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

Roie Levy[a,1], Andrew T. Magis[a], John C. Earls[a], Ohad Manor[b], Tomasz Wilmanski[a], Jennifer Lovejoy[a], Sean M. Gibbons[a,c], Gilbert S. Omenn[a,d], Leroy Hood[a,2], and Nathan D. Price[a,2]

[a]Institute for Systems Biology, Seattle, WA 98109; [b]Arivale, Seattle, WA 98104; [c]eScience Institute, University of Washington, Seattle, WA 98195; and [d]Department 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.

microbiome | multiomic | state transition | *Prevotella* | *Bacteroides*

Technological advances in molecular profiling and deep phenotyping of individual humans (i.e., measuring thousands of health-related biomarkers) are poised to transform biomedicine in coming years. Accordingly, numerous public and private institutions recently launched initiatives with the aim of determining how to translate deeply characterized phenotypes into improvements to health and health care. For example, the National Institutes of Health launched the Precision Medicine Initiative with the goal of creating a voluntary research cohort of one million individuals to identify genetic drivers of cancers and other diseases of unknown etiology (1), the Google Baseline study includes developing wearable technologies to profile biomolecules in real time (2), and Human Longevity, Inc., focuses on aging-associated diseases (3). Furthermore, integrating molecular profiling into ongoing longitudinal cohort studies, such as the Framingham Heart Study, has been successful in identifying genomic drivers of diseases like obesity (4).

In 2014, the Institute for Systems Biology launched the Pioneer 100 study (5) as a pilot for the longer-term 100K Wellness Project (6). As part of the Pioneer 100 study, we densely quantified the molecular profiles of 108 participants over 9 mo, producing thousands of measurements comprising genome, blood proteome, blood metabolome, gut microbiome, clinical chemistries, and activity monitoring (i.e., deep phenotyping). In contrast to the initiatives described above, we focused on optimizing general wellness as opposed to targeting specific disease phenotypes. Central to this focus, each participant's molecular profile was interpreted alongside a wellness coach (i.e., a qualified clinician-scientist) who identified actionable opportunities and incorporated individuals' goals to develop personalized regimens to optimize wellness (5, 7).

An integral element of the Pioneer 100 study was 16S profiling of the bacterial and archaeal component of the intestinal microbiome. The ecology of the gut microbiome directly affects its host by modulating metabolism (8–12) and influences many diseases, such as obesity (13), inflammatory bowel disease (14), and diabetes (11). Microbiome composition may influence how we metabolize certain foods and has led to calls for personalized diets (15). A major determinant of variation in the gut microbiome across people is the dominance of either *Prevotella* or *Bacteroides* (16), which influences the fermentative output of the microbiome (17) and can determine the outcome of dietary weight loss interventions (18, 19). However, while dietary intervention was able to modulate the abundance of *Bacteroides* relative to *Firmicutes* (20), specific dietary modulation of the *Bacteroides*/*Prevotella* ratio has not been thoroughly demonstrated. Specifically, in at least two studies, controlled, short-term dietary interventions were ineffective in pushing the microbiome between compositional states

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.

Author contributions: R.L., A.T.M., J.L., G.S.O., L.H., and N.D.P. designed research; R.L., A.T.M., O.M., J.L., S.M.G., and N.D.P. performed research; J.C.E. contributed new reagents/analytic tools; R.L., J.C.E., O.M., and T.W. analyzed data; and R.L., T.W., S.M.G., G.S.O., L.H., and N.D.P. wrote the paper.

Reviewers: E.B.C., The University of Chicago; and P.W., University of Luxembourg.

The authors declare no competing interest.

[1]Present address: Verily Life Sciences, South San Francisco, CA 94080.

[2]To whom correspondence may be addressed. Email: lhood@systemsbiology.org or nprice@systemsbiology.org.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1922498117/-/DCSupplemental.

First published May 29, 2020.

