Silverman *et al. Microbiome* (2018) 6:202
https://doi.org/10.1186/s40168-018-0584-3

Microbiome

RESEARCH Open Access

Dynamic linear models guide design and analysis of microbiota studies within artificial human guts

Justin D. Silverman[1,2,3], Heather K. Durand[5], Rachael J. Bloom[4], Sayan Mukherjee[1,6] and Lawrence A. David[1,3,4,5*]

Abstract

Background: Artificial gut models provide unique opportunities to study human-associated microbiota. Outstanding questions for these models' fundamental biology include the timescales on which microbiota vary and the factors that drive such change. Answering these questions though requires overcoming analytical obstacles like estimating the effects of technical variation on observed microbiota dynamics, as well as the lack of appropriate benchmark datasets.

Results: To address these obstacles, we created a modeling framework based on multinomial logistic-normal dynamic linear models (MALLARDs) and performed dense longitudinal sampling of four replicate artificial human guts over the course of 1 month. The resulting analyses revealed how the ratio of biological variation to technical variation from sample processing depends on sampling frequency. In particular, we find that at hourly sampling frequencies, 76% of observed variation could be ascribed to technical sources, which could also skew the observed covariation between taxa. We also found that the artificial guts demonstrated replicable trajectories even after a recovery from a transient feed disruption. Additionally, we observed irregular sub-daily oscillatory dynamics associated with the bacterial family Enterobacteriaceae within all four replicate vessels.

Conclusions: Our analyses suggest that, beyond variation due to sequence counting, technical variation from sample processing can obscure temporal variation from biological sources in artificial gut studies. Our analyses also supported hypotheses that human gut microbiota fluctuates on sub-daily timescales in the absence of a host and that microbiota can follow replicable trajectories in the presence of environmental driving forces. Finally, multiple aspects of our approach are generalizable and could ultimately be used to facilitate the design and analysis of longitudinal microbiota studies in vivo.

Keywords: Artificial gut, Bioreactor, Microbiome, Metagenomics, Compositional data, Bayesian statistics, Time series analysis

Background

Artificial gut models have been used for decades to replicate the human intestinal environment and study the dynamics of resident microbes [1–3]. These systems have the advantage of being sampled with arbitrary frequency, house environments that can be precisely controlled, and often face fewer ethical concerns than in-human studies [4]. Artificial gut models have therefore been used to discover the effect of nutritional supplements on infant gut microbiota [5], mechanisms by how commensals repress *Salmonella* virulence [6], and dose-dependencies between microbiota-targeting therapies and metabolite production [7].

Yet, despite the utility and long development of artificial gut models, fresh questions regarding their fundamental biology remain. Such questions include the rapidity with which the composition of these microbial communities varies both in the presence and absence of perturbations [1, 8]. While it is well known that in vivo microbiota may change on sub-daily timescales due to host forcing, it is

* Correspondence: lawrence.david@duke.edu
[1] Program in Computational Biology and Bioinformatics, Duke University, CIEMAS, Room 2171, 101 Science Drive, Box 3382, Durham, NC 27708, USA
[3] Center for Genomic and Computational Biology, Duke University, Durham, NC 27708, USA
Full list of author information is available at the end of the article

unclear whether such sub-daily dynamics may be seen in the absence of host effects [9]. The degree to which replicate ex vivo systems exhibit stochastic behavior, or conversely behave deterministically, remains another outstanding question [1, 10]. Finally, a deeper understanding of the reproducibility of artificial gut models may also have implications for our understanding of the relative importance of various factors in shaping the dynamics of host-associated microbiota [11].

Gaining greater insight into the biology of artificial gut models though requires addressing analytical and statistical challenges. One key challenge is the often unquantified impact of artificial sources of intra-study variation such as variation due to sequencing counting and technical variation from sample processing (e.g., unintended experimental errors and batch effects) [12–18]. In particular, while the impact of variation due to sequence counting is more commonly addressed and modeled [12, 19–21], the impact of technical variation from sample processing is often unquantified and less well understood. Such technical variation may alter inference of both the magnitude and direction of variation among bacterial taxa. In addition, it is well established that sequencing studies of microbiota provide information only on the relative amounts of taxa and not their absolute abundance [22–24]. The analysis of such relative (or compositional) data remains an open area of study and naive analyses can lead to a distorted view of the patterns of variation present in a community [21, 22, 25–27]. Notably, the variation due to sequence counting, technical variation from sample processing, and compositional effects are all challenges facing in vivo microbiota studies, in addition to artificial gut experiments.

The design of ex vivo microbiota studies is also confronted by the lack of appropriate benchmark datasets. For example, defining suitable sampling frequencies for artificial gut studies requires insight into the timescales on which human gut microbiota fluctuate [28]. Yet, even though gut microbiota dynamics in vivo and bacterial mock communities in vitro are known to behave on the timescale of hours [9, 29, 30], most longitudinal studies to date in ex vivo models are only sampled on the order of days [1, 8, 10, 31]. Additionally, while the collection of technical replicates could be used to quantify the effects of technical variation [32], such replicate sampling is generally not performed in longitudinal microbiota studies.

Here, we integrated model development and experimental design to address key challenges facing the analysis and design of longitudinal artificial gut studies. We collected longitudinal samples with up to hourly frequency from replicate artificial gut models over the course of 1 month. We combined this longitudinal sampling with the collection of technical replicates so that we could characterize the impact of technical sources of variation on observed microbiota dynamics. To isolate separate biological and technical sources of community variation in our dataset, we created a modeling framework called MALLARD that is based on a class of generalized dynamic linear models appropriate for microbiota time-series data. Together, our dataset and modeling framework allowed us to investigate the patterns and timescales of microbiota variation in an artificial human gut.

Results

Longitudinal modeling

To separate biological and technical variation in artificial gut time-series, we introduce an extension of dynamic linear models (DLMs) tailored for microbiota data. DLMs have widespread use including industrial applications such as commercial forecasting and engineering control systems [33]. At their core, DLMs model a system as a time-varying state that is observed through a noisy process. We extended DLMs to a class of multinomial logistic-normal dynamic linear models by building off of the work by Cargnoni C, Muller P, and West M [34]. We refer to this as the MALLARD class of models.

We analyzed the artificial gut dataset described below using a MALLARD model that is generative and assumes there exists an unobserved microbial composition (θ_t; the state) that evolves through time (Fig. 1a) due to stochastic biological variations (w_t). We regard the state sequence ($\theta_1, \ldots, \theta_t, \ldots, \theta_T$) as the true microbial dynamics in a time-series. Random technical variations (v_t) are then added to the true system state (θ_t) resulting in the composition η_t (Fig. 1b). We observe η_t through a multinomial counting process. This formulation is similar to the constant level model commonly used in Bayesian time-series analysis [34]. By separately modeling the process generating w_t and v_t with distinct covariance matrices (W and V, respectively), we can decouple biological and technical variations in artificial gut datasets (Fig. 1c). Visually, we found this model provided a good fit to artificial gut data (Additional file 1).

Daily and hourly gut microbiota time-series in an artificial human gut

We applied our model to an artificial gut that was constructed using continuous-flow anaerobic bioreactor systems that have been validated as models of human gut microbiota [1, 6, 14, 35, 36]. The same starting human fecal inoculum was seeded into replicate ex vivo vessels ($n = 4$) and cultured for 1 month (Fig. 2 and Additional files 2, 3, 4, and 5). Throughout the experiment, pH, temperature, media input rates, and oxygen concentration were all fixed ("Methods" section). To introduce microbial dynamics into our systems, a single bolus of *Bacteroides ovatus* isolated from the stool donor was administered to the system on Day 23 ("Methods" section). The *B. ovatus* bolus did not have discernable effects

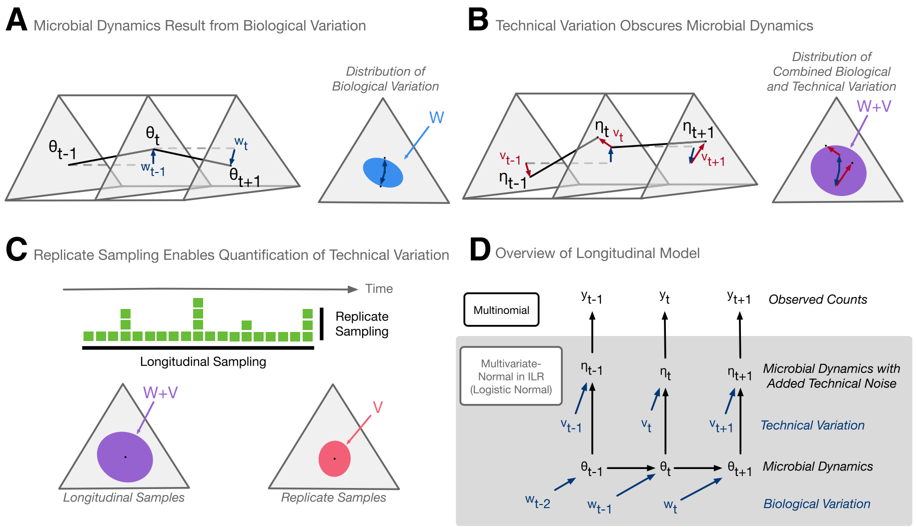
A Microbial Dynamics Result from Biological Variation
Distribution of Biological Variation
W
B Technical Variation Obscures Microbial Dynamics
Distribution of Combined Biological and Technical Variation
W+V
C Replicate Sampling Enables Quantification of Technical Variation
Time
Replicate Sampling
Longitudinal Sampling
W+V
V
Longitudinal Samples
Replicate Samples
D Overview of Longitudinal Model
Multinomial
Observed Counts
Multivariate-Normal in ILR (Logistic Normal)
Microbial Dynamics with Added Technical Noise
Technical Variation
Microbial Dynamics
Biological Variation

Fig. 1 A generative model for microbial dynamics obscured by technical variation. **a** Microbial dynamics result from biological variation. The time series ($\theta_1, , \ldots, \theta_t, \ldots, \theta_T$) defines the dynamics of a microbial community and results from biological variations ($w_1, , \ldots, w_t, \ldots, w_T$) which are assumed to be independent and identically distributed (i.i.d.) logistic-normal with mean zero and covariance W. **b** Technical variation obscures microbial dynamics. Technical variation ($v_1, , \ldots, v_t, \ldots, v_T$) from sample processing introduces noise into measurements of microbial dynamics and are assumed to be i.i.d. logistic-normal with mean zero and covariance V. **c** Replicate sampling enables quantification of technical variation. Hypothetical collected samples are denoted by green squares. Differences between longitudinal samples reflect both biological and technical sources of variation. In contrast, differences between technical replicates (samples from the same time point) should reflect only technical sources of variation and can be used to estimate V ("Methods" section). **d** Overview of longitudinal model. The microbial dynamics and the confounding technical variation (**a**, **b**) are modeled in an isometric log-ratio (ILR) space such that the logistic-normal distribution transforms to a multivariate normal distribution for mathematical convenience ("Methods" section). The observed count data is assumed to be distributed multinomial from the compositions ($\eta_1, , \ldots, \eta_t, \ldots, \eta_T$) ("Methods" section)

on microbial dynamics, but the media it was suspended in appeared to induce minor shifts in the relative abundances of select bacterial taxa that were visible with hourly sampling (Additional file 3). Additional microbial dynamics related to media input were generated by an inadvertent feed disruption in two vessels between days 11 and 13 of the study. Overall, similar to previous studies [1, 10], the artificial gut maintained much of the microbial diversity present in the inoculating stool: 91% of bacterial families present on days 1–5 of the study were detected between days 23 and 28. The 9% of bacterial families that did not persist represented only 0.06% of the total sequencing reads in the dataset.

To investigate the timescales on which microbial communities vary ex vivo and to characterize the impact of technical variation on our measurements, we developed a sampling scheme for our artificial gut that featured fine temporal resolution and numerous technical replicates (Additional file 4). As in previous studies, all four vessels of the artificial gut were sampled daily over the course of 1 month. To investigate potential sub-daily variation, we aimed to oversample the system and collected 120 sequential hourly samples, from each replicate vessel, during a 5-day period. To estimate the technical variation in our measurements, we collected 20 replicate samples from the final time-point of each artificial gut vessel (Fig. 1c; "Methods" section). As samples collected in the same vessel at the same time are expected to be biologically identical (i.e., feature no biological variation), any variation in the resulting community measurements between these technical replicates can be ascribed solely to technical sources (Fig. 1c). To ensure that the technical variation profile of the replicate sampled matched that of the longitudinal samples, all samples were randomized and processed together for sequencing.

