Directional Gaussian Mixture Models of the gut microbiome elucidate microbial spatial structure

Amey P. Pasarkar[a], Tyler A. Joseph[a], and Itsik Pe'er[a,b,c#]

[a] *Department of Computer Science, Columbia University, New York NY, USA*
[b] *Department of Systems Biology, Columbia University, New York NY, USA*
[c] *Data Science Institute, Columbia University, New York NY, USA*

#Address correspondence to Itsik Pe'er, itsik@cs.columbia.edu

Abstract: The gut microbiome is spatially heterogeneous, with environmental niches contributing to the distribution and composition of microbial populations. A recently developed mapping technology, MaPS-seq, aims to characterize the spatial organization of the gut microbiome by providing data about local microbial populations. However, information about the global arrangement of these populations is lost by MaPS-seq. To address this, we propose a class of Gaussian Mixture Models (GMM) with spatial dependencies between mixture components in order to computationally recover the relative spatial arrangement of microbial communities. We demonstrate on synthetic data that our spatial models can identify global spatial dynamics, accurately cluster data, and improve parameter inference over a naive GMM. We applied our model to three MaPS-Seq datasets taken from varying regions of the mouse intestine. On cecal and distal colon datasets, we find our model accurately recapitulates known spatial behaviors of the gut microbiome, including compositional differences between mucus and lumen-associated populations. Our model also seem to capture the role of a pH gradient on microbial populations in the mouse ileum and proposes new behaviors as well.

Importance: The spatial arrangement of the microbes in the gut microbiome is a defining characteristic of its behavior. Various experimental studies have attempted to provide glimpses into the mechanisms that contribute to microbial arrangements. However, many of these descriptions are qualitative. We developed a computational method that takes microbial spatial data and learns many of the experimentally validated spatial factors. We can then use our model to propose previously unknown spatial behaviors. Our results demonstrate that the gut microbiome, while exceptionally large, has predictable spatial patterns that can be used to help us understand its role in health and disease.

Code availability: `github.com/amepas/Spatial_Mbiome`

Abstract Word Count: 177

Text Word Count: 3367

1 Introduction

A defining characteristic of the gut microbiome community is its spatial structure. Nutrients and chemical conditions differ along the gastrointestinal (GI) tract, impacting the distribution of taxa that reside there (1, 2). This spatial arrangement of microbes within the gut microbiome likely contributes to major aspects of its dynamic behavior, including community stability and host-microbe interactions (3, 4).

Recently, a novel DNA technology, Metagenomic Plot Sampling by sequencing (MaPS-seq), was developed to offer insights into the spatial organization of the gut microbiome (5). In MaPS-seq, high-resolution segments ($\sim 20\mu$m squares) are extracted directly from along the gut. Segments are encapsulated in droplets with barcoded 16S rRNA amplification primers, such that sequencing reads with the same barcode originate from the same segment. Hence, MaPS-seq preserves localized information about the spatial structure of the microbiome, and is a valuable tool for investigating the biogeography of the gut microbiome. Yet, the assignment of barcodes to droplets is a random process: MaPS-seq does not preserve the global arrangement of droplets along the gut.

Known characteristics of the biogeography of the gut microbiome suggest it may be possible to reconstruct the global arrangement of MaPS-seq droplets. For example, antimicrobial peptides, oxygen levels, and acidity vary along the length of the small intestine. Consequently, bacterial loads increase along the longitudinal axis of the small intestine and lead to a more microbe-rich ileum (2). In the colon, the density of the mucus layer increases along its longitudinal and cross-sectional axes—creating environmental niches favored by different species (1). In principle, it should be possible to reconstruct some of these global patterns from the high-resolution sampling of MaPS-seq.

1.1 Our contribution

We developed a class of computational models to recover known characteristics of the biogeography of the gut microbiome from MaPS-seq data. Our models build upon the classical Gaussian Mixture Model (GMM, Figure 1). In a GMM, observations are mixtures of latent clusters, each of which is modeled as a multidimensional Gaussian random variable, independent of the others and with its own mean. We expand this framework by introducing spatial dependence between latent clusters. Specifically, clusters are arranged as a line (one-dimensional model) or grid (two-dimensional model) to investigate directional changes along the longitudinal axis only, or, respectively, both the longitudinal and radial axes of the gut.

A key question is whether our model can differentiate longitudinal from radial changes in the gut. We demonstrate on synthetic data that our model is capable of discriminating between one-dimensional and two-dimensional models. We apply our model to MaPS-seq mouse ileum, cecal, and distal colon datasets. We provide strong evidence for the presence of spatial structures across all datasets, with distinct regional characteristics. We show that our proposed model recovers known biological behaviors of microbes within the GI tract while also providing new insights into the spatial structure of the gut microbiome.

