

Spatial metagenomic characterization of microbial biogeography in the gut

Ravi U. Sheth^{1,2}, Mingqiang Li^{1,3}, Weiqian Jiang³, Peter A. Sims^{1,4,5}, Kam W. Leong^{1,3} and Harris H. Wang^{1,6*}

Spatial structuring is important for the maintenance of natural ecological systems^{1,2}. Many microbial communities, including the gut microbiome, display intricate spatial organization^{3–9}. Mapping the biogeography of bacteria can shed light on interactions that underlie community functions^{10–12}, but existing methods cannot accommodate the hundreds of species that are found in natural microbiomes^{13–17}. Here we describe metagenomic plot sampling by sequencing (MaPS-seq), a culture-independent method to characterize the spatial organization of a microbiome at micrometer-scale resolution. Intact microbiome samples are immobilized in a gel matrix and cryofractured into particles. Neighboring microbial taxa in the particles are then identified by droplet-based encapsulation, barcoded 16S rRNA amplification and deep sequencing. Analysis of three regions of the mouse intestine revealed heterogeneous microbial distributions with positive and negative co-associations between specific taxa. We identified robust associations between Bacteroidales taxa in all gut compartments and showed that phylogenetically clustered local regions of bacteria were associated with a dietary perturbation. Spatial metagenomics could be used to study microbial biogeography in complex habitats.

The local spatial organization of the gut microbiome influences various properties including colonization^{10,17–19}, metabolism¹¹, host-microbe and intermicrobial interactions²⁰, and community stability^{1,21,22}. However, current microbiome profiling approaches such as metagenomic sequencing require homogenization of input material, which means that underlying spatial information is lost. While imaging techniques can reveal spatial information, they rely on hybridization with short DNA probes of limited spectral diversity, yielding data with low taxonomic resolution, and often require extensive empirical optimization^{15,23}. Bacteria are densely packed in communities, limiting identification and analysis of individual cells using visual methods⁸. Although imaging approaches can simultaneously profile simple synthetic communities composed of a small number of cultivable species^{16,17} (for example, six in ref. ¹⁷), they are challenging to scale to complex and diverse natural microbiomes. Therefore, an unbiased method for high-taxonomic-resolution and micrometer-scale dissection of natural microbial biogeography is needed to better study the role of the gut microbiome in health and disease.

In macroecology, plot sampling is used to study the spatial organization of large ecosystems, which are otherwise impractical

to fully characterize. By surveying many smaller plots from a larger region, one can tractably delineate local distributions of species and statistically infer fundamental properties of global community organization and function. Inspired by this approach, we developed MaPS-seq, a multiplexed sequencing technique that analyzes microbial cells in their native geographical context to statistically reconstruct the local spatial organization of the microbiome (Fig. 1a).

To perform MaPS-seq, an input sample is first physically fixed and the microbiota is immobilized via perfusion and in situ polymerization of an acrylamide polymer matrix, which also contains a covalently linked reverse 16S ribosomal RNA (rRNA) amplification primer. The embedded sample is then fractured via cryo-bead beating, subjected to cell lysis and passed through nylon mesh filters for size selection to yield cell clusters or particles of desired and tunable physical sizes (by utilizing different mesh filter sizes). The resulting clusters contain genomic DNA immobilized in the original arrangement, preserving local spatial information. Next, a microfluidic device is used to co-encapsulate these clusters with gel beads, each containing uniquely barcoded forward 16S rRNA amplification primers. Primers are photocleaved from the beads and clusters, genomic DNA is released from the clusters by triggered degradation of the polymer matrix within droplets and PCR amplification of the 16S V4 region is performed. Droplets are then broken apart, and the resulting library is subjected to deep sequencing. Sequencing reads are filtered and grouped by their unique barcodes, which yield the identity and relative abundance (RA) of bacterial operational taxonomic units (OTUs) within individual cell clusters of defined size (Methods; Supplementary Figs. 1–4).

To rigorously test the feasibility of this spatial metagenomics approach, we first generated separate cluster communities from either homogenized mouse fecal bacteria or *Escherichia coli* (Methods; Supplementary Fig. 5) and profiled them with MaPS-seq. The resulting data revealed that the majority of the detected barcodes mapped uniquely to their respective initial communities with minimal mixing (Fig. 1b; 4.3% mixed) and negligible contamination introduced during sample processing (<0.2% of reads). In addition, the average abundance of taxa across individual fecal clusters obtained by enzymatic lysis and droplet PCR displayed good correlation with measurements from standard mechanical cell lysis and bulk 16S PCR (Fig. 1c; Pearson's correlation $r=0.76$). A replicate community-mixing experiment with new particles of a smaller size confirmed the technical performance of the approach

¹Department of Systems Biology, Columbia University Medical Center, New York, NY, USA. ²Integrated Program in Cellular, Molecular, and Biomedical Studies, Columbia University, New York, NY, USA. ³Department of Biomedical Engineering, Columbia University, New York, NY, USA. ⁴Sulzberger Columbia Genome Center, Columbia University Medical Center, New York, NY, USA. ⁵Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, NY, USA. ⁶Department of Pathology and Cell Biology, Columbia University Medical Center, New York, NY, USA.
*e-mail: hw2429@cumc.columbia.edu

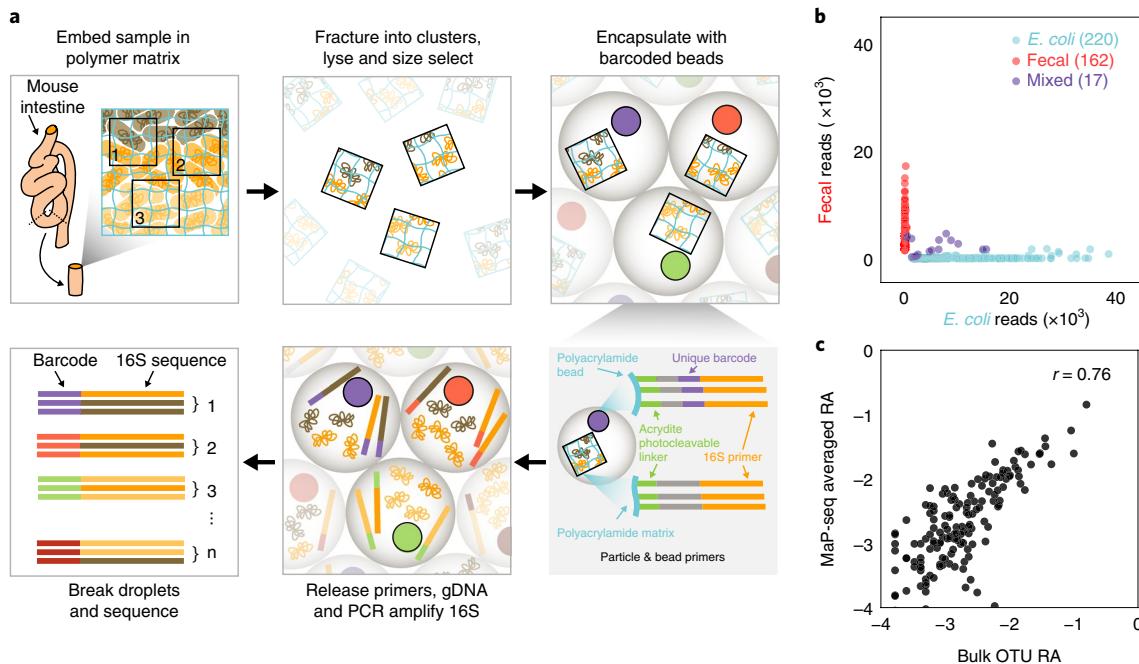


Fig. 1 | MaPS-seq and quality control. **a**, Schematic of the MaPS-seq technique for micrometer-scale plot sampling of microbiome samples. **b**, MaPS-seq profiling of a mixture of clusters prepared from homogenized fecal bacteria or *E. coli*. The number of reads for each of 399 barcodes belonging to either the *E. coli* OTU or fecal OTUs is displayed as a scatterplot. **c**, Correlation between OTU RA measurements obtained by standard bulk 16S sequencing of the same homogenized fecal community and MaPS-seq OTU RA measurements averaged across individual homogenized fecal clusters ($n=162$ clusters with $<10\%$ *E. coli* reads). All RAs are plotted on a \log_{10} scale; 152 OTUs with $RA > 0.01\%$ are displayed. r indicates Pearson's correlation.

(Supplementary Fig. 5g,h). Together, these results indicate that MaPS-seq accurately measures bacterial identity and abundance within individual spatially constrained cell clusters.

To explore the utility of spatial metagenomics in complex communities, we applied MaPS-seq to the mouse colonic microbiome. We generated and characterized cell clusters (median diameter of $\sim 30 \mu\text{m}$) from a segment of the distal colon (including both epithelium and digesta) of a mouse that had been fed a standard plant polysaccharide diet, yielding 1,406 clusters passing strict quality filtering criteria across two technical replicates (Methods; Fig. 2a and Supplementary Fig. 6a). In total, 236 OTUs were identified, and their prevalence across clusters was highly correlated with bulk abundance obtained by standard 16S sequencing, which implies that more abundant taxa are also physically dispersed over more space (Fig. 2b; Pearson's correlation $r=0.90$). The spatial distribution of taxa across clusters appeared mixed (median of nine OTUs per cluster), but some clusters contained only a few OTUs, indicating spatial aggregation or clumping in a fraction of the community (Fig. 2c). Moreover, this observed distribution of OTUs per cluster was significantly lower than that of clusters of the same size generated from homogenized fecal bacteria, which served as a control for a well-mixed community (Mann–Whitney U test, $P < 10^{-26}$). These results suggest that at the scale of tens of micrometers, individual taxa in the gut microbiome are neither fully mixed nor highly structured, but rather are heterogeneously distributed in mixed patches. Peristaltic mixing across the gut likely acts to decrease strong spatial segregation between taxa, but the weak but significant spatial structuring observed could nevertheless play an important role in the maintenance of high microbial diversity observed in the healthy gut^{1,22}.