The structure and magnitude of technical variation from sample processing

After fitting the MALLARD model to the resulting artificial gut dataset, we investigated the technical variation from sample processing (V) to our inference of variation due to biological sources (W). Variation from sample processing and variation due to sequence counting were represented as separate processes (Fig. 1d). Differing correlation structure between V and W would support our hypothesis that sources of technical variation could obscure biological forces acting on microbiota within an artificial gut. Indeed, a permutation analysis indicated it highly improbable that V and W had the same correlation structure (posterior probability < 1%; "Methods" section). Low-dimensional projections of the posterior distributions of V and W supported this conclusion and

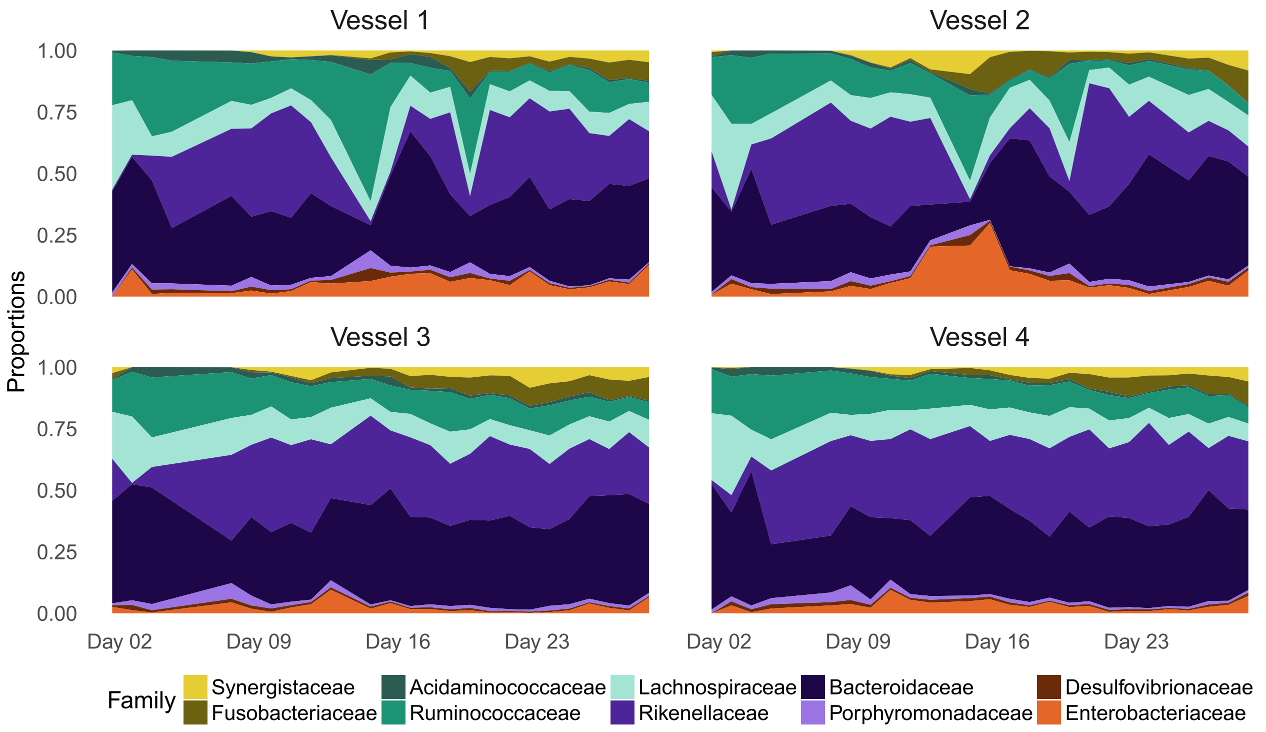

Fig. 2 Proportions of ten most abundant bacterial families over time. Four identical continuous-flow anaerobic bioreactor systems were each inoculated from a single human fecal specimen and cultured over the course of 1 month. Proportions of bacterial taxa at sampled time points were estimated by dividing observed read counts from 16S rRNA sequencing by the total number of counts observed for each sample. In addition to daily sampling (shown here), hourly samples were taken from each vessel (480 hourly samples total) as well as 20 technical replicate samples from the final time point of each vessel (Additional file 4). Proportions from the hourly sampling data are shown in Additional file 3. PCoA analysis using the Aitchison distance is shown in Additional file 5

revealed how overall variation patterns involving bacterial families like Lachnospiraceae, Fusobacteriaceae, and Bacteroidaceae more strongly resembled patterns of technical variation than biological variation (Fig. 3a). Thus, some patterns of covariation among taxa in artificial gut studies may be due to technical sources of variation.

We next investigated the relative magnitude of technical variation from sample processing and biological variation as a function of sampling frequency. Total variation is defined as the trace of a covariance matrix [37]. We therefore computed $Tr(W)/Tr(V)$, which is analogous to the signal-to-noise ratio used in signal processing [38]. We found that for our time-series analyzed on an hourly basis ("Methods" section), the median ratio of biological to technical variation was 0.30 (0.26–0.36 95% credible interval). We therefore estimate that only 24% (21%–26%) of the total variation in relative abundances was due to biological sources. This result was insensitive to perturbation of our priors (Additional file 6). Moreover, because variance is additive between time-points within our model, we could estimate biological and technical variation as a function of sampling interval (Fig. 3b; "Methods" section). We found that at a sampling interval of 3.5 h, the total biological and technical variation was approximately equal. At sampling frequencies faster than 3.5 h, technical variation outweighed biological variation; at slower sampling frequencies, signal associated with biological variation exceeded technical noise.

Timescales of microbial dynamics

In order to explore the timescale of microbial dynamics within replicate artificial gut vessels, we visualized microbial dynamics using PhILR balances. PhILR balances represent the log-ratio between phylogenetically neighboring clades of taxa, providing a phylogenetically informed way to study microbial dynamics without compositional artifacts [22]. We quantified the magnitude of changes in balance values in units of evidence information (*e.i.*), which are a measurement of compositional change [39]. A 1 *e.i.* change is equivalent to an approximately 4-fold change in the ratio of two bacterial taxa and a 2 *e.i.* change is equivalent to an approximately 17-fold change in the ratio of two taxa (see "Methods" section for further

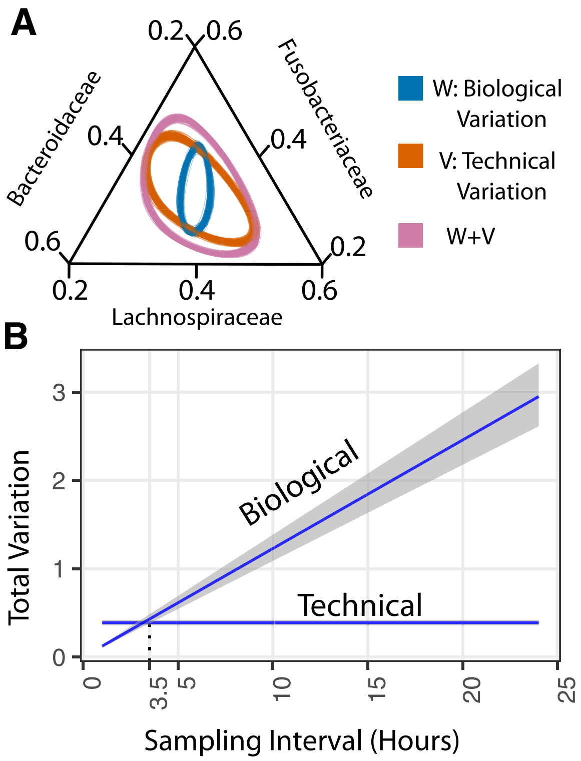

Fig. 3 Structure and magnitude of biological and technical variation. **a** Ternary plot showing the 95% probability regions of the logistic-normal distributions corresponding to *W*, *V*, and *W* + *V* along the Bacteroidaceae, Fusobacteriaceae, and Lachnospiraceae subcomposition. To visualize posterior uncertainty, 100 posterior samples of each of these quantities are plotted. **b** Mean and 95% credible interval for total biological (*Tr*(*W*)) and total technical (*Tr*(*V*)) variation as a function of sampling interval in hours

discussion). Multiple balances exhibited sub-daily dynamics (Additional files 7 and 8), most notably the ratio of bacteria from the phylum Bacteroidetes to bacteria from the phyla Proteobacteria and Fusobacteria (Fig. 4). The balance between Bacteroidetes and Proteobacteria/Fusobacteria appeared to fluctuate on timescales shorter than 1 day with an amplitude of approximately 1 *e.i.* (0.5–1.5, 95% credible interval; Fig. 4b). Balance dynamics did not correspond to recorded environmental or technical variations (e.g., media changes, identity of the researcher collecting the sample, sequencing batch number, *B. ovatus* supplementation, or the feed disruption of vessels 1 and 2) and did not display an exact 24-h periodicity (Additional file 9). Balance fluctuations were observed in all four replicate artificial gut models, but did not appear to be synchronized as would be expected if these dynamics were driven by a shared environmental factor (Additional file 8). We ultimately could not identify a technical or environmental cause of fluctuating balance dynamics on sub-daily timescales.

Exploratory analysis of biological variation

We next explored the biological variation captured in our model by investigating the pairwise variation between bacterial families in our dataset. Direct taxon-level analysis of the covariance matrix *W* is difficult though because the elements of this matrix are balances—not individual taxa. We therefore investigated the temporal variation of specific taxon pairs using a tool from compositional data analysis called the variation array [40]. A variation array represents the variance of the log-ratio between pairs of

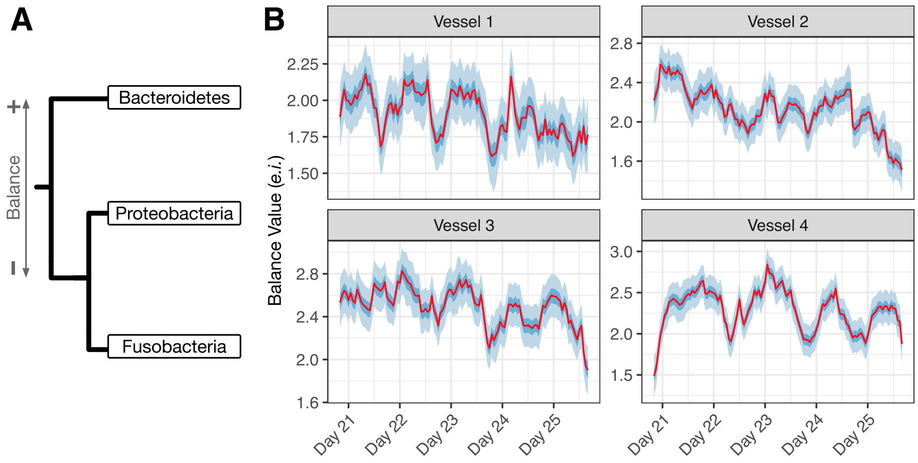

Fig. 4 Sub-daily microbiota dynamics within the artificial intestine. **a** An annotated phylogenetic tree defining PhILR balance between the phyla Fusobacteria, Proteobacteria, and Bacteroidetes. The balance reflects the scaled log-ratio of the geometric mean relative abundance of families in the Bacteroidetes phyla (numerator, +) to the geometric mean relative abundance of families in the Fusobacteria and Proteobacteria phyla (denominator, −) ("Methods" section). **b** Mean, 50%, and 95% credible interval of posterior distribution for θ (microbial dynamics in absence of technical variation) for the PhILR balance defined in (**a**). The full list of PhILR balances and corresponding posterior estimates are shown in Additional files 7 and 8

taxa ("Methods" section). When variance of a pairwise log-ratio is near zero, the two taxa positively covary; whereas when this variance is, high the two taxa exhibit either unlinked or exclusionary patterns [22, 41]. The resulting variation array (corresponding to *W*) revealed that most temporal variation was contained in log-ratios between four bacterial families: the Rikenellaceae, Synergistaceae, Enterobacteriaceae, and Fusobacteriaceae. By contrast, log-ratios between other bacterial families were approximately 1 order of magnitude smaller (Fig. 5a). Overall, we estimated that the Rikenellaceae, Synergistaceae, Enterobacteriaceae, and Fusobacteriaceae accounted for 72% (69–75%; CLR basis) of the total biological variation seen in the dataset (Additional files 10 and 11).

