Longitudinal analysis reveals transition barriers between dominant ecological states in the gut microbiome.

Roie Levy  
*Institute for Systems Biology, Seattle, WA 98109.*

Andrew T Magis  
*Institute for Systems Biology, Seattle, WA 98109.*

John C Earls  
*Institute for Systems Biology, Seattle, WA 98109.*

Ohad Manor  
Tomasz Wilmanski  
*Institute for Systems Biology, Seattle, WA 98109.*

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.psjhealth.org/publications](https://digitalcommons.psjhealth.org/publications)

Part of the Gastroenterology Commons

**Recommended Citation**
Levy, Roie; Magis, Andrew T; Earls, John C; Manor, Ohad; Wilmanski, Tomasz; Lovejoy, Jennifer; Gibbons, Sean M; Omenn, Gilbert S; Hood, Leroy; and Price, Nathan D, "Longitudinal analysis reveals transition barriers between dominant ecological states in the gut microbiome." (2020). *Articles, Abstracts, and Reports*. 3281.
[https://digitalcommons.psjhealth.org/publications/3281](https://digitalcommons.psjhealth.org/publications/3281)

This Article is brought to you for free and open access by Providence St. Joseph Health Digital Commons. It has been accepted for inclusion in Articles, Abstracts, and Reports by an authorized administrator of Providence St. Joseph Health Digital Commons. For more information, please contact digitalcommons@providence.org.
Authors
Roie Levy, Andrew T Magis, John C Earls, Ohad Manor, Tomasz Wilmanski, Jennifer Lovejoy, Sean M Gibbons, Gilbert S Omenn, Leroy Hood, and Nathan D Price
Longitudinal analysis reveals transition barriers between dominant ecological states in the gut microbiome

Roie Levy,a,1 Andrew T. Magisa, John C. Earlsa, Ohad Manorb, Jennifer Lovejoyb, Sean M. Gibbonsa,c, Gilbert S. Ommend, Leroy Hoodd,2 and Nathan D. Pricea,2

aInstitute for Systems Biology, Seattle, WA 98109; bArivale, Seattle, WA 98104; eScience Institute, University of Washington, Seattle, WA 98195; and dDepartment of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109

Contributed by Leroy Hood, March 25, 2020 (sent for review February 13, 2020; reviewed by Eugene B. Chang and Paul Wilmes)

The Pioneer 100 Wellness Project involved quantitatively profiling 108 participants’ molecular physiology over time, including genomes, gut microbiomes, blood metabolomes, blood proteomes, clinical chemistries, and data from wearable devices. Here, we present a longitudinal analysis focused specifically around the Pioneer 100 gut microbiomes. We distinguished a subpopulation of individuals with reduced gut diversity, elevated relative abundance of the genus Prevotella, and reduced levels of the genus Bacteroides. We found that the relative abundances of Bacteroides and Prevotella were significantly correlated with certain serum metabolites, including omega-6 fatty acids. Primary dimensions in distance-based redundancy analysis of clinical chemistries explained 18.5% of the variance in bacterial community composition, and revealed a Bacteroides/Prevotella dichotomy aligned with inflammation and dietary markers. Finally, longitudinal analysis of gut microbiome dynamics within individuals showed that direct transitions between Bacteroides-dominated and Prevotella-dominated communities were rare, suggesting the presence of a barrier between these states. One implication is that interventions seeking to transition between Bacteroides- and Prevotella-dominated communities will need to identify permissible paths through ecological state-space that circumvent this apparent barrier.

Significance

Deep molecular phenotyping of individuals provides the opportunity for biological insight into host physiology. As the human microbiome is increasingly being recognized as an important determinant of host health, understanding the host-microbiome relationship in a multiomics context may pave the way forward for targeted interventions. In this study, we analyze gut microbial composition of 101 individuals over the course of a year, alongside clinical markers and serum metabolomics. We establish association between specific gut compositional states and host health biomarkers (e.g., of inflammation). Finally, we provide evidence for an apparent transition barrier between these compositional states. A deeper understanding of microbiome dynamics and the associated variation in host phenotypes furthers our ability to engineer effective interventions that optimize wellness.
dominated by Bacteroides or Prevotella, despite the demonstrated association of these states with long-term diet (16, 21, 22). One possible explanation is that exclusionary interactions between these taxa or interactions with the host immune system establish a hysteresis; i.e., the behavior of the system depends not only on its input but also on its current and preceding states.