We next explored whether these observed spatial distributions reflect specific associations between individual taxa that may result from processes such as positive or negative interspecies interactions (for example, cooperative metabolism²⁴ and contact-dependent killing²⁰) or local habitat filtering¹¹. Across abundant and prevalent

OTUs ($RA > 2\%$ in $>10\%$ of clusters, $n=24$), we assessed whether pairwise co-occurrences were detected more or less frequently than expected as compared to a null model of independent random assortment of OTUs (Methods; Fisher's exact test, $P < 0.05$; false discovery rate (FDR) = 0.05). Application of this strategy to the cluster-mixing control experiment confirmed our ability to accurately detect the expected positive and negative spatial associations (Supplementary Fig. 5f). Of 276 possible pairwise combinations of taxa in the murine colon, we detected 75 statistically significant associations between diverse taxa, the majority of which were positive (72/75) but relatively weak in magnitude (Fig. 2d and Supplementary Fig. 6b,c). The strongest co-occurrence was a positive association between abundant Bacteroidaceae and Porphyromonadaceae taxa from the Bacteroidales order (odds ratio 3.9, $P < 10^{-23}$). In addition, a small number of negative associations were observed, which could reflect antagonistic processes such as production of inhibitory factors or competitive exclusion.

The number of detected associations increased as more of the dataset was sampled, implying that detection of weaker relationships between less abundant taxa can be improved by analyzing more clusters (Supplementary Fig. 6d). Nonetheless, the detected associations showed good correspondence between technical replicates (Supplementary Fig. 6e). Importantly, despite high microbiome variability between hosts, the nature of the associations (that is, sign, magnitude and number) and some strong associations could be recapitulated in MaPS-seq profiling of a second cohoused mouse, such as the co-occurrence of Bacteroidales taxa (Supplementary Fig. 7). This characterization implies that individual taxa in the colon are organized in distinct and reproducible spatial relationships.

To further investigate how the spatial organization of the microbiota is influenced by environmental context, we applied spatial metagenomics along the gastrointestinal (GI) tract. The mammalian GI tract is composed of distinct anatomical regions with different pH levels, oxygen concentrations, host-derived

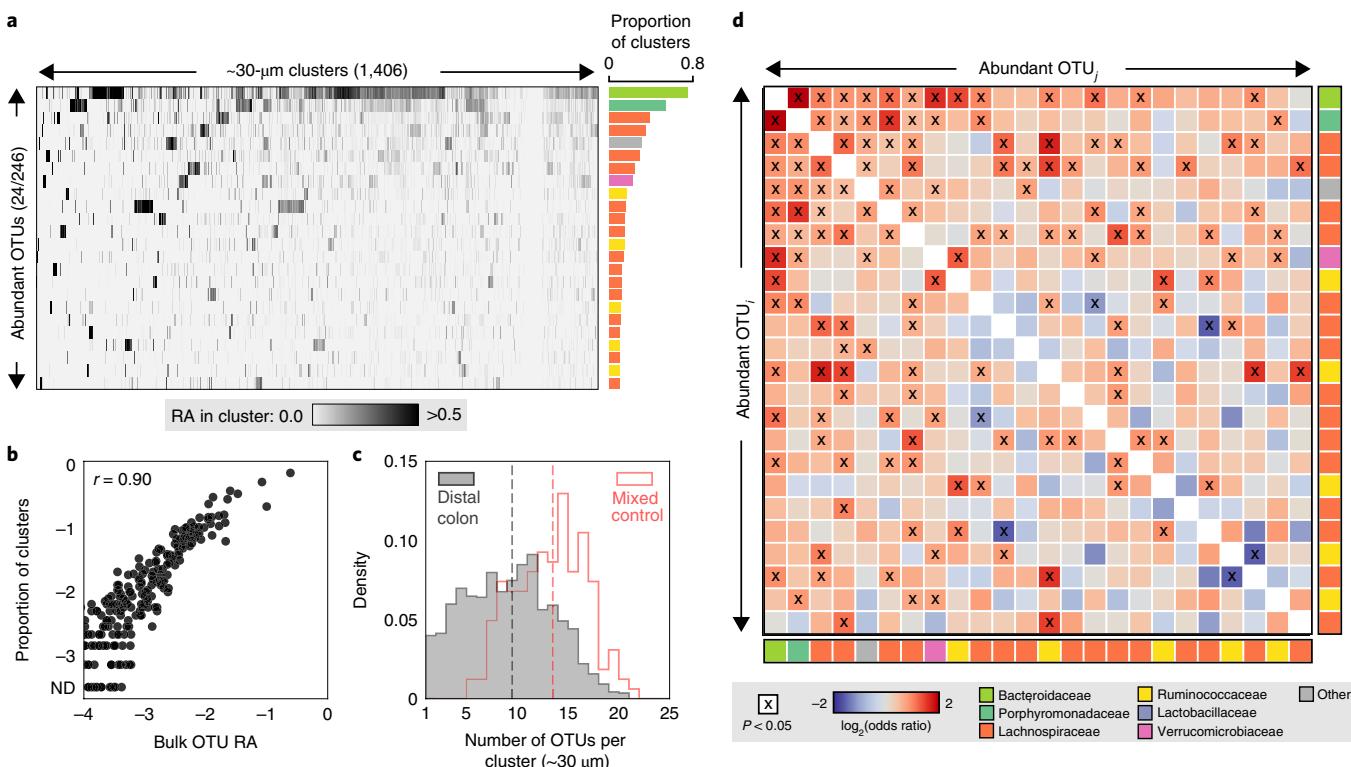


Fig. 2 | Spatial organization of the microbiota in the mouse distal colon. **a**, MaPS-seq profiling of distal colon clusters with a median diameter of $\sim 30 \mu\text{m}$. Raw RA data from MaPS-seq are displayed as a heat map; columns represent individual clusters ($n=1,406$) and rows represent abundant and prevalent OTUs (RA $> 2\%$ in $>10\%$ of all clusters; 24 of 246 detected OTUs) aggregated from two datasets from technical replicates of the same sample. Shading denotes the RA of individual OTUs in each cluster (linear scale); OTUs are sorted by decreasing prevalence (proportion of clusters in which each OTU has RA $> 2\%$), and clusters are clustered by Euclidean distance. The prevalence of each OTU across clusters is displayed to the right as a bar plot, and each bar is colored according to the OTUs assigned taxonomy at the family level (legend in **d**). **b**, Correlation between OTU RA measurements obtained by standard bulk 16S sequencing and the prevalence of OTUs in each cluster from the same sample (RA $> 2\%$ across $n=1,406$ clusters). ND, not detected (that is, RA $\leq 2\%$ in all clusters); 219 OTUs with $> 0.01\%$ RA as measured by bulk 16S sequencing are displayed. r indicates Pearson's correlation. **c**, Histogram of the number of OTUs per cluster (OTUs with RA $> 2\%$) shown for homogenized fecal clusters serving as a mixed control (red; $n=162$) and distal colon clusters (gray; $n=1,406$) of the same size. Dashed lines indicate the median value for each group. **d**, For each abundant and prevalent OTU pair (OTU_{ij} ; $n=24$ OTUs), spatial associations were calculated across $n=1,406$ clusters. Shading indicates the \log_2 -transformed odds ratio, and x indicates a statistically significant association (Fisher's exact test, two-sided, $P < 0.05$; FDR = 0.05). The colored boxes on the vertical and horizontal axes represent OTU taxonomy at the family level.

antimicrobials and transit times that together influence the local microbiota assemblage⁹. We first performed an adapted 16S community profiling approach along the murine GI tract that could also infer absolute OTU abundances²⁵ (Fig. 3a and Supplementary Fig. 8; Methods). This new mouse cohort (two cohoused mice) shared only $\sim 20\%$ of OTUs with the previous group (Supplementary Fig. 8b), illustrating the substantial microbiome heterogeneity between animals inherent to such studies. This further highlights challenges for other spatial profiling techniques such as 16S fluorescence in situ hybridization (FISH) imaging where probes must be designed in advance, as compared to MaPS-seq, which can be applied to measure diverse bacteria without advance specification. Analysis of the absolute abundance of the microbiota across the intestine revealed increased bacterial density (~ 16 -fold higher) and species richness in the large intestine as compared to the small intestine, with the cecum harboring the highest bacterial density and number of OTUs. We chose three separate GI regions that exhibited distinct microbiota assemblages for characterization by MaPS-seq: the ileum, cecum and distal colon. Given the high degree of species mixing previously observed at the $\sim 30\text{-}\mu\text{m}$ scale, we used smaller-sized clusters (median diameter of $\sim 20\text{-}\mu\text{m}$) to capture higher-resolution spatial associations.