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 metabolomic and clinical chemistries profiles. We identify within our cohort a subpopulation 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 impermissive region in microbiome state-space, 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 ($\rho = -0.66$, $P < 2.20 \times 10^{-16}$), as was the major intestinal phylum Firmicutes ($\rho = -0.74$, $q < 2.20 \times 10^{-16}$). Conversely, Bacteroidetes, the other major phylum, was positively correlated with this dimension ($\rho = 0.87$, $q < 2.20 \times 10^{-16}$). In contrast, this structure was not observed by NMDS of Bray–Curtis dissimilarity (BCD), which does not take into account phylogenetic relationships among taxa. Dimension 1 of the BCD NMDS revealed a nonmonotonic association with Bacteroidetes and Firmicutes (*SI Appendix*, Fig. S1), possibly indicating two different subclasses within the high-Bacteroidetes samples.

To further characterize these putative subclasses of high-Bacteroidetes samples, we compared an equivalent number of samples from both extremes of BCD NMDS dimension 1 ($n = 25$ per subsample) (*Methods*). These two subsamples differed significantly in α-diversity (Cohen's $d = -0.28$, $P < 0.018$), and thus were termed low diversity (LO) vs. high diversity (HI). We investigated which, if any, operational taxonomic units (OTUs) disproportionately represented LO vs. HI samples. We chose stringent selection criteria to preferentially weight more abundant representative OTUs (*Methods*). Using these criteria, we found that the OTUs resolving to *Prevotella* best represented the LO class ($d = 4.94$, false discovery rate [FDR] $< 6.70 \times 10^{-12}$), while those resolving to *Bacteroides* best represented the HI class ($d = -4.37$, FDR $< 6.70 \times 10^{-12}$) (*SI Appendix*, Fig. S2). Notably, a single OTU resolving to *Prevotella copri* dominated the LO class, while diversity was more evenly spread among multiple OTUs resolving to genus *Bacteroides*, likely driving the noted difference in α-diversity between these classes. Indeed, the *Prevotella copri* OTU represented 61 ± 18% (mean ± SD) of the LO samples, while the dominant *Bacteroides* OTU (which resolved to *B. uniformis*) represented only 8.5 ± 7.0% of the HI samples.

Table 1. Cohort demographics

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

Abbreviations: BMI, body mass index; HDL, high-density lipoprotein; IQR, interquartile range; TNFα, tumor necrosis factor α.

***Bacteroides* and *Prevotella* Correlate with Levels of Serum Metabolites and Clinical Chemistries.** Given their potential to partition microbiome samples, we next investigated the potential clinical relevance of these taxa. Specifically, we examined the potential association of *Bacteroides* and *Prevotella* with two molecular profiles from the Pioneer 100 study: clinical chemistries and metabolomes. We chose to focus on these two components for two reasons. First, they are most readily interpretable from a clinical perspective: Many features are already associated with pathways or phenotypes of interest. Second, like the microbiome, they vary in time: features have the potential to vary in response to intervention, on a per-individual basis. We observed a number of statistically significant pairwise correlations between metabolites and both genera (Dataset S1). Specifically, *Bacteroides* was anti-correlated with a number of intermediates of phenylalanine metabolism including p-cresol sulfate, consistent with previous results from our group (24). p-cresol sulfate is a product of microbial fermentation and a uremic toxin (25), which we previously found to be positively correlated to the families Verrucomicrobiaceae and Desulfovibrionaceae (5). *Bacteroides* was also negatively correlated with Verrucomicrobiaceae and Desulfovibrionaceae ($\rho = -0.20$, $P < 9.6 \times 10^{-4}$, and $\rho = -0.16$, $P < 0.01$, respectively). *Prevotella*, in turn, correlated negatively with omega-6 fatty-acid metabolism and carnitine intermediates, as well as thyroxine, a prohormone of the metabolism-regulating tri-iodothyronine (T3) thyroid hormone.