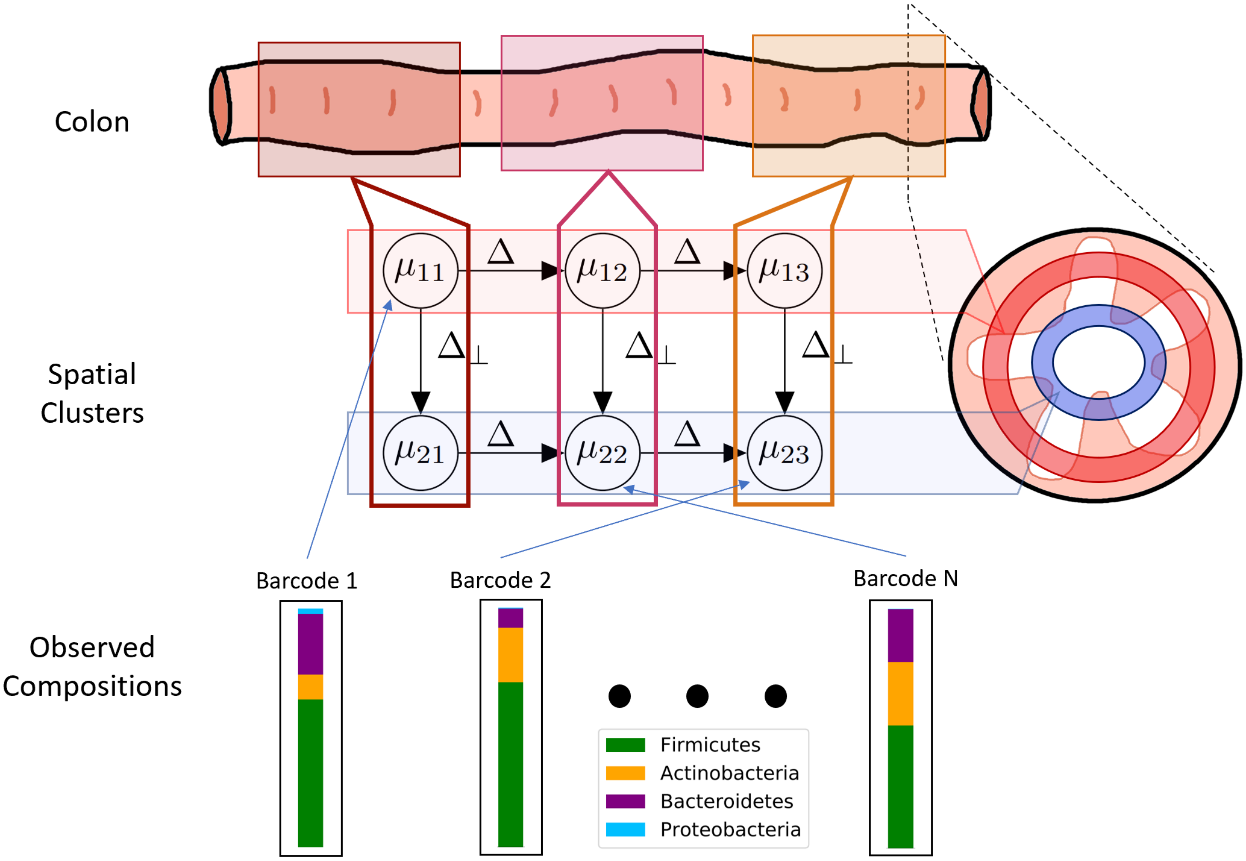

Figure 1: **Schematic Overview of directional Gaussian mixture model.** Given observed compositions from each barcode, the model simultaneously learns the community composition of each latent cluster (μ_i), and the assignment of each barcode to a latent cluster.

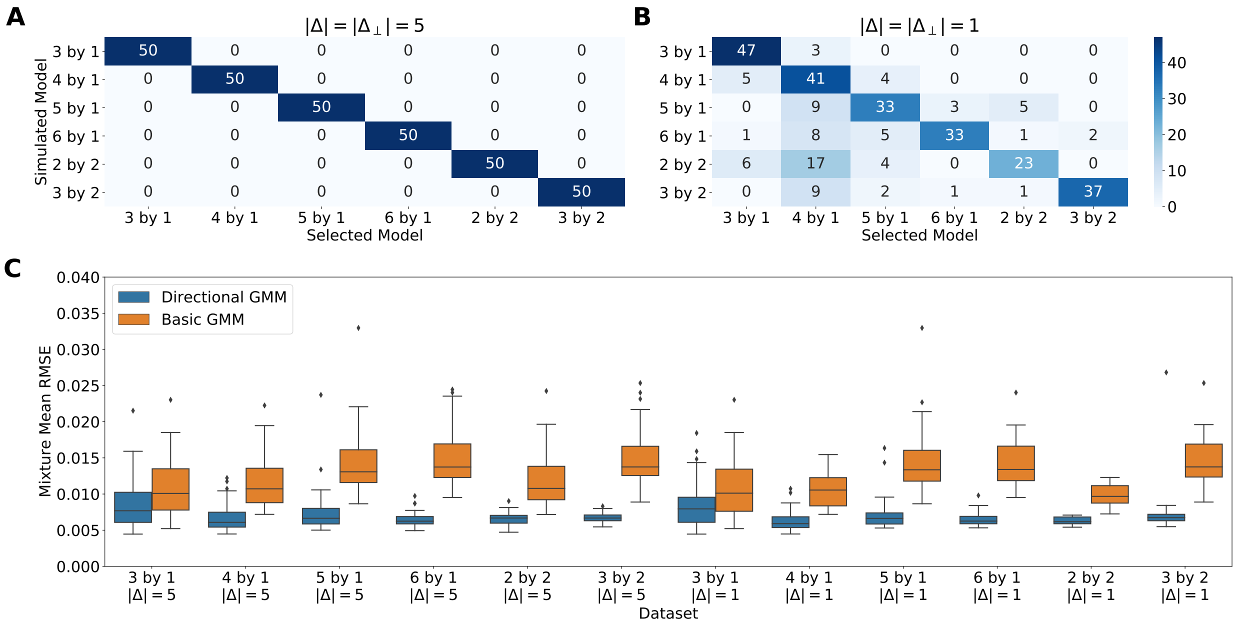

Figure 2: **Directional GMMs accurately select number of latent clusters and infer model parameters.** (A) Heatmap showing the accuracy of the selected model on various simulated datasets with $|\Delta| = |\Delta_{\perp}| = 5$ and within-cluster standard deviation of 1 (B) $|\Delta| = |\Delta_{\perp}| = 1$, keeping the same standard deviation (C) RMSE of learned cluster means on datasets with correct model selection. On all datasets, the directional GMMs are significantly improving parameter inference

2 Results

2.1 Simulation Results

We first evaluated whether our model can differentiate between one- and two-dimensional dynamics using simulated data. We simulated data under the one- and two-dimensional models (Methods 4.4), and asked if we could infer the number of latent clusters and their spatial arrangement. Using the Akaike Information Criterion (AIC), we found that our directional GMM is able to correctly determine the correct number of clusters (Figure 2A:B). Furthermore, the introduction of a dependence between latent clusters in the model also improved parameter inference compared to a naive GMM with no spatial structure (Figure 2C). For all dataset forms, the Wilcoxon signed-rank test p-value was less than 0.001.