We first assessed the distribution of OTUs per cluster to compare the spatial organization of taxa in the three regions (Fig. 3b). Clusters of $\sim 20\text{-}\mu\text{m}$ in size displayed lower numbers of OTUs per cluster than $\sim 30\text{-}\mu\text{m}$ clusters (median of three to four OTUs per cluster). The ileum had significantly fewer OTUs per cluster than the cecum or distal colon (Mann–Whitney U test, $P < 10^{-18}$ and $P < 10^{-14}$, respectively). In comparison, the cecum and colon displayed similar OTU distributions, while the cecum harbored more clusters with a large number of OTUs. This suggests that GI regions with more diverse microbiota also exhibit higher spatial diversity at microscopic scales. Colonic clusters of an even smaller size ($\sim 7\text{-}\mu\text{m}$) were also profiled, and contained a significantly lower number of OTUs per cluster as compared to $\sim 20\text{-}\mu\text{m}$ clusters, as may be expected (Mann–Whitney U test, $P < 10^{-6}$).

To understand how the local spatial organization of the microbiome may vary within and across different gut regions, we visualized the cell cluster data across the three gut regions using t -distributed stochastic neighbor embedding (t -SNE; utilizing the Bray–Curtis distance of OTU RAs within clusters), as well as the abundance of prevalent bacterial families in cell clusters across the resulting manifold (Methods; Fig. 3c and Supplementary Fig. 9). While some cell clusters from the ileum, cecum and distal colon separately projected

into distinct groups, other clusters from each site projected more broadly across the manifold. Interestingly, a subset of cell clusters from the cecum projected into a dense group without clusters from the ileum or distal colon and were compositionally dominated by Lachnospiraceae. When cell clusters from a second cohoused mouse were added to the t-SNE analysis, they were distributed in a similar manner to clusters from the first mouse across the manifold and displayed a similar cecum-specific Lachnospiraceae group, further strengthening these results (Supplementary Fig. 10). Our observations suggest that the spatial distributions of some taxa may have distinct local organizations in different GI regions, while other taxa may have similar local organization along the GI tract.

Next, we explored whether these different spatial distributions reflect distinct spatial co-associations between taxa at each GI site (Fig. 3d). The ileum harbored a network of positive and negative associations between the few taxa present. In contrast, the cecum exhibited a dense network of positively co-associated taxa, with co-associations primarily between abundant Lachnospiraceae, Ruminococcaceae and Porphyromonadaceae. Similarly to the cecum, the distal colon displayed only positive associations, including strong groupings between three abundant Porphyromonadaceae (OTUs 5, 8 and 9). Profiling the colon at an even smaller scale (~7 μm) confirmed strong positive associations between a subset of these three taxa, indicating that this spatial co-occurrence is robust at short local length scales. Species from these abundant Bacteroidales taxa often contain diverse carbohydrate-active enzymes²⁶ and are known to engage in cooperative metabolic cross-feeding^{24,27}, which could promote these spatial co-associations.

While the spatial association networks revealed by MaPS-seq differed across the three GI regions, some common co-associations (or lack of associations) were observed. For example, a positive association between Lachnospiraceae (OTU 10) and Lactobacillaceae (OTU 4) was found in both the cecum and colon; in contrast, Coriobacteriaceae (OTU 1), an abundant taxon at all sites, lacked co-associations with other taxa and was thus randomly assorted at all sites. Together, the differing spatial architectures observed across GI sites suggest that regional environmental factors can variably shape some local spatial structuring of the microbiota, while conserved spatial patterns across sites are more likely the result of robust ecological interactions not affected by environmental variations.

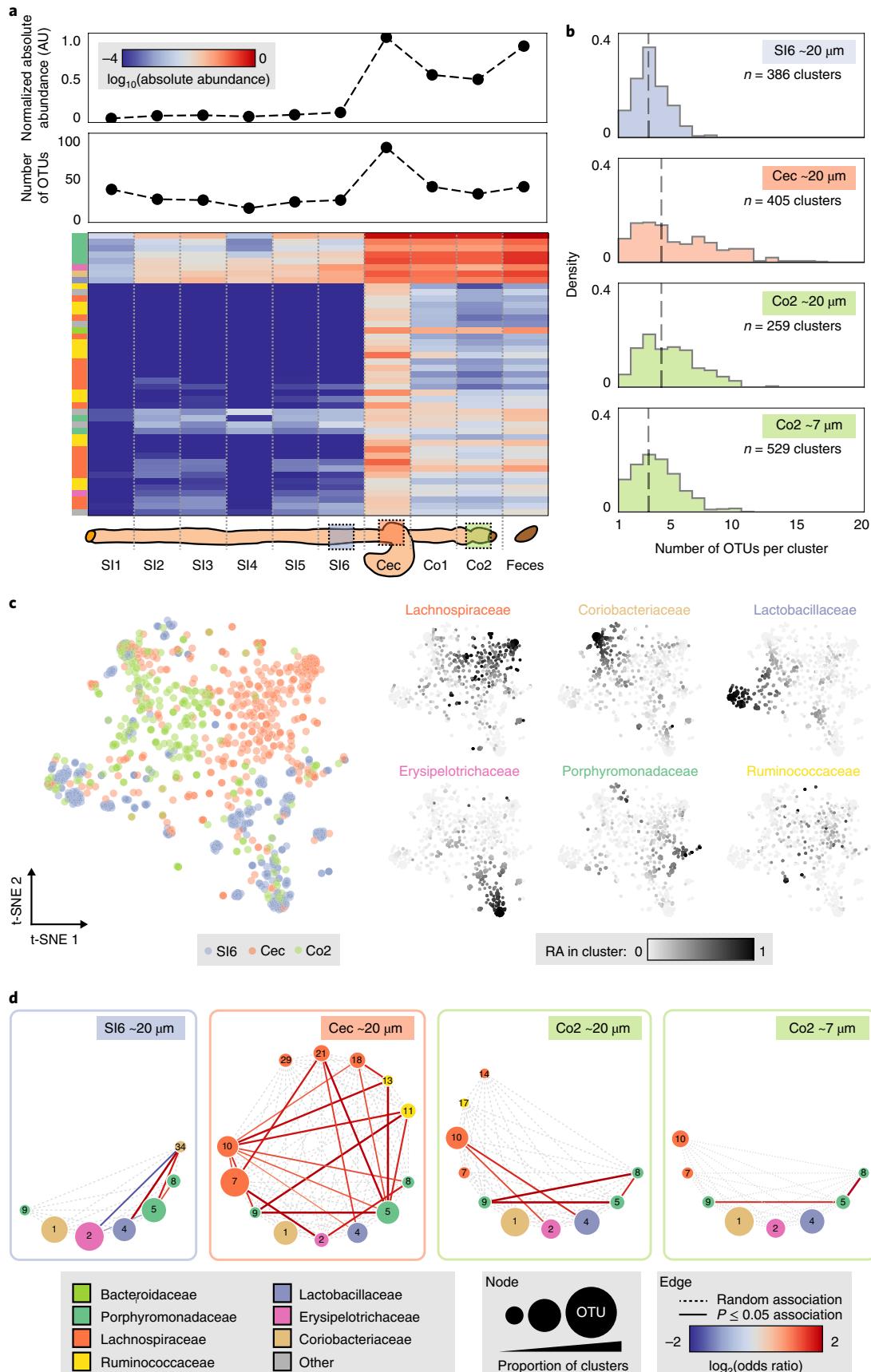
We further investigated whether MaPS-seq could identify individual taxa with unique or altered spatial patterns. While the cecum harbored the densest community and the highest degree of species mixing of the three sites (Fig. 3a,b), we hypothesized that specific taxa might self-aggregate to a higher degree than others, for example, by uniquely utilizing a specific metabolite¹¹. Assessing the aggregation of abundant taxa revealed a *Lachnospiraceae* (OTU 7; putatively of the genus *Dorea*, 60% confidence by Ribosomal Database Project (RDP) classifier) that had a clustering metric that was twofold higher than the average clustering metric for all taxa (Supplementary Fig. 11a). To validate this finding with an orthogonal approach, we performed 16S FISH on GI sections from

the same mouse sample using previously validated probes that target Lachnospiraceae (Erec482) as well as two other abundant taxa (for which FISH probes were available) that were predicted not to cluster to a similar degree (Coriobacteriaceae (Ato291) and Lactobacillaceae (Lab148); Methods). Strikingly, imaging confirmed that, while Lachnospiraceae were distributed across the cecum, they also formed large clustered aggregates that appeared to exclude other bacteria (Supplementary Figs. 11 and 12). Importantly, this result highlights that individual taxa in the gut can organize in unique and spatially varying micrometer-scale structures that can be revealed by MaPS-seq.

Having established the local spatial organization across the GI tract of mice fed a standard plant polysaccharide diet, we next sought to understand the extent to which diet might influence spatial structuring. Diet is known to play a major role in shaping the variation of gut microbiota across individuals^{28,29}. While diet shifts can rapidly alter microbiota composition within days³⁰, the detailed ecological mechanisms underlying these community-scale changes are not well understood. We thus took cohoused mice and split them into two cohorts, where one was maintained on the low-fat, plant-polysaccharide-based diet (LF; as in the previous cohorts) and one was switched to a high-fat, high-sugar diet (HF; commonly utilized in studies of dietary-induced obesity) to assess microbiota changes associated with these two diets representing distinct macronutrient profiles. After 10 d on the two diets, a considerable loss of species richness in the cecum and colon was observed in HF-fed mice as compared to LF-fed mice (Fig. 4a and Supplementary Fig. 13).