After multiple hypothesis correction, we observed no significant pairwise correlations between clinical chemistries and *Bacteroides* or *Prevotella*. Subsequently, we employed dbRDA (26) (*Methods*), a constrained ordination technique that determines how much variation in a set of observations can be described by a complementary set of features (i.e., chemistries). In contrast to other constrained ordination techniques such as canonical correlation analysis, dbRDA accommodates (dis)similarity metrics that are non-Euclidean (e.g., UniFrac), which are often more relevant to comparison of ecological communities. Along the first two dimensions, clinical chemistries accounted for 18.5% of total microbial β-diversity, and partitioned observations similarly to NMDS as described above. Specifically, dimension 1 (explaining 12.7% of β-diversity) separated samples high in Bacteroidetes from those high in Firmicutes, while among high-Bacteroidetes samples dimension 2 (explaining 5.8% of β-diversity)

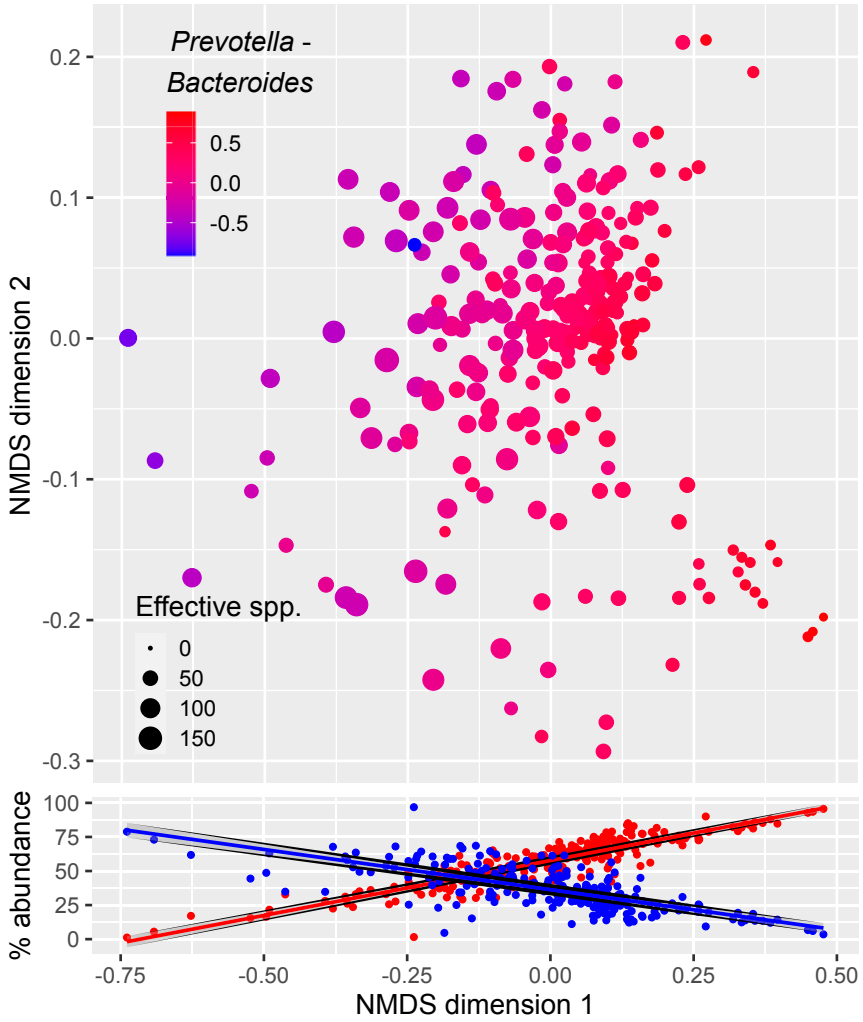

Fig. 1. β-Diversity of the Pioneer 100 microbiome. (*Top*) NMDS of the weighted UniFrac dissimilarity of microbiome samples. The first two of three dimensions are shown. Each sample's size corresponds to its Shannon diversity (larger size equals higher diversity), while its color corresponds to the difference between the relative abundance of Bacteroidetes and Firmicutes (red, higher Bacteroidetes; blue, higher Firmicutes). (*Bottom*) Scatterplot of the relative abundance of Bacteroidetes (red) and Firmicutes (blue) against NMDS dimension 1. The lines shown in the plot correspond to a $y \sim x$ regression line, with the shaded regions indicating the 95% confidence intervals for the slope of the line.

