

SAMPL-seq reveals micron-scale spatial hubs in the human gut microbiome

Received: 6 September 2023

Accepted: 12 December 2024

Published online: 3 February 2025

 Check for updates

Miles Richardson  ^{1,2}, Shijie Zhao ¹, Liyuan Lin ¹, Ravi U. Sheth ^{1,4}, Yiming Qu ^{1,2}, Jeongchan Lee ¹, Thomas Moody ^{1,2}, Deirdre Ricaurte ^{1,2}, Yiming Huang  ^{1,2}, Florencia Velez-Cortes ^{1,2}, Guillaume Urtecho ¹ & Harris H. Wang  ^{1,3} 

The local arrangement of microbes can profoundly impact community assembly, function and stability. However, our understanding of the spatial organization of the human gut microbiome at the micron scale is limited. Here we describe a high-throughput and streamlined method called Split-And-pool Metagenomic Plot-sampling sequencing (SAMPL-seq) to capture spatial co-localization in a complex microbial consortium. The method obtains microbial composition of micron-scale subcommunities through split-and-pool barcoding. SAMPL-seq analysis of the healthy human gut microbiome identified bacterial taxa pairs that consistently co-occurred both over time and across multiple individuals. These co-localized microbes organize into spatially distinct groups or ‘spatial hubs’ dominated by *Bacteroidaceae*, *Ruminococcaceae* and *Lachnospiraceae* families. Using inulin as a dietary perturbation, we observed reversible spatial rearrangement of the gut microbiome where specific taxa form new local partnerships. Spatial metagenomics using SAMPL-seq can unlock insights into microbiomes at the micron scale.

The human gut microbiome is stably colonized by hundreds to thousands of bacterial species¹, which when perturbed have been associated with numerous diseases². Beyond bulk compositional information, we know little about the micron-scale spatial assortment of microbes in the gut³. Microbes may spatially segregate due to metabolic and ecological interactions, ranging from cooperative sharing of niches to direct competition or antagonism⁴. As such, spatial organization can play a critical role in community makeup, function and stability^{1,5}. In general, a spatially structured ecosystem better maintains species diversity than a homogenized microbiome⁶. Nutrients can further tune species interactions^{7–9}. For example, dietary fibres are known to modulate short-chain fatty acid (SCFA) production by bacterial consortia in the colon¹⁰. Mapping the local spatial arrangement of the human gut microbiome could reveal rules governing its organization, diversity and resiliency in both healthy and diseased states.

Several high-resolution imaging-based approaches have been developed to map microbial spatial arrangements^{11–16}. These methods,

such as CLASI-FISH¹¹, HIPR-FISH¹³, SHM-seq¹⁶ and SEER-FISH¹⁵, rely on highly multiplexed barcoding and imaging setups to identify microbes in tissue sections. While these methods offer high spatial resolution and precise spatial coordinate information, they require previous metagenomic sequencing to obtain genomic information needed for probe design, need experimental validation of labelled bacterial taxa and demand sophisticated imaging setups. Nevertheless, these approaches have been used to profile the human oral microbiome and the mouse gut microbiome with success^{13,17,18}. However, the spatial organization of the human gut microbiome is more challenging to study due to its high taxonomic diversity and interpersonal heterogeneity, and imaging-based strategies have not been applied to the human gut microbiome.

We previously described a spatial metagenomic sequencing approach (MaPS-seq⁸) based on analysing ‘microbial plots’, which allows for the characterization of the bacteria present in hundreds of gut microbial subcommunities using metagenomic sequencing. However, the method required custom microfluidics, barcoded beads

¹Department of Systems Biology, Columbia University, New York, NY, USA. ²Integrated Program in Cellular, Molecular and Biomedical Studies, Columbia University, New York, NY, USA. ³Department of Pathology and Cell Biology, Columbia University, New York, NY, USA. ⁴Present address: Kingdom Supercultures, Brooklyn, NY, USA.  e-mail: harris.wang@columbia.edu

and emulsion PCR steps that greatly limited throughput, scalability and accessibility. Recent single-cell sequencing advances in combinatorial split-pool barcoding (of beads for microfluidic encapsulation¹⁹ or single cells directly^{20,21}) have streamlined the generation of a large number of barcode combinations that significantly increased throughput and reduced cost/time. These combinatorial barcoding strategies could be adopted to label each ‘microbial plot’ to achieve high-throughput sampling. However, these improvements have not been applied to complex microbial consortia, which would enable gut characterization at greatly increased scale to gain insights previously not possible. Here we introduce Split-And-pool Metagenomic Plot-sampling sequencing (SAMPL-seq), a streamlined spatial metagenomics method to analyse microbiome samples at micron-scale spatial resolution. By utilizing *in situ* amplification steps to combine micron-scale particle-level spatial information with bacterial abundance, this method integrates the community-sequencing approach of MaPS-seq with the high-throughput capacity of split-pool barcoding. With these innovations, SAMPL-seq provides the necessary order of magnitude increase in scale and ease of use to enable in-depth spatial studies of the human gut microbiome. To demonstrate these capabilities, we applied SAMPL-seq to human stool to reveal taxonomically distinct ‘spatial hubs’ of the human gut microbiota that were stable over time and conserved between people. In response to dietary changes, these hubs reorganized into alternative spatial arrangements in a reversible manner, highlighting the flexible spatial assortment of the gut microbiome based on nutritional availability and environmental conditions.

Results

Development of SAMPL-seq for microbial spatial metagenomics

SAMPL-seq utilizes the principle of microbial plot sampling to identify bacteria that co-localize across tens of micrometres in natural subcommunities within a microbiome. An input microbiome sample (for example, as little as ~3 mm³) is first embedded and solidified in an acrylamide polymer matrix to preserve its original spatial organization (Fig. 1a and Methods). This matrix contains acrydite linkers conjugated to a DNA adaptor to facilitate downstream split-pool barcoding. The embedded sample is then cryofractured via bead beating and the embedded bacteria are chemically lysed while their DNA remains trapped in the gel. Next, the particles undergo three rounds of split-pool primer extension¹⁹ to create barcoded 16S rRNA primers that are unique to each particle and are filtered to a desired size (for example, microbial plots of ~40 µm in diameter; Fig. 1b). An *in situ* PCR reaction is performed to amplify the 16S rRNA V4 region across all particles using the now uniquely barcoded primers (Extended Data Fig. 1a). The PCR product is then UV released from the particles, cleaned and concentrated (Methods). Sequencing and indexing adaptors are added by PCR and the library is sequenced on an Illumina platform. Reads thus contain both the 16S V4 sequence and a unique particle barcode (Fig. 1c,d).

SAMPL-seq sequencing reads undergo barcode identification filtering with an overall success rate of ~96.6% (Methods and Fig. 1e). Reads are then grouped by particle according to their unique barcode combination, and amplicon sequence variants (ASVs, defined in this study as 100% sequence identical operational taxonomic units) are assigned using denoising. By replacing bead-based co-encapsulation described previously⁸ with *in situ* split-pool barcoding and amplification, SAMPL-seq is substantially faster, scalable and easier to implement to profile >10⁴ particles per sample without the need for microfluidics or other complex setups (Extended Data Fig. 1b,c).

Characterizing SAMPL-seq performance using mixed communities

We first characterized SAMPL-seq performance including replicability, overall bulk correlation and throughput, along with spatial specificity, by determining background mixing rates of barcodes between particles.

Two sets of mixing experiments, M1 and M2, were performed. In the M1 experiment, a homogenized microbiome sample (M1A) was mixed with a pure *Sporosarcina pasteurii* culture (M1B) in two separate biological replicates, which were size filtered to control particle size (Fig. 2a and Extended Data Fig. 2a,b). *S. pasteurii* is an environmental bacterium not found in the gut. Resulting SAMPL-seq data showed high experimental consistency between the M1A community in each replicate ($r = 0.93$, Pearson’s correlation) and high correlation with bulk 16S relative abundance ($r = 0.88$, Pearson’s correlation) (Fig. 2b,c and Extended Data Fig. 2b,c). As expected, larger particle sizes tended to increase species diversity per particle (Extended Data Fig. 2d). Further, the libraries had an overall multiplet rate²² of 4.7%, suggesting low mixing between communities (Fig. 2d,e and Extended Data Fig. 2e). Together, these results confirm that SAMPL-seq has high technical performance and reproducibility, good consistency with bulk sequencing results and minimal methodological bias.

In mixing experiment M2, two separate defined microbial sources of known composition were prepared at two cell densities of 2×10^8 cells per µl (1× concentration) or 6×10^8 cells per µl (3× concentration), separately embedded and then mixed before cryofracturing (Extended Data Fig. 3a). The first source, M2A (Zymo, D6331), consisted of common gut bacterial taxa at defined concentrations to allow for comparison to a known reference, while the second, M2B (Zymo, D6320), consisted of a bacterial strain not present by design in M2A. After processing, each replicate yielded particles with a mean size 50 µm in diameter (~120–400 cells per particle) (Extended Data Fig. 3b). Reads from ~16,000 particles across 5 replicates passed quality filtering (Supplementary Table 1). Experimental replicates (1× versus 3× concentration) were highly correlated ($r = 0.84$, Pearson’s correlation) (Extended Data Fig. 3c). The particle prevalence of each species, defined as the percent of particles a species is found, also correlated well with its relative abundance as listed by the manufacturer ($r = 0.80$, Pearson’s correlation) (Extended Data Fig. 3d). Notably, at 3× bacterial input concentration, the species diversity per particle was higher (Extended Data Fig. 3e), which suggestss SAMPL-seq’s sensitivity to different biomass levels. The average particle capture rate was 16.2% across replicates, which is on par with other single-cell methods²³ (Methods). Importantly, only 1.4% of particles (177) contained mixed reads from both M2A and M2B sources. The overall multiplet rate, the mixing rate accounting for unobserved mixing, was 2.9% (Extended Data Fig. 3f,g and Supplementary Table 2), which is also comparable to current split-pool methods^{20,21,24}, with a low level of mixing between reads from different sources (Extended Data Fig. 3h).

Spatial metagenomics of the gut microbiome using stool

Most microbiome studies rely on faecal matter as a reliable representation of the gut microbiome²⁵. We sought to evaluate whether stool material can be used to assess the spatial architecture of the gut microbiome. SAMPL-seq was applied on three mouse gut compartments (small intestine, caecum, colon) along with the corresponding faecal pellets from the same mouse (Extended Data Fig. 4a). ASV overall relative abundance and prevalence among particles were most similar between colon and stool than any other samples ($r = 0.55, P < 2.2 \times 10^{-16}$, $r = 0.71, P < 2.2 \times 10^{-16}$ respectively, Pearson’s correlation) (Extended Data Fig. 4b,c). Consistent with our previous observations from the mouse gut microbiome⁸, the small intestine had a distinct set of spatially co-localized ASVs that persisted through the caecum and colon and remained co-localized in a subset of particles (Extended Data Fig. 4a); this spatial signal could not be delineated from just bulk measurements. Principal coordinate analysis (PCoA) on SAMPL-seq particles from all compartments showed clustering between stool and colon, and clear separation from small intestine-derived samples (Extended Data Fig. 4d). The caecum contained spatial signals from both small intestine and colonic communities. These results suggest possible spatial signals in stool samples that can be recovered with SAMPL-seq in a non-invasive manner to profile the *in vivo* colonic microbiome.

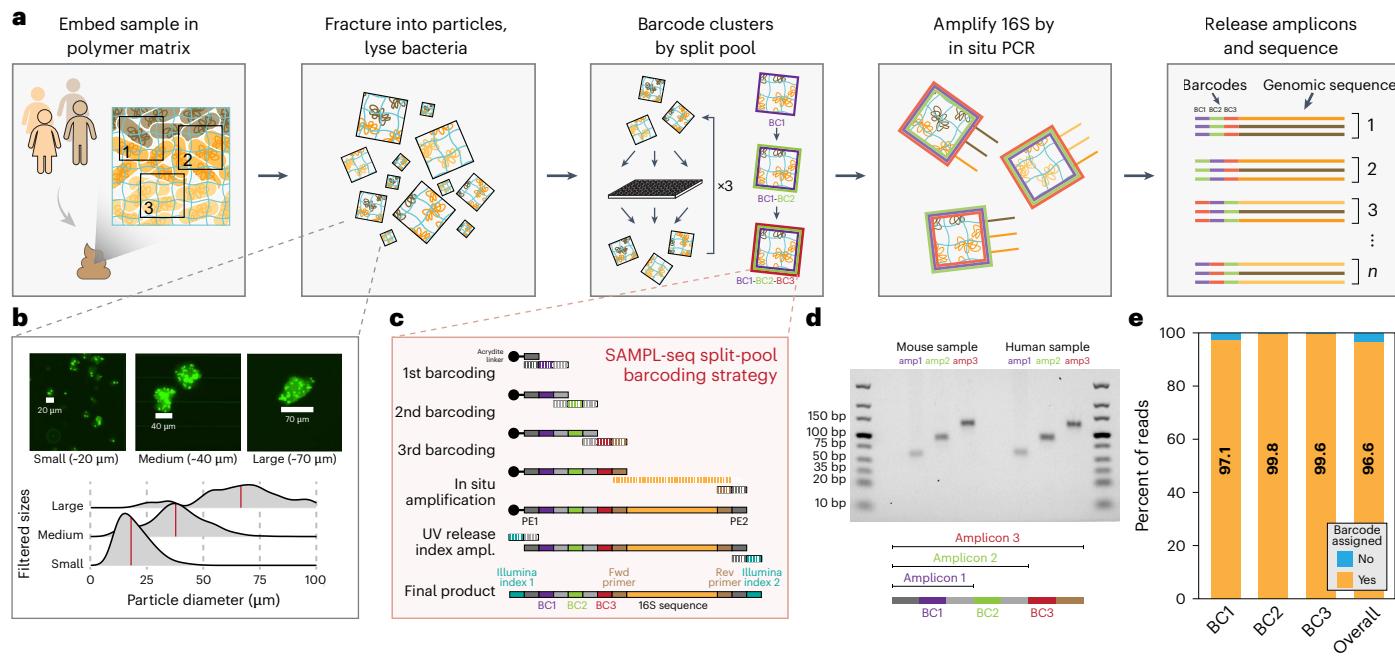


Fig. 1 | Spatial metagenomics of thousands of micron-sized communities using SAMPL-seq. a, Step-by-step outline of the SAMPL-seq method. **b,** Images of particles or ‘microbial plots’ and their corresponding size distributions. **c,** Schematic of split-pool barcoding steps that produce a barcoded primer and downstream steps to generate a final 16S amplicon library. **d,** Gel showing PCR

products of fully extended barcodes from murine and human samples using primers that bind to different parts of the primer barcode sequence. **e,** Barplot showing sequencing reads with successfully assigned barcodes across each barcoding step.

To explore the utility of SAMPL-seq for human gut microbiome studies, we applied the method to fresh stool from five healthy volunteers (H1, H10, H11, H18, H19), yielding data from >21,000 particles of ~40 μm in diameter (Extended Data Fig. 5a–f, and Supplementary Tables 1 and 3). In one individual (H11), we performed additional longitudinal SAMPL-seq for 5 consecutive days (H11-D1 to D5) to explore temporal variation, yielding 18,000 particles. Unique ASV–particle barcode combinations saturated for detecting highly prevalent ASVs (>0.01%) and ASV–ASV co-occurrences, indicating sufficient sequencing coverage (Extended Data Fig. 5g–i). Technical and biological SAMPL-seq replicates at Day 4 (H11-D4-R1 and R2) showed high correlation ($r = 0.92$, $P < 2.2 \times 10^{-16}$ and $r = 0.85$, $P < 2.2 \times 10^{-16}$ respectively, Pearson’s correlation), and longitudinal samples from H11 showed higher correlation than those from different donors (Extended Data Fig. 6a–e). ASVs in the disrupted sample were consistently more prevalent across particles than in the original intact sample, showing that mechanical disruption eliminated the previous microbial spatial structure (Extended Data Fig. 6a). The ASV abundance measured by SAMPL-seq and bulk 16S sequencing were highly correlated across all samples, indicating that taxonomic and compositional data were faithfully captured in these stool samples (Extended Data Fig. 6f,g). While the microbiome composition was relatively consistent in H11 over 5 days (Extended Data Fig. 7a), interpersonal samples exhibited greater compositional variation at the ASV level (Extended Data Fig. 7b, $W=1$, $P = 2.17 \times 10^{-5}$ by Wilcoxon rank-sum test) than the family level (Extended Data Fig. 7c). These results indicate that SAMPL-seq can be applied robustly to faecal samples despite natural variation in peoples’ microbiome, which allows further analysis of gut microbial spatial architecture.

