section of the results. All statistical analyses were done using R (40). Bray-Curtis distance matrix computation and analyses of similarity (ANOSIMs, 10,000 permutations) were carried out using the "vegan" package. Nonmetric multidimensional scaling (NMDS) was carried out using the "isoMDS" function in the MASS package. Exact tests for differences in means between two groups of negative binomially distributed counts (41–43) were performed using the edgeR package (44). Although originally developed for analysis of differential expression in RNA sequencing experiments, this method has been

shown to perform excellently in identifying enriched OTUs in 16S rRNA gene amplicon sequencing experiments as well (39). For these tests, we did not use common scaled data, as library size differences are accounted for as part of the statistical algorithm. Specific filtering criteria were used for exact tests in order to focus on the most abundant OTUs. For an OTU to be included in this analysis, it had to be observed at a relative abundance of at least 0.1% in at least three samples. This filtering procedure reduced the number of OTUs analyzed from 2,238 to 136. For the diversity measures, we did not perform filtering of the OTU tables other than library size scaling and singleton removal. This procedure removed any significant relationship between library size and observed OTU diversity. Time series modeling was performed as described in reference 9. Briefly, the dynamics of each OTU was modeled according to the equation, $x_{i,t+1} - x_{i,t} = \alpha_{i,j} + \beta_{i,j}x_{j,t}$, where x_i,t is the log relative abundance of taxon i at time t, $\alpha_{i,j}$ are intercept terms, $\beta_{i,j}$ are linear regression coefficients, and x_j,t are log relative abundances of taxon j at time t. The regression coefficients are then interpreted as describing the biotic interaction between OTUs i and j. The strength of the interaction is proportional to the size of the coefficient, with a positive coefficient signifying a cooperative or commensal interaction and a negative coefficient signifying competition or amensalism. Each element in the resulting interaction matrix is estimated from separate models (i.e., variable A as a function of variable B or B as a function of A) with different dependent and independent variables, resulting in a nonsymmetric matrix. This allows for identification of nonreciprocal interactions (commensalism or amensalism) or pairwise interactions with opposite signs (parasitism), as well as reciprocal pairwise relationships (cooperation and competition). The estimated within-OTU interactions (diagonal of the interaction matrix) should be interpreted with caution, as the independent variable term is also part of the dependent variable, and these coefficients are excluded from further analysis. The total number of equations in a system is equal to n^2, where n is the total number of taxa. This approach does not capture relationships that are strongly nonlinear that could be modeled, e.g., by generalized additive models, but previous work has shown linear regression to be a good approximation (11). Time series models were computed only for the most abundant OTUs (observed in at least 90% of samples), since the modeling technique is effective only for OTUs observed consistently and at levels that provide good estimations of relative abundance. Model P values were corrected for multiple hypotheses tested by applying the Benjamini-Hochberg adjustment in order to reduce the false-discovery rate (45). In order to test the robustness of the time series approach, the full set of models was recomputed using random permutations of the dependent term ($x_{j,t}$) 100 times, and P values were averaged over the replicated matrices. Cooccurrence modeling was conducted using two alternative methodologies. First, we computed pairwise Spearman correlation coefficients between relative OTU abundance profiles (25). Since studies have shown that simple Spearman correlations can produce spurious results when analyzing microbiome data (15), we also apply a more sophisticated approach. This is known as SparCC (15) and is designed for large, sparse, abundance matrices of the kind obtained in microbiome studies. We used the "sparcc" implementation in the mothur software suite (46), with default settings, in order to obtain a more robust estimation of the cooccurrence matrix. We use both of these cooccurrence modeling approaches to compute contemporaneous and spatial correlations. Contemporaneous correlations were computed for the longitudinally obtained fecal samples and provide information about which OTUs tend to cooccur over time. Spatial correlations were computed for the biopsy samples and provide information about which OTUs tend to reside in the same locations along the GI tract. In contrast to the time series modeling approach, correlation mapping forces a symmetrical pairwise interaction matrix and thus cannot identify nonreciprocal interactions or interactions with opposite signs.

Accession number(s). All sequence data used in this study have been made available at the NBCI Sequence Read Archive under the BioProject ID PRJNA387407.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mSystems.00086-17.

FIG S1, EPS file, 0.01 MB.
FIG S2, EPS file, 0.01 MB.
FIG S3, PDF file, 0.01 MB.
FIG S4, PDF file, 0.04 MB.
FIG S5, PDF file, 0.1 MB.
FIG S6, PDF file, 0.1 MB.
FIG S7, PDF file, 0.1 MB.
FIG S8, PDF file, 0.1 MB.
TABLE S1, XLSX file, 0.1 MB.