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 α (TNFα), 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% *Prevotella* 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 ∼0% 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 Bacteroidetes 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

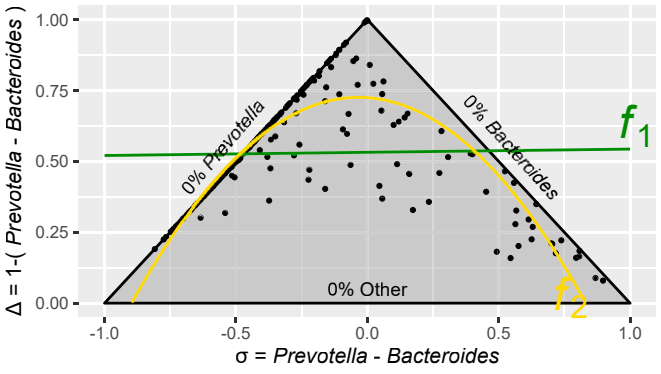

Fig. 2. Modeling the *Prevotella* and *Bacteroides* relationship. The space of possible relative abundances of taxa is a positive simplex, shown here as a ternary plot. Samples lying close to the leftmost vertex are dominated by *Bacteroides*, while those close to the right are dominated by *Prevotella* and those at the center top are depleted in both. Similarly, samples along the "northwest" edge are depleted in *Prevotella*, and those along the "northeast" are depleted in *Bacteroides*. Two ordinary least-squares fits of the transformed variables Δ vs. σ are shown: a linear fit (f_1) in green and a second-order fit (f_2) in gold. A second-order model more accurately reflects the relationship between the two taxa.

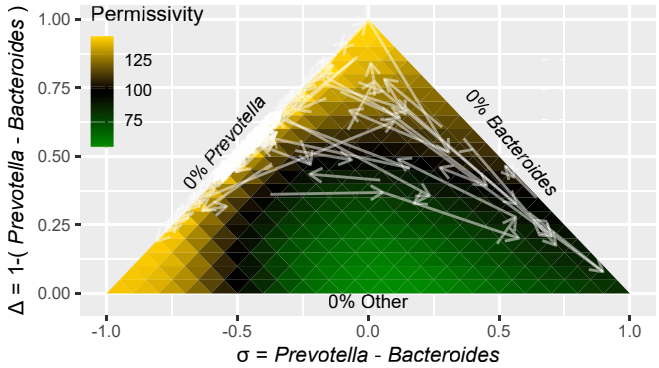

Fig. 3. Permissivity of *Prevotella*/*Bacteroides* state-space. Samples are plotted as in Fig. 2, but with arrows connecting consecutive observations of the same individual (i.e., trajectories). Spatial geometry of each plotted point corresponds to its abundance (samples in the "northwest" edge are depleted in *Prevotella* and those along the "northeast" are depleted in *Bacteroides*; samples in the center top are depleted in both), whereas color corresponds to permissivity. Specifically, the face of each triangular region corresponds to a 5% change in the relative abundance of one taxon. The color of the triangle represents the permissivity of that region (green, low permissivity; gold, high permissivity). More trajectories align with high-permissivity regions than with low-permissivity regions.

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 antimicrobial application) before subsequently administering a dietary and/or probiotic intervention to support the desired genus. Alternatively, diet could be persistently modified to support the opposite compositional state, and over time natural perturbations should lead to bottlenecks that allow the other genus to establish itself in its preferred niche [e.g., long-term high-fiber diet seems to support *Prevotella* dominance (46)]. A recent study characterizing gut microbiome changes associated with US migration from Thailand demonstrated that long-term lifestyle and dietary changes are able to induce a transition from a *Prevotella*-dominant to a *Bacteroides*-dominant state. However, these transitions took months of living in the United States to manifest themselves and were more pronounced in second-generation Thai Americans relative to Thai immigrants, indicating the importance of a persistent dietary/lifestyle modification in order to facilitate transition between these two genera (47). If we wish to engineer the gut microbiome to improve human health (48), we must first understand the forces that underlie its stability and resilience. In this study, we find that hysteresis can likely be overcome by mapping out permissive paths through microbiome state-space.