2.2 Spatial Structure of MaPS-Seq Data

We applied our directional GMM to three real MaPS-seq datasets from Sheth et al. (5) (Methods 4.5). The provided MaPS-seq data contains samples from 3 regions of a single mouse's GI tract: the cecum (n = 405 barcodes), the ileum (n = 386 barcodes), and the distal colon (n = 259 barcodes). On the Cecum and Distal Colon datasets, the best supported models were two-dimensional (4×2 and 3×2 respectively). On the Ileum dataset,

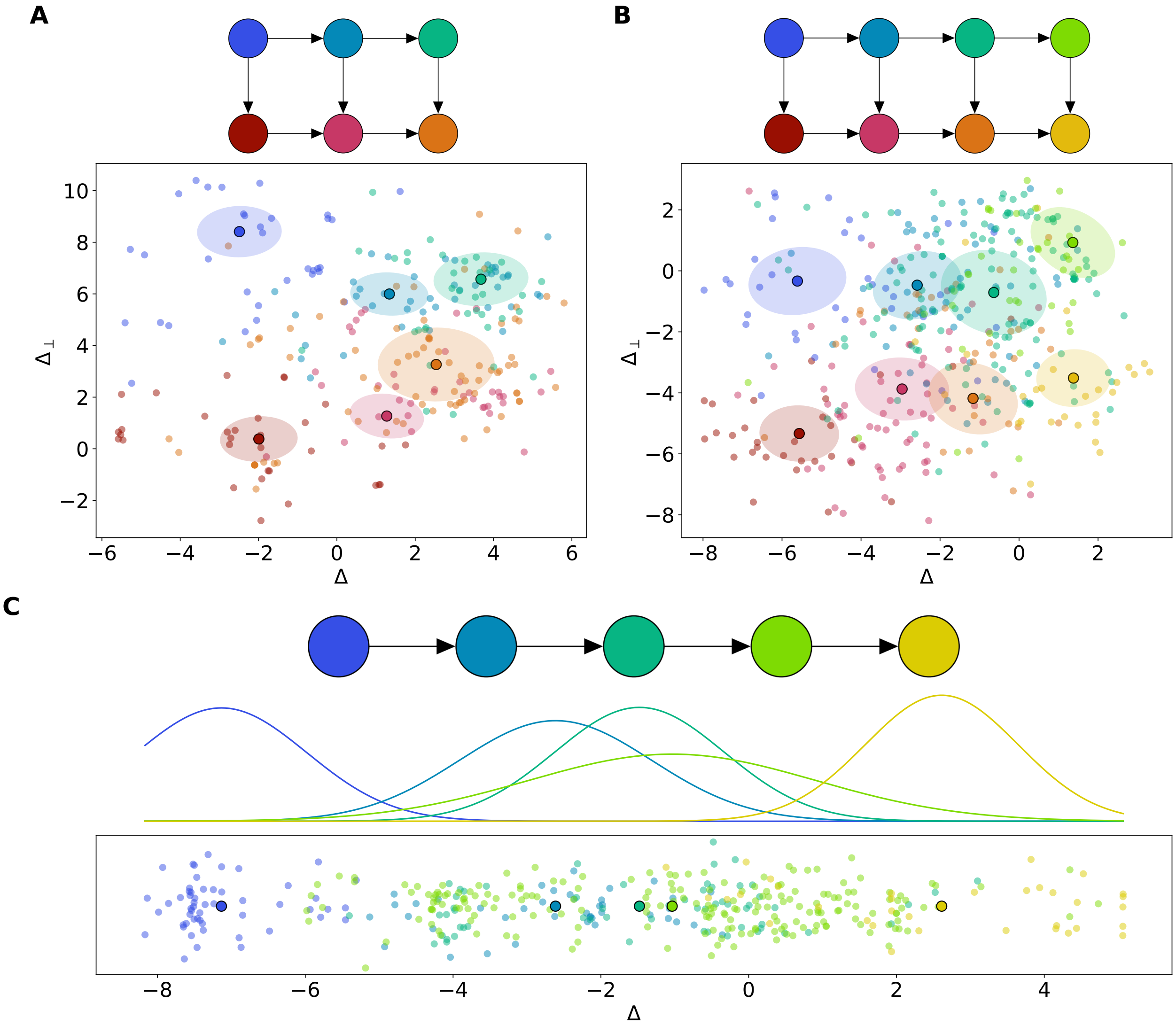

Figure 3: **Projections of MaPS-Seq Data**. (A) Distal Colon dataset. The selected model and MaPS-Seq data projected along the unit Δ and $\Delta_\perp$ axes. Colors correspond to samples belonging to a latent cluster. Ellipse radii represent the eigenvectors of the covariance matrix. (B) Cecum dataset. (C) Ileum dataset. Selected model and MaPS-Seq data projected along the unit Δ axis. Normal distribution represent density of covariance matrices around each cluster mean.

AIC Selections					
Dataset	One-Dimension		Two-Dimension		$AIC_{naive} - AIC_{structure}$
	Score	Mixture Model	Score	Mixture Model	
Ileum	**−37493**	5 × 1	-37340	3 × 2	954
Cecum	32714	8 × 1	**31930**	4 × 2	3832
Distal Colon	6636	6 × 1	**5588**	3 × 2	1168

Table 1: **AIC shows strong evidence for spatial structure across the GI tract.** (A) Best directional mixture model and its corresponding AIC score. Scores in bold indicate selected model. Comparison of directional GMMs to naive GMM by AIC metric show introduction of dependence between latent clusters significantly improves model fit. Full relative likelihoods calculated using AIC scores between models are also shown (Methods 4.3).