To determine whether a dietary shift could alter the spatial organization of the microbiota, which could contribute to the observed loss of species diversity, we performed MaPS-seq on distal colon samples from mice fed an LF or HF diet. We found that the distribution of unique OTUs per ~20-μm cluster was similar in both diets (Fig. 4b). This implies that species distributions at the local ~20-μm scale are governed by factors that are either common to, or not affected by, the two diets, for example spatial autocorrelation of bacterial growth. However, assessing diversity at the higher taxonomic family-rank revealed significantly higher diversity in HF clusters (HF, average of 4.0 families per cluster; LF, average of 2.7 families per cluster; Mann–Whitney *U* test, $P < 10^{-22}$; Fig. 4b), indicating that, while LF and HF clusters contained similar numbers of OTUs, the taxa within individual HF clusters were more phylogenetically diverse. Furthermore, positive co-associations were more frequently observed between diverse taxa for the HF diet than for the LF diet, which, in contrast, had co-associations mostly between Porphyromonadaceae or Lachnospiraceae (Supplementary Fig. 14). Interestingly, our observation of increased bacterial mixing at higher taxonomic levels has also been documented in mice fed with a plant-polysaccharide-deficient diet (as compared to an LF plant-polysaccharide-rich diet) using confocal imaging with 16S FISH probes of limited phylum-level specificity⁶, which further highlights the utility of examining spatial organization at the higher taxonomic resolution that is achievable by MaPS-seq.

Fig. 3 | Survey of spatial organization in the mouse gastrointestinal tract. **a**, Top, absolute abundance within gut intestinal regions calculated from spike-in sequencing (arbitrary units (AU), normalized to the maximum value) and the number of OTUs (that is, alpha diversity, the number of OTUs with RA > 0.1%). Bottom, absolute abundance of abundant OTUs (>1% of the maximum OTU absolute abundance in any sample) is shown as a heat map (\log_{10} scale); OTUs are clustered by Bray–Curtis dissimilarity. SI1–SI6, small intestine sections 1–6 (SI6 corresponds to the ileum); Cec, cecum; Co1, proximal colon; Co2, distal colon. **b**, Histogram of the number of OTUs per cluster (OTUs with RA > 2%). The number of clusters aggregated from two technical replicates is indicated (SI6: ~20 μm, $n = 386$; Cec: ~20 μm, $n = 405$; Co2: ~20 μm, $n = 259$; Co2: ~7 μm, $n = 529$), and the dashed line indicates median value. **c**, t-SNE visualization of 1,050 ~20-μm clusters from SI6, Cec and Co2 sites generated utilizing the Bray–Curtis dissimilarity of OTU RAs (subsampled to 314 reads across all clusters; the number of clusters is indicated). On the left, each cluster is colored by site of origin; on the right, each cluster is colored by the RA of the six most abundant families within each cluster (linear scale). **d**, Pairwise spatial associations for abundant and prevalent OTUs visualized as a circular graph; the number of clusters utilized was subsampled to the lowest number across the samples ($n = 259$ clusters). Nodes correspond to OTUs, with sizing proportional to the prevalence of OTUs across clusters and color representing OTU taxonomy at the family level; dotted edges denote all possible associations and shaded edges denote statistically significant associations (Fisher's exact test, two-sided, $P < 0.05$; FDR = 0.05).



Understanding the phylogenetic distribution of an ecosystem can provide important insights into ecological processes underlying community assembly^{31,32}. To better quantify possible changes in

phylogenetic diversity between the two diets, we calculated the net relatedness index (NRI) of clusters, a standardized effect size of the mean phylogenetic distance of taxa present within clusters against a

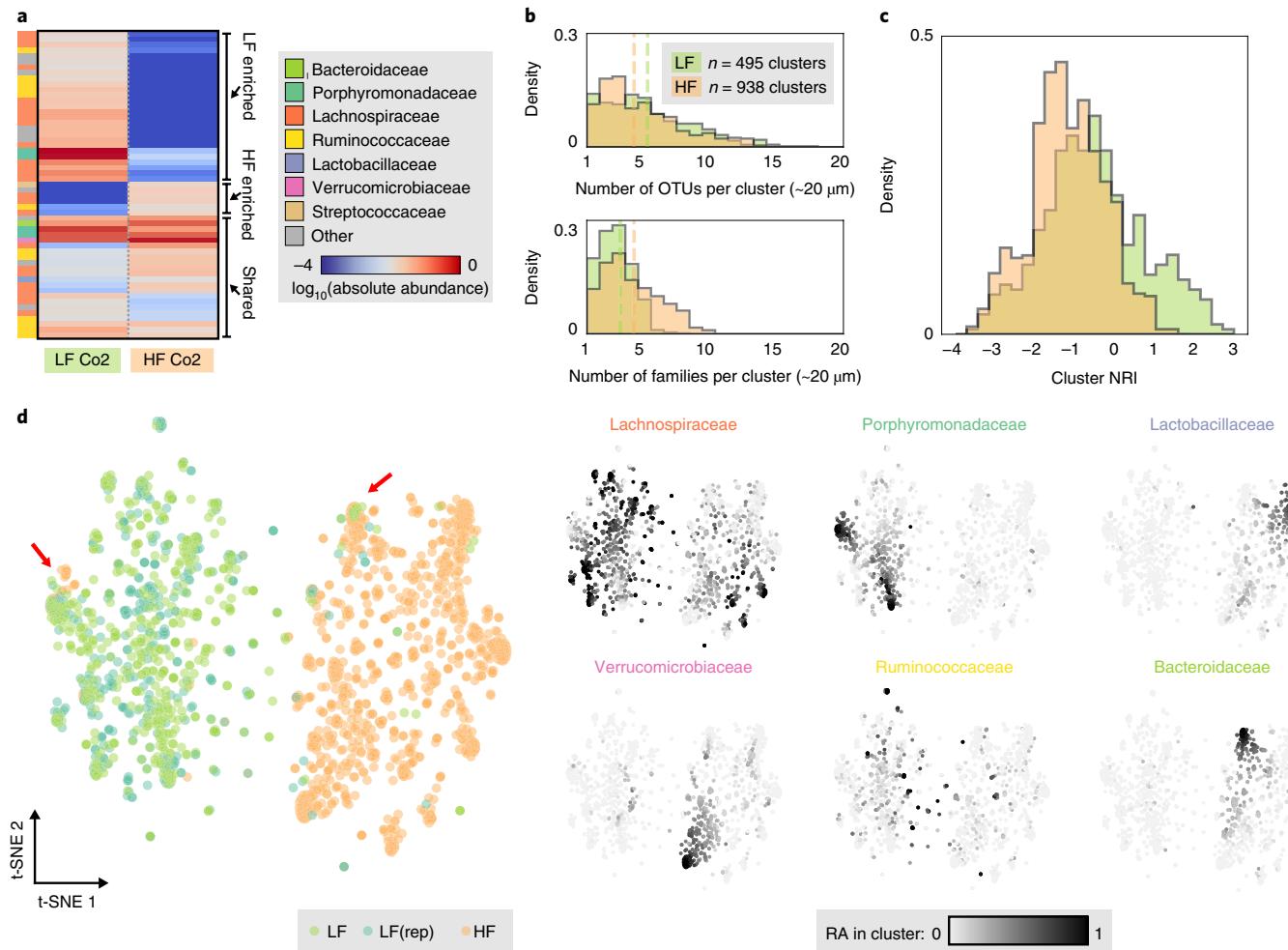


Fig. 4 | Spatial organization in the colon after dietary perturbation. **a**, Absolute abundance of dominant OTUs (>1% of the maximum OTU absolute abundance in any sample) in the distal colon of cohoused mice fed an LF or HF diet for 10 d is shown as a heat map (\log_{10} scale). Labels on the right indicate LF-enriched, HF-enriched and shared OTUs. **b**, Top, histogram of the number of OTUs per cluster (OTUs with RA > 2%). Bottom, histogram of the number of distinct families per cluster (families with RA > 2%). For both plots, green indicates LF clusters and orange indicates HF clusters, the dashed line indicates the median value and the number of clusters aggregated from two technical replicates is indicated (LF Co2, $n = 495$; HF Co2, $n = 938$). **c**, Histogram of NRI calculated for each cluster containing at least two OTUs; green indicates LF clusters and orange indicates HF clusters. **d**, t-SNE visualization of clusters utilizing the Bray-Curtis dissimilarity of OTU RAs (subsampled to 121 reads across all clusters). Left, clusters colored by site of origin: LF (green) and HF (orange); the number of clusters is indicated. In addition, a biological replicate from an adjacent colonic segment of the same LF-fed mouse is shown (LF(rep), dark green, $n = 359$ clusters). Red arrows indicate examples of cluster configurations observed in both diet conditions. Right, each cluster is colored by the RA of the six most abundant families within each cluster (linear scale).

null model of random sampling from the local species pool within each sample³¹ (Methods). For each microbiota cluster, a positive NRI value indicates phylogenetic clustering of its taxa, whereas a negative NRI indicates phylogenetic overdispersion. While most clusters had NRI values near 0, suggesting random phylogenetic distributions, samples from both LF- and HF-fed mice showed a subset of clusters with highly negative NRI values, suggesting a high degree of phylogenetic overdispersion in that subset. Interestingly, NRI values in LF clusters were significantly higher overall as compared to HF values (Mann-Whitney U test, $P < 10^{-18}$); this difference was driven by a subset of LF clusters with positive NRIs that were not observed in HF clusters (Fig. 4c and Supplementary Fig. 15). The phylogenetic clustering observed in this subset of LF clusters suggests that ecological habitat filtering due to factors associated with the LF diet (for example, complex plant polysaccharides) may be important in shaping the formation of these clusters at a length scale of ~20 μm (assuming that more phylogenetically similar taxa also have more similar phenotypes). A possible explanation for the loss of species

diversity when transitioning from an LF to an HF diet could therefore be the loss of this LF-specific local niche, which stably hosts these closely related taxa. Indeed, the same taxa (predominantly Lachnospiraceae) that were abundantly found in LF clusters with high NRI values were also those that were almost completely lost with the HF diet (Supplementary Fig. 15b).