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 gave written informed consent for analysis of their data. Blood, stool, and urine samples for all participants were collected during three separate 2-wk periods, which we refer throughout this manuscript as "rounds." Rounds were approximately 3 mo apart, and participants freely scheduled their own collections each round. A total of 101 of the 108 pioneers provided at least one stool sample for gut microbiome analysis, and hence were included in this study. Characteristics of the cohort are provided in Table 1.

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 genome), these profiles vary in time and in response to intervention. Features with more than 100 missing values were discarded: 3-deoxyglucosone hydroimidazolones, aminoadipic acid, bun/creatinine ratio, (carboxyethyl) lysine, carboxymethyl-lysine, glyoxal-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 quantify pairwise dissimilarity 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 dissimilarity in this reduced space is 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-Bacteroidetes samples along a single dimension (*SI Appendix*, Fig. S1). For simplicity in preliminary analysis, we used this ordination 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 abscissa 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, a difference in magnitude between groups, 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 Datasets 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:

$$\Delta_{tax} = A_i - A_j,$$

and their sum subtracted from 1:

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

This transformation accounts for an antisymmetry in linear regression (e.g., the regression of *Bacteroides* on *Prevotella* does not equal the regression of *Prevotella* on *Bacteroides*). After such a transformation, Δ_{tax} is weighted equally by both taxa, while σ_{tax} and any residuals are weighted by their sum; subtracting from 1 allows the plot of σ_{tax} versus Δ_{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:

$$R^2_{rel} = 1 - \frac{\sum(\Delta_{tax} - f_2)^2}{\sum(\Delta_{tax} - f_1)^2}.$$

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 state-space to permit transit. We term this 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 (Δ, σ) as follows:

$$P = \sum \left| \frac{v_{p\rightarrow}}{|v_{p\rightarrow}|} \cdot \frac{v_{t\rightarrow}}{|v_{t\rightarrow}|} \right|,$$

where $v_{t\rightarrow}$ represents the vector corresponding to a single individual's microbiome trajectory between consecutive timepoints, $(\Delta_{t+1} - \Delta_t, \sigma_{t+1} - \sigma_t)$, and $v_p \rightarrow$ represents the vector pointing to the point for which permissivity is being calculated,$(\Delta - \Delta_t, \sigma - \sigma_t)$. 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 (5) 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 (L.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.