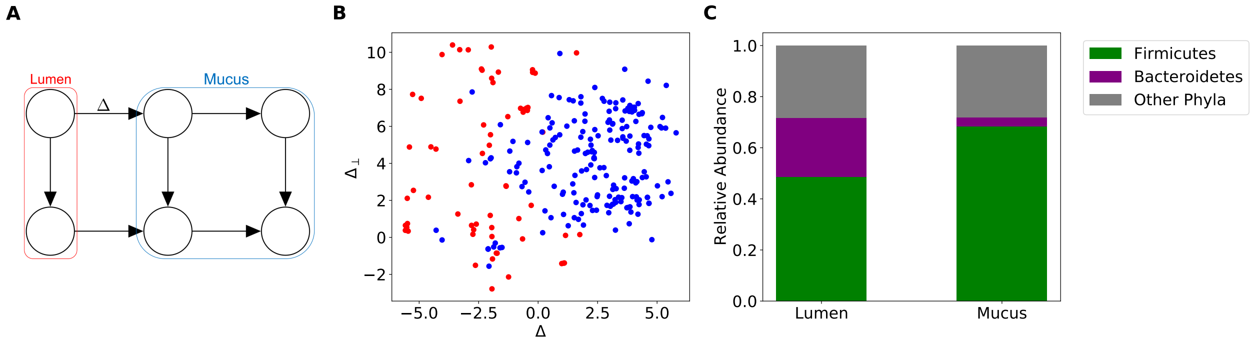

Figure 4: **Directional GMM recovers spatial dynamics in the distal colon** (A) Selected model and corresponding locations of clusters in the distal colon (B) Scatter plot of projected MaPS-seq samples assigned to lumen-associated clusters (red) and mucus-associated clusters (blue). (C) Clusters associated with the mucus are enriched in *Firmicutes* and those associated with the lumen display larger levels of *Bacteroidetes*.

the best supported model was a one-dimensional model with 5 clusters (5 × 1). Using the model parameters from the best supported model on each dataset, we created one- and two-dimensional visualizations depicting the directions learned by our model (Figure 3). Qualitatively, our model appeared to segregate barcodes into distinct clusters along the gut.

We also compared the support the selected directional model GMM to a naive GMM with no spatial structure. To compare models, we computed the AIC scores of our directional models to a naive GMM with the same number of latent clusters (Table 1). The naive GMMs have much larger AIC scores than the directional GMMs. Conventionally, models with scores that are larger by 10 or more are considered to have little support (6).

2.3 Recovery of GI Tract Biogeography

We also investigated learned model parameters for correspondence to some of the known spatial dynamics of the gut microbiome. Figure 4 illustrates the recovered dynamics on the Distal Colon dataset. Under the partition presented in Figure 4, we observe large differences in the average compositions of *Firmicutes* and *Bacteroidetes* between lumen- and mucus-

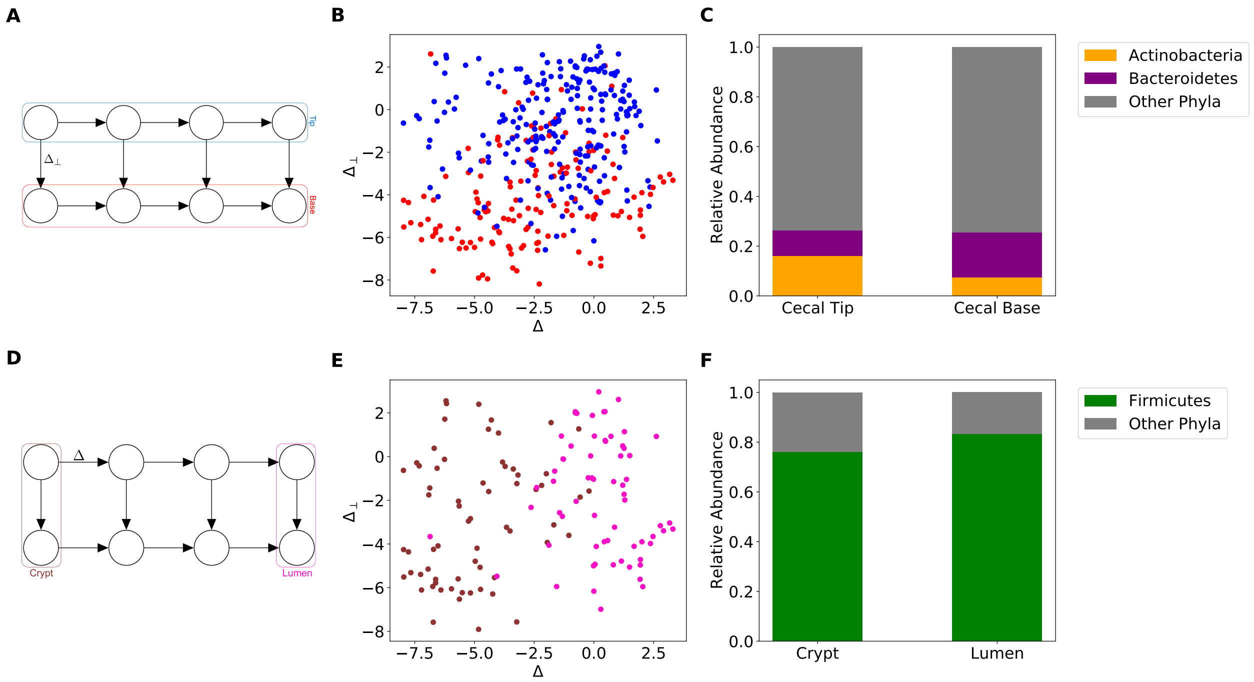

Figure 5: **Directional GMM recovers spatial dynamics in the cecum** (A) Selected model and corresponding locations of latent clusters in the cecum. (B) Scatter plot of projected MaPS-seq samples assigned to cecal tip-associated clusters (blue) and cecal base-associated clusters (red). (C) Clusters associated with the cecal tip have lower relative abundances of *Bacteroidetes* and higher relative abundances of *Actinobacteria* than the cecal base. (D) Selected model and correspoding locations of mixtures in the cecum. (E) Scatter plot of projected MaPS-seq samples assigned to cecal crypt-associated clusters (brown) and cecal lumen-associated clusters (pink). (F) *Firmicutes* are enriched in the lumen clusters compared to the crypt clusters.