We noticed an inverse relationship between the biological variation of taxa to their relative abundance in the starting inoculum. We observed this relationship by fitting a linear regression model to each posterior sample in a CLR-transformed space (95% posterior credible interval for regression slope: − 0.26 to − 0.18; Additional file 12; "Methods" section). Thus, our analyses suggested that more variable taxa in the artificial gut tended to be rarer at the time of inoculation. This result was insensitive to perturbation of our prior assumptions (Additional file 13), and there was only a 3.8% posterior probability that the observed inverse relationship was an artifact of low abundance families being more numerous than high abundance ones ("Methods" section).

To better understand the patterns of variation we observed from the above analysis, we manually created three balances which together highlighted patterns that we believe explain the high variability of the Rikenellaceae, Synergistaceae, Enterobacteriaceae, and Fusobacteriaceae families (Fig. 5b–d). Together, these three manually curated balances explained 73% of the biological variation in the dataset (69–76%, 95% credible interval). This manual curation was informed by the inspection of microbial dynamics within the PhILR basis and a hierarchical cluster analysis which suggested that the Fusobacteriacae and Synergistaceae shared similar dynamic patterns and should therefore be grouped together for analysis ("Methods" section, Additional file 14). One highly variable balance, which represented the ratio of Rikenellaceae to the other nine bacterial families, exhibited a 3 *e.i.* decrease during days 11–13 in two artificial gut replicate vessels before recovering over the subsequent week (Fig. 5b). This decrease corresponded with the transient feed disruption in those replicate vessels. A second highly variable balance, the ratio between the Fusobacteriaceae and Synergistaceae to all other bacterial families, increased by approximately 8 *e.i.* between days 2 and 7 and the end of the experiment (Fig. 5c). The nadir of this balance coincided with an initial

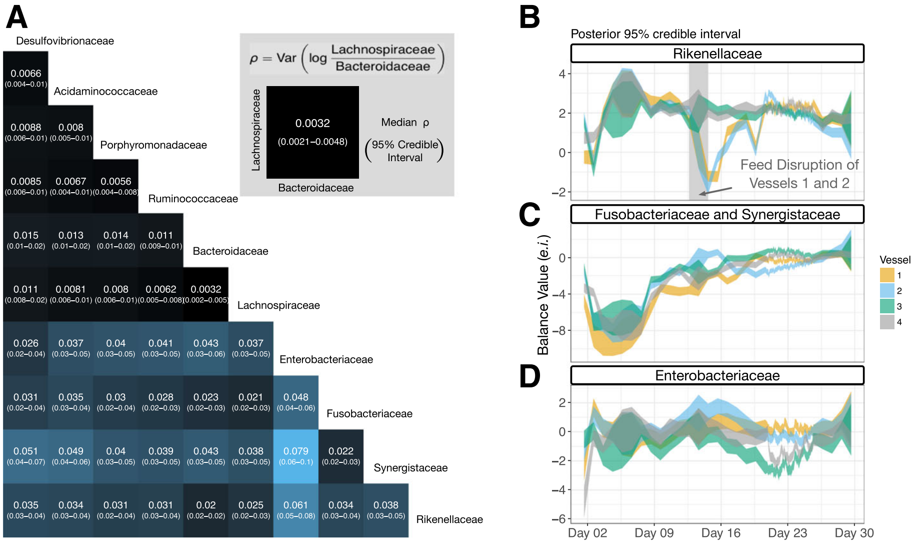

Fig. 5 The decomposition of biological variation among bacterial families. **a** Heatmap of posterior distribution of log-ratio variance (*ρ*) between pairs of bacterial families. Heatmap color is given by the median of the posterior distribution of *ρ*. Columns and rows refer to the bacteria in the numerator and denominator of the corresponding log-ratios respectively. A similar decomposition of technical variation is shown in Additional file 11. **b**–**d** The highest variance bacteria displayed distinct temporal patterns (95% posterior credible regions for *θ*). Hourly sampling results in smaller posterior credible intervals and more resolvable dynamics in the time around day 23. **b** The balance between the family Rikenellaceae versus all other bacterial families decreased substantially during the feed disruption of vessels 1 and 2. **c** The balance between the Fusobacteriaceae and Synergistaceae versus all other bacterial families slowly increased over the course of the study after an initial acclimation period. **d** The balance between the family Enterobacteriaceae versus all other bacterial families displayed fluctuating dynamics likely related to those shown in Fig. 4

microbiota adaptation phase that is repeatedly reported in artificial gut studies [1, 10]. A third balance, representing the ratio of Enterobacteriaceae to the other nine bacterial families, exhibited irregular sub-daily oscillatory behavior (Fig. 5d and Additional file 15). These balance dynamics are likely related to those observed earlier between the phyla Bacteriodetes and Fusobacteria/Proteobacteria (Fig. 4), given the membership of Enterobacteriaceae within the phylum Proteobacteria.

Discussion

Our findings in concert offer several insights that could be useful for the design and analysis of ex vivo studies of human gut microbiota. Our study suggests that technical variation introduced during sample processing can affect observations of microbiota dynamics not just in scale but also in the patterns of covariation among taxa. To date, knowledge has been limited on the effects of technical variation in ex vivo studies of human gut microbiota. Select exceptions include prior bioreactor studies that found technical variation to be limited and dominated by biological variation [10].

Our methods and results also have implications for the choice of sampling frequency in ex vivo artificial gut studies. While oversampling can waste resources, under-sampling can bias inference through a mechanism known as signal aliasing [28, 42, 43]. Here, to balance between resource-intensive exploration of sub-daily dynamics while still capturing longer multi-day dynamics, we made use of MALLARD and mixed rate sampling (e.g., sampling at both daily and hourly timescales). Our findings suggest that future studies interested in tracking the full range of ex vivo microbiota dynamics or that focus on community responses to rapidly changing external factors (e.g., single dose small molecule or prebiotic supplementation studies) should consider collecting sub-daily measurements. By contrast, studies interested solely in longer term changes (e.g., overall nutritional studies) may find daily sampling sufficient. Additionally, by combining replicate sampling with purposeful oversampling, we were able to determine an effective signal-to-noise ratio as a function of sampling frequency. This ratio could be used to estimate an upper-limit sampling frequency above which the benefits of increases samples are diminished by the relatively high levels of technical variation.

Another methodological approach we develop here for artificial gut studies is an exploratory technique for discovering temporal patterns among microbial taxa. As gut microbiota time-series often involve many taxa, methods of dimension reduction can aid data interpretation. Yet, many standard time-series tools for dimension reduction, such as dynamic principal component analysis, are not well suited for microbiota data in the form of relative taxa abundances. We overcome this challenge and identify distinct patterns in our dataset using a tool from compositional data analysis called the variation array [40]. Hierarchical cluster analysis and manual curation then allowed us to identify three balances (Fig. 5b–d), highlighting four bacterial families, that together accounted for over 70% of the variation in our dataset. These three balances revealed key dynamical patterns and aided in our interpretation of the forces acting on our artificial intestine.

Beyond methodology for designing ex vivo gut microbiota studies, our findings also have several implications for the underlying biology of these systems. First, we observed distinct differences in the replicability of dynamics between vessels at hourly compared to daily timescales. In keeping with previous reports, we found that replicate artificial guts display replicable dynamic patterns when analyzed on daily timescales [1]. Notably, even after a feed interruption, vessels 1 and 2 returned to a similar composition as the two continuously fed vessels after roughly 1 week and appeared to display similar overall dynamics thereafter (Fig. 5b and Additional file 7). Such observations suggest that these systems follow distinct and potentially predictable trajectories on the timescales of days. Conversely, dynamics appeared less synchronized at hourly timescales (Additional file 8). For example, balances involving the family Enterobacteriaceae (e.g., balance n12), which demonstrate oscillatory patterns in multiple replicate vessels, do not appear in phase with one another. Together, these observations suggest a conceptual model for our artificial gut systems in which strong deterministic forces such as resource availability drive predictable community dynamics on longer timescales while other, potentially more subtle and variable ecological forces, drive temporal variation on shorter timescales.

Another implication of our results to the biology of ex vivo gut microbiota involves sub-daily dynamics of bacterial families such as the Enterobacteriaceae. We acknowledge that despite our efforts to keep culture conditions constant in the artificial gut, we cannot exclude the possibility that these dynamics were caused by fluctuating and unmeasured aspects of our artificial gut setup (e.g., room temperature, ambient humidity, room lighting). Although, we consider it unlikely that such external driving forces would cause dynamics that were unsynchronized between replicate vessels (Additional file 9). Moreover, rhythmic dynamics within microbial communities have previously been observed or predicted in host-free environments [44–46], and bacteria have been speculated to harbor circadian clocks [47]. The dynamics we observed may have even underestimated the rate at which microbiota varied in our system as genomic material likely remains measurable for a period after cell death and our analysis at the family level could smooth dynamics at finer taxonomic scales. Thus, our results support hypotheses that dynamical properties inherent to microbiota could contribute to sub-daily oscillations observed in the mammalian gut [48].

We also found that rare bacterial taxa tended to harbor the greatest variation over time. Such variation may reflect a general property of artificial gut systems, in which dormant enteric microbes take advantage of differences between ex vivo and in vivo environments to expand [49, 50]. In particular, we found that in all four artificial gut vessels, the Fusobacteriaceae and the Synergistaceae underwent an approximately 4 *e.i.* drop in their relative abundance upon transplantation from the in vivo to ex vivo environment and then a subsequent increase to approximately 4 *e.i.* greater than their initial levels (Fig. 5c). These patterns may reflect a form of ecological succession in which bacterial families are well suited for the ex vivo environment, but are initially outcompeted by faster growing microbes. Such succession is well described in both in vivo and environmental ecosystems [51–53], and supports the general hypothesis that select theoretical frameworks from the field of environmental microbial ecology are also applicable to host-associated microbiota [54].

Still, insights from our study here into the methodology and underlying biology of artificial gut experiments have several limitations. First, some sources of technical variation, such as systematic biases introduced due to separation of samples into batches for DNA extraction, PCR, or sequencing, may have exploitable and reproducible structure that could be controlled for during modeling thus improving the resolution of longitudinal microbiome studies. Here, we have chosen to treat these sources of variation as random *en batch* and instead leave such extensions of our analysis for future work. We note however, that mixed effects models, which have proven to be a powerful method of accounting for such systematic biases or batch effects, and their dynamic extensions both represent special cases of dynamic linear models [38]. We therefore believe MALLARD will prove useful in modeling and controlling for such batch effects. Second, computational limitations restricted our analysis to relatively few dimensions. Due to this limitation, we analyzed only the ten most abundant bacterial families, which in turn could have masked dynamics occurring at finer taxonomic scales [28]. Two future refinements may improve computational efficiency: univariate filtering methods for multivariate time-series have been developed and would reduce the dimensionality of matrix manipulations used in MALLARD [55], and emulation methods that iteratively refine probabilistic models have been developed for high-dimensional count data and show promise [56]. Third, we chose to collect all technical replicates from the same time-point and relied on randomizing both longitudinal and replicate samples into batches to ensure that the replicate samples faithfully represented the technical variation profile in our study. As an alternative, we could have instead spread technical replicate samples out over the course of the experiment, collecting duplicate or triplicate samples at regular intervals. As the MALLARD framework enables technical replicate samples to be collected with any distribution within the study, distributed replicate samples may have estimated microbial dynamics with differing precision.

Despite our study's limitations though, we believe that our methodological insights are useful for the design of artificial gut experiments and, more broadly, designing in vivo longitudinal studies of host-associated microbial communities. Longitudinal studies in humans have provided unique insights into disease and therapy [28, 43], including optimal antibiotic treatment regimens [57], mechanisms underlying disease and recovery from acute secretory diarrhea [58] or intestinal cleanout [59], as well as identifying vaginal microbiota signatures associated with preterm birth [60]. Aspects of our modeling framework here could be applied to future longitudinal analyses in humans. Specifically, attention has been given recently to issues arising in the analysis of relative microbiome data [22–27, 41, 61, 62] and there is growing awareness of how count variability influences microbiota surveys [20, 21, 63]. Missing data are also a common challenge in longitudinal microbiome analyses, particularly when the temporal evolution of observed data is modeled. MALLARD is unique in working at the intersection of compositional analysis, count variability, and linear longitudinal models that can account for technical variation in datasets with many missing observations [64–67].