REFERENCES

1. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL, Rosenbaum M, Gordon JI. 2013. The long-term stability of the human gut microbiota. Science 341:1237439. https://doi.org/10.1126/science.1237439.
2. Donaldson GP, Lee SM, Mazmanian SK. 2016. Gut biogeography of the bacterial microbiota. Nat Rev Microbiol 14:20–32. https://doi.org/10.1038/nrmicro3552.
3. Sartor RB. 2015. Gut microbiota: optimal sampling of the intestinal microbiota for research. Nat Rev Gastroenterol Hepatol 12:253–254. https://doi.org/10.1038/nrgastro.2015.46.
4. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. 2005. Diversity of the human intestinal microbial flora. Science 308:1635–1638. https://doi.org/10.1126/science.1110591.
5. Aguirre de Cárcer D, Cuív PO, Wang T, Kang S, Worthley D, Whitehall V, Gordon I, McSweeney C, Leggett B, Morrison M. 2011. Numerical ecology validates a biogeographical distribution and gender-based effect on mucosa-associated bacteria along the human colon. ISME J 5:801–809. https://doi.org/10.1038/ismej.2010.177.
6. Zhang Z, Geng J, Tang X, Fan H, Xu J, Wen X, Ma ZS, Shi P. 2014. Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. ISME J 8:881–893. https://doi.org/10.1038/ismej.2013.185.
7. Hong PY, Croix JA, Greenberg E, Gaskins HR, Mackie RI. 2011. Pyrosequencing-based analysis of the mucosal microbiota in healthy individuals reveals ubiquitous bacterial groups and micro-heterogeneity. PLoS One 6:e25042. https://doi.org/10.1371/journal.pone.0025042.
8. Lavelle A, Lennon G, O'Sullivan O, Docherty N, Balfe A, Maguire A, Mulcahy HE, Doherty G, O'Donoghue D, Hyland J, Ross RP, Coffey JC, Sheahan K, Cotter PD, Shanahan F, Winter DC, O'Connell PR. 2015. Spatial variation of the colonic microbiota in patients with ulcerative colitis and control volunteers. Gut 64:1553–1561. https://doi.org/10.1136/gutjnl-2014-307873.
9. Trosvik P, de Muinck EJ. 2015. Ecology of bacteria in the human gastro-intestinal tract—identification of keystone and foundation taxa. Microbiome 3:44. https://doi.org/10.1186/s40168-015-0107-4.
10. Bashan A, Gibson TE, Friedman J, Carey VJ, Weiss ST, Hohmann EL, Liu YY. 2016. Universality of human microbial dynamics. Nature 534:259–262. https://doi.org/10.1038/nature18301.
11. Trosvik P, de Muinck EJ, Stenseth NC. 2015. Biotic interactions and temporal dynamics of the human gastrointestinal microbiota. ISME J 9:533–541. https://doi.org/10.1038/ismej.2014.147.
12. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, Collini S, Pieraccini G, Lionetti P. 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc Natl Acad Sci U S A 107:14691–14696. https://doi.org/10.1073/pnas.1005963107.
13. de Muinck EJ, Trosvik P, Gilfillan GD, Hov JR, Sundaram AYM. 2017. A novel ultra high-throughput 16S rRNA gene amplicon sequencing library preparation method for the Illumina HiSeq platform. Microbiome 5:68. https://doi.org/10.1186/s40168-017-0279-1.
14. Foster KR, Bell T. 2012. Competition, not cooperation, dominates interactions among culturable microbial species. Curr Biol 22:1845–1850. https://doi.org/10.1016/j.cub.2012.08.005.
15. Friedman J, Alm EJ. 2012. Inferring correlation networks from genomic survey data. PLoS Comput Biol 8:e1002687. https://doi.org/10.1371/journal.pcbi.1002687.
16. Cruaud P, Vigneron A, Lucchetti-Miganeh C, Ciron PE, Godfroy A, Cambon-Bonavita MA. 2014. Influence of DNA extraction method, 16S rRNA targeted hypervariable regions, and sample origin on microbial diversity detected by 454 pyrosequencing in marine chemosynthetic ecosystems. Appl Environ Microbiol 80:4626–4639. https://doi.org/10.1128/AEM.00592-14.
17. Albertsen M, Karst SM, Ziegler AS, Kirkegaard RH, Nielsen PH. 2015. Back to basics—the influence of DNA extraction and primer choice on phylogenetic analysis of activated sludge communities. PLoS One 10:e0132783. https://doi.org/10.1371/journal.pone.0132783.
18. Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ. 2012. Evaluation of methods for the extraction and purification of DNA from the human microbiome. PLoS One 7:e33865. https://doi.org/10.1371/journal.pone.0033865.
19. Rubin BE, Sanders JG, Hampton-Marcell J, Owens SM, Gilbert JA, Moreau CS. 2014. DNA extraction protocols cause differences in 16S rRNA amplicon sequencing efficiency but not in community profile composition or structure. Microbiologyopen 3:910–921. https://doi.org/10.1002/mbo3.216.