Here, we report a longitudinal analysis of the Pioneer 100 microbiome data and its relationship with metabolic and clinical chemistries profiles. We identify within our cohort a population distinguished by different levels of bacterial community diversity (i.e., α-diversity, the number of taxa and/or the evenness in their abundances within a sample) and by the dominance of either Bacteroides or Prevotella genera. The abundances of these taxa correlate strongly with serum metabolites, including medium- and long-chain fatty acids. Distance-based redundancy analysis (dbRDA) identified associations between the Bacteroides/Prevotella ratio and clinical chemistries including inflammation markers and cholesterol levels. Finally, longitudinal analysis of microbiome compositional trajectories indicates that while the microbiota may occasionally transition between Bacteroides- and Prevotella-dominated states, direct transitions are rare. We postulate that antagonistic interactions between these taxa and/or interactions with the host immune system forms an impermeable region in microbiome space-state, which tends to be circumnavigated rather than traversed during transitions between these two alternative stable states (23).

Results
Nonmetric Multidimensional Scaling Identifies Key Taxa Involved in Compositional Shifts of the Intestinal Microbiome. The Pioneer 100 pilot study comprised the broad molecular phenotyping of 108 individuals over three quarterly time points (referred to as rounds). This manuscript focuses on the characterization and dynamics of the stool microbiome of 101 participants who provided stool samples, as well as its association with serum metabolite and clinical chemistry profiles. Cohort characteristics are provided in Table 1. To begin characterizing the community composition of the Pioneer 100 intestinal microbiome, we applied nonmetric multidimensional scaling (NMDS) to β-diversity (i.e., differences in community composition between samples) as measured by weighted UniFrac dissimilarity (Methods and Fig. 1). α-Diversity was negatively correlated with NMDS dimension 1 (ρ = −0.66, P < 2.20 × 10−16), as was the major intestinal phylum Firmicutes (ρ = −0.74, q < 2.20 × 10−16). Conversely, Bacteroidetes, the other major phylum, was positively correlated with this dimension (ρ = 0.87, q < 2.20 × 10−16). In contrast, this structure was not observed by NMDS of Bray–Curtis dissimilarity (BCD), which does not take into account

Table 1. Cohort demographics

<table>
<thead>
<tr>
<th>P100 cohort (n = 101)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean (SD)</td>
<td>54.6 (13.6)</td>
</tr>
<tr>
<td>Sex, % female</td>
<td>41.6</td>
</tr>
<tr>
<td>Nonwhite, %</td>
<td>11.9</td>
</tr>
<tr>
<td>BMI, median [IQR]</td>
<td>24.6 [22.3–27.9]</td>
</tr>
<tr>
<td>Obese (BMI ≥ 30), %</td>
<td>12.9</td>
</tr>
<tr>
<td>Participants with data &gt;1 round, %</td>
<td>87.1</td>
</tr>
<tr>
<td>Participants with data for all 3 rounds, %</td>
<td>71.3</td>
</tr>
<tr>
<td>HDL, mg/dL, mean (SD)</td>
<td>61.1 (16.6)</td>
</tr>
<tr>
<td>% Glycated hemoglobin A1c, median [IQR]</td>
<td>5.6 [5.5–5.8]</td>
</tr>
<tr>
<td>Triglycerides, mg/dL, mean (SD)</td>
<td>96.7 (44.2)</td>
</tr>
<tr>
<td>C-reactive protein, mg/mL, median [IQR]</td>
<td>0.9 [0.4–1.9]</td>
</tr>
<tr>
<td>TNFα, pg/mL, median [IQR]</td>
<td>4.0 [2.9–5.1]</td>
</tr>
</tbody>
</table>

Abbreviations: BMI, body mass index; HDL, high-density lipoprotein; IQR, interquartile range; TNFα, tumor necrosis factor α.
separated those high in Bacteroides from those high in Prevotella (SI Appendix, Fig. S3).