Identifying patterns of microbial spatial co-localization

To determine which ASV pairs are more or less likely to spatially localize in the human gut, we applied a null model based on a fixed-fixed permutation method, which is commonly used to find co-association patterns in ecological studies²⁶ (Methods and Fig. 3a). The model randomizes

ASV presence across the dataset while preserving both the number of unique ASVs per particle and the prevalence of ASVs in the dataset. This model better accounts for the natural heterogeneity in particle-level ASV diversity compared with the Fisher’s exact test used previously⁸. With this null model, we could robustly detect the separation between M1A and M1B ASVs in the M1 mixing experiment, with minimal spurious associations (Fig. 3b). Using this approach on temporal SAMPL-seq data (H11-D1 to H11-D5), we identified on average 86 statistically significant positive or negative co-associated ASV pairs in each day across a total of 73 ASVs ($P < 0.05$, Benjamini–Hochberg (BH) false discovery rate (FDR) corrected) (Supplementary Table 4). As a control, SAMPL-seq on a mechanically disrupted faecal aliquot of the H11-D4 sample showed substantially fewer co-associations (31 significant ASV pairs in disrupted versus 77 and 89 in intact Day 4 samples) and co-associations found in the disrupted sample had low correlation with the intact samples (Extended Data Fig. 8a,b). Furthermore, we characterized the correlation in spatial associations of ASVs from three paired sets of fresh and frozen faecal samples and found high correlation between them ($R = 0.88, 0.77, 0.80$) (Extended Data Fig. 8c). These results indicate that SAMPL-seq could be performed on frozen stool samples without the need for additional cryopreservatives, which could allow retrospective analyses that leverage other existing stool biobanks²⁷. Analysis across additional samples H1, H10, H18, H19 revealed striking patterns of pairwise ASV spatial co-associations (Fig. 3c and Extended Data Fig. 8d).

Across the H11 longitudinal samples, we confirmed that the number of particles analysed sufficiently captured the underlying spatial co-localization patterns. Our subsampling analysis shows that the number of subcommunities sequenced provides sufficient number to reach robust inference. Such inference requires at least thousands of particles²⁸, which is only made possible with the throughput of our SAMPL-seq approach, which is superior to previous spatial metagenomic sampling methods (for example, MaPS-seq) (Fig. 4a). The spatial co-associations were consistent (that is, 89.6% having same co- or anti-associations), indicating that a robust and stable

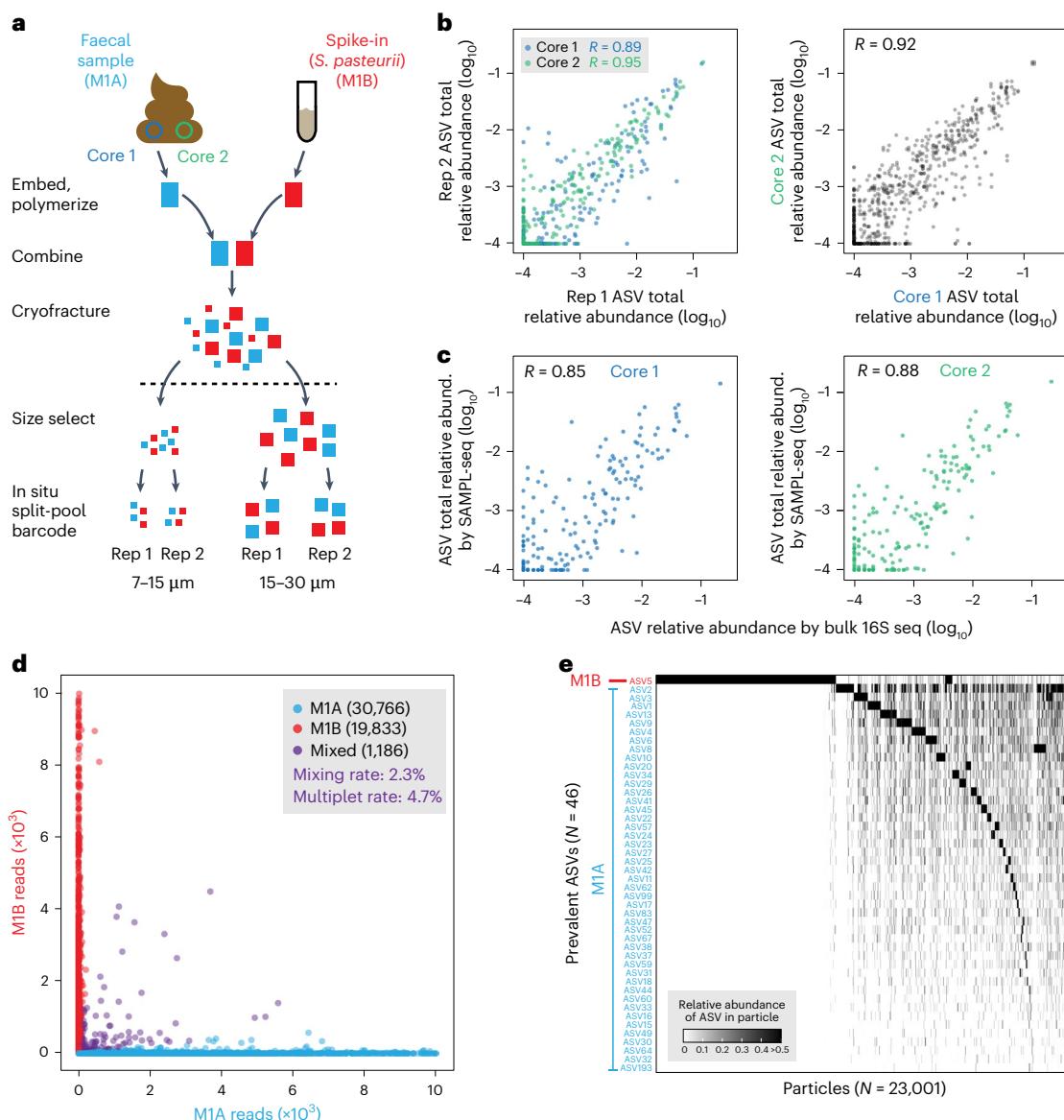


Fig. 2 | SAMPL-seq performance using mixing experiments. **a**, An outline of mixing experiment M1. A human faecal sample and *S. pasteurii* were separately embedded and polymerized. They were then combined during the cryofracturing step, size sorted, amplified and sequenced in aliquots of 10,000 particles. The homogenization procedure was repeated, for a total of 2 biological replicates. Two aliquots of 10,000 particles were sequenced as technical replicates. **b**, Correlation of ASV relative abundance for technical replicates

within each core and between cores at particle sizes of 15–30 µm. **c**, Correlation of ASV relative abundance by SAMPL-seq versus bulk 16S sequencing for cores 1 and 2 at particle sizes of 15–30 µm. **d**, Scatterplot of particles showing the relationship between mixing and read count. **e**, Heat map of particles filtered to >50 reads per particle and prevalent (>1% relative abundance) ASVs clustered by Bray–Curtis similarity and the Ward’s method at particle sizes of 15–30 µm.

spatial structure persisted over the 5-day sampling period (Fig. 4b and Extended Data Fig. 8e,f). To understand the overall spatial architecture in the longitudinal H11 dataset, we generated a co-association network using ASV pairs found across 2 or more days (Fig. 4c, Supplementary Table 5 and Methods). This spatial network of 33 ASVs could be grouped into four major clusters (L1–L4). Cluster L1 was composed of Gram-positive Ruminococcaceae and Lachnospiraceae, with *Faecalibacterium prauznitzi* (ASV2) acting as a central hub that linked with all other ASVs in the cluster. In contrast, cluster L3 contained mostly Lachnospiraceae with a denser subnetwork between *Fusicatenibacter saccharivorans* (ASV9), *Blautia massiliensis* (ASV13), *Blautia* sp. (ASV8), *Ruminococcus bromii* (ASV10), *Dorea longicatena* (ASV16) and *Agathobacter rectalis* (ASV1). Another distinct cluster L2 contained mostly Gram-negative Bacteroidaceae and Parabacteroidaceae, with *Bacteroides vulgatus/dorei* (ASV3) appearing as a

central hub. *B. vulgatus* and *B. dorei* could not be uniquely resolved due to high 16S V4 similarity. Finally, cluster L4 contained *Eubacterium coprastanoligenes* (ASV22), *Alistipes marseille* (ASV27) and *Ruminococcus bicirculans/champanellensis* (ASV4).

Across clusters, a strong inter-phylum co-association was observed between *B. vulgatus/dorei* (ASV3) of L2 and *A. rectalis* (ASV1) of L3. Moreover, *R. bicirculans/champanellensis* (ASV4) of L4 was co-associated with *F. prauznitzi* (ASV2) of L1 and anti-associated with several Lachnospiraceae from L3. To quantify the phylogenetic relatedness within spatial clusters, we calculated their respective net related indices (NRI), which showed clusters L2 and L3 individually having greater phylogenetic grouping than by chance ($Z = -2.57, -2.42, P = 0.003$ BH FDR corrected, for both L2 and L3; Extended Data Fig. 9a), and thus sharing more similar phylogenetic assortment of ASVs (Fig. 4d). Together, these results reveal a co-association network of temporally stable gut

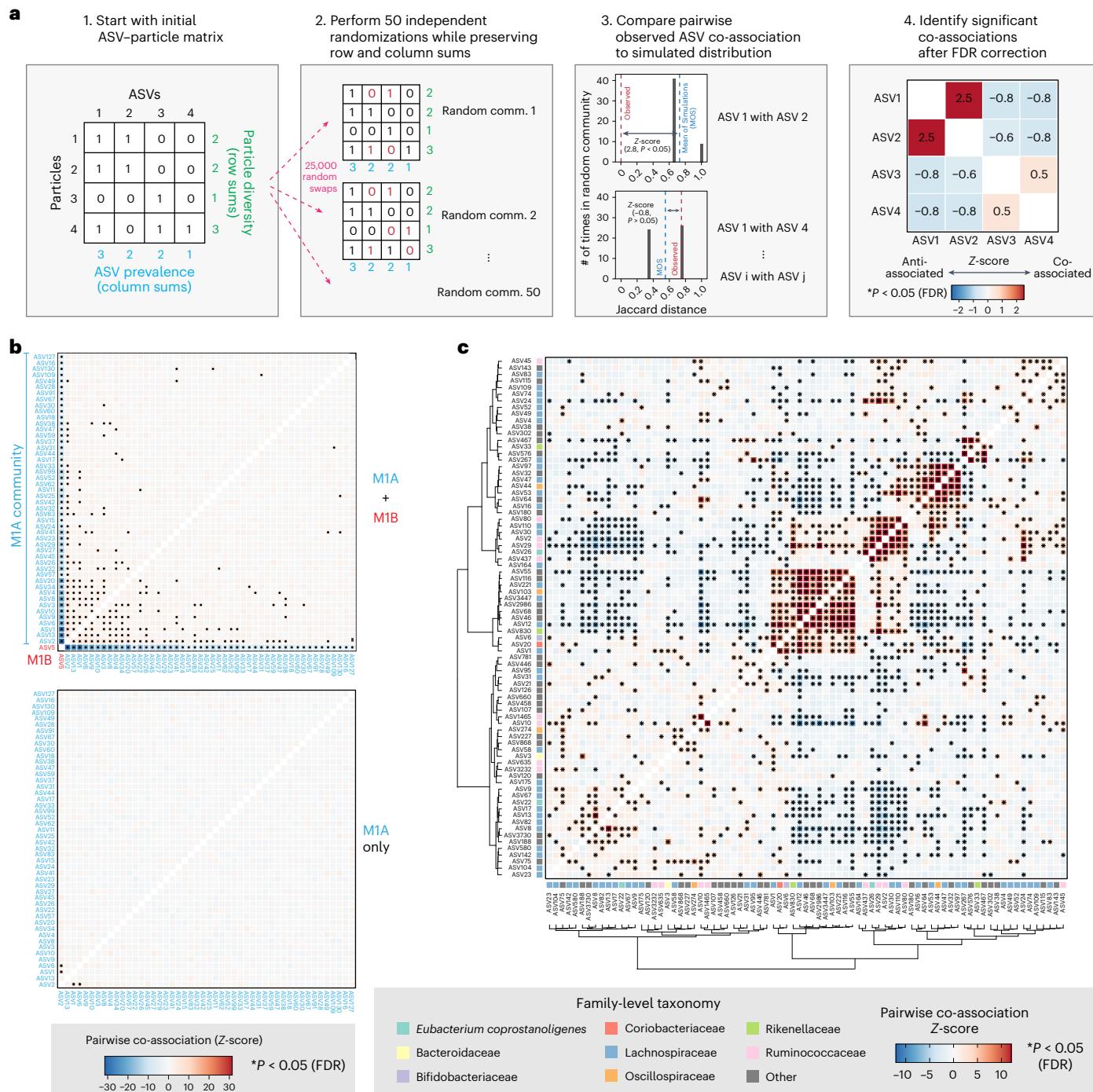


Fig. 3 | Co-localization analysis using SAMPL-seq data. **a**, Diagram of the null model analysis. Particle count data were binarized and subjected to the ‘sim9’ random swap algorithm. This was performed 50 times in parallel and the resulting randomized data were used to generate a null distribution of co-associations between ASVs. Then, a two-sided Z-score along with significance was calculated by comparing the observed co-association of an ASV pair to its null distribution. The *P* values of pairwise associations were FDR corrected

and significantly co-associated pairs of ASVs were identified. **b**, Pairwise co-association strength between ASVs in the homogenized M1 mixing experiment. Stars correspond to statistical significance of the two-sided Z-score ($P < 0.05$ FDR). Our method shows robust detection of the two separate communities M1A and M1B, along with minimal detection of significant associations in M1A alone. **c**, Example of pairwise ASV co-association patterns in human sample H1 using the co-localization analysis.

microbial assemblies that organizes into distinct ‘spatial hubs’ with varying levels of phylogenetic relatedness.

Conserved spatial hubs of gut microbiota across humans

We next sought to explore whether spatial co-association patterns were conserved across people. Even though many ASVs were unique to each person (that is, only 12 prevalent ASVs were found in all 5 individuals),

we identified a median of 261 significant co-associations across a median of 48 prevalent ASVs per individual. *F. prausnitzii* (ASV2) had the highest number of co-associations across the dataset (Fig. 5a). Other Ruminococcaceae, including ASVs 29, 45 and 80, were also highly co-associated, while Lachnospiraceae ASVs 8, 9 and 13 were frequently anti-associated. Approximately 85% of ASV pairs had consistent co- or anti-associations in two or more people (Fig. 5b, and Supplementary

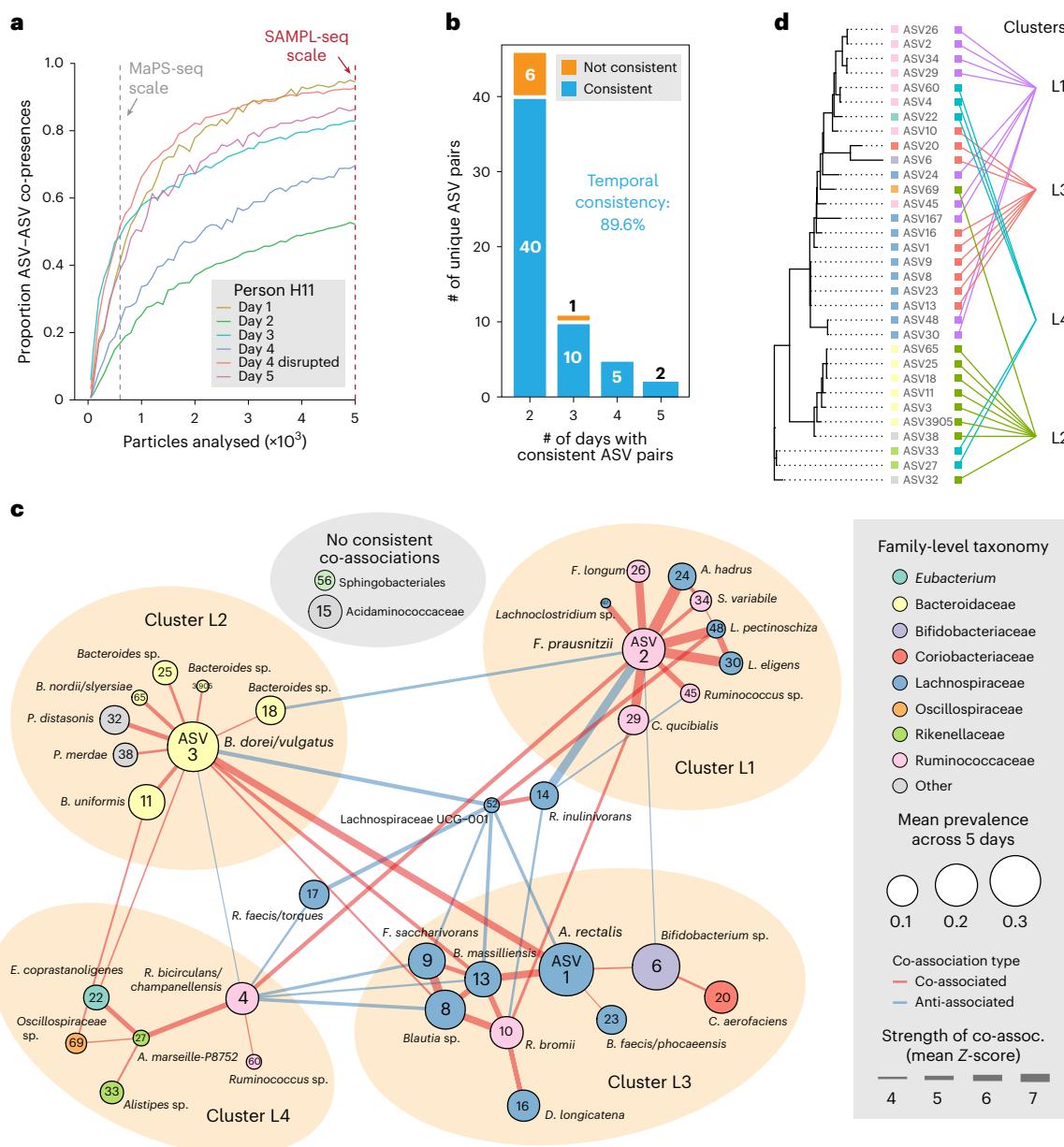


Fig. 4 | A longitudinal profile of gut microbiota spatial co-associations from one person. **a**, Rarefaction plot of unique ASV-ASV co-occurrence in a particle (mean diameter ~40 μm) (observed >3 times). SAMPL-seq scale shows the typical number of particles that could be sampled by SAMPL-seq, while MaPS-seq scale shows the typical number of particles that could be sampled by MaPS-seq. **b**, Barplot of pairs of ASVs shared across days coloured by the consistency of their

association. **c**, Network plot of ASV associations found on at least 2 days. Nodes are ASVs with size corresponding to mean prevalence across 5 days. Edges are association strengths, with colour representing type. ASVs without edges did not have consistent associations across multiple days. **d**, Phylogenetic tree of ASVs and their spatial hub cluster grouping.