Aspects of our artificial gut sampling design could also be applied to longitudinal microbiota studies in humans. Replicate sampling could be used to quantify the effects of technical variation for in vivo microbiota time-series. Moreover, choosing an appropriate sampling frequency remains an outstanding challenge in translational microbiome studies [68]. Pilot time-series experiments that intentionally sample in vivo systems more frequently than expected microbiota dynamics (i.e., oversampling) could be used to compute tradeoffs between technical and biological variation at higher sampling frequencies as we do here; aiming to sample at frequencies where technical variation does not exceed biological variation could help guide economical use of laboratory resources. Indeed, potentially oversampled longitudinal datasets for human gut microbiota already exist (albeit with limited replicate technical sampling) [69, 70]. Ultimately, by improving the design and analysis of microbial community dynamics in longitudinal in vivo *studies*, an improved understanding of the role of microbiota in human health and disease can be achieved.

Conclusions

Here, we created a modeling framework that allowed us to partition a densely sampled artificial human gut time-series into components associated with technical and biological sources of variation. Our results demonstrate that technical variation from sample processing can influence ex vivo microbiota dynamics, accounting for 76% of community variation on hourly timescales. Still, we observed evidence for bona fide microbiota dynamics on sub-daily timescales. Our integrated analyses also resulted in approaches for characterizing microbiota variation over time. By investigating the distribution of biological variation among taxa, we identified three distinct dynamic patterns that accounted for over 70% of the variation within our dataset. Together, our results contribute to our understanding of the dynamics of ex vivo artificial gut systems, as well as the design and analysis of future longitudinal microbiota studies.

Methods

Artificial intestine experiments

Collection and preparation of fecal inoculate

A fresh fecal sample was obtained from a healthy volunteer who provided written informed consent (Duke Health IRB Pro00049498). The sample was stored and prepared for inoculation into artificial gut systems within an anaerobic chamber (Coy). The fecal sample was weighed into 50 ml conical tubes, approximately 5 g per tube, and then pre-reduced MGM (McDonald Gut Media; [1]) was used to fill the tube. The fecal matter was homogenized briefly using a benchtop vortex and then centrifuged 10 min at a speed of 175×*g*. The supernatant was decanted into syringes for inoculation.

Artificial gut preparation

A four-vessel continuous flow artificial gut system (Multifors 2, Infors) was used to culture gut microbiota seeded from human stool. Vessels were sterilized and prepared with 300 ml of fresh MGM. Inoculation of reactors used 100 ml of fecal inoculate resulting in an overall volume of 400 ml. The media feed was started 24 h after inoculation at a constant rate of 400 ml per day to emulate the 24-h average passage time in the human gut. Media was changed 16 times throughout the course of the experiment, each time media was prepared fresh. On day 13, it was discovered that the feed line to vessels 1 and 2 was blocked; this blockage could have occurred any time after day 11.

In addition to media feed rate, oxygen, pH, temperature, and stir rate were controlled by the IRIS software (v6, Infors). Oxygen concentration in the vessels was kept below 1% via positive nitrogen pressure at 1 LPM. Oxygen concentration was measured continuously using Hamilton VisiFerm DO Arc 225 probes. The oxygen probes were calibrated using a two-point calibration performed with nitrogen flowing at 1 LPM as the zero-point calibration and room air flowing at 1 LPM as the 100% calibration point. pH was maintained between 6.9 and 7.1 using a 1 N HCl solution and a 1 N H3PO4 solution. pH was measured continuously with Hamilton EasyFerm Plus PH ARC 225 probes. The pH probes were calibrated with a two-point calibration with standardized pH buffers at 4.00 ± 0.1 and 10.00 ± 0.1 (BDH). Vessels were maintained at 37 °C via the Infors' onboard temperature control system. Vessels were continuously stirred at 100 rpm using magnetic impeller stir-shafts.

Bacteroides ovatus delivery

To study community dynamics in response to changes in a single bacterial taxa, we supplemented replicate vessels 1 and 4 with 2 ml of isolated *B. ovatus* at 10^{10} cells/ml (estimated by optical density) suspended in anaerobic blood heart infusion (BHI) agar, and vessels 2 and 3 with 2 ml of anaerobic BHI as a control on day 23. No evidence of *B. ovatus* increase was detected via community composition. Longitudinal modeling suggest effects of adding delivery media were limited to a transient 0.5 *e.i.* shift in the balance between the families Bacteroidaceae and Porphyromonadaceae (Additional file 8).

Sampling

For each time-point sampling of the four replicate vessels was done as follows. Prior to sampling, sampling ports were cleared with a sterile syringe and wiped clean with ethanol. Samples were collected in the following order: vessel 1, vessel 2, vessel 3, and vessel 4. Sampling consisted of the collection of 3 ml of active artificial gut culture via sterile syringe and then immediate storage in labeled and barcoded cryovials in a −80 °C fridge. The full list of sampled time-points including daily, hourly, and technical replicate samples is shown in Additional file 4.

DNA extraction, PCR amplification, and sequencing

For all samples, 16S rRNA gene amplicon sequencing was performed using custom barcoded primers targeting the V4 region of the gene [71] and published protocols [71–73]. All samples were randomized into 7 sets of 96 for all sample preparation steps. Extractions were performed using the MoBio PowerMag Soil DNA Isolation kit (p/n 27100-4-EP) adapted for use without robotic automation. Due to the large number of samples, samples were split randomly into 2 pools of 336 samples for sequencing to ensure adequate read depth per sample. The final DNA concentration for pool 1 was 35.5 ng/μl and for pool 2 was 34.5 ng/μl as assessed by Picogreen assay. Both sequencing runs were standardized to 10 nM and sequenced using an Illumina MiSeq with paired end 250 bp reads using the V3 chemistry kits at the Duke Molecular Physiology Institute core facilities.

Identifying sequence variants

We used DADA2 to identify and quantify sequence variants in our dataset [74]. To prepare data for denoising with DADA2, 16S rRNA primer sequences were trimmed from paired sequencing reads using Trimmomatic v0.36 without quality filtering [75]. Barcodes corresponding to reads that were dropped during trimming were removed using a custom python script. Reads were demultiplexed without quality filtering using python scripts provided with Qiime v1.9 [76]. Bases between positions 10 and 180 were retained for the forward reads and between positions 10 and 140 were retained for the reverse reads based on visual inspection of quality profiles. This trimming, as well as minimal quality filtering, of the demultiplexed reads was performed using the function *fastqPairedFilter* provided with the *dada2* R package (v1.1.6). Sequence variants were inferred by *dada2* independently for the forward and reverse reads of each of the two sequencing runs using error profiles learned from a random subset of 40 samples from each sequencing run. Forward and reverse reads were merged for each of the two sequencing runs. Bimeras were removed using the function *removeBimeraDenovo* with *tableMethod* set to "consensus." Finally, the two sequencing runs were merged together into a single count table.

Taxonomy assignment

Initially, taxonomy was assigned to sequence variants using a Naive Bayes classifier [77] trained using version 123 of the Silva database [78]. Initial taxonomic assignments were then augmented by searching for exact nucleotide matches to the Silva database. This resulted in 96% of sequence variants being classified at the family level, 85% at the genus level, and 15% at the species level.

Data preparation for modeling

After investigating the distribution of sample sequencing depth, we chose to retain only samples with more than 5000 read counts to remove outlying samples that may have been subject to library preparation or sequencing artifacts. This step retained 99.8% of total sequence variant counts. For computational tractability and to ensure a maximal number of retained sequence variant counts, we preformed our analysis at the family level and we retained only those families that were present with at least three counts in more than 90% of samples. While these filters yielded only ten bacterial families, they represented 97.7% of total sequence variant counts.

Construction of phylogenetic sequential binary partition

To use the PhILR transform [22], we manually created a sequential binary partition based on the phylogenetic relationships between the bacterial families in our dataset. This manual partition was created in accord with the phylogenetic relationships between bacterial families specified in Rajilic-Stojanovic M and de Vos WM [79]. The resulting sequential binary partition is given in Additional file 7.

The MALLARD framework

To accommodate time-series with replicate observations (as depicted in Fig. 1c), we refer to samples by sample index $k \in \{1, \ldots, K\}$ rather than time index $t \in \{1, \ldots, T\}$ such that $K \geq T$. Further, let the function ϕ provide a mapping between the sample index and the time index such that $\phi(k) = t$. As in standard time-series notation, we assume that these sample indices are temporally ordered such that $\phi(k) \leq \phi(k+1)$ for all k. With this notation, denote a typical longitudinal microbiome dataset as a matrix Y with element Y_{kd} representing the number of counts measured for taxa $d \in \{1, \ldots, D\}$ in sample k. We also denote the total sequence counts (sequencing depth) attributed to the kth sample as $n_k = \sum_{d=1}^{D} Y_{kd}$. In what follows, we will introduce the MALLARD framework in full generality and then demonstrate how features such as missing observations, multiple concurrent time-series, and replicate observations can be handled. After introducing the MALLARD framework, we will then introduce the MALLARD model used to analyze the artificial gut dataset in this work.

MALLARD overview

To introduce MALLARD, we first present the entire MALLARD framework and then discuss the motivation and intuition behind each component individually. MALLARD can be written as the following hierarchical model

$$Y_k \sim \text{Multinomial}(\pi_k, n_k) \tag{1}$$

$$\pi_k = \text{ILR}^{-1}(\eta_k) \tag{2}$$

$$\eta_k = F_k^{'\theta_k + \nu_k, \quad \nu_k \sim N(0, V_k)} \quad (3)$$

$$\theta_k = G_k \theta_{k-1} + w_k, \quad w_k \sim N(0, W_k) \tag{4}$$

$$\theta_0 \sim N(m_0, C_0) \tag{5}$$

$$V_1, \ldots, V_K, W_1, \ldots, W_k \sim p(\zeta) \tag{6}$$

where ILR^{-1} represents the inverse Isometric Log-ratio transform [22, 80] and $p(\zeta)$ is a placeholder for a variety of potential priors.

Equation 1 models the process of sequence counting. High-throughput DNA sequencing does not measure the total number of target DNA transcripts in a biological system but only a random subset of this total. The size of this subset is represented by the sequencing depth of a sample. This feature of DNA sequencing leads to a competition to be counted between transcripts in which more abundant transcripts can exclude observations of less abundant transcripts. To capture this behavior, MALLARD models DNA sequencing as a multinomial counting process where a D-dimensional vector of counts from each sample $Y_k = (Y_{k1}, \ldots, Y_{kD})$ provides a noisy measurement of the relative abundance of each taxa in the sequencing library ($\pi_k = (\pi_{k1}, \ldots, \pi_{kD})$ such that $\sum_d \pi_{krd} = 1$).

While the multinomial component of our model accounts for uncertainty due to the counting process underlying DNA sequencing, the multinomial component alone is insufficient to account for the other sources of technical and biological variation in longitudinal microbiome studies. To allow for this type of extra-multinomial variability, MALLARD treats the multinomial parameters (π) as distributed logistic-normal. We chose the logistic-normal distribution for two reasons. First, the logistic-normal has greater flexibility than the more common Dirichlet distribution allowing for both positive and negative covariation between taxa [40, 81]. Second, the logistic-normal distribution represents the central limiting distribution over the space of multinomial parameters assuming multiplicative errors [37, 82]. We chose to model with a multiplicative error structure due to the multiplicative nature of bacterial growth and DNA amplification.

Although the logistic-normal can be difficult to work with in terms of relative abundances (π_k), under the isometric log-ratio transform (ILR) [22, 80], the logistic-normal simplifies to a multivariate normal distribution on the transformed parameters η_k [37]. Additionally, while many log-ratio methods suffer from numerical problems when trying to take the log or ratio of zeros, by modeling the process of sequence counting and treating the relative abundances π_k as an unknown parameter to be inferred, zeros are handled by MALLARD without the need for pseudo-counts or multiplicative replacement schemes.