1. F. S. Collins, H. Varmus, A new initiative on precision medicine. *N. Engl. J. Med.* **372**, 793–795 (2015).
2. J. Kaiser, Google X sets out to define healthy human. Science. https://www.sciencemag.org/news/2014/07/google-x-sets-out-define-healthy-human. Accessed 4 November 2019.
3. S. Y. Rojahn, Microbes and metabolites fuel an ambitious aging project. MIT Technology Review. https://www.technologyreview.com/2014/03/11/173738/microbes-and-metabolites-fuel-an-ambitious-aging-project/. Accessed 4 November 2019.
4. A. Herbert *et al.*, A common genetic variant is associated with adult and childhood obesity. *Science* **312**, 279–283 (2006).
5. N. D. Price *et al.*, A wellness study of 108 individuals using personal, dense, dynamic data clouds. *Nat. Biotechnol.* **35**, 747–756 (2017).
6. L. Hood, J. C. Lovejoy, N. D. Price, Integrating big data and actionable health coaching to optimize wellness. *BMC Med.* **13**, 4 (2015).
7. N. Zubair *et al.*, Genetic predisposition impacts clinical changes in a lifestyle coaching program. *Sci. Rep.* **9**, 6805 (2019).
8. A. Mardinoglu *et al.*, The gut microbiota modulates host amino acid and glutathione metabolism in mice. *Mol. Syst. Biol.* **11**, 834 (2015).
9. S. Shoaie *et al.*; MICRO-Obes Consortium, Quantifying diet-induced metabolic changes of the human gut microbiome. *Cell Metab.* **22**, 320–331 (2015).
10. V. R. Velagapudi *et al.*, The gut microbiota modulates host energy and lipid metabolism in mice. *J. Lipid Res.* **51**, 1101–1112 (2010).
11. A. Koh *et al.*, Microbially produced imidazole propionate impairs insulin signaling through mTORC1. *Cell* **175**, 947–961.e17 (2018).
12. W. R. Wikoff, J. A. Gangoiti, B. A. Barshop, G. Siuzdak, Metabolomics identifies perturbations in human disorders of propionate metabolism. *Clin. Chem.* **53**, 2169–2176 (2007).
13. J. Zou *et al.*, Fiber-mediated nourishment of gut microbiota protects against diet-induced obesity by restoring IL-22-mediated colonic health. *Cell Host Microbe* **23**, 41–53.e4 (2018).
14. K. Lu, C. G. Knutson, J. S. Wishnok, J. G. Fox, S. R. Tannenbaum, Serum metabolomics in a *Helicobacter hepaticus* mouse model of inflammatory bowel disease reveal important changes in the microbiome, serum peptides, and intermediary metabolism. *J. Proteome Res.* **11**, 4916–4926 (2012).
15. D. Zeevi *et al.*, Personalized nutrition by prediction of glycemic responses. *Cell* **163**, 1079–1094 (2015).
16. M. Arumugam *et al.*; MetaHIT Consortium, Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
17. T. Chen *et al.*, Fiber-utilizing capacity varies in *Prevotella*- versus *Bacteroides*-dominated gut microbiota. *Sci. Rep.* **7**, 2594 (2017).
18. M. F. Hjorth *et al.*, *Prevotella*-to-*Bacteroides* ratio predicts body weight and fat loss success on 24-week diets varying in macronutrient composition and dietary fiber: Results from a post-hoc analysis. *Int. J. Obes.* **43**, 149–157 (2019).
19. M. F. Hjorth *et al.*, Pre-treatment microbial *Prevotella*-to-*Bacteroides* ratio, determines body fat loss success during a 6-month randomized controlled diet intervention. *Int. J. Obes.* **42**, 580–583 (2018).
20. L. A. David *et al.*, Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
21. G. D. Wu *et al.*, Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108 (2011).
22. H. M. Roager, T. R. Licht, S. K. Poulsen, T. M. Larsen, M. I. Bahl, Microbial enterotypes, inferred by the *Prevotella*-to-*Bacteroides* ratio, remained stable during a 6-month randomized controlled diet intervention with the new Nordic diet. *Appl. Environ. Microbiol.* **80**, 1142–1149 (2014).
23. B. E. Beisner, D. T. Haydon, K. Cuddington, Alternative stable states in ecology. *Front. Ecol. Environ.* **1**, 376–382 (2003).
24. T. Wilmanski *et al.*, Blood metabolome predicts gut microbiome α-diversity in humans. *Nat. Biotechnol.* **37**, 1217–1228 (2019).
25. R. Vanholder, E. Schepers, A. Pletinck, E. V. Nagler, G. Glorieux, The uremic toxicity of indoxyl sulfate and p-cresyl sulfate: A systematic review. *J. Am. Soc. Nephrol.* **25**, 1897–1907 (2014).
26. P. Legendre, M. J. Andersson, Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. *Ecol. Monogr.* **69**, 1–24 (1999).
27. D. Knights *et al.*, Rethinking "enterotypes.". *Cell Host Microbe* **16**, 433–437 (2014).
28. K. R. Theis *et al.*, Getting the hologenome concept right: An eco-evolutionary framework for hosts and their microbiomes. *mSystems* **1**, e00028-16 (2016).
29. S. R. Gill *et al.*, Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355–1359 (2006).
30. A. M. O'Hara, F. Shanahan, The gut flora as a forgotten organ. *EMBO Rep.* **7**, 688–693 (2006).
31. L. Dethlefsen, M. McFall-Ngai, D. A. Relman, An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature* **449**, 811–818 (2007).
32. E. K. Costello, K. Stagaman, L. Dethlefsen, B. J. M. Bohannan, D. A. Relman, The application of ecological theory toward an understanding of the human microbiome. *Science* **336**, 1255–1262 (2012).
33. J. C. Stegen *et al.*, Quantifying community assembly processes and identifying features that impose them. *ISME J.* **7**, 2069–2079 (2013).
34. J. Qin *et al.*; MetaHIT Consortium, A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
35. P. J. Turnbaugh *et al.*, A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
36. R. Levy, E. Borenstein, Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 12804–12809 (2013).
37. R. Levy, E. Borenstein, Metagenomic systems biology and metabolic modeling of the human microbiome: From species composition to community assembly rules. *Gut Microbes* **5**, 265–270 (2014).
38. S. Greenblum, P. J. Turnbaugh, E. Borenstein, Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 594–599 (2012).
39. B. Alper, B. Erdogan, M. Ö. Erdogan, K. Bozan, M. Can, Associations of trauma severity with mean platelet volume and levels of systemic inflammatory markers (IL1β, IL6, TNFα, and CRP). *Mediators Inflamm.* **2016**, 9894716 (2016).
40. B. T. Lee *et al.*, Association of C-reactive protein, tumor necrosis factor-alpha, and interleukin-6 with chronic kidney disease. *BMC Nephrol.* **16**, 77 (2015).