Fig. 8e and Table 6). For ASV pairs found in three or more individuals, the spatial network showed three dominant hubs (P1–P3) (Fig. 5c and Supplementary Table 7).

Hub P1 is highly connected, composed of Ruminococcaceae and Lachnospiraceae; *F. prausnitzii* was co-localized with all other cluster members (similarly to its hub architecture in L1), while *Cibicibacter quicibialis* (ASV29), *Lachnospira eligens* (ASV30) and *Faecalibacterium hattori* (ASV80) were also strongly co-associated. Hub P2 contained *Bacteroides*, including *B. dorei/vulgatus* ASV3, along with *A. rectalis* (ASV1) and *Collinsella aerofaciens* (ASV20). The *B. dorei/vulgatus* and *A. rectalis* co-association was the strongest across both longitudinal and interpersonal datasets (L2 and P2 hubs). Finally, hub P3 is composed purely of Lachnospiraceae, including *F. sacchivorans* (ASV9), *Blautia massiliensis* (ASV13) and *Blautia* sp. (ASV8). These P3 members were

found to also co-associate in longitudinal cluster L3 in H11; they also showed strong anti-association with *F. prausnitzii* from hub P1, suggesting spatial segregation. Members of hubs P1 and P3 were significantly more related within each cluster than by chance ($Z = -1.91, P = 0.034$; $Z = 0.21, P = 0.006$, both BH FDR corrected) (Fig. 5d and Extended Data Fig. 9b).

Conserved longitudinal and interpersonal spatial patterns showed strong agreement, with ASV co-association pairs agreeing in their magnitude and sign (that is, co- or anti-association) (Pearson's $r = 0.7$; Extended Data Fig. 9c,d). The spatial grouping of ASVs in longitudinal (L1–L4) and interpersonal (P1–P3) hubs also showed significant overlap, as ASVs were more likely to be found in the same hubs than by chance (chi-squared test, $\chi^2 = 22.4$, d.f. = 6, $P = 0.001$; Extended Data Fig. 9e). The overlapping membership of longitudinal

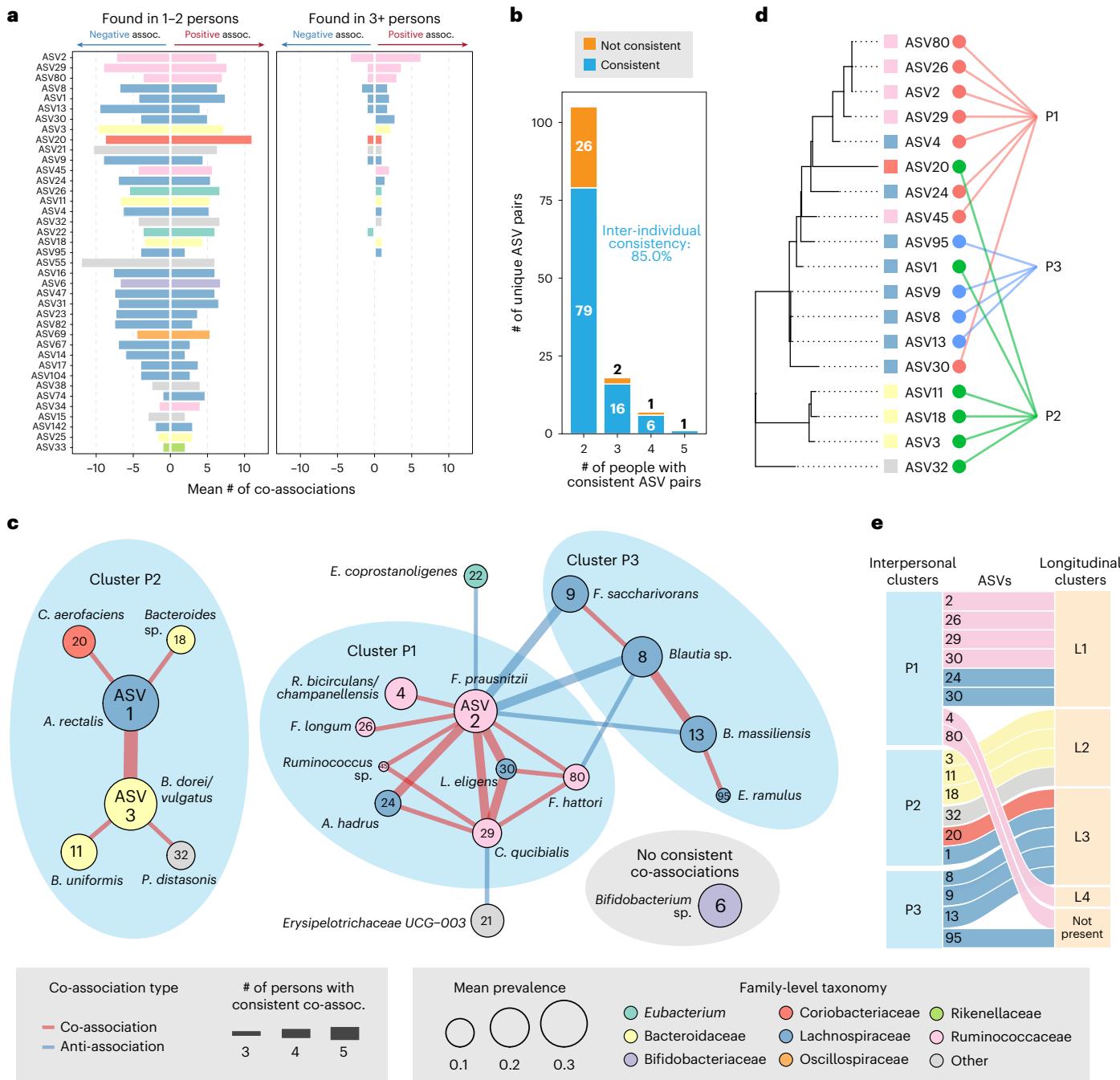


Fig. 5 | Consistent spatial hubs of gut microbiota found across humans. **a**, Barplot of the mean number of co-associations for prevalent ASVs across all participants (mean particle diameter ~40 µm). **b**, Barplot of pairs of ASVs shared across participants coloured by the consistency of their association. **c**, Network plot of ASV associations found in at least 3 participants. ASVs without edges

did not have consistent associations across multiple days. **d**, Phylogenetic tree of interpersonal cluster members, along with taxonomy and cluster group. **e**, Alluvial plot showing correspondence of ASVs from interpersonal spatial clusters (P1–P3) and longitudinal spatial clusters (L1–L4).

and interpersonal spatial hubs appears to be due to discrete sets of ASVs in both clusters; 6 ASVs present in P1 and L1, 4 ASVs present in P2 and L2, 3 ASVs from P3 and 2 ASVs from P2 forming L3 (Fig. 5e). The taxonomic composition of our observed spatial groups is also noteworthy, with L2 and P2 dominated by *Bacteroides*, one of the core guilds in the microbiome²⁹, while clusters L1, L3, P1 and P3 are dominated by Firmicutes, which belong to another main guild. Thus, spatial hubs present in both our interpersonal and longitudinal data-sets indicate a consistent spatial pattern that is stable between people and over time, and add to the evidence for conserved guilds in the human gut microbiome.

Spatial changes of the human gut microbiome during dietary perturbation

Diet can have a profound impact on the gut microbiome both in terms of its composition and metabolism. However, we do not know how dietary changes alter the spatial arrangement of bacteria in the human gut, although studies in mice show the potential for large spatial changes^{8,12}. We therefore applied SAMPL-seq to uncover possible micron-scale changes in the spatial organization of the gut microbiome following a dietary intervention. We chose inulin as the perturbation since inulin is a common food component not metabolized by human enzymes, correlates with short-chain fatty acid fermentation and can affect growth

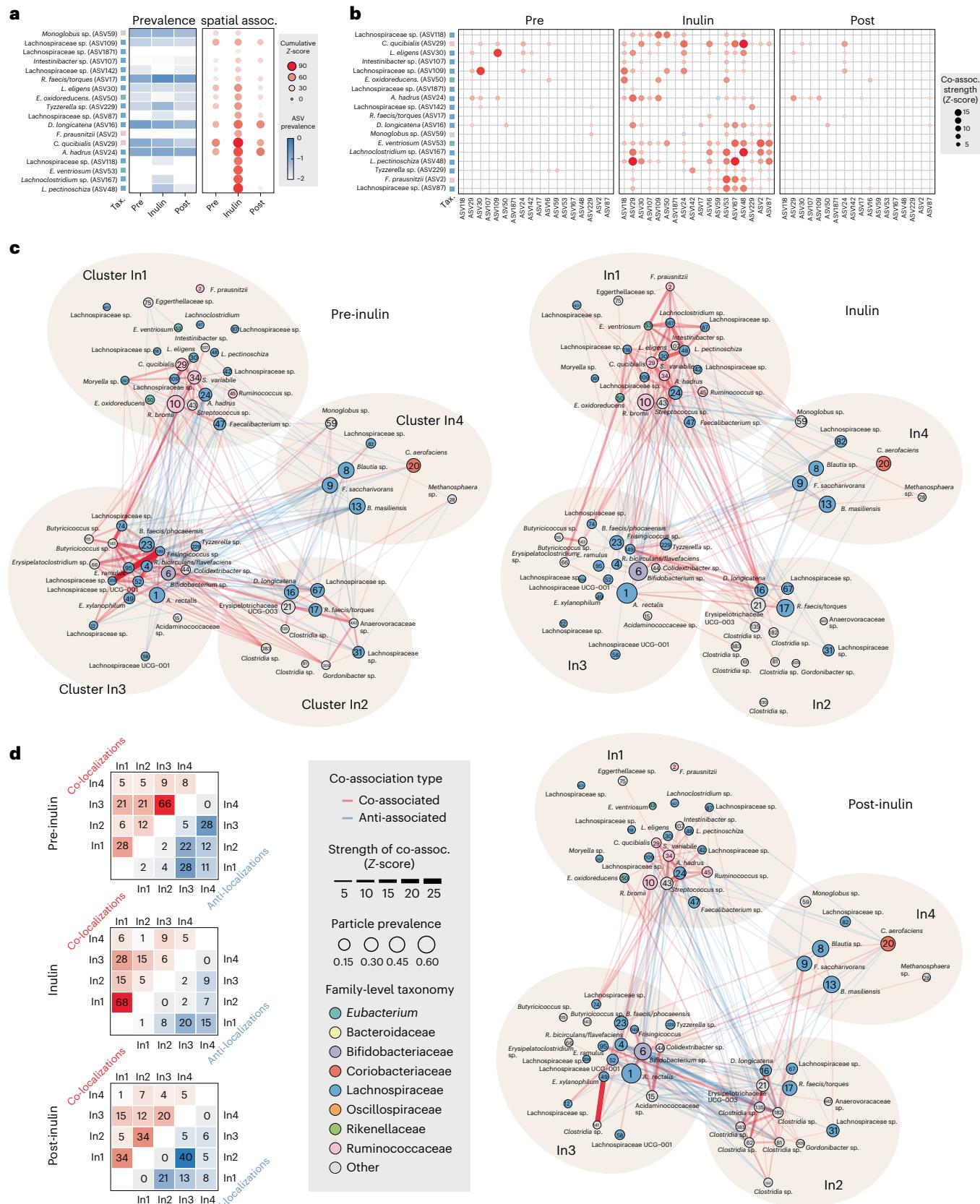


Fig. 6 | Spatial reorganization of the human gut microbiome in response to inulin supplementation. **a**, Comparison of ASV prevalence versus cumulative spatial association for 18 ASVs that strongly respond to inulin supplementation (mean particle diameter \sim 40 μm). **b**, A dotplot of pairwise spatial associations among 16 inulin-responsive ASVs before, during and post

inulin supplementation. **c**, Network plot of prevalent ASVs before, during and post inulin supplementation showing 4 major spatial clusters (In1–In4). **d**, Heatmaps summarizing the number of positive (red) and negative (blue) spatial localizations found between ASVs within and between clusters.

of beneficial commensal bacteria such as *Bifidobacterium*³⁰. We gave individual H11 oral inulin supplementation (20 g day⁻¹) in a 12-day study (Methods). Stool was obtained at baseline (4 days), during supplementation (4 days) and after discontinuation of supplementation (4 days). Both bulk 16S sequencing and SAMPL-seq were performed on these samples to assess compositional and spatial organizational changes.

Bulk 16S sequencing revealed no major alterations in the overall community composition (Extended Data Fig. 10), consistent with previous observations³⁰. With SAMPL-seq data, we first quantified the magnitude and total number of spatial interactions of an ASV by calculating its cumulative association Z-score (caZ-score) with all other ASVs, which showed large-scale spatial reorganization during inulin supplementation (Fig. 6a and Extended Data Fig. 10). While many ASVs had substantial caZ-score changes with inulin, including *F. prausnitzii* (ASV2) and *C. quicibialis* (ASV29), their abundance in the population did not change. This suggests that SAMPL-seq can identify alterations to the spatial organization of the microbiota that cannot be obtained via conventional bulk 16S analysis. For ASVs with the greatest overall changes in caZ-scores, we then assessed their pairwise spatial co-associations (Fig. 6b and Supplementary Table 8). With inulin, numerous ASVs had more spatial associations, suggesting the formation of new spatial pairings such as a notable triad of *L. pectinoschiza* (ASV48), *C. quicibialis* (ASV29) and *Lachnoclostridium* sp. (ASV167). When inulin is removed, these spatial structures also disappear, indicating an inulin-dependent change in the spatial organization.

