Spatial metagenomic characterization of microbial biogeography in the gut

Ravi U. Sheth1,2, Mingqiang Li3, Weiqian Jiang3, Peter A. Sims14,5, Kam W. Leong1,2 and Harris H. Wang1,6*

Spatial structuring is important for the maintenance of natural ecological systems14. Many microbial communities, including the gut microbiome, display intricate spatial organization15–18. Mapping the biogeography of bacteria can shed light on interactions that underlie community functions18–20, but existing methods cannot accommodate the hundreds of species that are found in natural microbiomes13–17. 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 colonization19–21, metabolism14, host–microbe and intermicrobial interactions22, and community stability12,23. 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 optimization18,22,23. Bacteria are densely packed in communities, limiting identification and analysis of individual cells using visual methods14. Although imaging approaches can simultaneously profile simple synthetic communities composed of a small number of cultivable species16,17 (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.
(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 ~30 μm) 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 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).

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...
antimicrobials and transit times that together influence the local microbiota assemblage. We first performed an adapted 16S community profiling approach along the murine GI tract that could also infer absolute OTU abundances (Fig. 3a and Supplementary Fig. 8; Methods). This new mouse cohort (two cohoused mice) shared only ~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.

We first assessed the distribution of OTUs per cluster to compare the spatial organization of taxa in the three regions (Fig. 3b). Clusters of ~20 μm in size displayed lower numbers of OTUs per cluster than ~30-μm 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, \( P < 10^{-14} \) 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 (~7 μm) were also profiled, and contained a significantly lower number of OTUs per cluster as compared to ~20-μm clusters, as may be expected (Mann–Whitney U test, \( P < 10^{-14} \)).

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 RA 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 enzymes and are known to engage in cooperative metabolic cross-feeding\(^7\)\(^8\), 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 metabolite\(^1\). 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 (At0291) 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\(^2\)\(^3\). While diet shifts can rapidly alter microbiota composition within days\(^4\), 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-level 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 phyllum-level specificity\(^5\), which further highlights the utility of examining spatial organization at the higher taxonomic resolution that is achievable by MaPS-seq.
Understanding the phylogenetic distribution of an ecosystem can provide important insights into ecological processes underlying community assembly\(^1,2\). 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
null model of random sampling from the local species pool within each sample\(^1\) (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 \(\sim 20 \mu m\) (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 localized 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 tagmentation 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/s41592-019-0467-y (2019).

Received: 22 February 2018; Accepted: 7 June 2019; Published online: 22 July 2019