41. Y. Heper *et al*., Evaluation of serum C-reactive protein, procalcitonin, tumor necrosis factor alpha, and interleukin-10 levels as diagnostic and prognostic parameters in patients with community-acquired sepsis, severe sepsis, and septic shock. *Eur. J. Clin. Microbiol. Infect. Dis.* **25**, 481–491 (2006).
42. K. Faust *et al*., Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.* **8**, e1002606 (2012).
43. C. Zhang *et al*., Ecological robustness of the gut microbiota in response to ingestion of transient food-borne microbes. *ISME J.* **10**, 2235–2245 (2016).
44. L. Dethlefsen, D. A. Relman, Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl. Acad. Sci. U.S.A.* **108** (suppl. 1), 4554–4561 (2011).
45. S. M. Gibbons, S. M. Kearney, C. S. Smillie, E. J. Alm, Two dynamic regimes in the human gut microbiome. *PLOS Comput. Biol.* **13**, e1005364 (2017).
46. K. Makki, E. C. Deehan, J. Walter, F. Bäckhed, The impact of dietary fiber on gut microbiota in host health and disease. *Cell Host Microbe* **23**, 705–715 (2018).
47. P. Vangay *et al*., US immigration westernizes the human gut microbiome. *Cell* **175**, 962–972.e10 (2018).
48. S. M. Kearney, S. M. Gibbons, Designing synbiotics for improved human health. *Microb. Biotechnol.* **11**, 141–144 (2018).
49. J. R. Rideout *et al*., Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ* **2**, e545 (2014).
50. T. Z. DeSantis *et al*., Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006).
51. J. G. Caporaso *et al*., QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
52. L. Jost, Entropy and diversity. *Oikos* **113**, 363–375 (2006).
53. C. E. Shannon, A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
54. J. Oksanen *et al*., vegan: Community ecology package (R package, Version 2.5-2, CRAN R, 2018).
55. J. R. Bray, J. T. Curtis, An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* **27**, 325–349 (1957).
56. C. A. Lozupone, M. Hamady, S. T. Kelley, R. Knight, Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. *Appl. Environ. Microbiol.* **73**, 1576–1585 (2007).
57. P. J. McMurdie, S. Holmes, phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217 (2013).
58. D. J. Stekhoven, P. Bühlmann, MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**, 112–118 (2012).
59. J. B. Kruskal, Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**, 1–27 (1964).
60. P. R. Minchin, An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio* **69**, 89–107 (1987).
61. Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
62. G. M. Sullivan, R. Feinn, Using effect size—or why the P value is not enough. *J. Grad. Med. Educ.* **4**, 279–282 (2012).
63. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
64. P. Legendre, L. Legendre, *Numerical Ecology*. Developments in Environmental Modelling (Elsevier, Amsterdam, 1988), vol. 24.