We next visualized the entire co-association network to better understand the global spatial changes of all ASV pairs (Fig. 6c and Supplementary Table 9). Four main inulin-mediated clusters emerged (In1, In2, In3 and In4), similar to the number of clusters previously found in H11 (L1–4). Cluster In1, comprising *Ruminococcaceae* and *Lachnospiraceae*, shared substantial overlap with previously observed clusters L1 and P1. We then assessed the number of positive and negative co-associations within and across the four spatial hubs (Fig. 6d). Before inulin exposure, cluster In3 exhibited the largest number of within-cluster positive associations (66). With inulin, these In3 associations mostly disappeared (dropped to 6) while In1 formed numerous new within-cluster spatial associations (totalling 68). When inulin was removed, within-In1 associations dropped back down to 34, but In3 associations did not fully recover to their pre-inulin levels (20 versus 66). Among these spatial changes, *L. pectinoschiza* (ASV48) was one of the major drivers of the observed spatial changes, with 20 new associations occurring only during inulin supplementation. *Lachnospira pectinoschiza* is an anaerobic gut bacteria known to utilize dietary fibres such as pectin³¹. Other key inulin-stimulated ASVs pairings in In1 involved *F. prausnitzii* (ASV2), *E. venturius* (ASV53), *C. quicibialis* (ASV29) and *Lachnoclostridium* sp. (ASV167) (Fig. 6c), which aligns with previous documented evidence of inulin metabolism by members of the *Lachnospiraceae* and *Ruminococcaceae* families³². Post inulin, we found more spatial associations in In2 than before, suggesting a spatial restructuring of the community. Nevertheless, the overall spatial patterns post inulin were more similar to those pre inulin, indicating reversible spatial restructuring by a dietary component. Collectively, these results highlight the coordinated spatial response of microbial hubs to the transient availability of a common dietary metabolite.

Discussion

Spatial metagenomics enabled by SAMPL-seq facilitates facile and high-throughput delineation of microbial co-localization at the micron scale. SAMPL-seq preserves spatial structure, as evidenced by low mixing rates, allows the profiling of tens of thousands of ‘microbial plots’ at a time, which is at least an order of magnitude improvement in scale over state-of-the-art methods in plot sampling, and is key for accurate estimation of microbial co-localization²⁸. The ability of SAMPL-seq to provide high taxonomic resolution and local spatial information nicely complements imaging-based methods that can give global spatial

positions of specific taxa^{11–13,16}. Application of SAMPL-seq to stool samples yields microbiome co-association data that reflect the spatial organization found in the large intestine, thus allowing for non-invasive and longitudinal analysis of the colon at steady state and during dietary or other environmental perturbations.

Both the pairwise associations and spatial hubs found in the human gut may be hallmark features of a stable and healthy microbiota³³, which when disturbed in disease states could lead to community-wide destabilization. Strains that grow together in spatial hubs may be metabolically coupled or share a similar niche preference³⁴. Previous work suggests that *Bacteroides* form a dominant ‘guild’ whose members are ecologically similar or metabolically complementary with one another in the Western adult gut²⁹, and our results showed that this group is also spatially organized as seen in clusters L2 and P2. *F. prausnitzii* is one of the most abundant butyrate-producing gut bacteria and its absence has been linked to disease-associated dysbiosis³⁵. The observed central role *F. prausnitzii* (ASV2) has in the P1 (and L1) spatial hub is particularly noteworthy, as it may indicate possible interspecies nutrient exchange. Indeed, past *in vitro* and *in vivo* experiments showed that *F. prausnitzii* grows better in the presence of other gut taxa^{36,37}. *Agathobacter rectalis* (ASV1) and *Bacteroides dorei/vulgatus* (ASV3) were observed to be the most consistent and significant ASVs across all individuals. Both ASVs have been observed to localize in the mucus layer, with *B. dorei* and *B. vulgatus* contributing to mucus degradation³⁸, and *A. rectalis* showing preferential mucosal colonization despite an inability to utilize mucosal sugars^{39,40}. Interestingly, *A. rectalis* can use sugars liberated by *Bacteroides* sp. to produce butyrate⁴¹, which may explain their strong co-localization in the gut. Additional studies are needed to better elucidate the nature of these relationships *in vivo*.

During inulin supplementation, we observed notable changes in spatial associations, including new spatial interactions between *L. pectinoschiza* (ASV48) and other *Ruminococcaceae* including *F. prausnitzii* and *C. quicibialis*, which are involved in SCFA production and microbiome stabilization. Interestingly, *A. rectalis* (ASV1) and *Bifidobacterium* sp. (ASV 6), which are known to consume inulin³², did not form additional spatial co-associations during inulin supplementation, indicating a spatially independent inulin metabolic process. Nevertheless, inulin supplementation has been shown to enhance SCFA production³⁰, which could be driven by the expanded co-associations within *Lachnospiraceae* and *Ruminococcaceae* families of cluster In1. Previous results in mice show that fibre depletion leads to increased homogenization throughout the lumen¹², and similarly, in our data, supplementation with inulin increased spatial structuring.

Further mechanistic experiments to probe the underpinnings that shape the observed microbial co-localizations could lead to better ways to modulate the gut microbiome and cultivate gut bacteria that have been recalcitrant to laboratory domestication³⁷. SAMPL-seq could be applied to other microbiomes such as those in soil or in foods to discover unseen spatially mediated microbial interactions⁴² and build more accurate community-scale metabolic models⁴³. With additional advancements, SAMPL-seq could evolve to encompass whole-genome sequencing and incorporate genomic information from host cells, enabling us to associate spatial interactions with microbial genes, pathways and microbiome–host spatial interactions.

Methods

Sample collection and study protocol

Written informed consent was obtained from individuals following approved protocol (IRB-AAAT4813, Columbia University Institutional Review Board) and no compensation was received for study participation. Participants ranged between 25 and 35 years of age, and all 5 individuals in the study were male. Bulk human faecal samples were extracted from intact faecal sample using a sterile loop, placed in a cryovial and stored at -80 °C until use. Samples used for strain isolation were extracted from an intact faecal sample using a sterile loop, added

to sterile pre-reduced PBS and then processed in an anaerobic chamber. Samples were disrupted by vortexing and then passed through a 40 µm filter. The resulting slurry was then diluted 1:1 with 50% glycerol in PBS and stored at -80 °C until use. SAMPL-seq human faecal cores were derived from intact faecal samples. Using the wide-diameter end of a P20 filter tip (Rainin), pieces of faecal sample were 'cored' and then immediately placed in tubes containing methacarn (60% methanol, 30% chloroform, 10% acetic acid). After 1 day of fixation, samples were removed from the P20 tip and allowed to fix for an additional 12–24 h. Then samples were transferred to 70% ethanol and stored at 4 °C until use. Samples were used within 1 month. Mouse small intestine, caecum, large intestine and faecal samples were collected from a single 12-week-old female C57BL6/J mouse from Envigo laboratories (Protocol AABD4554, Columbia University Institutional Animal Care and Use Committee). The mouse was individually caged and kept on a 12 h day/night cycle. Samples were extracted and placed in methacarn for 24 h. Once fixed, sections were cut to 3 × 3 mm and used for downstream processing.

Detailed SAMPL-seq protocol

Sample embedding. Faecal cores were cut to no larger than 3 × 3 mm with a sterile razor to ensure full polymerization and placed in a sterile PCR tube. Disrupted faecal samples were generated by bead beating a 5-mm-diameter faecal pellet with 0.1 mm glass beads for 1 min at 4 °C. Cores were then washed twice with 200 µl 1× PBS, 200 µl permeabilization solution (1× PBS, 0.1% Triton X-100 (v/v)) was added to the tubes, and samples were then incubated for 5 min. Then, all excess solution liquid was removed from the tube and samples were placed in a drying oven set to 90 °C for 10 min. Once removed from the oven, samples were placed on ice to cool before embedding. The embedding solution contained 1× PBS, 10% (w/w) acrylamide, 0.25% (w/w) bisacrylamide, 5 µM primer (pe1) (Supplementary Table 10), 0.2% (w/w) 4-hydroxy-2,2,6,6-tetramethylpiperidin-1-oxy and 0.2% (w/w) tetramethylethylenediamine. The PE1 primer contains an acrydite group to enable adhesion to the gel and a photocleavable spacer to allow for release using UV light. Samples were then covered with embedding solution to completely cover the sample (~20–30 µl) and remained on ice for 5 min. The excess embedding solution was removed and an additional fresh solution was added to cover the sample. Samples were then incubated on ice for 6–12 h to ensure full perfusion. For final polymerization, excess embedding solution was removed and samples were incubated on ice for 1 h. After incubation, samples were placed in a 95 °C oven for up to 30 min to ensure polymerization. Once embedded, samples were extracted from the PCR tube, excess polymer was trimmed using a sterile razor and then washed with 1 ml PBS.

Particle fracturing. Samples were first placed in a stainless-steel microvial (Biospec, 2007). Next, samples were frozen using liquid nitrogen for 2 min without submerging the vial. Before proceeding, samples were shaken to ensure that the sample could move freely. Next, a single 6.35 mm stainless-steel bead (Biospec, 11079635ss) was added to the vial and the vial was plugged with a silicone rubber plug cap (Biospec, 2008). The sample was then placed in liquid nitrogen for at least 2 min. Immediately thereafter, the sample was transferred to a bead beater (Biospec, 11201) and beaten for 10 s at 3,800 rcf. Samples were then resuspended in 1 ml PBS. The suspended samples were then passed through a 100-µm cell strainer (Greiner Bio One, 542100) into a new sterile tube. Particles were then washed twice more with 1× PBS. Washes were performed by spinning the sample down at 20,000 rcf, removing excess PBS without disturbing the particle pellet and then adding 1 ml PBS.

Particle lysis. Particles were resuspended in 500 µl lysis buffer (10 mM Tris HCl (pH 8), 1 mM EDTA, 100 mM NaCl) along with 375 U µl⁻¹ lysozyme (Epicentre, R1810M), and incubated at 37 °C for 1 h. Next,

samples were resuspended in 500 µl digestion buffer (30 mM Tris HCl pH 8.0, 1 mM EDTA, 0.5% Triton X-100, 800 mM guanidine HCl) and 0.1 µg µl⁻¹ proteinase K (Epicentre, MPRK092). Samples were then incubated at 65 °C for 15 min, and then at 95 °C for 5 min to inactivate proteinase K. Particles were then washed 3 times with TET (10 mM Tris HCl pH 8.0, 1 mM EDTA, 0.1% Tween 20). If not proceeding to the next step, samples were brought to 15% glycerol and frozen at -20 °C until further use.

Barcode of particles via primer extension. This protocol uses a modified version of the procedures in ref. 19 to barcode primers present in the particles. The embedded primers in each particle were iteratively extended by primer extension over three rounds. All particle washes were done as follows: sample pellet was resuspended in 1 ml of washing solution and then spun down at 20,000 rcf for 1 min. The supernatant was removed. For each sample, a 96-well PCR plate was prepared with 1 µl of unique primer (Supplementary Table 11) distributed to each well (pe1, pe2, pe3 primer sets). Samples were then washed 3 times with wash buffer (10 mM Tris HCl pH 8.0, 0.1 mM EDTA, 0.1% Tween 20) and adjusted to a volume of ~833 µl. A volume of 110 µl 10× isothermal amplification buffer (NEB), 33 µl 10 mM dNTPs [0.3 mM final] (NEB) and 14 µl 8,000 U ml⁻¹ Bst2.0 [100 U ml⁻¹ final] (NEB) were added to the sample. Then, 9 µl particle/Bst2.0 mix was distributed to each well, either using a pipette or a Mantis liquid handler (Formulatrix). Plates were sealed and incubated at 60 °C for 30 min. Then 20 µl of STOP25 (10 mM Tris HCl pH 8.0, 25 mM EDTA, 0.1% Tween 20, 100 mM KCl) was added to each well and plates were incubated at r.t. for 5 min. Plates were then pooled into a 5 ml Eppendorf tube and the total volume brought to 5 ml with STOP25 to completely stop the reaction. The conical tube was then spun down at 20,000 rcf for 2 min, the supernatant was removed and the pellet transferred to a 1.5 ml tube. The pellet was then washed 3 times with STOP10 (10 mM Tris HCl pH 8.0, 10 mM EDTA, 0.1% Tween 20, 100 mM KCl). To ensure that primers were single stranded for the next barcoding reaction, 1 ml freshly made DENATURE (0.5% Brij35, 150 mM NaOH) was used to resuspend the particles and this was incubated at r.t. for 10 min. The particles were then washed 3 times with DENATURE and once with NEUTRALIZE solution (100 mM Tris HCl pH 8.0, 10 mM EDTA, 0.1% Tween 20, 100 mM NaCl). This protocol was then repeated at the wash steps for each round of barcoding. If the protocol was stopped between barcoding rounds, the particles were washed 3 times with TET, brought to 10% glycerol and frozen at -20 °C until the process was continued. Once barcoding was complete, incompletely extended primers needed to be removed. This was accomplished using hybridization to protect complete primers and then Exo1 digestion to remove the rest. Samples were washed 3 times with wash buffer and once with HYBRIDIZE (10 mM Tris HCl pH 8.0, 0.1 mM EDTA, 0.1% Tween 20, 330 mM KCl). Then the volume of the sample was adjusted to 300 µl with HYBRIDIZE and 7.5 µl 1 mM M16S_515f_RC primer [-20 µM final] was added. This solution was incubated at 50 °C for 1 h to hybridize. Then, 50 µl 10× Exo1 buffer [1× final], 112.5 µl nuclease-free water and 7.5 µl Exo1 [0.3 U µl⁻¹ final] were added and the solution incubated at 37 °C for 1 h. Then the tube was filled with STOP25, mixed and incubated at r.t. for 5 min. This was then washed 3 times with STOP 10, incubated with DENATURE for 10 min at r.t., washed 3 times with DENATURE, once with NEUTRALIZE and 3 times with TET, similar to the above barcoding. If stopped here, the solution was brought to 10% glycerol and stored at -20 °C.

Size filtering. To ensure consistent sizing, cell strainers were used. Samples were washed 3 times with PBS and resuspended in 1 ml PBS. PBS was used as other buffers would impede flow through the filter. Samples were first passed through a 40-µm cell strainer (GBO, 542140), the strainer was washed with an additional 3 ml PBS and the samples allowed to flow into the same tube. To recover particles larger than the filter, the strainer was inverted and placed onto a new tube. PBS (1 ml)

was then passed through the strainer. This procedure was then repeated using the smaller filtered fraction and a 20- μ M cell strainer (GBO, 542120). Once collected, all 5 ml tubes were spun down at 20,000 rcf, the supernatant removed and particles put into 1.5 ml tubes. Then all samples were washed 3 times with TET, brought to 10% glycerol and frozen at -20 °C until the process was continued. Particle concentrations and sizes were determined by microscopy using a haemocytometer (Bulldog Bio, DHC-N420). Particles were stained with SYBR green I [1x final] and imaged using a Nikon TI2 microscope. Particles were identified using the binary/define threshold function and the equivalent diameter calculated using the NIS-Elements software.

In situ PCR. Once quantified, particles were aliquoted for PCR reactions. Between 1,000 and 10,000 particles were amplified at a time. Particles were washed 3 times with TET, volume adjusted to 22.5 μ l and transferred to PCR tubes. PCR was then set up with the following reagents: 2.5 μ l 10 μ M pe2 816r REV primer [0.5 μ M final] and 25 μ l of KAPA Hifi 2 \times Readymix (Roche, KK2601). It was then cycled with the following parameters: 98 °C for 30 s, 15 cycles; 98 °C for 10 s; 55 °C for 30 s; 65 °C for 60 s; extension at 65 °C for 2 min. Particles were then washed 3 times with TET. This process was then repeated twice, for a total of 45 cycles.

UV release and magnetic bead cleanup. To release DNA from the particles, particle aliquots were washed 3 times with diffusion buffer (0.1% SDS, 1 mM EDTA, 500 mM ammonium acetate), brought to a volume of 100 μ l with diffusion buffer and transferred to PCR tubes. Particles were then placed on ice and treated with UV radiation for 15 min to break the photocleavable spacer. Aliquots were then incubated at 50 °C to allow for DNA diffusion into solution. Aliquots were then mixed at a 1:1 ratio with magnetic beads (Speedbeads, Cytiva, 65152105050250) and cleaned using a standard protocol. Cleaned DNA was eluted into 22 μ l.

Indexing PCR. Purified PCR product (10 μ l) was transferred to a new PCR tube with the following reaction setup: 12.5 μ l KAPA Hifi 2 \times Readymix (Roche, KK2601), 2 mM SYTO9 (Thermo-Fisher, S34854), 1 μ l forward index primer, 1 μ l reverse index primer, 10.5 μ l of in situ PCR product. The samples were then run using a qPCR (BioRad, CFX96) with the following programme: 98 °C for 45 s, 30 cycles; 98 °C for 10 s; 68 °C for 20 s; 65 °C for 30 s; repeat at 65 °C for 120 s; 10 °C hold. Samples were removed during the extension phase if they appeared to leave the linear phase of PCR (usually between 14–20 cycles) and then replaced during the final extension. The resulting PCR product was assessed using a 2% acrylamide gel, the ~490 bp band was extracted and purified (NEB Monarch, T1020L), and then stored at -80 °C until use.