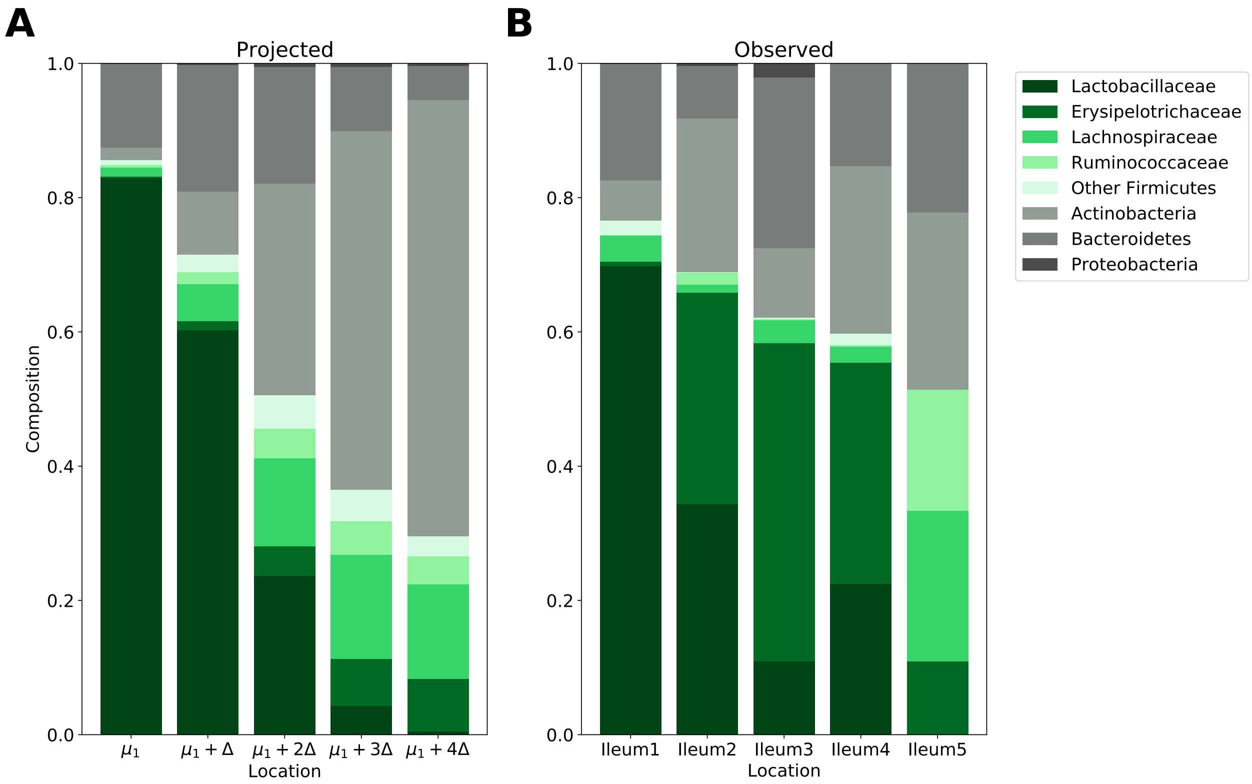

Figure 6: **Projected and observed Ileum dynamics** (A) Projected compositions moving along Δ axis (B) Observed compositions in learned model clusters. Green bars correspond to observed families in the *Firmicutes* phylum and gray correspond to other observed phyla

associated clusters.

On the Cecum dataset, we observed compositional differences along both axes. Figure 4A:C shows a cecal tip and base partition that has a noticeable compositional difference in the abundances of *Actinobacteria* and *Bacteroidetes*. The clusters on the two ends of the model have differences in the abundances of *Firmicutes* that correspond to the cecal crypt and lumen (Figure $5D : F$).

On the Ileum dataset, we compared microbial population relative abundances across each latent cluster. Along the length of the ileum, we observed a general decreasing trend in *Lactobacillaceae* and increases in both *Ruminococcaceae* and *Lachnospiraceae* (Figure 6B). Our model's choice of Δ seems to capture some of these dynamics. Moving along the Δ axis shows decreases in *Lactobacillaceae* and increases in *Lachnospiraceae* (Figure 6A). Some discrepancy is observed, most noticeably with the behavior of *Actinobacteria*.

3 Discussion

Novel experimental methods focused on the gut microbiome's spatial organization have provided new datasets for computational analysis. Here, we developed directional GMMs with dependent mixtures to infer spatial behaviors of phyla within the gut microbiome. We demonstrated the accuracy of the proposed directional GMMs on simulated data in terms of ability to infer model parameters, and to differentiate one-dimensional from two-dimensional spatial structure. On MaPS-seq data, we demonstrated the presence of spatial structure in distinct regions of the mouse GI tract. Encouragingly, our model recapitulated well known spatial phenomena on the Distal Colon and Cecum datasets.

In the distal colon, it has been shown that *Bacteroidetes* is enriched in the lumen, while *Firmicutes* are enriched in the mucus layer and crypts (2, 1). We observe these compositional differences, suggesting that our model is recovering the radial dynamics of the distal colon. The presence of four distinct clusters representing the mucus layer is not surprising because mucosal communities vary significantly over lengths as small as 1cm (7).