References
1. Reichenbach, T., Mobilia, M. & Frey, E. Mobility promotes and jeopardizes biodiversity in rock-paper-scissors games. Nature 448, 1046–1049 (2007).
2. MacArthur, R. H. & Wilson, E. O. The Theory of Island Biogeography (Princeton University Press, 1967).
3. Cordero, O. X. & Datta, M. S. Microbial interactions and community assembly at microscales. Curr. Opin. Microbiol. 31, 227–234 (2016).
4. Swidinska, A., Loening Baucke, V., Verstraeten, H., Ouwsova, S. & Doerfler, Y. Biostucture of fecal microbiota in healthy subjects and patients with chronic idiopathic diarrhea. Gastroenterology 135, 568–579 (2008).
5. Yasuda, K. et al. Biogeography of the intestinal mucosal and luminal microbiome in the Rhesus Macaque. Cell Host Microbe 17, 385–391 (2015).
6. Earle, K. A. et al. Quantitative imaging of gut microbiota spatial organization. Cell Host Microbe 18, 478–488 (2015).
7. Mark Welch, J. L., Rossetti, B. J., Rieken, C. W., Gordon, J. I. & Borisy, G. G. Geobiography of a human oral microbiome at the micron scale. Proc. Natl Acad. Sci. USA 113, E791–800 (2016).
8. 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–E9117 (2014).
9. Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. Nat. Rev. Microbiol. 14, 20–32 (2016).
10. Lee, S. M. et al. Bacterial colonization factors control specificity and stability of the gut microbiota. Nature 501, 426–429 (2013).
11. Nagara, Y., Takada, T., Nagata, Y., Kado, S. & Kushiro, A. Microscale spatial analysis provides evidence for adaptive monopolization of dietary nutrients by specific intestinal bacteria. PLoS ONE 12, e0175497 (2017).
12. 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).
13. Nava, G. M., Friedrichsen, H. J. & Stepannenk, T. S. Spatialization of intestinal microbiota in the mouse ascending colon. ISME J. 5, 627–638 (2010).
14. Pedron, T. et al. A crypt-specific core microbiota resides in the mouse colon. mBio 3, e00116-12 (2012).
15. Vals, 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).
16. 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).
17. Whitaker, J. W., Shepherd, E. S. & Sonnenburg, J. L. Tunable expression tools enable single-cell strain distinction in the gut microbiome. Cell 169, 538–546 (2017).
18. Pereira, F. C. & Berry, D. Microbial nutrient niches in the gut. Environ. Microbiol. 19, 1366–1378 (2017).
19. Donaldson, G. P. et al. Gut microbiota utilize immunoglobulin A for mucosal colonization. Science 360, 795–800 (2018).
20. Westler, 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).
21. 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).
22. Cotye, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: networks, competition, and stability. Science 350, 663–666 (2015).
23. Aman, R. & Fuchs, B. M. Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. Nat. Rev. Microbiol. 6, 339–348 (2008).
24. Rakoff-Nahoum, S., Covey, M. J. & Comstock, L. E. An ecological network of pectic saccharide utilization among human intestinal symbionts. Curr. Biol. 24, 40–49 (2014).
25. Jin, 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).
26. Ormerod, K. L. et al. Genomic characterization of the uncultured Bacteroides family S24-7 inhabiting the guts of homoeothermic animals. Microbiome 4, 36 (2016).
27. Rakoff-Nahoum, S., Foster, K. R. & Comstock, L. E. The evolution of cooperation within the gut microbiota. Nature 533, 255–259 (2016).
28. Carmody, R. N. et al. Diet dominates host genotype in shaping the murine gut microbiota. Cell Host Microbe 17, 72–84 (2015).
29. Sonnenburg, E. D. et al. Diet-induced extinctions in the gut microbiota compound over generations. Nature 529, 212–215 (2016).
30. David, L. A. et al. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505, 559–563 (2014).
31. Webb, C. O., Ackerly, D. D., McPeek, M. A. & Donoghue, M. J. Phylogenies and community ecology. Annu. Rev. Ecol. Syst. 33, 475–505 (2002).
32. Cavender-Bares, J., Kozak, K. H., Fine, P. V. A. & Kembl, 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. Rababan and D. Vitkop for technical advice and discussions, and D. Peér and the CUMC Pathology Department for access to sequencing instruments. H.H.W. acknowledges specific funding from the NIH (1R01AI132403, 1R01DK118044), ONR (N00014-15-1-2704) and Burroughs3 Welcome Fund (P1016691) 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 microbial 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 (colonic analysis shown in Fig. 1) 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 Devices were fabricated utilizing standard SU-8 Microfluidic device fabrication.

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 (colonic analysis shown in Fig. 1) 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 Devices were fabricated utilizing standard SU-8 Microfluidic device fabrication.

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 (colonic analysis shown in Fig. 1) 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 Devices were fabricated utilizing standard SU-8 Microfluidic device fabrication.
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/vt) surfactant in HFPE-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 bead clean-up by adding 50 μl of 1× ExoI buffer with 1 U UlA1l-ExoI (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 emulsion 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_5’ and p7_5’; 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 20 s for 2 min) was run on a real-time PCR machine (Bio-Rad, EFX96) 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 (typically ~10 cycles were used). Products were quantified via fluorometric quantitation (Thermo Fisher, Q32854), pooled and sequenced with an Illumina MiSeq 300 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. 43) 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 OTUs. Taxonomy was assigned to OTUs using the RDP classifier46. This yielded an OTU table consisting of individual barcodes (that is, putative clusters) as samples.

**Cluster mixing quality control test.** 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 E. coli is not expected in...
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 λ ×40 objective. The extended depth of focus module was applied to result in stacking 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. et al. 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).
56. Harmsen, H., Elfferich, P. & Schut, E. 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).
57. Harmsen, H. et al. Development of 16S rRNA-based probes for the Coriobacterium group and the Atopobium cluster and their application for enumeration of Coriobacteriaceae in human feces from volunteers of different age groups. Appl. Environ. Microbiol. 66, 4523–4527 (2000).
58. Amann, R. I. et al. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl. Environ. Microbiol. 56, 1919–1924 (1990).
Corresponding author(s): Harris Wang
Last updated by author(s): May 20, 2019

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.

- 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 Qime 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, sci-kit 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

- n/a Involved in the study
- ☑ Antibodies
- ☑ Eukaryotic cell lines
- ☑ Palaeontology
- ☑ Animals and other organisms
- ☑ Human research participants
- ☑ Clinical data

### Methods

- n/a Involved in the study
- ☑ ChIP-seq
- ☑ Flow cytometry
- ☑ MRI-based neuroimaging

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.