Sequencing and read processing. Samples were sequenced on the Nextseq 550 system (Illumina) using the 150 bp mid-output or high-output kit, depending on the number of samples, with a 30% phiX spike in. Over 10 million reads per particle library were targeted to ensure sufficient particle coverage. The resulting sequencing reads needed additional processing to identify the particle barcode sequence. After demultiplexing, reads were demultiplexed using a custom BASH script (<https://github.com/wanglabcumc/SAMPL-seq>). Reads were first filtered using USEARCH10 (ref. 44), with a cut-off of less than 1 expected error and minimum length of 150 bp. Then, the particle barcode was identified and extracted from each read using ULTRAPLEX⁴⁵ and a custom barcode mapping. The 16S primers were then stripped, particle names modified using SeqKit⁴⁶ to allow for recognition by USEARCH/VSEARCH⁴⁷, reads of less than 69 bp were removed and remaining reads were truncated to 69 bp using SeqKit. The resulting reads correspond to a 69-bp 16S V4 region. Then, all samples except for the M2 mixing experiment were pooled for denoising using UNOISE3 (ref. 48), and reads were mapped to ASVs using VSEARCH. For the M2 mixing experiment, reads were mapped directly to the reference 16S provided by

the manufacturer. Since particle barcodes can contain errors, particle barcodes were extracted and subjected to error correction using the DNABarcodes⁴⁹ package in R using a custom script (<https://github.com/wanglabcumc/SAMPL-seq>). Our barcode set allows for error correction of 1 base error, so barcodes with hamming distance larger than 1 were considered uncorrectable and removed. Approximately 96% of all barcode sequences were either correct or correctable. The resulting corrected ASV table was then used for subsequent analysis. For species-level identification, 16S ASV sequences were matched to cultured strains from H1. If not present in the dataset, strains were matched to the refseq 16S database at 100% identity. If no match was found, ASVs were labelled using the most specific taxonomic level available. The SINA Aligner⁵⁰ was used to create 16S rRNA alignments, which were then used to generate a phylogenetic tree with FastTree⁵¹. Taxonomy was also assigned with the SINA search and classify tool, and SILVA⁵² taxonomy was used for downstream analysis.

Detailed SAMPL-seq data analysis

Statistics and reproducibility. Rarefaction curves and estimates from the literature²⁸ were used to determine the number of particles needed for co-association analysis. All individuals who participated in this study were used to look at natural variability in the spatial structure of the gut microbiome, but no statistical method was used to predetermine sample size. Data were not assumed to be normal, hence non-parametric tests were used across the paper. The experiments were not randomized, as there was $N=1$ for perturbation experiments, and other data were used as a whole. Data collection and analysis were not performed blind to the conditions of the experiments

Filtering. To remove potential read-through between particles, ASVs must be present at greater than 2% relative abundance within each particle to be considered ‘present’. Particles with less than 25 reads were removed from analysis. For visualization and co-association analysis, particles with 2 or fewer ASVs were removed, as it cannot be distinguished whether they represent a failed amplification or a monolithic community.

Rarefaction. Rarefaction was performed on individual amplification replicates for the subset of ASVs with >1% particle prevalence across each amplification replicate. Unique ASV–particle pairs were used as the measurement as they represent ‘new’ ASVs being found in new particles. Using reads from filtered particles, reads were sampled 10 times at a given level and the resulting number of unique read–particle pairs averaged at that point. This was repeated until the maximum number of reads was reached.

Co-localization. Co-association was quantified using a custom implementation of the SIM9 algorithm²⁰, chosen for its low false positive rate, as implemented in the ‘sim9_single’ function in the EcoSimR package. The script used and an example are included (<https://github.com/wanglabcumc/SAMPL-seq>). In brief, on each set of particles from one individual, a binarized (presence–absence) ASV table was subjected to a random swap, which preserved the ASV prevalence and particle diversities. Since this step only swapped a subset of values, it was performed 25,000 times to generate a ‘randomized’ community on the basis of the original diversity of the dataset. Fifty of these randomized communities were generated to generate a null distribution of ASV co-localization. Then, the observed co-localization was compared to the distribution using a Z-test, and the resulting significance was subjected to FDR correction using the Benjamini–Hochberg procedure, with significance being an FDR-corrected $P < 0.05$.

ASV association networks. The longitudinal association graph was generated by subsetting to ASVs pairs found to be significantly associating on 2+ days and averaging the Z-score over that time. The

interpersonal association graph was generated by subsetting to ASVs pairs found to be significantly associating in 3+ donors and averaging the Z-score over that time. ASVs in each graph were then clustered using the spinglass clustering method as implemented in igraph⁵³.

Net relatedness. Net relatedness was calculated using the ‘ses.mpd’ function of the R package Picante⁵⁴. The taxalabels of each cluster were randomized 10,000 times and the random MPD distribution was used to calculate the *P* value. The *P* values were then corrected using the Benjamini–Hochberg FDR correction.

Interpersonal distance at the ASV or family level. Bray–Curtis distances were calculated between individual donors, using either ASV relative abundances or aggregated family-level relative abundances, and compared using a Wilcoxon rank-sum test to determine whether distances were significantly higher when looking at the family level.

Plotting. Plotting for most graphs was performed using ggplot2 (ref. 55). Correlation was added to plots using ggpublisher⁵⁶. Statistical tests were performed using R 4.0. ASV association graphs were generated using ggraph⁵⁷. Particle-level heat maps were generated using the geom_tile() function in ggplot. Particles were clustered using the Simpson overlap at the sample level, and ASVs were clustered by their Jaccard overlap across all particles in the heat map.

Barcode validation experiment. To validate the presence of barcodes after barcoding, aliquots of 10,000 barcoded but unamplified particles were obtained and subjected to UV release, as described above. Then, the purified DNA was subjected to PCR using primers targeting anchor regions of the primers, with primer PE1 serving as the forward primer. For the reverse primers, Anchor 1-RC targeted the first extension, Anchor 2-RC the second and 515RC the full length of the primer (Supplementary Table 8). Reactions were set up with 5 µl KAPA HiFi 2× Readymix (Roche, KK2601), 1 µl forward primer (0.3 µM), 1 µl reverse primer (0.3 µM), 2 µl of cleaned primer DNA and 1 µl of nuclease-free water. Cycling was performed at 98 °C for 3 min, 30 cycles; 98 °C for 20 s; 60 °C for 20 s; 65 °C for 20 s; repeat at 65 °C for 120 s.

Mixing experiments. Mixing rate calculations were performed in two ways. In the first case, two bacterial communities were assembled: a homogenized faecal sample and a pure culture of *S. pasteurii*, an environmental bacteria not found in the gut. The homogenized faecal community M1A was generated by bead beating a 5-mm-diameter faecal pellet with 0.1 mm glass beads for 1 min at 4 °C. The resulting solution was passed through a 40-µm cell strainer, fixed in methacarn, washed with PBS, and subjected to the same embedding and polymerization protocol described above. Once polymerized and washed, samples were then subjected to cryofracturing together in replicate with equal volumes of each polymerized community. Once co-fractured, the mixed community was treated as a single community and subjected to the rest of the protocol described above. In the second case, a mixed community was generated using two defined communities: the ZymoBIOMICS Gut Microbiome Standard (Zymo, D6331) and ZymoBIOMICS Spike-in Control I (D6320). Cell concentrations were matched between them (2×10^6 and 6×10^6 cells per µl). Each was embedded separately in equal volumes of embedding solution. As above, equal volumes were mixed during cryofracturing and processed according to the protocol described above. The mixing rate was calculated using the percentage of the particle assigned to the spike-in community, either the *S. pasteurii* or ZymoBIOMICS spike-in control. Particles were considered mixed if they contained between 10–90% of the spike in. The multiplet rate was calculated as previously described³². The particle capture rate was calculated by dividing the number of particles after quality control (QC) per library by either 10,000 or the number of particles identified

before QC, whichever was greater. The estimated number of particles added to the sample for sequencing was 10,000, quantified using a haemocytometer.

Bulk 16S sequencing. Bulk 16S samples were acquired by chemical or physical lysis. For chemical lysis, 3 × 3 mm faecal samples were washed twice with PBS. Next, samples were homogenized by vortexing in 500 µl lysis buffer (10 mM Tris HCl pH 8, 1 mM EDTA, 100 mM NaCl). Lysozyme was then added to the sample (final concentration \sim 375 U µl⁻¹), and the sample was vortexed and incubated at 37 °C for 1 h. Subsequently, 500 µl digestion buffer (50 mM Tris HCl pH 8.0, 1 mM EDTA, 1% Triton X-100, 1,600 mM guanidine HCl) was added along with proteinase K to 0.1 µg µl⁻¹. The sample was vortexed again and placed at 65 °C for 15 min. Then, 100 µl of lysate was removed, subjected to a 1× bead cleanup and resuspended in 22 µl of nuclease-free H₂O. Physical lysis was performed using our established sequencing pipeline without spike in⁵⁸. Dual indexing amplification was performed using a modified protocol⁵⁹. TruSeq 16S versions of the Earth Microbiome 515F and 806R⁶⁰ matching those used in the SAMPL-seq protocol were used for the first round of amplification, and standard TruSeq indices were used for the second round of amplification. Both rounds were performed using a qPCR, with samples removed before the end of linear amplification, usually between 8–12 cycles. Bulk samples were then pooled with SAMPL-seq libraries for sequencing.

Frozen-sample processing. Faecal sample cores were taken from intact faecal samples as described earlier. These cores were divided in half, and one half was fixed immediately in methacarn at r.t. and processed as described earlier. The other half was immediately placed in a –80 °C freezer and kept frozen for up to 1 week. When ready for SAMPL-seq processing, the frozen sample was placed into pre-chilled (–20 °C) methacarn, and fixation proceeded at 4 °C for 24 h. Once fixed, the sample was processed as described earlier.

Reporting summary

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

Data availability

Raw sequencing reads are available from PRJNA1196417. The Refseq 16S database (https://www.ncbi.nlm.nih.gov/refseq/targetedloci/16S_process/) was used for species-level taxonomic identification, and SILVA 138.1 (<https://www.arb-silva.de/documentation/release-138/>) was used for taxonomic assignment of all ASVs. Source data are provided with this paper.

Code availability

Scripts for read processing are implemented in BASH and R. They are available from GitHub at <https://github.com/wanglabcumc/SAMPL-seq> (ref. 61).

References

- Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14**, 20–32 (2016).
- Gilbert, J. A. et al. Current understanding of the human microbiome. *Nat. Med.* **24**, 392–400 (2018).
- 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).
- Ho, P.-Y., Good, B. H. & Huang, K. C. Competition for fluctuating resources reproduces statistics of species abundance over time across wide-ranging microbiotas. *eLife* **11**, e75168 (2022).
- Wu, F. et al. Modulation of microbial community dynamics by spatial partitioning. *Nat. Chem. Biol.* **18**, 394–402 (2022).

6. Reichenbach, T., Mobilia, M. & Frey, E. Mobility promotes and jeopardizes biodiversity in rock–paper–scissors games. *Nature* **448**, 1046–1049 (2007).
7. 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).
8. Sheth, R. U. et al. Spatial metagenomic characterization of microbial biogeography in the gut. *Nat. Biotechnol.* **37**, 877–883 (2019).
9. Riva, A. et al. A fiber-deprived diet disturbs the fine-scale spatial architecture of the murine colon microbiome. *Nat. Commun.* **10**, 4366 (2019).
10. Koh, A., Vadder, F. D., Kovatcheva-Datchary, P. & Bäckhed, F. From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. *Cell* **165**, 1332–1345 (2016).
11. Valm, A. M. et al. Systems-level analysis of microbial community organization through combinatorial labeling and spectral imaging. *Proc. Natl Acad. Sci. USA* **108**, 4152–4157 (2011).
12. Earle, K. A. et al. Quantitative imaging of gut microbiota spatial organization. *Cell Host Microbe* **18**, 478–488 (2015).
13. Shi, H. et al. Highly multiplexed spatial mapping of microbial communities. *Nature* **588**, 676–681 (2020).
14. Mondragón-Palomino, O. et al. Three-dimensional imaging for the quantification of spatial patterns in microbiota of the intestinal mucosa. *Proc. Natl Acad. Sci. USA* **119**, e2118483119 (2022).
15. Cao, Z. et al. Spatial profiling of microbial communities by sequential FISH with error-robust encoding. *Nat. Commun.* **14**, 1477 (2023).
16. Lötstedt, B., Stražar, M., Xavier, R., Regev, A. & Vickovic, S. Spatial host–microbiome sequencing reveals niches in the mouse gut. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-023-01988-1> (2024).
17. Welch, J. L. M., 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* **114**, E9105–E9114 (2017).
18. Grodner, B. et al. Spatial mapping of mobile genetic elements and their bacterial hosts in complex microbiomes. *Nat. Microbiol.* <https://doi.org/10.1038/s41564-024-01735-5> (2024).
19. Zilionis, R. et al. Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protoc.* **12**, 44–73 (2017).
20. Blattman, S. B., Jiang, W., Oikonomou, P. & Tavazoie, S. Prokaryotic single-cell RNA sequencing by *in situ* combinatorial indexing. *Nat. Microbiol.* **5**, 1192–1201 (2020).
21. Kuchina, A. et al. Microbial single-cell RNA sequencing by split-pool barcoding. *Science* **371**, eaaba5257 (2021).
22. Bloom, J. D. Estimating the frequency of multiplets in single-cell RNA sequencing from cell-mixing experiments. *PeerJ* **6**, e5578 (2018).
23. Yamawaki, T. M. et al. Systematic comparison of high-throughput single-cell RNA-seq methods for immune cell profiling. *BMC Genomics* **22**, 66 (2021).
24. Ding, J. et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* **38**, 737–746 (2020).
25. Knight, R. et al. Best practices for analysing microbiomes. *Nat. Rev. Microbiol.* **16**, 410–422 (2018).
26. Gotelli, N. J. Null model analysis of species co-occurrence patterns. *Ecology* **81**, 2606–2621 (2000).
27. Ryan, M. J. et al. Development of microbiome biobanks – challenges and opportunities. *Trends Microbiol.* **29**, 89–92 (2021).
28. Blanchet, F. G., Cazelles, K. & Gravel, D. Co-occurrence is not evidence of ecological interactions. *Ecol. Lett.* **23**, 1050–1063 (2020).
29. Frioux, C. et al. Enterosignatures define common bacterial guilds in the human gut microbiome. *Cell Host Microbe* **31**, 1111–1125.e6 (2023).
30. Le Bastard, Q. et al. The effects of inulin on gut microbial composition: a systematic review of evidence from human studies. *Eur. J. Clin. Microbiol. Infect. Dis.* **39**, 403–413 (2020).
31. Cornick, N. A., Jensen, N. S., Stahl, D. A., Hartman, P. A. & Allison, M. J. *Lachnospira pectinoschiza* sp. nov., an anaerobic pectinophile from the pig intestine. *Int. J. Syst. Evol. Microbiol.* **44**, 87–93 (1994).
32. Riva, A. et al. Identification of inulin-responsive bacteria in the gut microbiota via multi-modal activity-based sorting. *Nat. Commun.* **14**, 8210 (2023).
33. Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489**, 220–230 (2012).
34. Wu, G., Zhao, N., Zhang, C., Lam, Y. Y. & Zhao, L. Guild-based analysis for understanding gut microbiome in human health and diseases. *Genome Med.* **13**, 22 (2021).
35. Miquel, S. et al. *Faecalibacterium prausnitzii* and human intestinal health. *Curr. Opin. Microbiol.* **16**, 255–261 (2013).
36. Kim, H., Jeong, Y., Kang, S., You, H. J. & Ji, G. E. Co-culture with *Bifidobacterium catenulatum* improves the growth, gut colonization, and butyrate production of *Faecalibacterium prausnitzii*: *in vitro* and *in vivo* studies. *Microorganisms* **8**, 788 (2020).
37. Huang, Y. et al. High-throughput microbial culturomics using automation and machine learning. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-023-01674-2> (2023).
38. Hoskins, L. C., Boulding, E. T., Gerken, T. A., Harouny, V. R. & Kriaris, M. S. Mucin glycoprotein degradation by mucin oligosaccharide-degrading strains of human faecal bacteria. Characterisation of saccharide cleavage products and their potential role in nutritional support of larger faecal bacterial populations. *Microb. Ecol. Health Dis.* **5**, 193–207 (1992).
39. Van den Abbeele, P. et al. Butyrate-producing *Clostridium* cluster XIVa species specifically colonize mucins in an *in vitro* gut model. *ISME J.* **7**, 949–961 (2013).
40. Ge, X. et al. SRS-FISH: a high-throughput platform linking microbiome metabolism to identity at the single-cell level. *Proc. Natl Acad. Sci. USA* **119**, e2203519119 (2022).
41. Mahowald, M. A. et al. Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proc. Natl Acad. Sci. USA* **106**, 5859–5864 (2009).
42. Wolfe, B. E. & Dutton, R. J. Fermented foods as experimentally tractable microbial ecosystems. *Cell* **161**, 49–55 (2015).
43. Levy, R. & Borenstein, E. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proc. Natl Acad. Sci. USA* **110**, 12804–12809 (2013).
44. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
45. Wilkins, O. G., Capitanchik, C., Luscombe, N. M. & Ule, J. Ultraplex: a rapid, flexible, all-in-one fastq demultiplexer. *Wellcome Open Res.* **6**, 141 (2021).
46. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **11**, e0163962 (2016).
47. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
48. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).