In the Cecum dataset, correspondence with other *in vivo* experiments suggest that we recover dynamics in both the radial and longitudinal directions (Figures 5A:F). Zaborin et al. (8) suggested that in the mouse cecum, *Bacteroidetes* increases in relative abundance from the cecal tip to base. It should be noted that in their experiment, this trend did not reach statistical significance. Our model seems to identify this compositional difference, in addition to a distinction in the relative abundances of *Actinobacteria* (Figure 5A:C). Zaborin et al. (8) did find a statistically significant difference between the levels of *Firmicutes* in the lumen compared to cecal crypts. We observe a similar difference at the two ends of our model (Figure 5D:F).

Within the ileum, we select a model with only a single direction of change. There is evidence that our choice of model is biologically accurate: unlike in the cecum and distal colon, the small intestine mucus layer is largely uninhabited due to the presence of antimicrobial peptides (9). However, along the length of the small intestine, oxygen concentrations and pH gradients vary (2). Among the learned clusters, we observe a stark decrease in the relative abundance of *Lactobacillaceae* (6). Along the flow of the digesta, the ileum becomes more alkaline. Because *Lactobacillaceae* are known to contribute to highly acidic environments, it is unsurprising that we observe this compositional differences along the length of the ileum. The pH gradient seems to be embedded in the Δ our model learns: cluster means along the Δ axis show a decrease in *Lactobacillaceae* similar to the observed compositions. The presence of discrepancies on phyla like *Actinobacteria* suggest that there potentially exist other sources of microbial dynamics in the ileum as well. To our knowledge, there are not any experimental studies that describe the microbiome's spatial dynamics within the ileum. This demonstrates the utility of our model: not only can we computationally confirm known aspects of the gut biogeography, but we can also propose new microbial spatial behaviors.

A limitation of the present approach is the resolution of the resulting clusters. Our directional GMM was able to capture global spatial patterns in the gut microbiome. Specifically, given that MaPS-seq samples are approximately 20 μm apart, the clusters from the best supported models on the Ileum, Distal Colon, and Cecum datasets correspond to approximately 1 cm regions. It would be interesting to investigate if a finer resolution change be achieved. Future work should focus on investigating this possibility of high-resolution mapping of MaPS-seq samples.

A valuable next step would be designing MaPS-seq experiments with ground truth labels denoting spatial locations. With coarse-grained labels from various adjacent segments of the GI tract, we could better confirm our model's ability to identify the microbiome's spatial structure, and also the spatial scale recovered by the model. Nonetheless, the present work provides strong evidence that global spatial patterns can be reconstructed from MaPS-seq

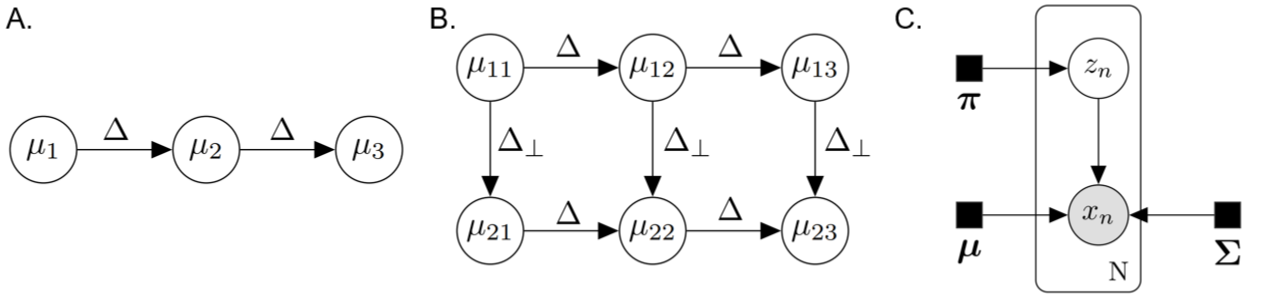

Figure 7: **A directional Gaussian mixture model** (A) Graphical depiction of relationships between latent clusters in a one-dimensional model. (B) Relationships in a two-dimensional model, where changes from left-to-right are described by Δ and perpendicular changes are described by $\Delta_\perp$. (C) GMM used to model sampling noise of observed samples x_n. μ_i represent latent clusters with relationships given in (A) and (B).

data that will only be improved with more detailed collection.

4 Methods

4.1 Directional Gaussian Mixture Models

Our approach uses a Bayesian network to describe the relationship of spatially arranged clusters in the gut (Figure 7A-B). In detail, given a spatial configuration, the goal is to simultaneously learn community states for each latent cluster and assign barcoded MaPS-seq droplets to a cluster (Figure 1). The nodes of the Bayesian network, $\{\mu_s | s \in \mathcal{S}\}$, represent composition vectors of archetypal communities in respective clusters.

In the present work, we are interested in changes along one- or two-dimensions. Studies suggest the presence of two natural directions in the gut microbiome (1). One dimension moves along the flow of the digesta, while the other moves orthogonally along the radial axis (inward out).

This motivates the following definition for our model. We define a one dimensional model where $\mathcal{S} = \{i | 1 \leq i \leq K\}$ for K latent clusters (Figure 7A). Let Δ represent directional changes between adjacent community compositions.