To model the relationship between samples, we consider that the specific multivariate normal model relating the parameters $\eta_1, \ldots, \eta_K$ is a class of linear Gaussian state-space models (Eqs. 3 and 4) often referred to as dynamic linear models (DLMs) [38]. The DLM can be thought of as modeling an unobserved system parameterized by a p-dimensional vector θ_k that evolves through time based on a deterministic linear model ($\theta_k = G_k\theta_{k-1}$) where G_k is a $p \times p$ matrix of covariates. Additionally, the state evolution has a random component which is modeled as mean zero multivariate normal random perturbations w_k with covariance W_k. Furthermore, the DLM models that the state parameters θ_k is translated into the parameters η_k through a similar, though independent, deterministic linear model with added multivariate normal variation ($\eta_k = F'_k\theta_k + \nu_k$) where η_k and ν_k are $D-1$ dimensional vectors and F_k is a $p \times (D-1)$ matrix of covariates. In this work, we use this later component to model technical variation. Flexibility in the specification of F_k, G_k, W_k, and V_k, allows the MALLARD framework to encompass many different types of models as special cases including dynamic and static mixed effects or factor models, models with seasonal or polynomial trends, and even dynamic regression models [38]. A thorough review of the use of dynamic linear models, and MALLARD models by extension, is given in West M and Harrison J [38].

We specified two types of prior beliefs for parameters in the MALLARD likelihood model, the priors over the state vectors θ and the priors over the covariance components (V_1, …, V_K and W_1, …, W_k). For the state vector component, Eq. 5 enables prior knowledge regarding the value of the state vector 1 time-step prior to the first observation to be encoded as a normal distribution with mean m_0 and covariance C_0. In contrast to the state vector component, we allow greater flexibility in prior form regarding V_1, …, V_K and W_1, …, W_k. Classically, priors for these terms are based on the inverse Wishart distribution; however, in this work, we instead modeled using a decomposition of these covariance matrices for numerical stability (see subsection "MALLARD model specifications for analysis of the artificial gut dataset"). Finally, flexibility in the specification of these covariance terms allows MALLARD to model time-varying covariance structures as is commonly done in stochastic volatility models [38].

Handling multiple concurrent time-series with MALLARD

Many longitudinal microbiome studies involve multiple concurrent time-series either from different individuals or, as in this study, different artificial gut vessels. Denoting each of R concurrent time-series by an index $r \in \{1, \ldots, R\}$, we introduce a state expansion as one method for modeling concurrent time-series with MALLARD. First, let $\theta_k = [\theta_k^{(1)\prime}, \ldots, \theta_k^{(R)\prime}]'$ where $\theta_k^{(r)}$ denotes a p-vector of state parameters for time-series r at sampling point k. Similarly, we can define $\eta_k = [\eta_k^{(1)\prime}, \ldots, \eta_k^{(R)\prime}]'$. Using this state expansion, we can write the MALLARD model for multiple concurrent time-series just as in Eqs. 1, 2, 3, 4, 5 and 6 but with the replacement of Eqs. 1 and 2 by

$$Y_k^{(r)} \sim \text{Multinomial}\left(\pi_k^{(r)}, n_k\right)$$

$$\pi_k^{(r)} = \text{ILR}^{-1}\left(\eta_k^{(r)}\right).$$

With this specification, θ_k is a vector of length Rp and η_k is a vector of length $R(D-1)$. This state expansion also enables flexible modeling of covariation between and within concurrent time-series through an induced expansion of components V_1, …, V_K, W_1, …, W_K. For example, consider that if θ_k denotes an Rp vector of the combined states of R concurrent time-series, then the off-diagonal blocks of W_k model the covariation between concurrent time-series while the diagonal blocks represent the covariation within each time-series.

Posterior inference

The inference goal for the above model is to sample from the posterior distribution $p(\theta, \eta, V, W| D, B)$ where $D = \{Y, F, G\}$, $B = \{m_0, C_0, \zeta\}$ and we have denoted sets of parameters by dropping the associated subscripts (e.g., V represents the set $\{V_1, …, V_K\}$). The inference method we describe is based on the conditional decomposition of this posterior as

$$p(\theta, \eta, V, W|D, B) = p(\theta|\eta, V, W, D, B)p(\eta, V, W|D, B).$$

In particular, we show that the density $p(\eta, V, W| D)$ can be efficiently computed and sampled using MCMC in combination with the Kalman filter to marginalize over the state parameters θ and that, conditional on having samples from the conditional posterior $p(\eta, V, W| D)$, we can efficiently sample from $p(\theta| \eta, V, W, D)$ using recursive samplers (i.e., the Kalman smoother or the backwards sampling algorithm; see below). Importantly, this approach removes all state parameters θ from the MCMC and instead samples them directly from the conditional posterior using more efficient recursive samplers. This is in contrast to prior work with multinomial conditionally Gaussian dynamic linear models which used a Metropolis-within-Gibbs sampling scheme but did not make use of such marginalization [34]. Additionally, in contrast to prior work, our sampling scheme allows us to use MCMC methods that incorporate posterior adaptation further improving MCMC efficiency.

The term $p(\eta, V, W| D)$ can be efficiently computed using the Kalman filter as follows. Using Bayes rule and the conditional independence relationships $Y \perp B$, V, W, F, G | η, $\eta \perp \zeta$ | V, W, and V, $W \perp m_0$, C_0, F, G | ζ, we can write

$$p(\eta, V, W|D, B) \propto p(Y|\eta)p(\eta|V, W, F, G, m_0, C_0)p(V, W|\zeta).$$

Importantly, this first term is easily calculable and is given by $p(Y|\eta) = \prod_{k=1}^{K} \text{Multinomial}(y_k|\pi_k)$. The third term is also easily calculable as it is the density of the prior for the covariance matrices evaluated at V and W. Letting $\Xi = \{V, W, F, G, m_0, C_0\}$ for notational convenience and by noting the first-order Markov structure of the DLM, we can simplify the second term as

$$p(\eta|\Xi) = p(\eta_1|\Xi)\prod\nolimits_{k=2}^{K} p(\eta_k|\eta_{k-1}, …, \eta_1, \Xi).$$

This relation is directly calculable as the product of 1-step ahead predictive densities in the Kalman filter (see Additional file 16). As the above represents an efficient method of calculating the density of $p(\eta, V, W| D, B)$ up to a proportionality constant, sampling from this density can be accomplished via MCMC. In this work, we choose to use adaptive Hamiltonian Markov Chain Monte Carlo (HMCMC) provided by the Stan modeling language to simulate from the density $p(\eta, V, W| D, B)$ [83, 84]. Finally, the use of the Kalman filter also enables simple and efficient handling of missing observations (see Additional file 16).

Given samples from $p(\eta, V, W| D, B)$, we now describe an efficient method for sampling from the conditional posterior distribution $p(\theta| \eta, V, W, D, B)$. Using the conditional independence relationship $\theta \perp Y$, ζ | H, V, W, we can simplify this conditional density to $p(\theta| H, V, W, F, G, m_0, C_0)$ which we show in Additional file 16 can be sampled from directly using either the Kalman smoother or the backwards sampling algorithm (see Additional file 16).

MALLARD model specifications for analysis of the artificial gut dataset

We describe the specific MALLARD model used in this work as a special case of the general MALLARD framework. We introduce these simplifications in four parts relating to model structure, prior specifications, handling of missing data, and posterior inference. Regarding model structure, here we analyzed four concurrent time-series (from four artificial gut vessels) using the state expansion mentioned above. As these vessels were physically isolated from each other, we modeled the four vessels as being conditionally independent (given shared covariance components) of each other such that we could rewrite Eqs. 3 and 4 as

$$\eta_k = [I_R \otimes F_k^{'] \theta_k + v_k, \quad v_k \sim N(0,[I_R \otimes V_k])}$$

$$\theta_k = [I_R \otimes G_k]\theta_{k-1} + w_k, \quad w_k \sim N(0, [I_R \otimes W_k]).$$

where $\otimes$ represents the Kronecker product. This specification also had computational advantages as $p(\eta| \Xi)$ could then be written as $\prod_{r=1}^{R} p(\eta^{(r)}|\Xi^{(r)})$ and therefore sampled using R independent p-dimensional Kalman filters rather than a single Rp dimensional Kalman filter over the expanded state. Additionally, we model the state of each vessel as being identified with an unobserved microbial composition before the addition of confounding technical variation. With this specification, the dimension of the state space p is equal to $4(D - 1)$. As our primary interest was in retrospective inference, we chose to model the state evolution as a simple random walk (or constant level) such that $F'_k = G_k = I_{(D-1)\times(D-1)}$. Furthermore, as we randomized our samples through all steps of preprocessing, we assumed that the technical noise profile of each sample is identical such that we could write $V_k = V$. Finally, to simplify the model and mitigate overfitting, we assumed that there was a single biological variation profile that was identical across all non-equal time-points such that

$$W_k = \begin{cases} W, \text{if } \phi(k) \neq \phi(k-1) \\ 0, \text{if } \phi(k) = \phi(k-1). \end{cases} \tag{7}$$

The conditional relationship in Eq. (7) allowed us to account for the lack of temporal evolution between replicate samples. Importantly, while this assumption of a single biological variation profile does not directly model complex dynamic patters such as linear trend or oscillation, it is flexible enough to infer such patterns if they are strongly supported by the data (e.g., Figs. 4 and 5). Finally, we chose the phylogenetic basis defined in Silverman JD, Washburne AD, Mukherjee S, and David LA [22], without tip or branch weights as a default basis for all posterior computations. A table showing the total number of modeled parameters induced by these choices is given in Additional file 17.

Regarding the prior specifications, we specified two types of prior beliefs for parameters in the MALLARD likelihood model, the priors over the state vectors θ and the priors over the covariance components V and W. For the state vector component, we specified a distribution over the true composition of each of the replicate vessels 1-h prior to the first observed sample such that for $r \in \{1, \ldots, 4\}$

$$\theta_0^{(r)} \sim N(m_0, C_0)$$

with $m_0 = 0_{(D-1)}$ and $C_0 = 25 \cdot I_{(D-1)\times(D-1)}$. With this specification, there is an approximately 66% probability (1 standard deviation) that no single taxa was greater than approximately 200 times more abundant than the geometric mean of the remaining taxa. As we are only modeling bacterial families that are present with at least three counts in at least 90% of samples, we believed that such concentration of our prior about zero was warranted.

To quantify our uncertainty in the covariance components of our model, we also specified a prior distribution for V and W. While this is most commonly done using inverse Wishart distributions due to their conjugacy with the multivariate normal distribution, here we chose to use a reparameterization of the covariance matricies V and W with non-conjugate priors to improve numerical stability. A covariance matrix $\Sigma \in \{V, W\}$ can be parameterized as

$$\Sigma = \sigma \Lambda \sigma'$$

where Λ denotes the correlation matrix corresponding to Σ, and σ represents the diagonal matrix with positive diagonal entries $(\sigma_1, \ldots, \sigma_{D-1})$ which dictate the scale of the covariance matrix. Thus, for the covariance matrices V and W**,** we specified the components $(\sigma_1^V, \ldots, \sigma_{D-1}^V)$, Λ^V, $(\sigma_1^W, \ldots, \sigma_{D-1}^W)$, and Λ^W respectively. As the representation of Λ^V and Λ^W depends on the chosen ILR basis and we had no prior knowledge regarding how the technical and biological variation will decompose in our chosen basis, we chose to use a uniform prior over the space of symmetric positive semi-definite matrices such that

$$\Lambda^V \sim \text{LKJ}\left(\zeta^V\right)$$

$$\Lambda^W \sim \text{LKJ}\left(\zeta^W\right)$$

with $\zeta^V = \zeta^W = 1$ and where LKJ represents copula-based distribution over correlation matrices introduced by Lewandowski D, Kurowicka D and Joe H [85]. With regards to the terms $(\sigma_1^V, \ldots, \sigma_{D-1}^V)$ and $(\sigma_1^W, \ldots, \sigma_{D-1}^W)$, we chose independent log-normal priors to ensure that these terms were strictly positive and to allow parameterization of uncertainty with respect to multiplicative fold-changes. In particular, for each $i \in \{1, \ldots, D - 1\}$, we specified

$$\sigma_i^V \sim \text{Log–Normal}\left(\xi_i^V, \tau_i^V\right)$$

$$\sigma_i^W \sim \text{Log–Normal}\left(\xi_i^W, \tau_i^W\right)$$

with $\xi_i^V = 1$, $\tau_i^V = 2$, $\xi_i^W = 0$, and $\tau_i^W = 2$. This specification reflects a 95% probability that the ratio of total technical ($\sum_i [\sigma_i^V]^2$) to total biological variation ($\sum_i [\sigma_i^W]^2$) is between 10^{-2} and 10^2 with an expected value of approximately 10.