20. Li G, Yang M, Zhou K, Zhang L, Tian L, Lv S, Jin Y, Qian W, Xiong H, Lin R, Fu Y, Hou X. 2015. Diversity of duodenal and rectal microbiota in biopsy tissues and luminal contents in healthy volunteers. J Microbiol Biotechnol 25:1136–1145. https://doi.org/10.4014/jmb.1412.12047.
21. Stearns JC, Lynch MD, Senadheera DB, Tenenbaum HC, Goldberg MB, Cvitkovitch DG, Croitoru K, Moreno-Hagelsieb G, Neufeld JD. 2011. Bacterial biogeography of the human digestive tract. Sci Rep 1:170. https://doi.org/10.1038/srep00170.
22. Wang L, Christophersen CT, Sorich MJ, Gerber JP, Angley MT, Conlon MA. 2013. Increased abundance of Sutterella spp. and Ruminococcus torques in feces of children with autism spectrum disorder. Mol Autism 4:42. https://doi.org/10.1186/2040-2392-4-42.
23. Williams BL, Hornig M, Parekh T, Lipkin WI. 2012. Application of novel PCR-based methods for detection, quantitation, and phylogenetic characterization of Sutterella species in intestinal biopsy samples from children with autism and gastrointestinal disturbances. mBio 3:e00261-11. https://doi.org/10.1128/mBio.00261-11.
24. Ze X, Duncan SH, Louis P, Flint HJ. 2012. Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. ISME J 6:1535–1543. https://doi.org/10.1038/ismej.2012.4.
25. Faust K, Raes J. 2012. Microbial interactions: from networks to models. Nat Rev Microbiol 10:538–550. https://doi.org/10.1038/nrmicro2832.
26. Levy R, Borenstein E. 2013. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. Proc Natl Acad Sci U S A 110:12804–12809. https://doi.org/10.1073/pnas.1300926110.
27. Darwin C. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London, United Kingdom.
28. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Doré J, MetaHIT Consortium, Weissenbach J, Ehrlich SD, Bork P. 2011. Enterotypes of the human gut microbiome. Nature 473:174–180. https://doi.org/10.1038/nature09944.
29. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L, Li H, Bushman FD, Lewis JD. 2011. Linking long-term dietary patterns with gut microbial enterotypes. Science 334:105–108. https://doi.org/10.1126/science.1208344.
30. Ley RE. 2016. Gut microbiota in 2015: Prevotella in the gut: choose carefully. Nat Rev Gastroenterol Hepatol 13:69–70. https://doi.org/10.1038/nrgastro.2016.4.
31. David LA, Materna AC, Friedman J, Campos-Baptista MI, Blackburn MC, Perrotta A, Erdman SE, Alm EJ. 2014. Host lifestyle affects human microbiota on daily timescales. Genome Biol 15:R89. https://doi.org/10.1186/gb-2014-15-7-r89.
32. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
33. Bushnell B. 2016. BBMap short read aligner. SourceForge Media, La Jolla, CA. https://sourceforge.net/projects/bbmap.
34. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:10–12. https://doi.org/10.14806/ej.17.1.200.
35. Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957–2963. https://doi.org/10.1093/bioinformatics/btr507.
36. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584. https://doi.org/10.7717/peerj.2584.
37. Edgar RC. 2010. Search and clustering orders of magnitude faster than

BLAST. Bioinformatics 26:2460–2461. https://doi.org/10.1093/bioinformatics/btq461.

38. Edgar RC. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10:996–998. https://doi.org/10.1038/nmeth.2604.

39. McMurdie PJ, Holmes S. 2014. Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput Biol 10:e1003531. https://doi.org/10.1371/journal.pcbi.1003531.

40. R Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

41. Robinson MD, Smyth GK. 2007. Moderated statistical tests for assessing differences in tag abundance. Bioinformatics 23:2881–2887. https://doi.org/10.1093/bioinformatics/btm453.

42. Robinson SJ, Parkin IA. 2008. Differential SAGE analysis in Arabidopsis uncovers increased transcriptome complexity in response to low temperature. BMC Genomics 9:434. https://doi.org/10.1186/1471-2164-9-434.

43. Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11:R25. https://doi.org/10.1186/gb-2010-11-3-r25.

44. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140. https://doi.org/10.1093/bioinformatics/btp616.

45. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B Stat Methodol 57:289–300.

46. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541. https://doi.org/10.1128/AEM.01541-09.