49. Buschmann, T. DNABarcodes: an R package for the systematic construction of DNA sample tags. *Bioinformatics* **33**, 920–922 (2017).
50. Pruesse, E., Peplies, J. & Glöckner, F. O. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823–1829 (2012).
51. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
52. Yilmaz, P. et al. The SILVA and 'All-species Living Tree Project (LTP)' taxonomic frameworks. *Nucleic Acids Res.* **42**, D643–D648 (2014).
53. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJ. Complex Syst.* **1695**, 9 (2006).
54. Kembel, S. W. et al. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464 (2010).
55. Wickham, H. *ggplot2. Wiley Interdiscip. Rev. Comput. Stat.* **3**, 180–185 (2011).
56. Kassambara, A. *ggpubr: 'ggplot2' based publication ready plots. R package version 0.6.0* <https://rpkgs.datanovia.com/ggpubr/> (2018).
57. Pedersen, T. L. *ggraph: an implementation of grammar of graphics for graphs and networks. R package version 2.2.1.9000* <https://github.com/thomasp85/ggraph> (2022).
58. Ji, B. W. et al. Quantifying spatiotemporal variability and noise in absolute microbiota abundances using replicate sampling. *Nat. Methods* **16**, 731–736 (2019).
59. Gohl, D. M. et al. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat. Biotechnol.* **34**, 942–949 (2016).
60. Caporaso, J. G. et al. *EMP 16S Illumina Amplicon Protocol*. <https://doi.org/10.17504/protocols.io.nuudeww> (2018).
61. Richardson, M. *SAMPL-Seq, V1.0* (2025); <https://doi.org/10.5281/zenodo.14619156>

Acknowledgements

We thank P. Sims, D. Vitkup, L. Dietrich, K. Tchourine for immensely helpful insights and discussions; G. Gerber and T. Gibson for strengthening the computational rigour of the method; and members of the Wang Laboratory for providing a supportive environment. H.H.W. acknowledges funding support from the NSF (MCB-2025515), NIH (2R01AI132403, 1R01DK118044, 1R01EB031935, 1R21AI146817), ONR (N00014-18-1-2237), Burroughs Wellcome Fund (1016691), ARO (W911NF-22-2-0210), DARPA (HR0011-23-2-0001), Irma T. Hirsch Trust, and the Schaefer Research Award. M.R., R.U.S. and F.V.-C. were supported by the NSF Graduate Research Fellowship Program (DGE-1644869). R.U.S. was supported by the Fannie and John Hertz Foundation Fellowship. This study was supported in part by the Columbia University Digestive and Liver Disease Research Center

(funded by NIH grant 5P30DK132710) through use of its bioinformatic and single-cell analysis core.

Author contributions

M.R., R.U.S. and H.H.W. conceived the project. R.U.S., M.R. and T.M. developed and validated the protocol. M.R., R.U.S., D.R., Y.H., L.L., J.L. and G.U. performed experiments. M.R., S.Z., Y.Q. and F.V.-C. analysed the data. M.R. and H.H.W. generated and edited figures. H.H.W. supervised the overall project. M.R., H.H.W. and S.Z. wrote the manuscript with input from co-authors. All authors reviewed and approved the manuscript.

Competing interests

H.H.W. is a scientific advisor of SNIPR Biome, Kingdom Supercultures, Fitbiomics, VecX Biomedicines, Genus PLC, and a scientific co-founder of Aclid and Foli Bio, all of which are not involved in the study. R.U.S. is a co-founder of Kingdom Supercultures. The remaining authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41564-024-01914-4>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41564-024-01914-4>.

Correspondence and requests for materials should be addressed to Harris H. Wang.

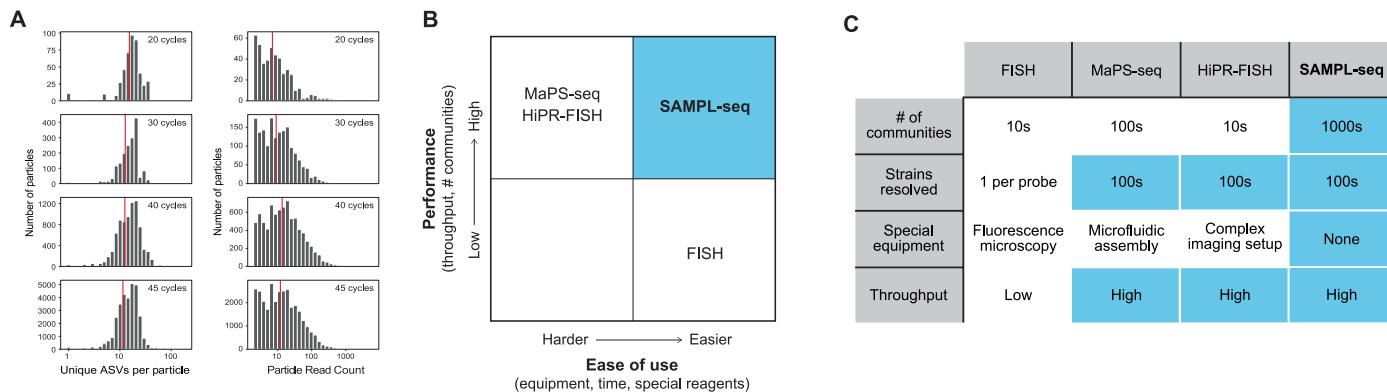
Peer review information *Nature Microbiology* thanks the anonymous reviewers for their contribution to the peer review of this work.

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

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

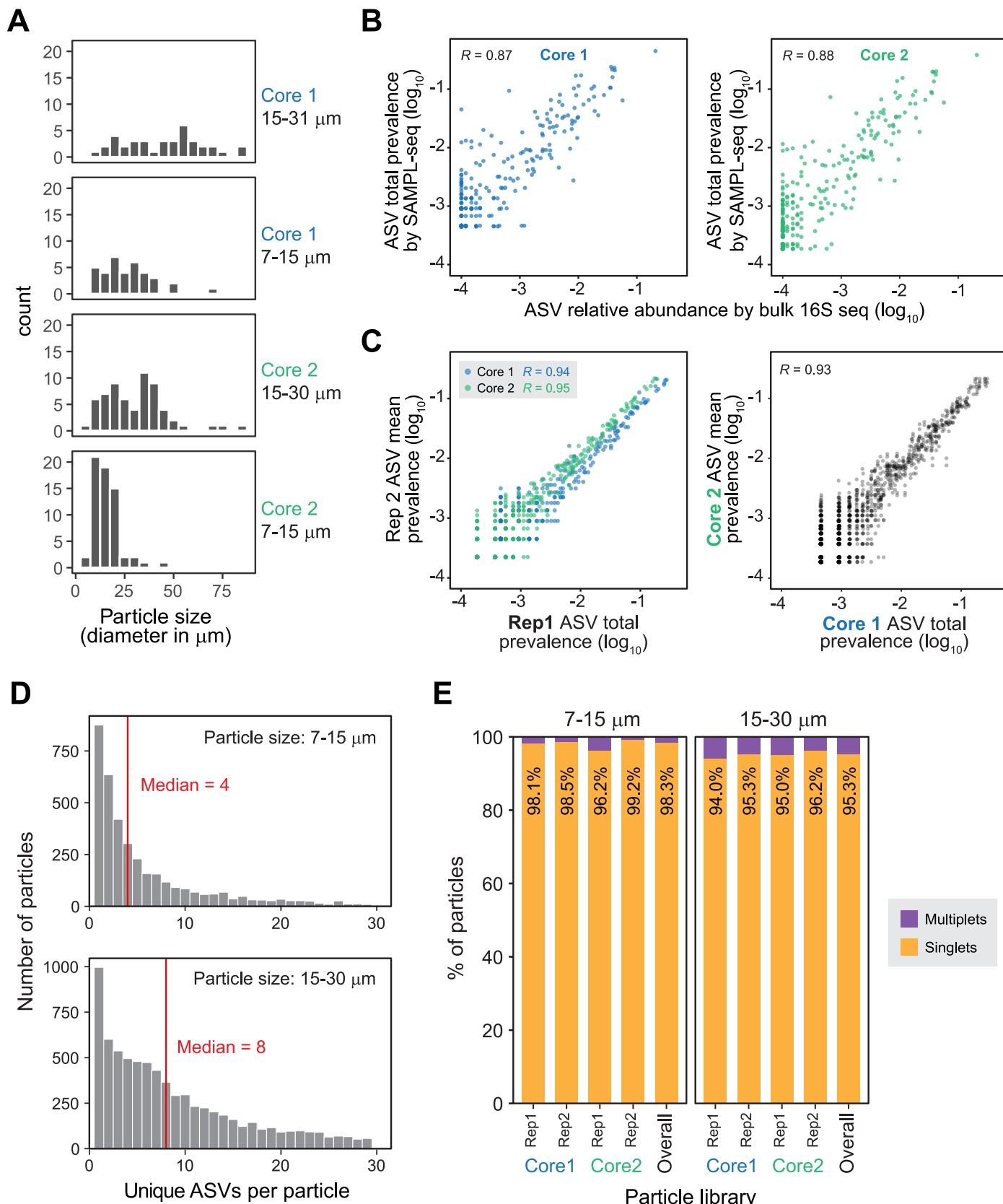
Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2025, corrected publication 2025



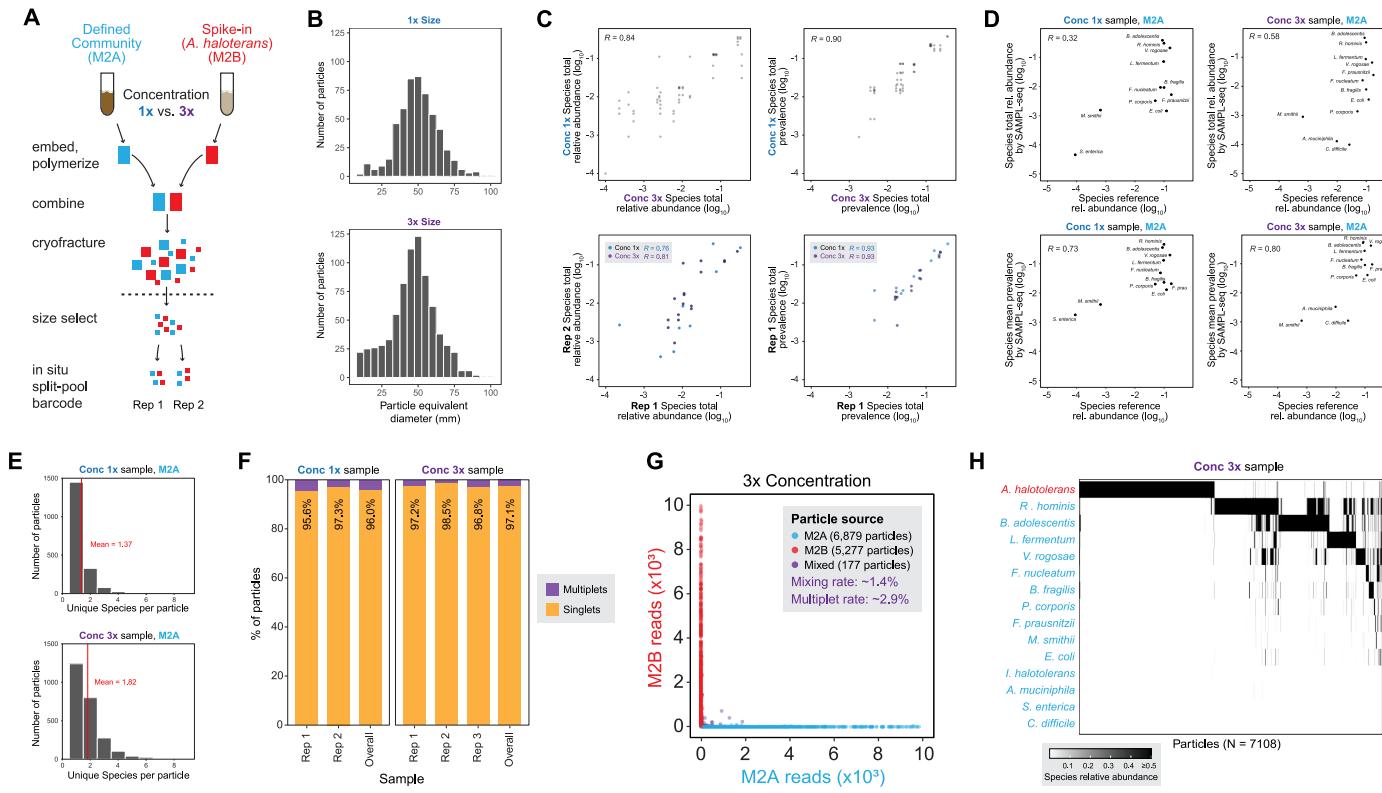
Extended Data Fig. 1 | Characterization of SAMPL-seq steps and comparison with other methods. (a) Histograms showing the effect of the number of in situ PCR cycles on both the ASVs per particle and reads per particle based on

different number of PCR cycles. (b) Plot summarizing the overall ease of use and performance of various microbial spatial analysis methods. (c) Table comparing the performance of different spatial analysis methods.



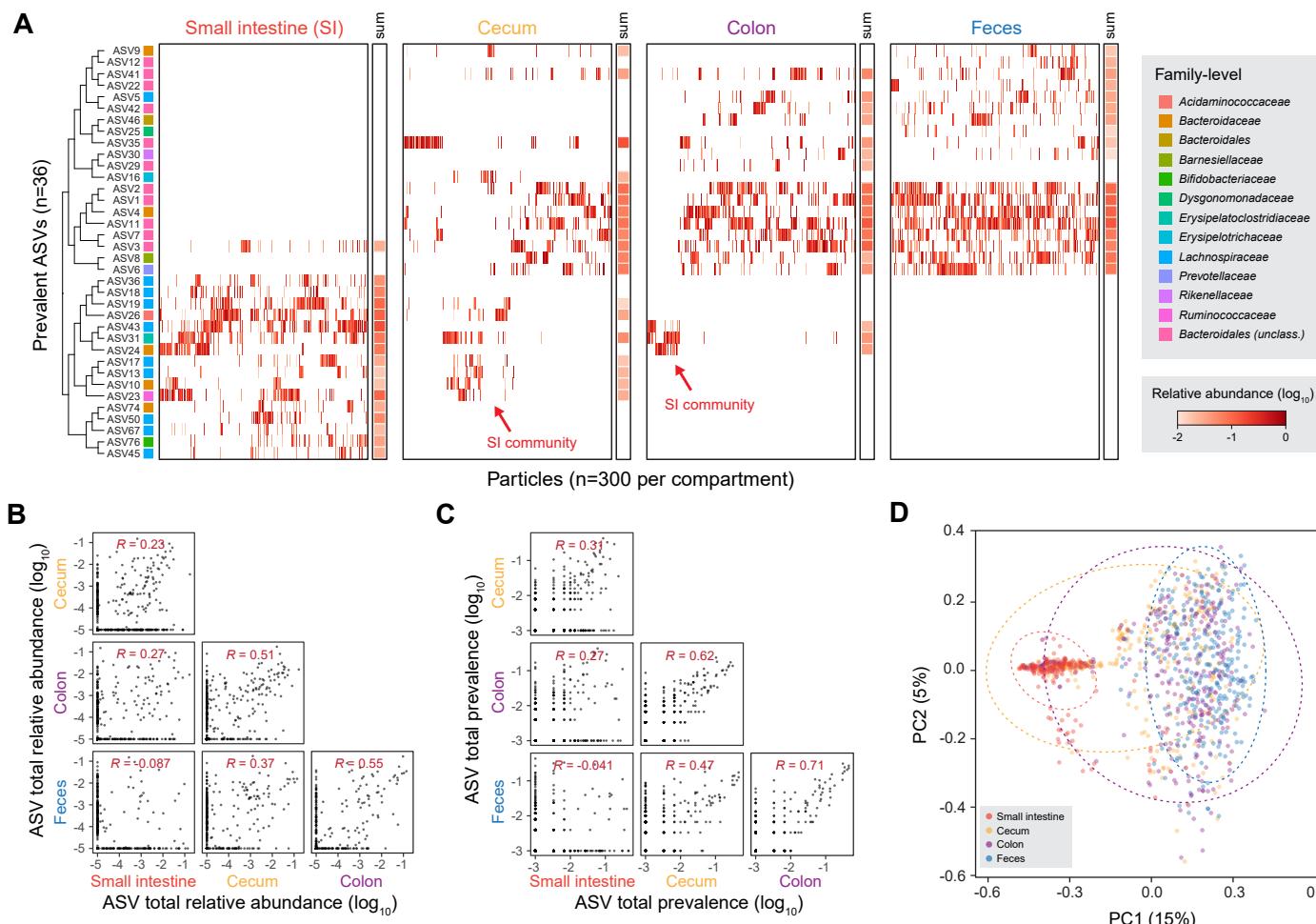
Extended Data Fig. 2 | Homogenized fecal mixing experiment (M1). (a) Histograms of particle sizes for the replicates. (b,c) Correlation of ASV prevalence by SAMPLE-seq and ASV relative abundance by bulk 16S sequencing for Core 1 at particle size 15–30 μm (b) and Core 2 (c). SAMPL-seq abundances are

averaged between replicates (excluding Spike-in). (d) Histogram of the ASV per particle distribution by size (excluding Spike-in). (e) Barplot of the singlet rate of each replicate, grouped by particle size.