The loadings of clinical chemistries along the first two dimensions are provided in Dataset S2. Along dimension 2, the chemistry most aligned with Prevotella was tumor necrosis factor \(\alpha\) (TNF\(\alpha\)), a marker of systemic inflammation. Conversely, three of the five chemistries most aligned with Bacteroides were chloride, sodium, and saturated fat, reiterating the association between this genus and the high-fat, high-sodium “westernized” diet (21). A number of other associations are discussed below.

Because there are many explanatory variables in the chemistries data, we additionally repeated this analysis using stepwise feature selection (SI Appendix). Furthermore, because the number of metabolites profiled exceeded the number of samples \((n < m)\), full metabolomes did not constrain ordination; the multiple regression problem is overdetermined by having more explanatory variables than observations to fit. Analysis of loadings along dimension 2 confirmed a number of correlations reported above (Dataset S3). Specifically, intermediates of phenylalanine metabolism such as phenylacetate aligned with Prevotella (opposite Bacteroides), and thyroxine with Bacteroides (opposite Prevotella). In addition, a number of tocopherols (class of vitamin E compounds) aligned positively with Prevotella. We previously reported these compounds forming a coherent module of covariance with plasma lipids and low-density lipoprotein (LDL) cholesterol (5), effectively adjoining this taxon to this module despite weaker pairwise correlation scores.

**Microbiome Trajectories Reveal Barriers to Transition.** Using unsupervised learning to cluster microbiome samples in high dimensions led researchers to suggest that the intestinal microbiome occupies only a small set of discrete states (termed enterotypes), and that Bacteroides and Prevotella strongly influence this clustering (16). In contrast, direct analysis of the abundances of only these genera suggested that they vary in a relatively continuous manner, contradicting the claim that microbiome composition varies discretely (27). Irrespective of whether these states are discrete or continuous in nature, subsequent experiments associated long-term dietary patterns with Bacteroides-vs. Prevotella-dominated states (21). Intriguingly, despite this association with long-term diets, short-term dietary interventions have not been successful in mediating transitions between these two states (21, 22).

Given the established recalcitrance of the Bacteroides-to-Prevotella ratio to short-term dietary intervention, we leveraged the...
longitudinal nature of the Pioneer 100 dataset to investigate potential Bacteroides–Prevotella transitions. Not all regions of state-space were equally occupied (Fig. 2). Most samples fell close to the boundary spanning 0% Bacteroides abundance, representative of this taxon’s relative rarity in the intestine. Nonetheless, in rare cases, up to ~90% of the relative abundance of a sample was composed of Prevotella spp. Finally, while a continuous distribution of points was observed from Bacteroides to “Other” (i.e., 1 = [Bacteroides + Prevotella]; see Methods) and from “Other” to Prevotella, the space representing codominance of these genera was essentially unoccupied.

To quantify this phenomenon, we compared linear and polynomial regression models of Bacteroides/Prevotella relative abundances (Methods). We found that a second-order model of Bacteroides and Prevotella abundance (which allows for curvature about this “empty” region) explained 75% of the variance, compared to ~10% in the linear model. We overlaid on this space the trajectories of each individual’s microbiome over time (Fig. 3). We observed that individual trajectories followed this curvature: While indirect transitions between Prevotella- to Bacteroides-dominated regions were observed, direct transitions between these spaces were all but absent. Finally, to quantify this tendency, we calculated the local “permissivity” of all regions in this state-space. Regions with high permissivity more easily allow the microbiome to transition directly through them. The region of state-space dominated by Bacteroides, and that dominated by “Other” both had high permissivity, indicating both their frequent occupancy and the ease with which the microbiome can transition between these states. In contrast, the high-Prevotella region revealed less permissivity. Most critically, permissivity analysis identified a particularly low-permissivity region between high Bacteroides and Prevotella regions, representing an apparent barrier to direct transitions between these genera (Fig. 3 and SI Appendix). In contrast to the results described above, we did not observe a low-permissivity barrier between Bacteroidites and Firmicutes (SI Appendix, Fig. S4).