Next, to compare spatial organization of taxa across the two diets, we visualized clusters using tSNE (Fig. 4d and Supplementary Fig. 16). Cell clusters from the two diets each formed highly distinct groups with minimal overlap, indicating that the spatial organization in the distal colon was substantially altered by the dietary shift. Despite this overall separation, we observed examples of cluster configurations that were shared between the two diets. For example, HF clusters were observed in a predominantly LF region marked by high abundance of a Porphyromonadaceae taxon (OTU 5), and LF clusters were observed in a predominantly HF region marked by high abundance of a Bacteroidaceae taxon (OTU 6) (Supplementary Fig. 16d). These shared cluster regions could represent spatial niches

that may be independent of diet (for example, mucus layers secreted by the host). Taking these findings together, MaPS-seq profiling of a diet perturbation enabled mechanistic analysis of ecological processes associated with community shifts and loss of diversity.

Spatial metagenomics enables the high-throughput characterization of microbial biogeography through plot sampling of colocalized nucleic acids at tunable length scales. Our approach could be applied to multiple perturbations in the gut, for example, diet, antibiotics and fecal microbiota transplantation, in other microbiota such as on skin or genitalia, or in diverse environmental niches including soils or biofilms. MaPS-seq enables in-depth analysis of these processes at previously inaccessible and ecologically meaningful local length scales within individual microbiomes. A variety of established spatial ecology tools and emerging computational and analytical approaches could be applied to this new type of high-dimensional microbiome dataset, which will require more rigorous evaluation and further development. Our approach could be modified to capture metagenomic information by introduction of additional capture primers or use of random priming or fragmentation strategies, which could enable profiling of interactions between bacteria and eukaryotes (for example, epithelial cells or fungi). Plot sampling of biological structures at microscopic scales opens up new directions of research that employ spatial ecology tools to study these complex systems.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at <https://doi.org/10.1038/s41587-019-0183-2>.

Received: 22 February 2018; Accepted: 7 June 2019;

Published online: 22 July 2019

References

- Reichenbach, T., Mobilia, M. & Frey, E. Mobility promotes and jeopardizes biodiversity in rock–paper–scissors games. *Nature* **448**, 1046–1049 (2007).
- MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton University Press, 1967).
- Cordero, O. X. & Datta, M. S. Microbial interactions and community assembly at microscales. *Curr. Opin. Microbiol.* **31**, 227–234 (2016).
- Swidsinski, A., Loening Baucke, V., Verstraelen, H., Osowska, S. & Doerffel, Y. Biostructure of fecal microbiota in healthy subjects and patients with chronic idiopathic diarrhea. *Gastroenterology* **135**, 568–579 (2008).
- Yasuda, K. et al. Biogeography of the intestinal mucosal and luminal microbiome in the Rhesus Macaque. *Cell Host Microbe* **17**, 385–391 (2015).
- Earle, K. A. et al. Quantitative imaging of gut microbiota spatial organization. *Cell Host Microbe* **18**, 478–488 (2015).
- Mark Welch, J. L., Rossetti, B. J., Rieken, C. W., Dewhirst, F. E. & Borisy, G. G. Biogeography of a human oral microbiome at the micron scale. *Proc. Natl Acad. Sci. USA* **113**, E791–800 (2016).
- Mark Welch, J. L., Hasegawa, Y., McNulty, N. P., Gordon, J. I. & Borisy, G. G. Spatial organization of a model 15-member human gut microbiota established in gnotobiotic mice. *Proc. Natl Acad. Sci. USA* **21**, E9105–E9114 (2017).
- Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14**, 20–32 (2016).
- Lee, S. M. et al. Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature* **501**, 426–429 (2013).
- Nagara, Y., Takada, T., Nagata, Y., Kado, S. & Kushiro, A. Microscale spatial analysis provides evidence for adhesive monopolization of dietary nutrients by specific intestinal bacteria. *PLoS ONE* **12**, e0175497 (2017).
- Tropini, C., Earle, K. A., Huang, K. C. & Sonnenburg, J. L. The gut microbiome: connecting spatial organization to function. *Cell Host Microbe* **21**, 433–442 (2017).
- Nava, G. M., Friedrichsen, H. J. & Stappenbeck, T. S. Spatial organization of intestinal microbiota in the mouse ascending colon. *ISME J.* **5**, 627–638 (2010).
- Pedron, T. et al. A crypt-specific core microbiota resides in the mouse colon. *mBio* **3**, e00116–12 (2012).
- Valm, A. M., Welch, J. L. M. & Borisy, G. G. CLASI-FISH: principles of combinatorial labeling and spectral imaging. *Syst. Appl. Microbiol.* **35**, 496–502 (2012).
- Geva-Zatorsky, N. et al. In vivo imaging and tracking of host–microbiota interactions via metabolic labeling of gut anaerobic bacteria. *Nat. Med.* **21**, 1091–1100 (2015).
- Whitaker, W. R., Shepherd, E. S. & Sonnenburg, J. L. Tunable expression tools enable single-cell strain distinction in the gut microbiome. *Cell* **169**, 538–546 (2017).
- Pereira, F. C. & Berry, D. Microbial nutrient niches in the gut. *Environ. Microbiol.* **19**, 1366–1378 (2017).
- Donaldson, G. P. et al. Gut microbiota utilize immunoglobulin A for mucosal colonization. *Science* **360**, 795–800 (2018).
- Wexler, A. G. et al. Human symbionts inject and neutralize antibacterial toxins to persist in the gut. *Proc. Natl Acad. Sci. USA* **113**, 3639–3644 (2016).
- Kim, H. J., Boedicker, J. Q., Choi, J. W. & Ismagilov, R. F. Defined spatial structure stabilizes a synthetic multispecies bacterial community. *Proc. Natl Acad. Sci. USA* **105**, 18188–18193 (2008).
- Coyte, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: networks, competition, and stability. *Science* **350**, 663–666 (2015).
- Amann, R. & Fuchs, B. M. Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nat. Rev. Microbiol.* **6**, 339–348 (2008).
- Rakoff-Nahoum, S., Coyne, M. J. & Comstock, L. E. An ecological network of polysaccharide utilization among human intestinal symbionts. *Curr. Biol.* **24**, 40–49 (2014).
- Ji, B. W. et al. Quantifying spatiotemporal variability and noise in absolute microbiota abundances using replicate sampling. *Nat. Methods* <https://doi.org/10.1038/s41592-019-0467-y> (2019).
- Ormerod, K. L. et al. Genomic characterization of the uncultured Bacteroidales family S24-7 inhabiting the guts of homeothermic animals. *Microbiome* **4**, 36 (2016).
- Rakoff-Nahoum, S., Foster, K. R. & Comstock, L. E. The evolution of cooperation within the gut microbiota. *Nature* **533**, 255–259 (2016).
- Carmody, R. N. et al. Diet dominates host genotype in shaping the murine gut microbiota. *Cell Host Microbe* **17**, 72–84 (2015).
- Sonnenburg, E. D. et al. Diet-induced extinctions in the gut microbiota compound over generations. *Nature* **529**, 212–215 (2016).
- David, L. A. et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
- Webb, C. O., Ackerly, D. D., McPeek, M. A. & Donoghue, M. J. Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* **33**, 475–505 (2002).
- Cavender-Bares, J., Kozak, K. H., Fine, P. V. A. & Kembel, S. W. The merging of community ecology and phylogenetic biology. *Ecol. Lett.* **12**, 693–715 (2009).

Acknowledgements

We thank A. Kaufman for technical assistance, R. Rabadian and D. Vitkup for technical advice and discussions, and D. Pe'er and the CUMC Pathology Department for access to sequencing instruments. H.H.W. acknowledges specific funding from the NIH (R01AI132403, R01DK118044), ONR (N00014-15-1-2704) and Burroughs Wellcome Fund PATH (1016691) for this work. K.W.L. is partially supported by NIH R01GM110494. P.A.S. acknowledges support from NIH/NIBIB K01EB016071. R.U.S. is supported by a Fannie and John Hertz Foundation Fellowship and an NSF Graduate Research Fellowship (DGE-1644869).

Author contributions

R.U.S. and H.H.W. developed the initial concept; R.U.S. developed the technique, performed experiments and analyzed data with input from P.A.S. and H.H.W.; and M.L., W.J. and K.W.L. assisted with prototypes of the microfluidic device. R.U.S. and H.H.W. wrote the manuscript. All authors discussed results and commented on and approved the manuscript.

Competing interests

H.H.W. and R.U.S. are inventors on a patent application filed by the Trustees of Columbia University in the City of New York regarding this work.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41587-019-0183-2>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to H.H.W.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

Methods

Materials and reagents. All primers and FISH probes were ordered from Integrated DNA Technologies. Primers containing modifications were HPLC purified by the manufacturer. Photocleavable primers were protected from unnecessary light exposure throughout.

Animal procedures. All mouse procedures were approved by the Columbia University Medical Center Institutional Animal Care and Use Committee (protocol AC-AAAR1513) and complied with all relevant regulations. Six- to 8-week-old female C57BL/6J mice were obtained from Taconic (colon analysis shown in Fig. 2) or Jackson (analysis across the GI tract shown in Fig. 3 and dietary perturbation shown in Fig. 4) and fed a plant-polysaccharide-based diet (LabDiet, 5053). Mice were allowed to adjust to the animal facility for 2 weeks before all studies, and all mice analyzed within each cohort were cohoused within the same cage. Dietary perturbation was performed by splitting four cohoused mice into two cages; one cage received the same plant-polysaccharide-based diet and one cage received the HF diet (Teklad, TD.06414).