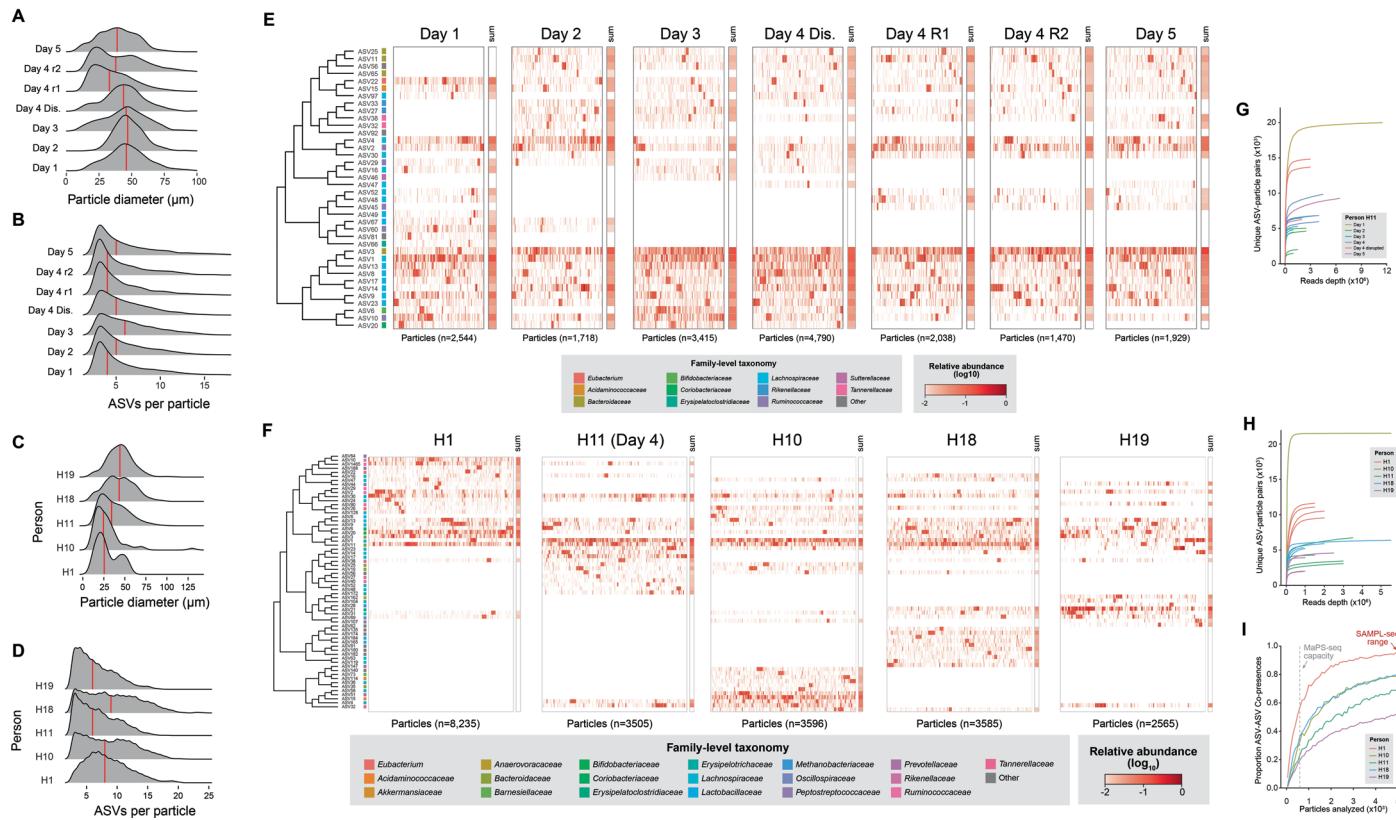
Extended Data Fig. 3 | Defined community mixing experiment (M2). (a)
An outline of the process for producing the fecal mixing library. Zymo Gut Microbial Standard and Zymo High Concentration Spike-in are separately embedded at equal cell ratios at 1x or 3x concentration replicates. They are then combined during the cryofracturing step, and are then size sorted, amplified and sequenced in aliquots of 10,000 particles, of average size ~50 μ m, which were used for further analysis **(b)** Histograms of particle sizes for the replicates. **(c)** Scatterplots of technical (amplification) and biological (concentration) replicates, using both the relative abundance based on summed reads, and

ASV prevalence among particles, which is the percentage of particles an ASV is found (excluding the Spike-in). **(d)** Scatterplot of ASV relative abundance and prevalence compared to absolute reference provided by the manufacturer. SAMPL-seq abundances are averaged between replicates (excluding Spike-in). **(e)** Histograms of the ASV per particle distribution by concentration (excluding Spike-in). **(f)** Barplot of the multiplet rate of each replicate, grouped by concentration. **(g)** Plot showing mixing rates of two defined communities (M1A and M1B), with each colored dot corresponding to a classified particle. **(h)** Heatmap of particles clustered by Bray-Curtis similarity and the Ward's method.



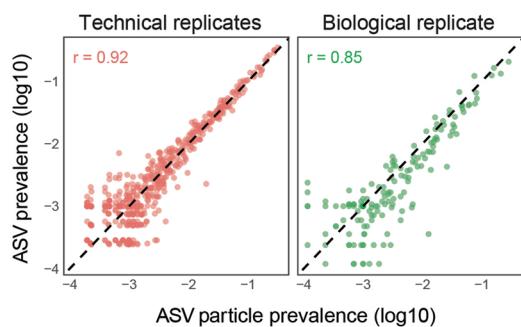
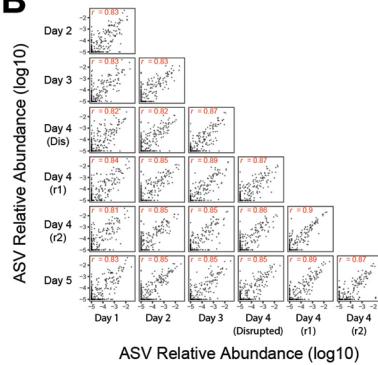
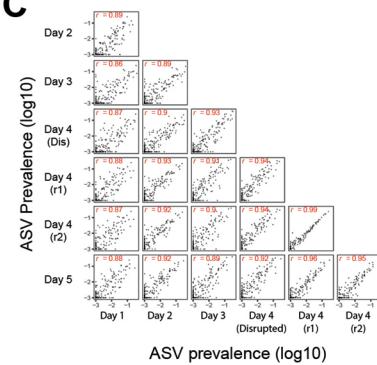
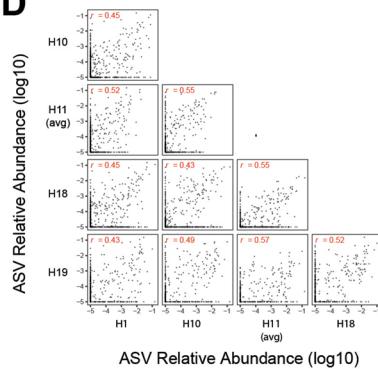
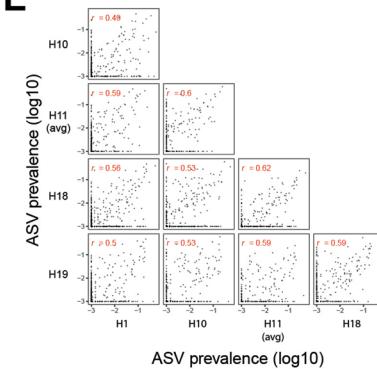
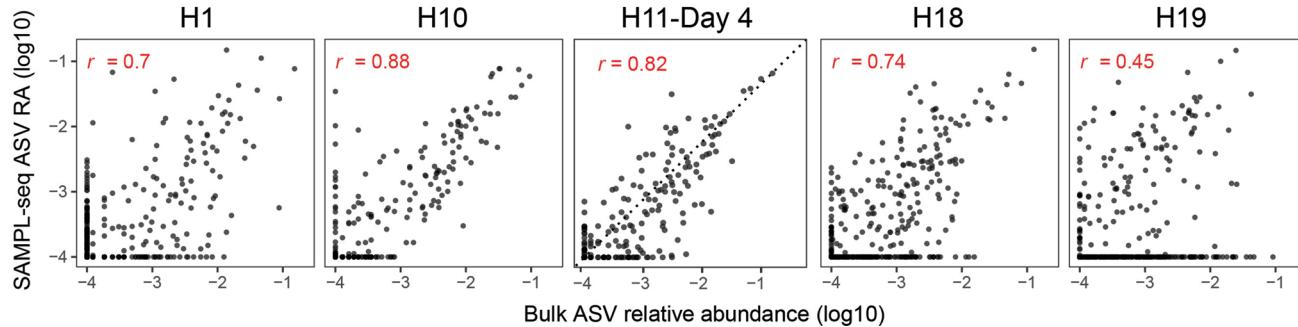
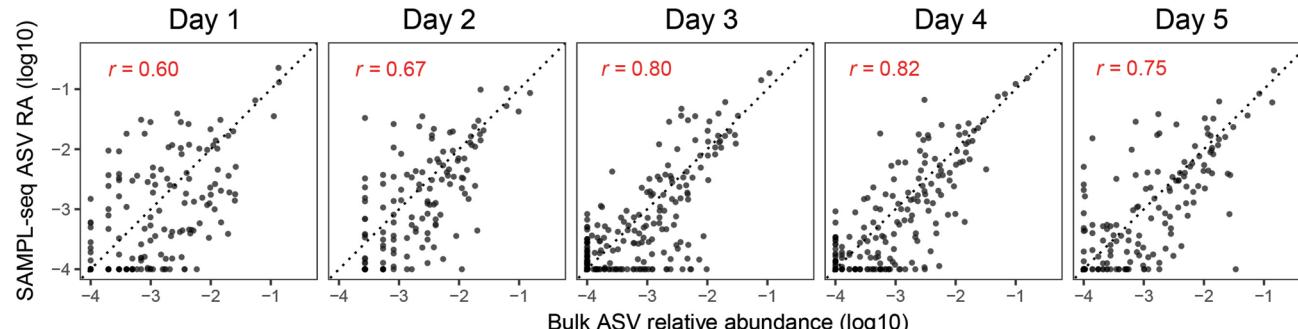
Extended Data Fig. 4 | Correlation between ASVs from different murine gut compartments. (a) Clustered heatmap of prevalent mouse ASVs grouped by gut compartments, with summed abundances at the end of the row. ASVs are clustered by Jaccard overlap across the dataset. (b, c) Correlation between ASV relative abundance (b) and prevalence among particles (c) between mouse gut

compartments. Colon and feces samples showed the highest correlation among samples. (d) Principal Coordinate Analysis (PCoA) plot of particles derived from different mouse gut compartments using Simpson distance, colored by gut compartments. Dashed circles correspond to the 95% confidence interval for each compartment using the multivariate t-distribution.



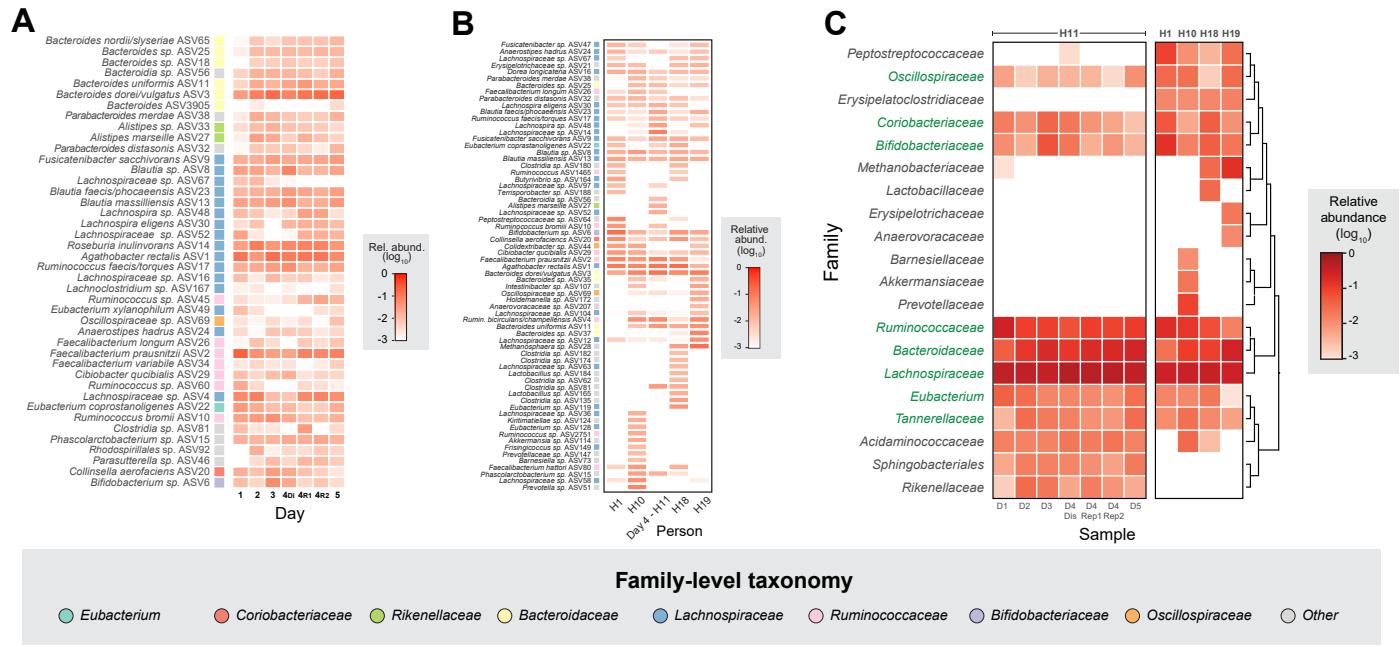
Extended Data Fig. 5 | Particle-level data of the human gut microbiome from stool profiling. (a, b) Using filters for 20–40 μm , distributions of particle sizes (a) and ASVs per particle (b) for longitudinal human stool samples from H11 are shown. (c, d) Using filters for 20–40 μm , distributions of particle size (c) and ASVs per particle (d) for interpersonal samples are shown. (e, f) Particles from longitudinal (e) or interpersonal (f) are clustered within each day using

the Simpson overlap, and ASVs are clustered using their Jaccard overlap across all days. (g) Rarefaction plot for longitudinal samples of unique ASV-particle pairs for prevalent ASVs (>1% prevalence in particles). (h, i) Rarefaction plots for interpersonal samples of unique ASV-particle pairs (h) for prevalent ASVs (>1% particle prevalence) or unique ASV-ASV co-presence (i) in a particle (observed >3 times).