We can define

$$\begin{aligned} f(\mu_1) &= \mathcal{N}(\mu_1 | \bar{\mu}_0, Q_0) \\ f(\mu_i | \mu_{i-1}) &= \mathcal{N}(\mu_i | \mu_{i-1} + \Delta, Q) \quad \text{for } i = 2..K \end{aligned}$$

We also define a two-dimensional model where $\mathcal{S} = \{(i,j) | 1 \leq i \leq K, 1 \leq j \leq 2\}$ for $2K$ latent clusters (Figure 7B). Let $\Delta_\perp$ represent the direction along the second dimension, such that $\Delta \cdot \Delta_\perp = 0$. We define

$$\begin{aligned} f(\mu_{11}) &= \mathcal{N}(\mu_{11}|\bar{\mu}_0, Q_0) \\ f(\mu_{1i}|\mu_{1(i-1)}) &= \mathcal{N}(\mu_{1i}|\mu_{1(i-1)} + \Delta, Q) \quad \text{for } i = 2..K \\ f(\mu_{21}) &= \mathcal{N}(\mu_{21}|\mu_{11} + \Delta_\perp, Q) \\ f(\mu_{2i}|\mu_{1i}, \mu_{2(i-1)}) &= \mathcal{N}\left(\mu_{2i}\middle|\frac{1}{2}\left(\mu_{1i} + \mu_{2(i-1)} + \Delta + \Delta_\perp\right), Q\right) \quad \text{for } i = 2..K \end{aligned}$$

MaPS-seq outputs read counts for each of the operational taxonomic units (OTUs) in each barcoded droplet. However, the total number of reads is independent of overall community size. Therefore, the sequencing counts only provide information about the proportions of each OTU in the community. Recent work has advocated using such compositional data transformations to model microbiome data (10). We transformed read counts to relative abundances, and then applied the PhILR transformation: an isometric log-transform (ILR) with a phylogenetically derived basis (11). Each coordinate for the PhILR transformed data measures the relative proportions of two clades in a phylogeny. Phylogenetic trees were generated using QIIME (12), and provided as input to the PhILR R package. Given D taxa, the latent community states are $D-1$ dimensional vectors $\mu_s \in \mathbb{R}^{D-1}$. Zeros are handled using multiplicative replacement with $\delta = 1/D^2$ for D-taxa (13).

The reads in a particular barcoded droplet provide noisy observations from a latent cluster (Figure 7C). Thus, we can think of the data generation process as first selecting a latent community state per barcode, then generating a noisy observation from that community state. Let $\mathcal{B}$ index the set of barcodes, $x_b \in \mathbb{R}^{D-1}$ for $b \in \mathcal{B}$ be PhILR computed from the observed sequencing reads for that barcode, and $\pi_\mathcal{S} = (\pi_s)_{s\in\mathcal{S}}$ be the probability that a barcode originated from each cluster $s \in \mathcal{S}$. Let ρ_s be the set of direct ancestors of μ_s. We have

$$\begin{aligned} p(z_b) &= \text{Categorical}(z_b|\pi_\mathcal{S}) \\ p(x_b|z_b, \mu_{z_b}) &= \mathcal{N}(x_b|\mu_{z_b}\Sigma_{z_b}) = \prod_{s\in\mathcal{S}} \left[\mathcal{N}(x_b|\mu_s, \Sigma_s)\right]^{\mathbb{1}(z_b=s)} \end{aligned}$$

Altogether, the complete likelihood of the model can be written

$$\begin{aligned} p(\mu_\mathcal{S}, z_\mathcal{B}, x_\mathcal{B}) &= \prod_{s\in\mathcal{S}} f(\mu_s|\mu_{\rho_s}) \prod_{b\in\mathcal{B}} p(x_b|z_b, \mu_{z_b})p(z_b) \\ &= \prod_{s\in\mathcal{S}} f(\mu_s|\mu_{\rho_s}) \prod_{b\in\mathcal{B}} \prod_{s\in\mathcal{S}} \left[p(x_b|z_b = s, \mu_s)p(z_b = s)\right]^{\mathbb{1}(z_b=s)} \end{aligned}$$

4.2 Parameter Inference

In both models, we seek to optimize $p(\mu_\mathcal{S}, z_\mathcal{B}, x_\mathcal{B}|\theta)$ where $\theta = (\pi_\mathcal{S}, \Sigma_\mathcal{S}, \Delta_*, Q, Q_0, \mu_0)$. This optimization is performed through an Expectation-Maximization (EM) algorithm. Under this algorithm, parameters are inferred by alternating between two steps:

- **E step:** Given the current estimates of community states $\mu^t_\mathcal{S}$, model parameters θ^t, compute the posterior expectation of each cluster assignment: $\mathbb{E}[\mathbb{1}(z_b = s)|\mu^t_\mathcal{S}, \theta^t]$

- **M step:** Maximize the expected complete log-likelihood $\log p(\mu_{\mathcal{S}}, z_{\mathcal{B}}, x_{\mathcal{B}})$:

$$(\mu_{\mathcal{S}}^{t+1}, \theta^{t+1}) = \arg\max_{(\mu_{\mathcal{S}},\theta)} \sum_{s\in\mathcal{S}} \log f(\mu_s|\mu_{\rho_s}) + \sum_{b\in\mathcal{B}}\sum_{s\in\mathcal{S}} \mathbb{E}[\mathbb{1}(z_b = s)|\mu_{\mathcal{S}}^t, \theta^t]\left[\log p(x_b|z_b = s, \mu_s) + \log p(z_b = s)\right]$$

Thus we take maximum a posteriori estimates of $\mu_{\mathcal{S}}$ and maximum likelihood estimates of the remaining parameters. Model parameters are initialized using a basic GMM with independent clusters trained on the same data. On simulated data, 20 initializations are used. On real data, 200 are used. Inference terminates following 5 consecutive steps where the expected complete log-likelihood increases by $< 10^{-4}$ of the previous step.

4.3 Model Selection

The Akaike Information Criterion (AIC) was used to evaluate models:

$$\text{AIC}(k) = -2\ln(\hat{L}) + 2p_k$$

where $\hat{L}$ denotes the likelihood of the data under the fitted model and p_k is the number of parameters for the model k. We used the complete log likelihood as a surrogate for the log likelihood of the data since it is a lower bound. When comparing models with different numbers of latent clusters, we choose the model with the *minimum* AIC score (6).

In the case of models with the same number of latent clusters (i.e. 4 clusters arranged in a line vs. clusters arranged in a 2 by 2 grid), we can directly compare the complete likelihoods $p(\mu_{\mathcal{S}}, z_{\mathcal{B}}, x_{\mathcal{B}})$ of either model.