Discussion

There is a growing interest in determining the role the microbiome plays in defining human health. Although the choice of terminology varies by source, the microbiome is now typically described as a crucial constituent of the human body, rather than accessory to it (28–30). Accordingly, efforts have shifted from simply identifying specific pathogens toward community-ecological approaches (31–33), which associate positive and negative health states with variation in the composition or functional structure of a commensal community (34, 35), or with specific health-related interactions between particular taxa or genes (36–38). Taking such an approach, we identified the genera Bacteroides and Prevotella as key determinants of community composition and diversity for our studied population. Relative abundance of these taxa correlated with fatty-acid metabolic intermediates, and formed an ecological gradient associated with inflammation and cholesterol markers. Finally, longitudinal analysis revealed a barrier to direct transition between Bacteroides- and Prevotella-dominated compositional states.

We identified subclasses of the Pioneer 100 cohort distinguished by community diversity levels, and subsequently by the relative abundance of the genera Bacteroides and Prevotella. While this cohort did not represent a case-control study, we associated levels of physiologically relevant metabolites and clinical chemistries with relative abundance of these key genera. Specifically, samples high in either Bacteroides or Prevotella were also high in LDL cholesterol, potentially underscoring the influence of cholesterol on the microbiome, or possibly the influence of cholesterol-lowering medications on the microbiome. Furthermore, samples high in Prevotella relative to Bacteroides were elevated in TNFα, adiponectin, and HDL cholesterol and reduced in saturated fats and C-reactive protein (CRP). TNFα and CRP are both inflammation markers, but they aligned opposite to one another along this dimension. Previous investigations demonstrated that TNFα but not CRP levels correlate with severity of trauma (39) and chronic kidney disease (40), and are a predictor of morbidity due to sepsis (41), potentially indicating Prevotella taxa associate with different inflammatory states.

Given that the abundances of these taxa correlated with wellness markers, we investigated the tendency of individuals to transition between high-Bacteroides and high-Prevotella states. We observed transitions between these states (in either direction), but with a tendency to first pass through a population bottleneck in which both are relatively depleted. This is of particular note given the discussion surrounding these genera. Bacteroides and Prevotella, despite being phylogenetically related, exhibit marked
exclusionary occurrence across intestinal habitats (42). They are at the center of the enterotype model of microbiome community assembly, which posits that communities occupy discrete regions of compositional state-space (16). Conversely, arguments against this model attest these genera themselves do not vary in abundance in a discrete manner (27). Our results demonstrate a potential reconciliation between these two arguments: While microbiota composition generally varies in a continuous manner, exclusionary interactions maintain quasi-discrete states dominated by either Bacteroides or Prevotella.

More broadly, the stable high-Bacteroides or high-Prevotella states may be thought of as attractors, or basins in an energy landscape representing microbiota composition (31, 32). Once the system has settled into a basin, microbe–microbe and microbe–host interactions can prevent transition into the alternate state unless they first traverse other transitional states. These ecological basins could be responsible for long-term robustness observed in the microbiome (43–45). Our analysis suggests that the Bacteroides- and Prevotella-dominated states can only be traversed through a phylum-level Bacteroidetes bottleneck, where either genus must be depleted for the other to invade and establish itself. This is compatible with the observation that short-term dietary interventions were insufficient to initiate transitions between enterotypes (18, 20–22). A potentially successful strategy might involve a two-stage approach to first diminish Bacteroidetes (e.g., via targeted anti-

### Methods

**Overview of the Pioneer 100 Study.** All sample collection and quantification was performed as part of the Pioneer 100 Wellness project at the Institute for Systems Biology, and approved by the Western Institutional Review Board (IRB Protocol Number 20121979) (5). All participants recruited for this study were insufficient to initiate transitions between enterotypes (18, 20–22). A potentially successful strategy might involve a two-stage approach to first diminish Bacteroidetes (e.g., via targeted anti-