In this work, we chose to distinguish two types of missing observations based on whether they were to be imputed or marginalized over during posterior sampling. To ensure each of the four replicate vessel time-series remains synchronized, we considered any sample point that is missing from one vessel but observed in at least one other vessel as a missing value to be imputed by augmenting HMCMC simulations with a corresponding vector $\eta_k^{(r)}$. Conversely, we considered sample points that are present in none of the four replicate vessels as missing values that were marginalized over using the Kalman filter (see Additional file 16). We padded periods of daily sampling with missing values so that the entire dataset could be analyzed at an hourly base interval.

For this model, posterior inference was performed using the No-U-Turn Sampler (NUTS; a variant of HMCMC) provided in the Stan modeling language [83, 84] using 4 chains run in parallel each with 1000 transitions for warmup and adaptation and 1000 iterations collected as posterior samples. Preliminary results suggested that the time required for the NUTS sampler to converge to the typical set could be quite sensitive to entirely random parameter initializations. To address this computational limitation, the following parameters were manually specified: η was set by first adding a pseudo-count of 0.65 to each observed counts and then normalizing the counts of each sample to sum to 1, Λ^V and Λ^W were each initialized to the $p \times p$ dimensional identity matrix, and all other parameters were randomly initialized. To mitigate the potential bias introduced by fixing these parameters during initialization, approximate posterior samples from the model were first drawn using a variational algorithm [86] and then four randomly selecting posterior samples from this approximate posterior sample were used to initialize four parallel HMCMC chains. Convergence of the chains was determined both by manual inspection of sampler trace plots and through inspection of the split $\hat{R}$ statistic [87, 88]. All sampled parameters had an $\hat{R}$ value less than 1.01. Posterior intervals for all calculations derived from directly sampled quantities were calculated by preforming the necessary computations on each posterior sample independently and then summarizing the resulting distribution over calculated quantities.

Comparing the correlation structure of technical and biological variation

To quantitatively compare the correlation structure of technical and biological variation, we analyzed the probability that posterior samples of the correlation matrices corresponding to V and W came from the same distribution using a permutation scheme and a distance metric on the space of square symmetric positive semi-definite matricies. To isolate our inferences to only involve the correlation structure of V and W and not the magnitude of the variation, we transformed sampled covariance matrices into corresponding correlation matrices which we denote V^c and W^c. We took as a measure of distance between two correlation matrices S_1 and S_2 the Riemannian metric on the space of square symmetric positive definite matrices defined by

$$d_R(S_1, S_2) = \left\| \log\left(S_1^{-\frac{1}{2}} S_2 S_1^{-\frac{1}{2}}\right) \right\|$$

as described in [89] and calculated using the function *distcov* with option "*Riemannian*" in the R package *shapes* [90]. Using this distance metric, we calculated a distance matrix between 500 posterior samples of V^c and W^c each. Let D represent the resultant 1000 by 1000 distance matrix such that D_{ij} represents the distance between correlation matrix i and correlation matrix j. Let l represent the vector of labels of the correlation matrices such that $l_i \in \{V^c, W^c\}$ for $i \in \{1, \ldots, 1000\}$. We define a statistic δ as the ratio of the within to between group distances in D as

$$\delta = \frac{\sum_{l_i \in W^c, l_j \in W^c} D_{ij} + \sum_{l_i \in V^c, l_j \in V^c} D_{ij}}{2\sum_{l_i \in W^c, l_j \in V^c} D_{ij}}.$$

Thus, low values of δ indicate that most of the distance between the correlation matrices is attributable to intergroup differences in correlation structure, whereas larger values suggest that most of the pairwise distances come from differences within groups. We built a distribution of δ under a model in which the sampled correlation matrices all came from the same distribution by permuting the labels l and recomputing δ 1000 times. The probability that posterior samples of V^c and W^c came from the same distribution was calculated by calculating the probability of a test statistic more extreme than that which we observed under the created permutation distribution.

Computation of total technical and biological variation

Denoting the trace of a matrix as $Tr(\cdot)$, total biological and total technical variation were calculated as $Tr(W)$ and $Tr(V)$ respectively. The proportion of variation attributable to biological sources was calculated as $Tr(W)/(Tr(V) + Tr(W))$**.** Based on the linear systems assumption underlying our model, we choose to calculate the total biological variation as a function of sampling interval L as $L \cdot Tr(W)$ (Fig. 3b). Technical variation does not vary with sampling interval and was therefore held constant as a function of sampling interval.

Changing representations of posterior state estimates and covariance matrices

While we chose the PhILR basis [22] for all posterior computations as well as initial analysis (Fig. 4 and Additional files 7 and 8), we also made use of the isometric properties of the ILR transform to represent our posterior state estimates and our posterior samples of covariance matrices W and V with respect to alternative coordinates. In what follows, we denote vectors or matrices represented with respect to something other than the PhILR coordinates with an associated superscript. We used the inverse PhILR transform to represent any vector $\boldsymbol{x} = (x_1, \ldots, x_{D-1})$ in terms of raw relative abundances such that $\boldsymbol{x}^* = \mathrm{PhILR}^{-1}(\boldsymbol{x})$ [22]. Any other log-ratio quantities could then be calculated from $\boldsymbol{x}^*$. Transforming a covariance matrix Σ between representations was done by making use of the following identities [37]

$$\Sigma^{ILR} = \Psi^{ILR}\left(\Psi'\Sigma\Psi\right)\Psi^{(ILR)'} \quad (7)$$

$$\Sigma^{CLR} = \Psi'\Sigma\Psi \quad (8)$$

where Ψ denotes the contrast matrix of the PhILR transform [22] and Ψ^{ILR} represents the contrast matrix of an arbitrary ILR transform. In addition, we made use of the variation matrix representation of a covariance matrix in which the covariance matrix is represented as a matrix describing the variance of all pairwise log-ratios [40]. Letting ρ_{ij} represent the entry in the *i*th row and *j*th column of Σ^{CLR}, we compute the variation matrix T element-wise as [37]

$$t_{ij} = \rho_{ii} + \rho_{jj} - 2\rho_{ij}. \quad (9)$$

Hierarchical clustering of variation matrix

To develop a sparse representation of the principal directions of biological variation in our dataset, we make use of an algorithm for determining a sequence of orthonormal balances that maximize successively the explained variance in a dataset (principal balances) [91]. As T^W, the variation matrix calculated from W is proportional to the Aitchison distance between the bacterial families in our dataset [91, 92], applying Ward clustering to the matrix T^W results in an approximate solution to the problem of determining principal balances [91, 93]. For each posterior sample of W, we calculated T^Wusing Eqs. (8) and (9). We then calculated a sequential binary partition from each posterior sample of T^W using Ward clustering [94]. To summarize this posterior sample of sequential binary partitions, we computed the majority-rule consensus tree and the frequency with which a given bipartition occurred in our posterior sample using the functions *consensus* and *prop.part* from the R package *ape* [95].

Exploring the relation between starting composition and biological variation

To investigate the relationship between starting community composition and biological variation in our replicate vessels, we made use of the CLR representation of both the community state at the first observed time-point ($\theta_1^{(r)}$) and the biological variation (W). The mean composition in the first observed sample ($\hat{\theta}_1 = 1/R \cdot \sum_{r=1}^{R}\theta_1^{(r)}$) represented in the CLR basis was used as a measurement of the starting community. As we were primarily interested in the relative degree to which bacterial families vary in our dataset, we normalized the diagonal elements of the CLR representation of the biological variation matrix W to sum to 1 forming a composition and then used the CLR transform of these normalized variances as our measure of the relative biological variation of each bacterial family. We represent the resulting CLR transformed relative biological variation of each bacterial family by the vector $\omega = (\omega_1, \ldots, \omega_D)$. For each posterior sample, a univariate linear regression model given by the relation $\omega_i \sim \beta\hat{\theta}_{1i} + \beta_0$ where β denotes the slope of fitted model and β_0 represents an intercept was fit resulting in a posterior sample over β. Probability contours over the joint posterior distribution of each pair $(\omega_i, \hat{\theta}_{1i})$ was calculated using kernel density estimates computed by the R package *ks* [96].

We constructed a permutation distribution to investigate whether the negative relationship between $\hat{\theta}_1$ and ω suggested by this posterior distribution of β could be trivially due to there being more low abundance families than high abundance families in our starting community or a feature of working with such CLR transformed variables. This permutation distribution was constructed by randomly permuting the labels of the vector $\hat{\theta}_1$ and recomputing the posterior mean of β 1000 times. We computed the probability of our observed posterior mean of β under the constructed permutation distribution to estimate the probability, conditioned on our prior beliefs that the observed inverse relationship was simply due to the rank abundance distribution of our starting community or a feature of working with such CLR transformed variables. This probability is different than a traditional frequentist p value in that it is conditioned on our prior specifications and our observed data.

Sensitivity analyses

The sensitivity of our results to our prior specifications was assessed by rerunning our posterior inference and computations under perturbed priors. We identified two key quantities in our analysis, the regression slope β and the percentage of total variation attributable to biological sources $Tr(W)$, and looked at the change of the associated posterior intervals under the perturbed priors. The results and specifications of the modified priors are shown in Additional files 6 and 13.

Calculation of fold changes from balance values

Changes in balances values are a measure of evidence information [39]. Given a composition x with D taxa, we can denote the balance that separates a group of r taxa from another group of s taxa as

$$y = \sqrt{\frac{rs}{r+s}}\log\frac{g(x^+)}{g(x^-)}$$

where we denote the geometric mean of the elements of x that correspond to each of the two groups as $g(x^{+})$ and $g(x^{-})$ respectively. In this form, the equivalent fold change between geometric means of the two groups can be calculated as

$$\frac{g(x^{+})}{g(x^{-})} = e^{\left(\sqrt{\frac{rs}{r+s}}\right)^{-1} y}.$$

Data simulation

To test our implementation and explore the behavior of the MALLARD model we used to analyze our artificial gut dataset, we simulated a toy microbial community time-series of three bacterial taxa (Additional file 18 panels A and B) based on the likelihood component of this model. We simulated a single vessel over 200 time-points with an extra 25 replicate samples taken on the final time-point of the series. We also modeled missing observations ($n = 3$), sparsity (23% of counts were zeros), and many low counts (53% of counts were less than or equal to 10). We created a random ILR basis denoted by the contrast matrix Ψ^{true} by using the functions *named_rtree*, *phylo2sbp*, and *buildilrBasep* from the R package *philr* [97]. With respect to Ψ^{true}, we simulated data in accordance with our likelihood model using the following "true" values,

$$W^{true} = \begin{bmatrix} 0.05 & 0.01 \\ 0.01 & 0.05 \end{bmatrix}$$

$$V^{true} = \begin{bmatrix} 0.2 & -0.1 \\ -0.1 & 0.2 \end{bmatrix}$$

$$\theta_0^{true} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}.$$

After simulation, we removed samples from time-points 15, 16, and 20 to simulate missing observations. To demonstrate that our inferences were not sensitive to our choice of basis, we modeled the resulting dataset using a second random ILR basis Ψ. Our results in comparison to the true values are shown in Additional file 18 with respect to the basis Ψ.

Analysis of this simulated dataset showed that estimates for the unobserved compositions η and θ were closer to the true simulated values than a standard modeling approach of normalizing read counts to proportions (Additional file 18 panels C and D). This result is clearest in regimes rich in low and zero read counts, which we expected because of our Bayesian approach to modeling the read counting process. In addition, we found that the model correctly estimated distinct technical and biological variation patterns (Additional file 18 panels E and F). Thus, our simulations suggested that our model implementation could successfully decompose longitudinal microbiota data into component processes and characterize technical and biological variation.

Visualizations

All visualization were produced using the R programming language (v3.4.0) with the addition of the following packages: *ggplot2* [98], *ggtree* for plotting phylogenetic trees [99], and *compositions* [100].