Microfluidic device fabrication. Devices were fabricated utilizing standard SU-8 soft lithography. Silanized SU-8 silicon wafer molds were fabricated by FlowJEM with a feature height of ~40 µm. Polydimethylsiloxane (PDMS) (Dow Corning, Sylgard 184) was mixed for 5 min at a base:curing agent ratio of 10:1, degassed under house vacuum for 30 min and poured over the wafer. The PDMS mixture was cured at 80 °C for 1 h, allowed to cool to room temperature and removed from the wafer. Individual devices were cut from the PDMS slab and ports were punched utilizing a 1-mm biopsy punch (World Precision Instruments, 504646). A PDMS device and precleaned glass slide (Fisher, 12-550-A3) were then treated in a plasma cleaner (Harrick Plasma, PDC-32G) for 40 s, and the device was bonded to the slide and cured at 80 °C for 30 min. Devices were then treated to make channel surfaces fluorophilic³³. In brief, Aquapel solution was injected into the device and left for 30 s, and the device was flushed with air, FC-40 (Sigma-Aldrich, F9755) and air again, and baked at 80 °C for 10 min.

Design and construction of uniquely barcoded beads. We designed custom barcoded hydrogel beads containing one of 884,736 unique barcoded primers per bead as well as a partial sequencing adaptor and 16S V4 primer 515f (refs. ^{34,35}). Theoretically, around 17,500 clusters can be captured per sample with a 1% multiple barcoding rate³⁶. Barcoded primer sequences were constructed via a split-and-pool primer-extension strategy^{36,37} with three rounds of barcode extension. Each barcode position contained 96 possible sequences, and sets of barcodes were selected such that each had a Hamming distance of at least 3 base pairs from the other barcodes in the set (allowing for 1-base-pair error correction). The first barcode position was 7 to 9 base pairs in length (allowing for dephasing of reads to improve sequencing quality) while the second and third positions were 8 base pairs in length.

Construction of the barcoded beads followed procedures from Zilionis et al.³⁸ with minor modification for our barcoding scheme. In brief, acrylamide beads (6% (wt/wt) acrylamide, 0.18% (wt/wt) N,N'-methylenebisacrylamide (Sigma-Aldrich, 146072), 20 µM primer (acry_pc_pe1; Supplementary Table 1)) were generated using a custom microfluidic droplet device. Resulting beads were ~20–25 µm in diameter. Batches of ~20 million beads were then subjected to three rounds of primer extension using the three sets of 96 barcode sequences (pe1, pe2 and pe3 primer extension sets; Supplementary Table 2). For each round, beads and primers were distributed into the wells of a 96-well PCR microplate and primers were annealed to the beads by incubation. Bst polymerase reaction master mix (NEB, M0537L) was then distributed to each well and incubated to allow for extension. Finally, the reaction was quenched with EDTA and pooled for cleanup steps. The beads were then subjected to denaturing of the extension primers by sodium hydroxide and washing, and the extension protocol was repeated. These procedures were automated on a Biomek 4000 liquid-handling robot where possible. After the final extension step, a primer targeted to the terminal 515f primer sequence (515f_RC; Supplementary Table 1) was annealed, and Exo1 enzymatic cleanup (NEB, M0293L) was utilized to remove extension intermediates. Resulting barcoded beads were subjected to a final denaturing and washing step and stored at 4 °C in TET (10 mM Tris-HCl pH 8.0, 1 mM EDTA and 0.1% Tween-20).

Sample fixation and in situ polymerization. Intact tissue segments (from the colon, cecum or small intestine as noted) were obtained by dissection and immediately fixed in methacarn solution (60% methanol, 30% chloroform and 10% acetic acid) for 24 h (ref. ³⁹). The fixed tissue was trimmed with a sterile razor into segments no larger than 3 mm in length, and segments containing digesta were selected. Thus, all input samples for MaPS-seq analysis contained undisturbed epithelial tissue and luminal digesta contents. The trimmed sample was then incubated in PBS for 5 min and was permeabilized in PBS with 0.1% (vol/vol) Triton X-100 for 5 min. Next, a matrix-embedding solution^{40,41} containing a reverse sequencing primer with 16S V4 primer 806rB^{35,42} and acrydite and photocleavable linker groups was prepared on ice by mixing concentrated stocks of the following components in order: 1× PBS, 10% (wt/wt) acrylamide (Sigma-Aldrich, A9099), 0.4% (wt/wt) N,N'-bis(acryloyl)cystamine (BAC; Alfa Aesar, 44132-03), 5 µM

primer (acry_pc_pe2_816r, Supplementary Table 1), 0.01% (wt/wt) 4-hydroxy-2,2,6,6-tetramethylpiperidin-1-oxyl (Sigma-Aldrich, 176141), 0.2% (wt/wt) tetramethylethylenediamine (Sigma-Aldrich, T7024) and 0.2% (wt/wt) ammonium persulfate (Sigma-Aldrich, A3678). The BAC crosslinker enables gel degradation upon exposure to reducing conditions. The sample was dabbed dry with a sterile Kimwipe, placed in a PCR tube with excess matrix-embedding solution (~50 µl per segment) and incubated on ice for 5 min. Excess embedding solution was removed by pipetting and replaced, and the sample was subsequently incubated on ice for >1 h for perfusion. Excess embedding solution was removed, and samples were placed in a 37 °C incubator in an anaerobic chamber (Coy Laboratory Products) for >3 h. Gel-embedded samples were removed, excess polymer matrix was trimmed from the sample with a sterile razor, and the sample was washed twice with PBS and once with TET and stored in TET at 4 °C.

Sample fracturing, lysis and size selection. Samples were placed in a stainless steel vial (Biospec, 2007) along with a 6.35-mm stainless steel bead (Biospec, 11709635ss), and the vial was sealed with a silicone rubber plug cap (Biospec, 2008). The vial was placed in liquid nitrogen for >2 min, vigorously shaken to dislodge the sample from the vial wall and quickly transferred to a bead beater (Biospec, 112011) where it was subjected to beating for 10 s. PBS was added to the vial and vortexed, and clusters in PBS were then removed and washed twice with PBS by centrifugation at 15,000 r.p.m. for 1 min (Eppendorf, 5424). Next, embedded cells were lysed⁴³, and clusters were resuspended in 500 µl of lysis buffer (10 mM Tris HCl pH 8.0, 1 mM EDTA and 100 mM NaCl) with 75 µl⁻¹ lysozyme (Epicentre, R1810M) and incubated at 37 °C for 1 h. Clusters were then resuspended in 500 µl of digestion buffer (30 mM Tris HCl pH 8.0, 1 mM EDTA, 0.5% Triton X-100 and 800 mM guanidine hydrochloride (Sigma-Aldrich, G9284)) with 0.1 µg µl⁻¹ proteinase K (Epicentre MPRK092) and incubated at 65 °C for 15 min. Finally, clusters were incubated at 95 °C for 5 min to inactivate proteinase K and washed three times with TET.

Samples were next subjected to size selection. Clusters were first passed through a 40-µm cell strainer (Fisher, 22-363-547) to remove large particulate matter. Next, nylon mesh filters (Component Supply Company: 7 µm, U-CMN-7-A; 15 µm, U-CMN-15-A; and 31 µm, U-CMN-31-A) were cut to size using a half-inch hole punch and two filter punches were placed in a holder (EMD Millipore, SX0001300) for each size. Clusters were passed through the 31-µm filter, 15-µm filter and 7-µm filter sequentially using a 3-ml syringe (BD Biosciences, 309657); for each filter, clusters were passed through three times, and retained clusters on filters were washed once with TET. Clusters were washed off the 15-µm (large, ~30 µm median diameter) and 7-µm (medium, ~20 µm median diameter) filters or collected from the pass-through from the final 7-µm filter (small, ~7 µm median diameter). The concentration of clusters was quantified by counting on a hemocytometer (INCYTO, DHC-N01), and clusters were stored at 4 °C in TET for processing within ~2 d.

Coencapsulation of beads and clusters. A microfluidic coencapsulation strategy was utilized with three syringe pumps (Harvard Apparatus Pump 11 Elite) and observations were performed under a microscope (Nikon Eclipse Ti2). First, 300 µl of HFE-7500 (3M) with 5% wt/wt surfactant (RAN Biotechnologies, 008-FluoroSurfactant) was loaded into a 1-ml low-dead-volume syringe (Air-Tite Products A1), and the syringe was fitted with a needle (BD Biosciences, 305122) and polyethylene tubing (Scientific Commodities, BB31695-PE/2) and primed on a syringe pump.

Packed barcoded beads (30 µl) were then removed and washed twice with wash buffer (10 mM Tris HCl pH 8.0, 0.1 mM EDTA and 0.1% Tween-20) and twice with bead buffer (10 mM Tris HCl pH 8.0, 0.1% Tween-20, 50 mM KCl and 10 mM fresh dithiothreitol (utilized to degrade clusters within droplets)) by addition of buffer and centrifugation at 15,000 r.p.m. for 1 min. After four washes, remaining buffer supernatant was removed with a gel-loading tip (Fisher, 02-707-139). Packed beads (~5 µl) were loaded into polyethylene tubing and primed with a 1-ml syringe (BD Biosciences, 309626) backfilled with 500 µl of HFE-7500. The tubing was protected from light with a black tubing sheath (McMaster-Carr, 5231K31) and primed on a syringe pump with the needle facing upward.