### Microbiome Data Collection and Processing

Stool sample preparation and 16S rRNA (V4) sequencing were performed by Second Genome. Once per round, participants collected personal stool samples at home, using standard Second Genome collection kits. The 250-bp paired-end MiSeq profiling of the 16S v4 region was performed; ~200,000 ± 58,500 reads (median ± median absolute deviation) were generated per sample. Forward reads were trimmed to 150 bp, and any reads not reaching this length were discarded; reverse reads were not utilized in this analysis. Open reference OTU picking (49) was performed against the Greengenes database (50) (version 13.8) using Qiime (51) (version 1.9.1). Rare OTUs, defined here as those not representing 0.01% of at least one sample, were removed. Remaining OTU counts were unit normalized. α-Diversity, a measure of the number of OTUs observed within an individual sample as well as the evenness of their distributions, was quantified by the effective number of taxa (52) from Shannon's index (53, 54). β-Diversity, a measure of the diversity distinguishing two or more samples, was quantified by the Bray–Curtis (54, 55) and the weighted UniFrac dissimilarities (56, 57).

### Molecular Profiles of Wellness Markers

Two separate molecular profiles were analyzed: clinical chemistries and serum metabolomes. As described in the text, these profiles were chosen for their clinical relevance and interpretability, and because like the microbiome (and in contrast to the gene), these profiles vary in time and in response to intervention. Features with more than 100 missing values were discarded: 3-deoxyglucosone hydroimidazolones, aminoacidic acid, bun/creatinine ratio, (carboxyethyl) lysine, carboxymethyl-lysine, glycoxyl-derived hydroimidazolone G-H1, homocysteine, and methionine-sulfoxide. eGFR (non-African American) was discarded as it was redundant (Pearson's r > 0.99 with eGFR [African American]). After filtering, 203 clinical chemistries and 257 metabolites were included in subsequent analyses. Features were independently standard normalized. Remaining missing values were imputed using a nonparametric random forest approach (58). Because standard normalization produces negative values and ecological (dis)similarities are interpretable in the positive domain, the Euclidean distance was used to measure similarity between molecular profiles. For any association of microbiome to molecular profiles, only samples with matching microbiome, metabolome, and clinical chemistries were analyzed.

**Ordination of β-Diversity.** Initial ordination was performed using NMDS (59). In contrast to metric dimensional scaling (principal coordinate analysis), NMDS attempts to embed observations in a space of arbitrary dimensionality such that pairwise dissimilarities in this reduced space are monotonically related to original dissimilarities and is more robust to curvilinear distortion (60). Analysis of the stress-dimension plot revealed an elbow at dimension 3 with a stress value of ~0.010 (SI Appendix, Fig. S5).

**Defining and Characterizing Microbiome Subclasses.** Ordination of BCD separated high-Bacteroides samples along a single dimension (SI Appendix, Fig. S1). Visual simplicity in preliminary analysis, we used a random forest approach (56, 57) to define which set of samples belonged to which class (rather than select along two dimensions via ordination of UniFrac dissimilarity). Specifically, we selected samples above 1.0 (n = 25) on the absissa as the “LO” samples. To compare balanced classes, we took an equal number of samples from the opposite end (25 samples less than ~0.60 along NMDS dimension 1). Difference in α-diversity across subclasses was tested by the Wilcoxon rank sum test.

We sought to identify taxa that were not only differentially abundant across sample classes but were categorically representative of those classes. To that end, we employed the two-sided Wilcoxon rank sum test with Benjamini–Yekutieli multiple hypothesis correction (61) (FDR < 0.05), and further selected only those taxa with Cohen’s d of magnitude greater than or equal to 4.0. Whereas the P value (and by association, the FDR) represents the confidence that two samples come from different distributions, Cohen’s d is a measure of effect size; it is in magnitude comparable to effect sizes in many other fields and more directly assesses the magnitude change of relative abundance (62). d values greater than 1.0 typically signify extremely strong effects; our threshold was chosen ad hoc to identify differential dominant taxa. Furthermore, to investigate whether subclasses as defined indeed represent distinct breakpoints of dominant taxa, we plotted relative abundance across NMDS dimension 1 (SI Appendix, Fig. S2). While Bacteroides abundance trended downward over the entire span, Prevotella appeared to elbow at ~0.5. Therefore, we infer that the specific choice of cutoff is not absolutely critical to associate these specific taxa with this dimension.