Additional files

Additional file 1: Model fits to the observed data. Posterior mean (red), 50% (dark blue) and 95% credible (light blue) intervals for θ_t in terms of log transformed proportions (black). The raw count data are shown in black for comparison. A pseudo-count of 0.65 was added to the raw data prior to normalization and log-transformation to avoid taking the log of zero values. (PDF 163 kb)

Additional file 2: Artificial gut setup. (1) Reactor vessels; (2) flow meters controlling gas inputs; (3) pump-fed media; (4) acid and base to regulate pH; (5) central controller; (6) snorkel for exhaust. Note that only two of four replicate vessels are illustrated in this photograph. (PNG 151 kb)

Additional file 3: Proportions of most abundant bacterial families estimated from count data during hourly sampling period. Proportions were estimated by dividing observed counts by the total number of counts observed for each sample. The time-point corresponding to *B. ovatus* supplementation is depicted as a black line. (PDF 37 kb)

Additional file 4: Samples were collected over a one month period with both daily and hourly sampling intervals. Daily samples were collected at 15:00 ± 00:30 h, hourly samples were collected within ± 10 min of depicted time. Samples that were either not collected or filtered from analysis due to low sequencing depth are shown in white, samples that were included in analyses are shown in blue. Samples from days 5, 6, and 13 were not collected due to holiday. In addition to standard hourly and daily longitudinal sampling, 20 samples were collected from the final time-point of each replicate vessel. The number of replicate samples that were included in the analysis are depicted in blue. (PDF 30 kb)

Additional file 5: PCoA based on Aitchison distance applied to most abundant bacterial families. To avoid taking the log or ratio of zero counts, a pseudo-count of 0.65 was added to all counts prior to calculation of the Aitchison distance. In panel (A) samples are labeled by collection time since the start of the experiment with technical replicates labeled in black. In panel (B) samples are labeled by sequencing batch and show no clear separation between batches. In panel (C) samples are labeled by artificial gut vessel. All results in Panels A, B, and C are shown with respect to the same two principle coordinates and are thus directly comparable across plots and panels. (PDF 87 kb)

Additional file 6: Posterior estimates for the percent of total variation attributable to biological sources is not sensitive to modification of prior parameters. The "Base" prior parameter values refer to the values specified throughout the "Methods" section. In addition, the complete model was rerun with 14 separate prior parameters settings, each deviating from the Base values with respect to one parameter. Posterior 95% credible intervals and mean are shown for each set of prior parameters. (PDF 5 kb)

Additional file 7: Posterior 95% credible regions for bacterial dynamics (θ) in the PhILR Basis. (Left) The hierarchical tree of phylogenetic relationships between the bacterial families with PhILR balances (n1-n16, non-consecutive numbering) depicted ("Methods" section). Balance n12 is highlighted in Fig. 4. Branch lengths are not to scale. (+) and (−) refer to which subclade is found in the numerator or denominator of the balance respectively. (Right) Posterior 95% credible regions for the bacterial dynamics for each PhILR balance is depicted. The time-point corresponding to *B. ovatus* supplementation is depicted as a black line. (PDF 100 kb)

Additional file 8: Posterior 95% credible regions for bacterial dynamics (θ) for hourly samples in the PhILR Basis. Balances are defined in the left panel of Additional file 7, here highlighting the dynamics observed during the hourly sampling period. Balance n12 is highlighted in Fig. 4. Balance n14 shows a decrease in the ratio of the families Bacteroidaceae to Porphyromonadaceae. It is likely that this effect was due to the effects of fresh delivery media as balance shifts appear strongest in the control vessels that received sham treatment (media alone) (#2 and #3) compared to the treatment vessels that received media and *B. ovatus* (#1 and #4). (PDF 77 kb)

Additional file 9: Irregular sub-daily oscillation observed in PhILR balance n12 does not correlate with known external factors. As in Fig. 4b, the posterior mean and 95% credible interval of the microbial dynamics (θ) for balance n12 is shown during hourly sampling. The posterior mean is colored with the ID of the researcher who obtained each corresponding sample. Samples that were dropped from analysis due to low sequencing depth are denoted by NA for researcher ID. Times at which media feed bottles were changed are indicated with red dashed lines. Time-points corresponding to the daily sampling regimen are indicated by dark gray lines. (PDF 20 kb)

Additional file 10: Four bacterial families comprise most the biological variation in our study. Relative biological variation for each family was taken as the corresponding diagonal entry of the CLR transformed Biological variation matrix (W^{CLR}; "Methods" section). The median and 95% posterior credible interval are depicted for each bacterial family. (PDF 20 kb)

Additional file 11: The decomposition of technical variation among bacterial families. Posterior distribution for log-ratio variance (ρ) between pairs of bacterial families for technical variation (V). Heatmap color is given by the median of the posterior distribution of ρ. Each cell also gives the median and 95% credible region for the log-ratio variance (ρ) for the corresponding bacterial families. Columns and rows refer to the bacteria in the numerator and denominator of the corresponding log-ratios respectively. (PDF 5 kb)

Additional file 12: An inverse relationship between biological variation and initial relative abundance. 5, 25, 50, 75, and 95% highest posterior density regions of the posterior distribution of mean relative abundances on day 0 and biological variation of the ten most abundant bacterial families. Both axes are CLR-transformed. Posterior mean and 95% credible regions are also shown for the regression between these variables ("Methods" section). (PDF 84 kb)

Additional file 13: Posterior estimates for the regression slope between biological variation and bacterial family starting relative abundance is not sensitive to modification of prior parameters. The "Base" prior parameter values refer to the values specified throughout the "Methods" section. In addition, the complete model was rerun with 14 separate prior parameters settings, each deviating from the Base values with respect to one parameter. Posterior 95% credible intervals and mean are shown for each set of prior parameters. (PDF 5 kb)

Additional file 14: Posterior 95% credible regions for bacterial dynamics (θ) in the Ward Basis. (Left) The consensus tree created by Ward clustering of bacterial families with Ward balances (w1-w9) is depicted ("Methods" section). Balances nearer the root of the tree display higher variance than balances nearer the tips ("Methods" section). (+) and (–) refer to which subgroup is found in the numerator or denominator of each Ward balance respectively. The proportion of samples from the posterior distribution in which a given bipartition was present is denoted under the corresponding balance name ("Methods" section). (Right) Posterior 95% credible regions for the bacterial dynamics for each Ward balance is depicted. The time-point corresponding to *B. ovatus* treatment is depicted as a black line. (PDF 89 kb)

Additional file 15: Dynamics inferred in the balance between the Enterobacteriaceae and all other taxa does not correlate with known external factors. As in Fig. 5d, the posterior 95% credible interval of the microbial dynamics (θ) for the balance between the Enterobacteriaceae and all other taxa shown. The posterior mean is colored according to the ID of the researcher who obtained each corresponding sample. Samples that were dropped from analysis due to low sequencing depth are denoted by NA for researcher ID. Time-points corresponding to the daily sampling regimen are shown in black. (PDF 42 kb)

Additional file 16: Summary of the Kalman Filter, Kalman Smoother, and Backwards Sampling algorithm. (PDF 295 kb)

Additional file 17: Table of parameters in MALLARD model used to analyze artificial gut dataset. For the artificial gut dataset: $R = 4$, the number of artificial gut vessels; $D = 10$, the number of bacterial families analyzed; $T^{(1)} = 158$, the number of sample points at which η is to be inferred; and $T^{(2)} = 138$, the number of time-points at which θ is to be inferred. $T^{(1)}$ and $T^{(2)}$ differ due to technical replicates and time-points which lacked measurement but where inference of θ was still desired. Only the parameters η, Λ^V, Λ^W, σ^V, and σ^W are sampled using HMCMC. The parameters θ are sampled directly from the posterior using the Kalman smoother. (PDF 22 kb)

Additional file 18: Analysis of a toy simulated microbial community demonstrates the advantages of accounting for technical noise and uncertainty due to counting. (A) A 3 taxon (t1, t2, t3) microbial community was simulated according to the likelihood model used to analyze the artificial gut dataset ("Methods" section). Data from time-points 15, 16, and 20 were removed to simulate the effects of missing data on inferences. (B) A simulated phylogeny with annotated PhILR balances (n1, n2) used to analyze the simulated dataset. (+) and (–) refers to taxa in the numerator and denominator of associated balances. Pseudo-count based (PC) estimates for the multinomial parameters are obtained by adding 0.65 to all counts and then dividing each count by the sequencing depth of its associated sample and are shown as reference in (C-D). (C) Posterior mean and 95% credible interval for the multinomial parameters η. (D) Posterior mean and 95% credible interval for the unobserved microbial dynamics θ. PC estimates for the covariance of the multinomial parameters was obtained as the covariance of the first difference of the PC parameter estimates and are shown as reference in (E-F). 100 samples from the posterior distribution of the biological variation (W, E) and technical variation (V, F) depicted as 95% probability regions of the Logistic Normal distribution centered at the point in the simplex where each bacterial taxon is equally abundant. Black points represent the simulated biological variations (w_t). (PDF 9820 kb)

Acknowledgements

We thank Rachel Silverman, Alex D. Washburne, Lionel Watkins, Firas Midani, Max Villa, Zachary Holmes, Brianna Petrone, B. Jesse Shapiro, Jonathan Friedman, and Susan Holmes for their manuscript comments and insights. This work used a high-performance computing facility partially supported by grant 2016-IDG-1013 ("HARDAC+: Reproducible HPC for Next-generation Genomics") from the North Carolina Biotechnology Center.

Funding

JDS and LAD acknowledge support from the Duke University Medical Scientist Training Program (GM007171), the Global Probiotics Council, a Searle Scholars Award, the Hartwell Foundation, an Alfred P. Sloan Research Fellowship, the Translational Research Institute through Cooperative Agreement NNX16AO69A, the Damon Runyon Cancer Research Foundation, the Hartwell Foundation, and NIH 1R01DK116187-01. SM would like to acknowledge the support of grants NSF IIS-1546331, NSF DMS-1418261, NSF IIS-1320357, NSF DMS-1045153, and NSF DMS1613261.

Availability of data and materials

Demultiplexed sequencing data were submitted to the NCBI SRA database under accession number SRP140877 or accessed through BioProject record PRJNA450809 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA450809). All code has been made available at https://github.com/LAD-LAB/MALLARD-Paper-Code.

Authors' contributions

JDS developed the MALLARD modeling framework and analyzed the data. JDS, HD, LAD, and SM designed the study. JDS, HD, LAD, and RJB performed

the study and collected samples. HD and RJB prepared samples for DNA sequencing. JDS, LAD, and SM prepared the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

A fresh fecal sample was obtained from a healthy volunteer who provided written informed consent (Duke Health IRB Pro00049498).

Competing interests

The authors declare that they have no competing financial, professional, or personal interests that might have influenced the performance or presentation of the work described in this manuscript.

Publisher's Note

Author details

[1]Program in Computational Biology and Bioinformatics, Duke University, CIEMAS, Room 2171, 101 Science Drive, Box 3382, Durham, NC 27708, USA. [2]Medical Scientist Training Program, Duke University, Durham, NC 27708, USA. [3]Center for Genomic and Computational Biology, Duke University, Durham, NC 27708, USA. [4]University Program in Genetics and Genomics, Duke University, Durham, NC 27708, USA. [5]Department of Molecular Genetics and Microbiology, Duke University, Durham, NC 27708, USA. [6]Departments of Statistical Science, Mathematics, Computer Science, Biostatistics & Bioinformatics, Duke University, Durham, NC 27708, USA.