Next, a cluster stock was vortexed for 1 min, ~2,500 clusters were removed and washed three times in wash buffer, and the remaining buffer was removed as above. A 45-µl encapsulation mix was prepared (25 µl NEBNext Q5 Hot Start HiFi PCR Master Mix (NEB, M0543L), 4 µl Nycoprep Universal (Accurate Chemical and Scientific, AN1106865), 5 µl 10% (wt/vol) Pluronic F-127 (Sigma-Aldrich, P2443), 1.25 µl 20 mg ml⁻¹ BSA (NEB, B9000S) and 9.75 µl nuclease-free water) and clusters were resuspended in the mix and vortexed for >10 s. A 1-ml low-dead-volume syringe was backfilled with 500 µl of HFE-7500, and the encapsulation mix was added directly into the tip of the syringe. A needle and polyethylene tubing were fitted to the syringe, protected from light with a black tubing sheath, and primed on a syringe pump with the needle facing upward.

Tubing was connected for the carrier bead and cluster encapsulation mix channels to a new microfluidic device. Pumps were primed for the carrier, beads and cluster encapsulation mix channels in order and once stable bead packing was observed, set to final flow rates of 2 µl min⁻¹ for carrier, 0.3 µl min⁻¹ for beads and 2.7 µl min⁻¹ for cluster encapsulation mix. Once stable droplet formation was

observed, polyethylene tubing was connected to the outlet port and emulsion was collected in a PCR tube (Axygen, PCR-02-L-C) prefilled with 10 µl of 30% (wt/wt) surfactant in HFE-7500 and 50 µl of mineral oil. Under these conditions, generated droplets were ~35–45 µm in diameter with a bead occupancy of ~25–50% (packed bead ordering enables loading above expected Poisson encapsulation statistics⁴⁴) and an extremely low cluster occupancy of <0.1% (cluster aggregation and channel clogging is a limiting factor at higher concentrations).

Emulsion PCR, library preparation and sequencing. The carrier phase underneath the emulsion was removed and replaced with 30 µl of 30% (wt/wt) surfactant in HFE-7500 to ensure droplet stability during PCR cycling. Tubes were placed on ice under a 365-nm UV light (Ted Pella Blak-Ray) and exposed for 10 min to release amplification primers. The emulsion was then subjected to PCR cycling (10 °C for 2 h; 98 °C for 30 s; 30 cycles of 98 °C for 10 s, 55 °C for 20 s, 65 °C for 30 s; and 65 °C for 2 min) with lid heating turned off. Coalesced droplet fraction, if present, was removed by pipetting and the carrier phase and mineral oil were removed. Droplets were broken by addition of 20 µl of 1H,1H,2H,2H-perfluoro-1-octanol (Sigma-Aldrich, 370533) and brief centrifugation in a microfuge tube. The aqueous phase was extracted, passed through a 0.45-µm spin column (Corning, 8162) and subjected to an Exo1 clean-up by adding 50 µl of 1× Exo1 buffer with 1 U µl⁻¹ Exo1 (NEB, M0293L) and incubating at 37 °C for 30 min. The mixture was then subjected to 1× SPRI bead cleanup (Beckman Coulter, A63881) per the manufacturer's protocol with addition of 1 volume of beads and elution in 20 µl of 10 mM Tris HCl pH 8.0.

The resulting products were then subjected to a second PCR to add sample indexes and Illumina P5 and P7 adaptors. Clean-up product (10 µl) was used as template for a 50-µl reaction with 1× NEBNext Q5 Hot Start HiFi PCR Master Mix, 0.5 µM of each of the indexing primers (p5_X and p7_X; Supplementary Table 3) and 0.1× SYBR Green I (Invitrogen, S7567). PCR (98 °C for 30 s; cycle: 98 °C for 10 s, 68 °C for 20 s, 65 °C for 30 s; and 65 °C for 2 min) was run on a real-time PCR machine (Bio-Rad, CFX96) to stop reactions during exponential amplification (typically ~10 cycles were used). Products were assessed on an agarose gel (2% E-gel; Thermo Fisher, G501802) to confirm the expected ~490-base-pair amplicon and were subjected to 1× SPRI bead clean-up as above. Resulting libraries were quantified via fluorometric quantitation (Thermo Fisher, Q32854), pooled and sequenced with an Illumina MiSeq 500 cycle v2 kit (read 1, 254 base pairs; read 2, 254 base pairs) at a 12 pM loading concentration with 20% PhiX spike in.

Sequence filtering and 16S analysis. For MaPS-seq data, a custom Python script was utilized to demultiplex reads on the basis of barcode identity and strip primer sequences from reads. Reads were merged and filtered using USEARCH 9.2.64 (ref. ⁴⁵) with maximum expected errors of one. The resulting sequences were then dereplicated and de novo clustered to OTUs at 97% identity, and reads were mapped to OTUs⁴⁶. Taxonomy was assigned to OTUs using the RDP classifier⁴⁷. This yielded an OTU table consisting of individual barcodes (that is, putative clusters) as samples.

Cluster mixing quality control experiment. Two bacterial communities were assembled; the first contained a single strain (for example, *E. coli* NEB 10-beta), while the second contained homogenized fecal bacteria. *E. coli* is not expected in the mouse gut at high abundances⁴⁸. To generate homogenized fecal bacteria, fecal pellets were subjected to bead beating (Biospec, 1001) with 0.1-mm glass beads in PBS for 1 min and passed through a 40-µm cell strainer. The two communities were fixed in methacarn, resuspended in a volume of matrix-embedding solution approximately equal to the volume of the fixed pellet and subjected to cluster generation per the MaPS-seq protocol described above. The homogenized fecal clusters therefore contained lower cell densities as compared to the actual fecal samples, owing to the dilution in embedding solution. The resulting size-selected clusters were then mixed in equal quantity and subjected to encapsulation and sequencing.

Analysis of MaPS-seq data. An overview of all MaPS-seq datasets generated in this study can be found in Supplementary Table 5. The resulting dataset contained a large number of barcodes (that is, clusters) with varying numbers of reads. A conservative threshold cutoff for considering clusters to be genuine was set as the total number of reads in a sample divided by 2,500 (that is, the number of clusters that were utilized as input during microfluidic encapsulation, assuming an equal read distribution for each cluster). Reactions yielding an extremely low number of clusters passing this threshold (<50) were conservatively excluded as they may represent failed encapsulation or amplification reactions.

Clusters were first preprocessed to remove a small number of clusters displaying highly similar OTU abundance profiles within a single technical replicate, which appeared to represent technical artifacts (that is, clusters encapsulated into droplets containing multiple barcoded beads or beads erroneously containing multiple barcodes) that could confound association detection. The pairwise Pearson's correlation of all clusters was calculated, and highly correlated sets of clusters ($r > 0.95$) that were dominated by a single technical replicate and large in size (>90% belonging to a single technical replicate and cluster constitutes >1% of the overall dataset) were removed. These artifacts

constituted a small amount of the overall dataset. For analysis of the presence or absence of species within a cluster, a threshold of RA > 2% within clusters was utilized, owing to the observation of a small amount of background read-through across clusters and to ensure that at least two reads (and not singletons) were required to denote the presence of a species.

To determine pairwise associations, prevalent and abundant OTUs within filtered clusters (RA > 2% in >10% of clusters) were identified, and two-by-two contingency tables of appearance (RA > 2%) were calculated for all pairs of OTUs. Fisher's exact test was then used to calculate the probability of pairs occurring together more or less frequently than expected (that is, a null model of random assortment of the two species, assuming equiprobable occupancy at all sites), and resulting P values were adjusted via the Benjamini–Hochberg procedure (FDR = 0.05).

For t-SNE analysis⁴⁹, reads for each cluster were subsampled to the lowest number for all clusters in the dataset (as specified in the text) as raw RA values were analyzed (that is, the RA > 2% threshold used for other analyses was not applied). The Bray–Curtis distance between the RAs of taxa within clusters was calculated, and this resulting distance matrix was utilized as the input for tSNE analysis.

The NRI was calculated as previously described⁵¹ adapting code from the relatedness_library.py script from Qiime 1.9.1 (ref. ⁵⁰), which implements the same calculation as phylocom 4.2 (ref. ⁵¹). In brief, species presence and absence across clusters was defined using the same RA > 2% threshold, and clusters containing only one OTU were omitted from analysis. OTU sequences were aligned and a neighbor-joining tree was constructed using MUSCLE 3.8.31 (ref. ⁵²). The NRI was calculated as a standardized effect size for each cluster: $NRI = -1 \times \frac{MPD_{cluster} - MPD_{null}}{s.d.(MPD_{null})}$ where $MPD_{cluster}$ denotes the mean phylogenetic distance (MPD), and $MPD_{cluster}$ and $s.d.(MPD_{null})$ indicate the mean MPD and the standard deviation of the MPD over 1,000 iterations of a null model, respectively. The null model, calculated for each cluster, was random draws for the number of OTUs present in the sample (that is, preserving cluster OTU richness) from the sample pool (that is, any OTU observed at least once in any cluster in the sample) without replacement. The null model therefore preserves the OTU richness of each cluster but randomizes the OTUs present from the set of OTUs occurring in the sample.