### Multivariate Analysis of Microbiota and Molecular Profiles

We used the nonparametric Spearman correlation coefficient with Benjamini–Hochberg multiple hypothesis correction (63) (FDR < 0.05) to determine which analytes correlated with Bacteroides and which with Prevotella. We further employed dbRDA to associate β-diversity with molecular profiles (26, 54). We used the weighted UniFrac dissimilarity with a minor additive constant to adjust negative eigenvalues (64). Because dbRDA does not perform feature selection, in the main text we focus on the features with the most extreme loadings along the second dimension; the full tables are provided in Data sets S2 and S3. Subsequently, we performed stepwise feature selection according to the Akaike information criterion (AIC) (54) (SI Appendix).

Specifically, bidirectional elimination was implemented using function ordistep in the vegan package with default parameters; at each step, each
feature's AIC is tested by permutation; those with $P < 0.05$ are added to the model and with $P > 0.1$ are removed; model selection terminates when no features can be added or removed or (as in this case) after 50 steps.

**Exclusionary Analysis of Taxa.** We used regression to quantify the degree to which a linear relationship could or could not describe the relationship between pairs of taxon abundances $i$ and $j$. We first transformed the relative abundances of taxa into their respective difference:

$$\Delta \text{tax} = A_i - A_j,$$

and their sum subtracted from 1:

$$\sigma_{\text{tax}} = 1 - (A_i + A_j).$$

This transformation accounts for an antisymmetry in linear regression (e.g., this regression of *Bacteroides* on *Prevotella* does not equal the regression of *Prevotella* on *Bacteroides*).

After such a transformation, $\Delta \text{tax}$ is weighted equally by both taxa, while $\sigma_{\text{tax}}$ and any residuals are weighted by their sum; subtracting from 1 allows the plot of $\sigma_{\text{tax}}$ versus $\Delta \text{tax}$ to correspond with typical ternary plots. We used ordinary least-squares regression to fit a straight line ($f_1$) and a second-order polynomial ($f_2$) to these plots. We calculated the percent of variance explained by the second-order model relative to the first-order from the relative coefficient of determination:

$$R^2_{\text{adj}} = 1 - \frac{\sum (\Delta \text{tax} - f_2)^2}{\sum (\Delta \text{tax} - f_1)^2}.$$

In analogy to the standard interpretation of $R^2$, this corresponds to the amount of additional variance accounted for by the inclusion of a parabolic term, as opposed to both a constant offset as well as a linear slope.

**Calculation of Permissivity.** We used the trajectories of individuals' microbiota to calculate the relative tendency of regions of space-state to permit transit. We evaluated the property permissivity, in alignment with related concepts delineating the microbiota's ability to permit or resist variation (44). We define the permissivity of a point in state-space $(\Delta, \sigma)$ as follows:

$$P = \sum \frac{V_{\text{tax}}}{V_{\text{tax}} - 1},$$

where $V_{\text{tax}}$ represents the vector corresponding to a single individual's microbiota trajectory between successive timepoints, $(\Delta_1, \sigma_1, \tau_{\text{tax}} - \tau_1)$, and $V_{\sigma}$ represents the vector pointing to the point for which permissivity is being calculated, $(\Delta - \Delta_0, \sigma - \sigma_0)$. In other words, it is the absolute value of the cosine of the angle formed between these two vectors, summed over all such vector pairs. In this analysis, the state-space was subdivided into 400 equally sized regions corresponding to 5% differences in relative abundance of taxa along a given face, and the permissivity was calculated at the centroid of these triangular regions.

**Data Availability.** All data collected as part of the Pioneer 100 project (S) are available from dbGaP with accession ID phs001363.v1.p1 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001363.v1.p1).

**ACKNOWLEDGMENTS.** We are grateful to our pioneers for enabling, and participating in, this study. We thank Daniel McDonald for assistance reprocessing 16S data. This work was supported by the M. J. Murdock Charitable Trust (U.H. and N.D.P.). S.M.G. was supported by a Washington Research Foundation Distinguished Investigator Award and by startup funds from the Institute for Systems Biology. T.W. was supported by the K. Carole Ellison Fellowship in Bioinformatics.

10. V. R. Velagapudi et al., The gut microbiota modulates host energy and lipid metabolism in mice. J. Lipid Res. 51, 1101–1112 (2010).