On both simulated and real data, we test up to 8 clusters for the one and two dimensional models, or until the average community state size is 50 samples, whichever comes first. For the one dimensional model, clusters were arranged in a line from a 2 × 1 model up to an 8 × 1 model. For the two dimensional model, clusters were arranged in a grid from a 2 × 2 model up to a 4 × 2 model.

4.4 Simulation Analysis

This is an unsupervised learning problem, so we first evaluated our model on simulated data. To this end, we create simulated datasets under the two proposed models. First we sample two clusters means from a Pareto distribution with $\alpha = 1$, normalize to the relative abundance space with $D = 47$ taxa, sort taxa in decreasing order, and then transform to the ILR space. The difference in the ILR space between these two means is defined to be Δ. The Δ parameter is then scaled to our desired magnitude. Our method for sampling Δ allows for larger dynamics to be observed on more abundant taxa. For two-dimensional models, we sample $\Delta_\perp$ from a standard multivariate normal distribution and then orthogonalize relative to Δ. The remaining clusters means are arranged around one of the two original cluster means as per the two models (arranged in the ILR space in a line or in a grid).

Then, we randomly sample cluster covariance matrices Σ from an Inverse-Wishart distribution with $\nu = D + 1$ and $\Psi = \frac{I_{D-1}}{D-1}$. Finally, a total of 360 artifical MaPS-seq samples are drawn evenly and independently from each cluster.

We analyzed two aspects of model performance on simulated data: 1) selection of the correct number of latent clusters, and 2) parameter estimation accuracy. In order to evaluate our model selection framework, we train both the one and two directional models with varying amounts of latent clusters on simulated data. We used the aforementioned model selection criteria to determine the optimal model.

Next, the accuracy of our parameter inference is determined by calculating the average RMSE of the learned cluster means. Although our proposed model assigns labels to clusters to reflect their spatial arrangements, other unsupervised clustering algorithms assign arbitrary labels. Therefore, to compare RMSE of model parameters, we look at our proposed model's RMSE and the best RMSE of all label permutations of a naive GMM.

4.5 MaPS-seq data analysis

We used the publicly available data from Sheth et al. The Cecum, Ileum, and Distal Colon datasets were each extracted from 3cm segments of their respective regions. MaPS-seq clusters are the same size in all datasets (20 μm). Each sample is a vector of the relative abundances of all OTUs. We focused on the most abundant taxa that constitute 95% of all relative abundance across the three datasets. This corresponded to 47 taxa. Relative abundances were then renormalized. Using the provided fasta files, we generated phylogenetic trees in QIIME (12). Data is then transformed using the PhILR R package.

Acknowledgements

We thank Harris Wang and Miles Richardson for sharing the data and useful conversations. Research was supported by a seed grant from the Data Science Institute at Columbia, and a Driving Biological Project award from the CaST institute U54CA209997

References

1. Tropini C, Earle KA, Huang KC, Sonnenburg JL. 2017. The Gut Microbiome: Connecting Spatial Organization to Function. eng. Cell host & microbe 21: S1931-3128(17)30120-8[PII],433–442. DOI: 10.1016/j.chom.2017.03.010.

2. Donaldson GP, Lee SM, Mazmanian SK. 2016. Gut biogeography of the bacterial microbiota. Nature Reviews Microbiology 14:20–32. DOI: 10.1038/nrmicro3552.

3. Coyte KZ, Schluter J, Foster KR. 2015. The ecology of the microbiome: Networks, competition, and stability. Science 350:663. DOI: 10.1126/science.aad2602.

4. Nagara Y, Takada T, Nagata Y, Kado S, Kushiro A. 2017. Microscale spatial analysis provides evidence for adhesive monopolization of dietary nutrients by specific intestinal bacteria. PloS one 12:e0175497. DOI: 10.1371/journal.pone.0175497.

5. Sheth RU, Li M, Jiang W, Sims PA, Leong KW, Wang HH. 2019. Spatial metagenomic characterization of microbial biogeography in the gut. Nature Biotechnology 37:877–883. DOI: 10.1038/s41587-019-0183-2.

6. Burnham KP, Anderson DR. 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociological methods & research 33:261–304.

7. Riva A, Kuzyk O, Forsberg E, Siuzdak G, Pfann C, Herbold C, Daims H, Loy A, Warth B, Berry D. 2019. A fiber-deprived diet disturbs the fine-scale spatial architecture of the murine colon microbiome. Nature communications 10:1–11.

8. Zaborin A, Bernabe BP, Keskey R, Sangwan N, Hyoju S, Gottel N, Gilbert JA, Zaborina O, Alverdy JC. 2020. Spatial Compartmentalization of the Microbiome between the Lumen and Crypts Is Lost in the Murine Cecum following the Process of Surgery, Including Overnight Fasting and Exposure to Antibiotics. Msystems 5: DOI: 10.1128/mSystems.00377-20.

9. Vaishnava S, Yamamoto M, Severson KM, Ruhn KA, Yu X, Koren O, Ley R, Wakeland EK, Hooper LV. 2011. The antibacterial lectin RegIIIγ promotes the spatial segregation of microbiota and host in the intestine. Science 334:255–258.

10. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. 2017. Microbiome datasets are compositional: and this is not optional. Frontiers in microbiology 8:2224.

11. Silverman JD, Washburne AD, Mukherjee S, David LA. 2017. A phylogenetic transform enhances analysis of compositional microbiota data. Elife 6:e21887. DOI: 10.7554/eLife.21887.

12. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature biotechnology 37:852–857.

13. Martín-Fernández JA, Barceló-Vidal C, Pawlowsky-Glahn V. 2003. Dealing with zeros and missing values in compositional data sets using nonparametric imputation. Mathematical Geology 35:253–278.