Bulk 16S sequencing and spike-in for absolute abundance calculation. An overview of all bulk 16S sequencing datasets generated in this study can be found in Supplementary Table 6. The bulk sequencing protocol followed our established spike-in sequencing pipeline⁵³. In brief, genomic DNA (gDNA) extraction was performed using a custom liquid-handling protocol on a Biomek 4000 robot based on the Qiagen MagAttract PowerMicrobiome DNA/RNA Kit (Qiagen, 27500-4-EP) but adapted for lower volumes. Samples were subjected to bead beating for a total of 10 min. For samples processed with the spike-in sequencing approach for absolute abundance calculation, the sample added was weighed on an analytical balance and 10 µl of a frozen spike-in strain concentrate (*Sporosarcina pasteurii*, an environmental bacterium not found in the gut microbiome; ATCC, 11859) was added during gDNA preparation. Resulting gDNA was amplified and the 16S V4 region was sequenced following a dual-indexing scheme⁵³; updated 515f and 806R primers were used as in the MaPS-seq technique. A 20-µl PCR amplification was performed (1 µM forward and 1 µM reverse barcoded primers, 1 µl prepared gDNA, 10 µl NEBNext Q5 Hot Start HiFi Master Mix and 0.2× final concentration SYBR Green I). PCR (98 °C for 30 s; cycle: 98 °C for 20 s, 55 °C for 20 s, 65 °C for 60 s; and 65 °C for 5 min) was run on a real-time PCR machine to stop reactions during exponential amplification. Amplicon products were quantified and pooled, the expected 390-base-pair product was gel-extracted, and paired-end sequencing was performed with an Illumina MiSeq 300 cycle v2 kit (read 1, 154 base pairs; read 2, 154 base pairs; custom sequencing primers spiked into sequencing kit) at a 10-pM loading concentration with 20% PhiX spike-in. Resulting sequences were processed with USEARCH as above. The absolute bacterial density for a sample (A) was calculated by utilizing the weight of sample added (w) and the proportion of reads mapping to the spike-in strain (p_s) using the following formula: $A = \frac{1-p_s}{p_s \times w}$. The absolute density of individual OTUs was calculated by rescaling the absolute density of the total sample by the RA of sample OTUs.

16S fluorescence in situ hybridization and imaging. Samples were fixed as with the MaPS-seq protocol and embedded within paraffin blocks, and 4-µm-thick luminal sections were cut and deparaffinized. 16S FISH was performed as previously described^{7,17}. In brief, previously validated FISH probes targeting abundant taxa present in the sample were obtained with the following conjugated fluorophores suitable for multiplex imaging: Erec482_a488 or Erec482_cy3 (ref. ⁵⁴) targeting Lachnospiraceae, Lab158_cy3 (ref. ⁵⁵) targeting Lactobacillaceae and Enterococcaceae, Ato291_cy5 (ref. ⁵⁶) targeting Coriobacteriaceae, Eub338_cy5 (ref. ⁵⁷) targeting Bacteria and a Non338_cy5 (ref. ⁵⁸) control probe (Supplementary Table 4). Sections were incubated with probes at 10 ng µl⁻¹ in FISH hybridization buffer (0.9 M NaCl, 20 mM Tris-HCl pH 7.5, 0.01% SDS and 10% formamide) at 47 °C for 4 h. Sections were then incubated in preheated FISH wash buffer (0.9 M NaCl and 20 mM Tris-HCl pH 7.5) for 10 min, washed three times in PBS, incubated with 10 µg ml⁻¹ DAPI in PBS for 10 min and washed three times in PBS. Sections were then mounted in mounting medium (Vector Laboratories, H1000).

Images were acquired on a Nikon Eclipse Ti2 epifluorescence microscope with a SOLA-SE2 illuminator and Andor Zyla 4.2 plus camera controlled by Nikon Elements AR software. DAPI, FITC/GFP, RFP and CY5 filter cubes (Nikon, 96359, 96362, 96364 and 96366, respectively) were utilized. Large-area, four-color fluorescence scans with three 0.6- μm z stacks within the 4- μm section were performed with a Plan Apo $\lambda \times 40$ objective. The extended depth of focus module was applied to resulting z stacks to obtain a focused image across the stack, and images across the entire section were stitched together.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All sequencing data are available from the NCBI Sequence Read Archive under accession [PRJNA541181](#).

Code availability

The code utilized in this study as well as microfluidic device designs and OTU tables can be accessed at <http://github.com/ravisheth/mapsseq>.

References

33. Mazutis, L. et al. Single-cell analysis and sorting using droplet-based microfluidics. *Nat. Protoc.* **8**, 870–891 (2013).
34. Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **18**, 1403–1414 (2016).
35. Walters, W. et al. Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *mSystems* **1**, e00009-15 (2016).
36. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
37. Bose, S. et al. Scalable microfluidics for single-cell RNA printing and sequencing. *Genome Biol.* **16**, 120 (2015).
38. Zilionis, R. et al. Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protoc.* **12**, 44–73 (2017).
39. Johansson, M. E. V. & Hansson, G. C. Preservation of mucus in histological sections, immunostaining of mucins in fixed tissue, and localization of bacteria with FISH. *Methods Mol. Biol.* **842**, 229–235 (2012).
40. Chung, K. et al. Structural and molecular interrogation of intact biological systems. *Nature* **497**, 332–337 (2013).
41. Chen, F., Tillberg, P. W. & Boyden, E. S. Expansion microscopy. *Science* **347**, 543–548 (2015).
42. Apprill, A., McNally, S., Parsons, R. & Weber, L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* **75**, 129–137 (2015).
43. Spencer, S. J. et al. Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *ISME J.* **10**, 427–436 (2016).
44. Abate, A. R., Chen, C.-H., Agresti, J. J. & Weitz, D. A. Beating Poisson encapsulation statistics using close-packed ordering. *Lab Chip* **9**, 2628–2631 (2009).
45. Edgar, R. C. & Flyvbjerg, H. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics* **31**, 3476–3482 (2015).
46. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).
47. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
48. Xiao, L. et al. A catalog of the mouse gut metagenome. *Nat. Biotechnol.* **33**, 1103–1108 (2015).
49. Maaten, L. V. D. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
50. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
51. Webb, C. O., Ackerly, D. D. & Kembel, S. W. Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics* **24**, 2098–2100 (2008).
52. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
53. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).
54. Franks, A. H. et al. Variations of bacterial populations in human feces measured by fluorescent *in situ* hybridization with group-specific 16S rRNA-targeted oligonucleotide probes. *Appl. Environ. Microbiol.* **64**, 3336–3345 (1998).
55. Harmsen, H., Elfferich, P. & Schut, F. A 16S rRNA-targeted probe for detection of lactobacilli and enterococci in faecal samples by fluorescent *in situ* hybridization. *Microb. Ecol. Health Dis.* **11**, 3–12 (1999).
56. Harmsen, H. et al. Development of 16S rRNA-based probes for the *Corynebacterium* group and the *Atopobium* cluster and their application for enumeration of *Corynebacteriaceae* in human feces from volunteers of different age groups. *Appl. Environ. Microbiol.* **66**, 4523–4527 (2000).
57. Amann, R. I. et al. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* **56**, 1919–1925 (1990).
58. Wallner, G., Amann, R. & Beisker, W. Optimizing fluorescent *in situ* hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. *Cytometry* **14**, 136–143 (1993).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Illumina sequencing data was collected with MiSeq Control Software v2.6. Images were acquired with Nikon Elements AR 5.11.00.
Data analysis	USEARCH 9.2.64 was used for sequencing data analysis. Taxonomic classification of OTUs was performed with RDP classifier release 11.5. Net Relatedness Index was calculated using code adapted from Qiime 1.9.1 relatedness_library.py script and input neighbor joining tree was calculated with MUSCLE 3.8.31. Custom code utilized to demultiplex barcoded MaP-seq data is available at https://github.com/ravisheth/mapseq . Jupyter 1.0.0, Python 2.7, Pandas 0.19.2, NumPy 1.15.4, Matplotlib 2.2.3, seaborn 0.9.0, scikit-bio 0.4.2, SciPy 1.1.0, NetworkX 1.11, and PyGraphviz 1.3.1 were utilized for data analysis and figure generation.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about availability of data

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All sequencing data is available at NCBI SRA under PRJNA541181. OTU tables can be accessed at <http://github.com/ravisheth/mapseq>.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculations were performed as the number clusters analyzed depended on the yield from the experiment; sample sizes are listed in figure legends where applicable.
Data exclusions	Data exclusion was based on sequencing coverage or cluster yield with predetermined criteria to remove technical artifacts as described in the Methods section.
Replication	All mouse samples were profiled in technical replicate. To assess reproducibility of the technique, technical replicates were analyzed (Supplementary Figure 6). Biological replicates were also performed and analyzed for various mouse samples (Supplementary Figure 7, 10, 11, 16, 17 etc.). Further information on MaP-seq datasets can be found in Supplementary Table 5. All attempts at replication were successful.
Randomization	Where relevant mice were allocated randomly to different experimental groups.
Blinding	Blinding was not possible during experiments as HF and LF diets visually appear different. All analyses of MaP-seq data were performed with the same parameters and criteria across different conditions, minimizing bias.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		

Animals and other organisms

Policy information about [studies involving animals](#); ARRIVE guidelines recommended for reporting animal research

Laboratory animals	6-8 week-old female C57BL6/J mice were utilized from Taconic or Jackson as indicated in the text.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve field-collected samples.
Ethics oversight	All mouse procedures were approved by the Columbia University Medical Center Institutional Animal Care and Use Committee (protocol AC-AAAR1513) and complied with all relevant regulations.

Note that full information on the approval of the study protocol must also be provided in the manuscript.