A**B****C****D****E****F****G****Extended Data Fig. 6 | Technical validations of H11 Day 4 SAMPL-seq**

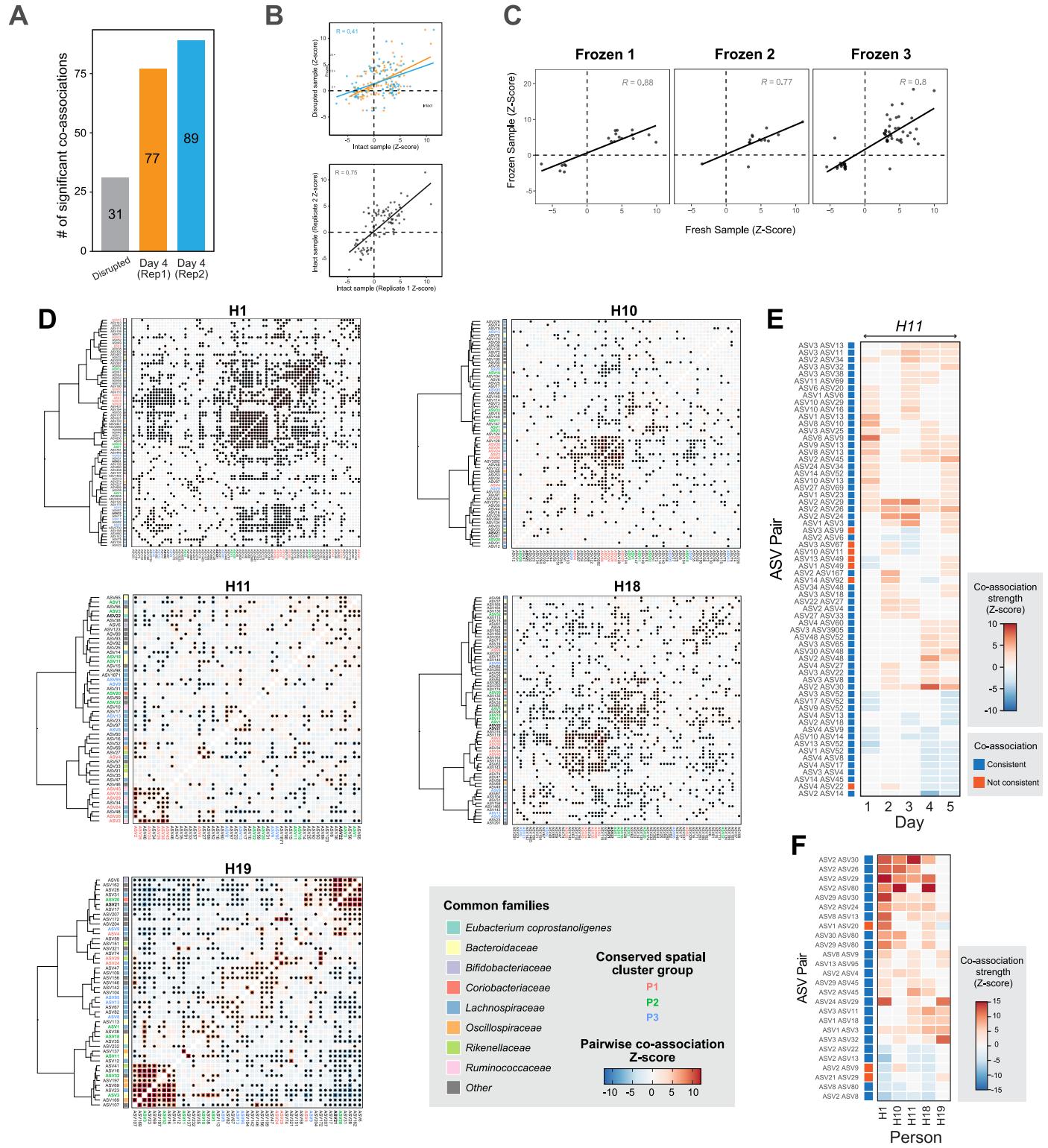
Libraries. (a) Scatterplots of amplification (technical) replicates showed high correlation. Correlation between spatial (biological) replicates also showed high correlation. Homogenized sample showed high correlation, but increased particle prevalence relative to intact libraries. (b, c) Correlation of bulk ASV

relative abundance (b) or ASV prevalence (c) between longitudinal SAMPL-seq libraries. (d, e) Correlation of bulk ASV relative abundance (d) or ASV prevalence (e) between interpersonal SAMPL-seq libraries. (f, g) Correlation of ASV relative abundance between interpersonal (f) or H11 longitudinal (g) samples with their corresponding bulk measurements.



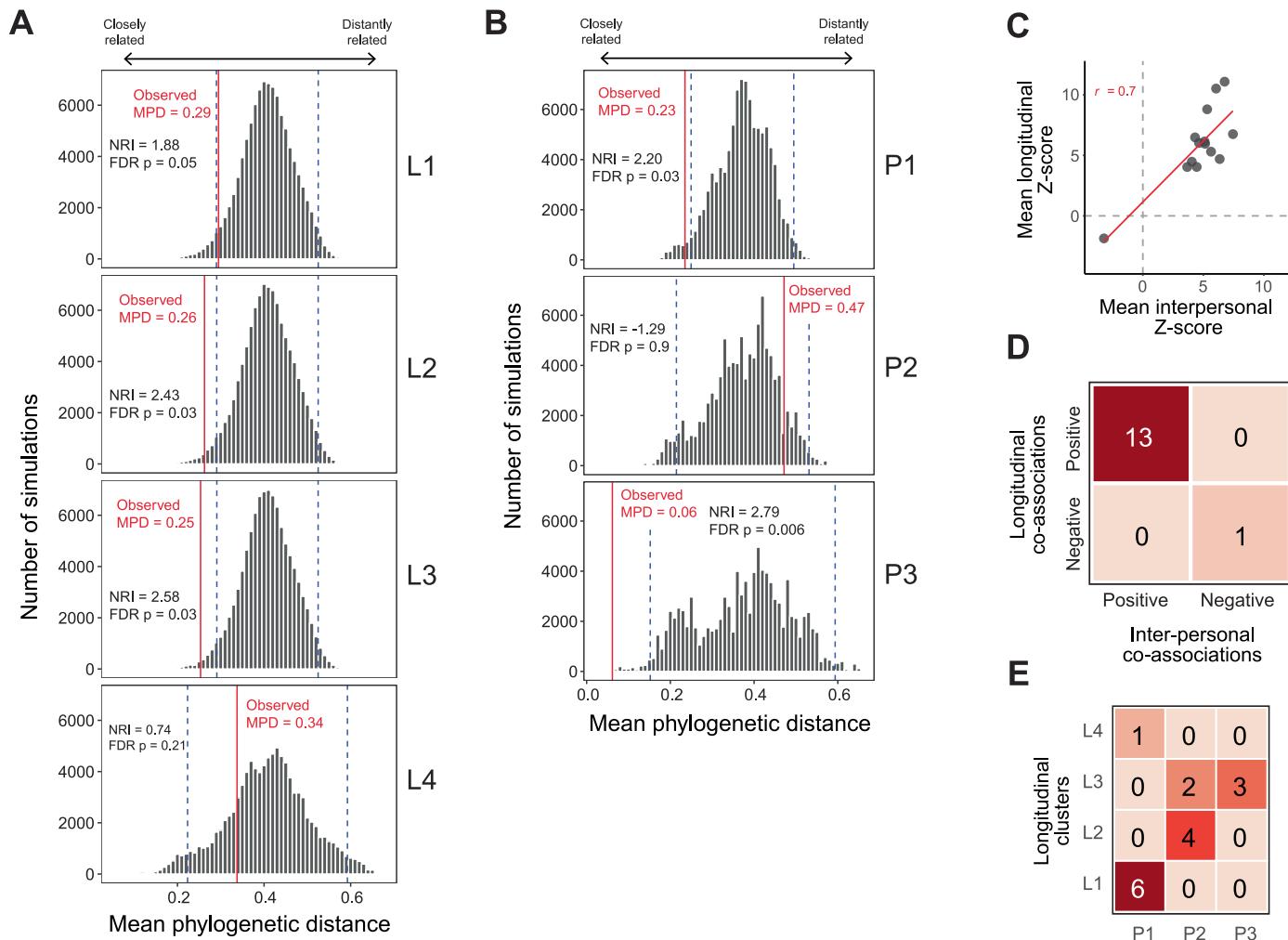
Extended Data Fig. 7 | Large scale ASV compositional patterns. (a) Heatmap of overall ASV abundance in the dataset of prevalent ASVs (>1%), clustered by Jaccard overlap. (b) Heatmap of overall ASV abundance of prevalent ASVs (>1%) across 5 humans (H1, H10, H11, H18, H19), clustered by Bray-Curtis distance.

(c) Heatmap of family-level relative abundance of human fecal samples. Families are clustered using the Jaccard overlap, and families conserved across all individuals are indicated in green.



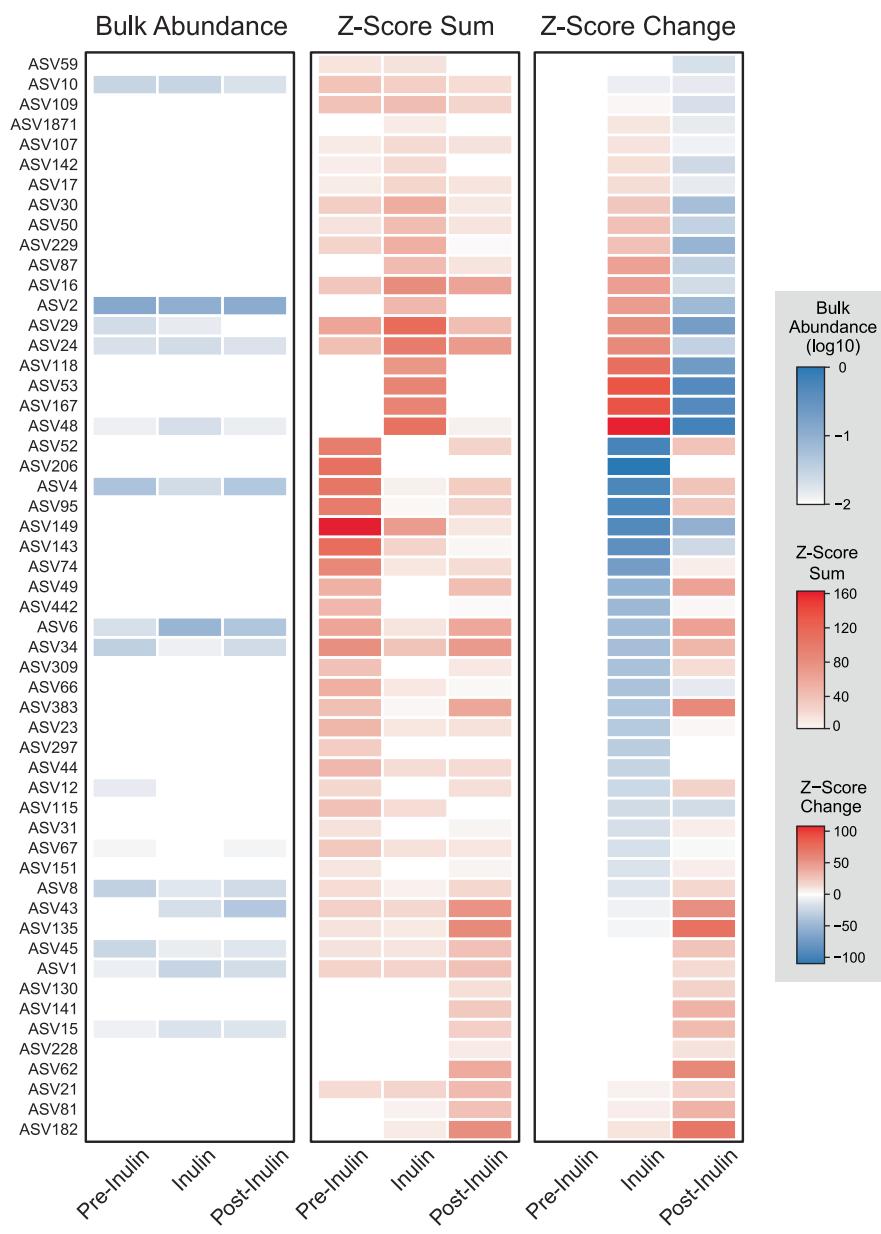
Extended Data Fig. 8 | Pairwise ASV colocalization analysis. (a) Barplot of the number of significant ASV pairs found in each sample. (b) Scatterplot between ASV-pair Z-Scores between intact and disrupted samples. (c) Scatterplots of ASV-Pair Z-scores between fresh and frozen samples. (d) Pairwise ASV spatial associations in five people. Each heatmap shows all statistically significant spatial associations between pairs of ASVs for each individual (H1, H10, H11, H18, H19). Colors in the heatmap correspond to Z-scores and stars correspond to statistical

significance ($p < 0.05$ BH FDR Corrected). ASVs are labeled in 3 possible colors (red, green, blue) if they belong to a conserved spatial cluster group (P1, P2, P3) found across 3 or more individuals. Common taxonomic families are labeled next to each ASV label on the y-axis. (e) Heatmap of significant co-associations found on 2 or more days in H11. (f) Heatmap of significant co-associations found in 3 or more people.


Extended Data Fig. 9 | Phylogenetic Distance Distributions and Relationships

Between Clusters. Histograms of simulated MPD distributions for L1-L4 (a) or P1-P3 (b) spatial hubs. The red line indicates the observed MPD in the cluster, while blue dashed lines indicate the 95% confidence interval around the mean of simulations. P-value was calculated using the two-sided Z-score. (c) Scatterplot

of Z-score for associations found in both for longitudinal and interpersonal samples with the corresponding correlation. (d) Contingency table of the sign of longitudinal versus interpersonal associations. (e) Contingency table of ASV presence across the clusters. Chi-squared test of independence ($\chi^2 = 22.4$, df = 6, $p = 0.001$).



Extended Data Fig. 10 | Inulin supplementation. Heatmaps of bulk relative abundance, Z-score sum, and change in total Z-score for ASVs that had a total Z-score change >10 over the course of inulin supplementation.

Corresponding author(s): Dr. Harris Wang

Last updated by author(s): Nov 23, 2024

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our [Editorial Policies](#) 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 No software was used for data collection.

Data analysis Scripts for read processing are implemented in BASH and R. They are available from <https://github.com/wanglabcumc/SAMPL-seq>. R packages used include EcoSimR 0.1.0, ggplot2 3.5.1, ggpublish 0.6.0, ggraph 2.1.0, data.table 1.14.10, picante 1.8.2, stringr 1.5.1, patchwork 1.2.0, ape 5.7-1, igraph 2.0.1.1, and graphlayouts 1.1.0

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Raw sequencing reads are available from PRJNA996899. Refseq 16S database (https://www.ncbi.nlm.nih.gov/refseq/targetedloci/16S_process/) was used for species-level taxonomic identification, and SILVA 138.1 (<https://www.arb-silva.de/documentation/release-1381/>) was used for taxonomic assignment of all ASVs.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Sex and gender were not considered in the study design for this manuscript, as diet and other factors are far more significant determinants of gut microbiome composition, which are the focus of this research.

Reporting on race, ethnicity, or other socially relevant groupings

Sex and gender were not considered in the study design for this manuscript, as diet and other factors are far more significant determinants of gut microbiome composition, which are the focus of this research.

Population characteristics

Members of the study cohort are young (under 40) and undiagnosed for GI disease, as it was essential to study healthy individuals for this study.

Recruitment

Individuals were recruited by word of mouth, and this does limit the breadth of this study, as the diet of the individuals studied is not a random sample. Self-selection bias may be present, as those most available to participate in this study are not necessarily representative of the larger population.

Ethics oversight

Columbia University Institutional Review Board approved this study under IRB-AAAT4813.

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

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

The sample size for particle number was determined using rarefaction and estimates drawn from Blanchet et al 2020 Ecology Letters. The number of individuals surveyed was designed to give a sampling of individual microbial spatial diversity, but power analysis was not performed.

Data exclusions

No data was excluded.

Replication

Extensive replication was done to ensure the repeatability of our method- Figure 2, and Suppl. Figures 2, 3, 6, and 8 all show biological and technical replication of SAMPL-seq using additional fecal samples. The specific fecal samples obtained cannot be completely replicated, as each is unique, but additional samples could be collected from the individuals of interest, which would bear close resemblance to those we obtained.

Randomization

Participants were not randomized into groups, as there was no experimental intervention on a multi-individual level. Differences between/within individuals was the covariate of interest, which cannot be randomized. Individual H11 was placed in the experimental group due to their availability and served as their own control. Randomization was not relevant to these experiments because the single experimental group was their own control.

Blinding

Investigators were blinded to the samples during processing, and during analysis, samples were only known by arbitrary individual names.

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
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology and archaeology
<input type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	Dual use research of concern
<input checked="" type="checkbox"/>	Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals

A single female C57BL6/J mouse (Envigo labs) was used for this study. The mouse was individually caged, and the room was kept at a 12hr day/night cycle.

Wild animals

No wild animals were used in this study.

Reporting on sex

Sex was not considered, as diet is far more important for microbiome composition.

Field-collected samples

No field-collected samples were used in this study.

Ethics oversight

The Columbia University Institutional Review Board oversaw the protocol for this study, under AABD4554.

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

Plants

Seed stocks

N/A

Novel plant genotypes

N/A

Authentication

N/A