Received: 21 May 2018 Accepted: 23 October 2018
Published online: 12 November 2018

References

1. McDonald JA, Schroeter K, Fuentes S, Heikamp-Dejong I, Khursigara CM, de Vos WM, Allen-Vercoe E. Evaluation of microbial community reproducibility, stability and composition in a human distal gut chemostat model. J Microbiol Methods. 2013;95:167–74.
2. Duncan SH, Scott KP, Ramsay AG, Harmsen HJ, Welling GW, Stewart CS, Flint HJ. Effects of alternative dietary substrates on competition between human colonic bacteria in an anaerobic fermentor system. Appl Environ Microbiol. 2003;69:1136–42.
3. Molly K, Vande Woestyne M, Verstraete W. Development of a 5-step multi-chamber reactor as a simulation of the human intestinal microbial ecosystem. Appl Microbiol Biotechnol. 1993;39:254–8.
4. Macfarlane GT, Macfarlane S. Models for intestinal fermentation: association between food components, delivery systems, bioavailability and functional interactions in the gut. Curr Opin Biotechnol. 2007;18:156–62.
5. Gamage H, Tetu SG, Chong RWW, Ashton J, Packer NH, Paulsen IT. Cereal products derived from wheat, sorghum, rice and oats alter the infant gut microbiota in vitro. Sci Rep. 2017;7:14312.
6. Antunes LC, McDonald JA, Schroeter K, Carlucci C, Ferreira RB, Wang M, Yurist-Doutsch S, Hira G, Jacobson K, Davies J, et al. Antivirulence activity of the human gut metabolome. MBio. 2014;5:e01183–14.
7. De Preter V, Falony G, Windey K, Hamer HM, De Vuyst L, Verbeke K. The prebiotic, oligofructose-enriched inulin modulates the faecal metabolite profile: an in vitro analysis. Mol Nutr Food Res. 2010;54:1791–801.
8. Aguirre M, Eck A, Koenen ME, Savelkoul PH, Budding AE, Venema K. Diet drives quick changes in the metabolic activity and composition of human gut microbiota in a validated in vitro gut model. Res Microbiol. 2016;167:114–25.
9. Liang X, Bushman FD, FitzGerald GA. Rhythmicity of the intestinal microbiota is regulated by gender and the host circadian clock. Proc Natl Acad Sci U S A. 2015;112:10479–84.
10. Auchtung JM, Robinson CD, Britton RA. Cultivation of stable, reproducible microbial communities from different fecal donors using minibioreactor arrays (MBRAs). Microbiome. 2015;3:42.
11. Payne AN, Zihler A, Chassard C, Lacroix C. Advances and perspectives in in vitro human gut fermentation modeling. Trends Biotechnol. 2012; 30:17–25.
12. McIntyre LM, Lopiano KK, Morse AM, Amin V, Oberg AL, Young LJ, Nuzhdin SV. RNA-seq: technical variability and sampling. BMC Genomics. 2011;12:293.
13. Rosenthal M, Aiello AE, Chenoweth C, Goldberg D, Larson E, Gloor G, Foxman B. Impact of technical sources of variation on the hand microbiome dynamics of healthcare workers. PLoS One. 2014;9:e88999.
14. Sinha R, Abnet CC, White O, Knight R, Huttenhower C. The microbiome quality control project: baseline study design and future directions. Genome Biol. 2015;16:276.
15. Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, Schwager E, Crabtree J, Ma S, Microbiome Quality Control Project C, et al. Assessment of variation in microbial community amplicon sequencing by the microbiome quality control (MBQC) project consortium. Nat Biotechnol. 2017;35:1077–86.
16. Goodrich JK, Di Rienzi SC, Poole AC, Koren O, Walters WA, Caporaso JG, Knight R, Ley RE. Conducting a microbiome study. Cell. 2014;158:250–62.
17. Voigt AY, Costea PI, Kultima JR, Li SS, Zeller G, Sunagawa S, Bork P. Temporal and technical variability of human gut metagenomes. Genome Biol. 2015;16:73.
18. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet. 2010;11:733–9.
19. Zhou J, Wu L, Deng Y, Zhi X, Jiang YH, Tu Q, Xie J, Van Nostrand JD, He Z, Yang Y. Reproducibility and quantitation of amplicon sequencing-based detection. ISME J. 2011;5:1303–13.
20. McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput Biol. 2014;10:e1003531.
21. Gloor GB, Macklaim JM, Vu M, Fernandes AD. Compositional uncertainty should not be ignored in high-throughput sequencing data analysis. Aust J Stat. 2016;45:73.
22. Silverman JD, Washburne AD, Mukherjee S, David LA. A phylogenetic transform enhances analysis of compositional microbiota data. eLife. 2017;6: e21887.
23. Gloor GB, Wu JR, Pawlowsky-Glahn V, Egozcue JJ. It's all relative: analyzing microbiome data as compositions. Ann Epidemiol. 2016;26:322–9.
24. Tsilimigras MC, Fodor AA. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. Ann Epidemiol. 2016;26:330–5.
25. Gloor GB, Reid G. Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. Can J Microbiol. 2016;62: 692–703.
26. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. PLoS Comput Biol. 2012;8:e1002687.
27. Lovell D, Müller W, Taylor J, Zwart A, Helliwell C. Proportions, percentages, ppm: do the molecular biosciences treat compositional data right. In: Pawlowsky-Glahn V, Buccianti A, editors. Compositional data analysis: Theory and Applications. West Sussex: Wiley; 2011. p. 193–207.
28. Gerber GK. The dynamic microbiome. FEBS Lett. 2014;588:4131–9.
29. Trosvik P, Rudi K, Naes T, Kohler A, Chan KS, Jakobsen KS, Stenseth NC. Characterizing mixed microbial population dynamics using time-series analysis. ISME J. 2008;2:707–15.
30. Trosvik P, Rudi K, Straetkvern KO, Jakobsen KS, Naes T, Stenseth NC. Web of ecological interactions in an experimental gut microbiota. Environ Microbiol. 2010;12:2677–87.
31. van de Wiele T, Boon N, Possemiers S, Jacobs H, Verstraete W. Inulin-type fructans of longer degree of polymerization exert more pronounced in vitro prebiotic effects. J Appl Microbiol. 2007;102:452–60.
32. Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, Hall AB, Brady A, Creasy HH, McCracken C, Giglio MG, et al. Strains, functions and dynamics in the expanded human microbiome project. Nature. 2017;550:61–6.
33. Prado R, West M. Time series: modeling, computation, and inference. Boca Raton: CRC Press; 2010.
34. Cargnoni C, Muller P, West M. Bayesian forecasting of multinomial time series through conditionally Gaussian dynamic models. J Am Stat Assoc. 1997;92:640–7.
35. Belik J, Shifrin Y, Arning E, Bottiglieri T, Pan J, Daigneault MC, Allen-Vercoe E. Intestinal microbiota as a tetrahydrobiopterin exogenous source in hph-1 mice. Sci Rep. 2017;7:39854.
36. Yen S, McDonald JA, Schroeter K, Oliphant K, Sokolenko S, Blondeel EJ, Allen-Vercoe E, Aucoin MG. Metabolomic analysis of human fecal microbiota: a comparison of feces-derived communities and defined mixed communities. J Proteome Res. 2015;14:1472–82.
37. Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R: Modeling and analysis of compositional data. West Sussex: Wiley; 2015. p. 67–71, 114–121, 226–227.

38. West M, Harrison J: Bayesian forecasting and dynamic models. 2nd edn. New York: Springer; 1997. Pgs 32–61, 97–122, 178–423.
39. Egozcue JJ, Pawlowsky-Glahn V. Evidence information in Bayesian updating. In: Proceedings of the 4th International Workshop on Compositional Data Analysis; 2011. p. 1–13.
40. Aitchison J. The statistical analysis of compositional data. London; New York: Chapman and Hall; 1986. Pgs 68-71. p. 112–37.
41. Lovell D, Pawlowsky-Glahn V, Egozcue JJ, Marguerat S, Bahler J. Proportionality: a valid alternative to correlation for relative data. PLoS Comput Biol. 2015;11:e1004075.
42. Åström KJ. On the choice of sampling rates in parametric identification of time series. Inf Sci. 1969;1:273–8.
43. Faust K, Lahti L, Gonze D, de Vos WM, Raes J. Metagenomics meets time series analysis: unraveling microbial community dynamics. Curr Opin Microbiol. 2015;25:56–66.
44. Yurtsev EA, Conwill A, Gore J. Oscillatory dynamics in a bacterial cross-protection mutualism. Proc Natl Acad Sci U S A. 2016;113:6236–41.
45. Huisman J, Weissing FJ. Biodiversity of plankton by species oscillations and chaos. Nature. 1999;402:407–10.
46. Beninca E, Huisman J, Heerkloss R, Johnk KD, Branco P, Van Nes EH, Scheffer M, Ellner SP. Chaos in a long-term experiment with a plankton community. Nature. 2008;451:822–5.
47. Lenz P, Sogaard-Andersen L. Temporal and spatial oscillations in bacteria. Nat Rev Microbiol. 2011;9:565–77.
48. Thaiss CA, Zeevi D, Levy M, Zilberman-Schapira G, Suez J, Tengeler AC, Abramson L, Katz MN, Korem T, Zmora N, et al. Transkingdom control of microbiota diurnal oscillations promotes metabolic homeostasis. Cell. 2014; 159:514–29.
49. Lennon JT, Jones SE. Microbial seed banks: the ecological and evolutionary implications of dormancy. Nat Rev Microbiol. 2011;9:119–30.
50. Epstein SS. Microbial awakenings. Nature. 2009;457:1083.
51. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE. Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci U S A. 2011;108(Suppl 1):4578–85.
52. Shade A, Jones SE, Caporaso JG, Handelsman J, Knight R, Fierer N, Gilbert JA. Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. MBio. 2014;5:e01371–14.
53. Fierer N, Nemergut D, Knight R, Craine JM. Changes through time: integrating microorganisms into the study of succession. Res Microbiol. 2010;161:635–42.
54. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial community variation in human body habitats across space and time. Science. 2009;326:1694–7.
55. Koopman SJ, Durbin J. Fast filtering and smoothing for multivariate state space models. J Time Ser Anal. 2000;21:281–96.
56. Chen X, Irie K, Banks D, Haslinger R, Thomas J, West M. Scalable Bayesian modeling, monitoring and analysis of dynamic network flow data. J Am Stat Assoc. 2018;113:519–533.
57. Meredith HR, Lopatkin AJ, Anderson DJ, You L. Bacterial temporal dynamics enable optimal design of antibiotic treatment. PLoS Comput Biol. 2015;11: e1004201.
58. David LA, Weil A, Ryan ET, Calderwood SB, Harris JB, Chowdhury F, Begum Y, Qadri F, LaRocque RC, Turnbaugh PJ. Gut microbial succession follows acute secretory diarrhea in humans. MBio. 2015;6:e00381–15.
59. Fukuyama J, Rumker L, Sankaran K, Jeganathan P, Dethlefsen L, Relman DA, Holmes SP. Multidomain analyses of a longitudinal human microbiome intestinal cleanout perturbation experiment. PLoS Comput Biol. 2017;13: e1005706.
60. DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, Sun CL, Goltsman DS, Wong RJ, Shaw G, et al. Temporal and spatial variation of the human microbiota during pregnancy. Proc Natl Acad Sci U S A. 2015;112:11060–5.
61. Washburne AD, Silverman JD, Leff JW, Bennett DJ, Darcy JL, Mukherjee S, Fierer N, David LA. Phylogenetic factorization of compositional data yields lineage-level associations in microbiome datasets. PeerJ. 2017;5: e2969.
62. Morton JT, Sanders J, Quinn RA, McDonald D, Gonzalez A, Vazquez-Baeza Y, Navas-Molina JA, Song SJ, Metcalf JL, Hyde ER, et al. Balance trees reveal microbial niche differentiation. mSystems. 2017;2(1):e00162–16.
63. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. Microbiome. 2014;2:15.
64. Aijo T, Mueller CL, Bonneau R. Temporal probabilistic modeling of bacterial compositions derived from 16S rRNA sequencing. In: bioRxiv; 2016.
65. Grantham NS, Reich BJ, Borer ET, Gross K. MIMIX: a Bayesian Mixed-Effects Model for Microbiome Data from Designed Experiments. ArXiv e-prints. 2017; 1703:arXiv:1703.07747.
66. Xia F, Chen J, Fung WK, Li H. A logistic normal multinomial regression model for microbiome compositional data analysis. Biometrics. 2013;69:1053–63.
67. Li Z, Lee K, Karagas MR, Madan JC, Hoen AG, Li H. A multivariate zero-inflated logistic model for microbiome relative abundance data. ArXiv e-prints. 2017; 1709:arXiv:1709.07798.
68. Shankar J. Insights into study design and statistical analyses in translational microbiome studies. Ann Transl Med. 2017;5(12).
69. David LA, Materna AC, Friedman J, Campos-Baptista MI, Blackburn MC, Perrotta A, Erdman SE, Alm EJ. Host lifestyle affects human microbiota on daily timescales. Genome Biol. 2014;15:R89.
70. Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, Knights D, Gajer P, Ravel J, Fierer N, et al. Moving pictures of the human microbiome. Genome Biol. 2011;12:R50.
71. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A. 2011; 108(Suppl 1):4516–22.
72. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012;6:1621–4.
73. Maurice CF, Haiser HJ, Turnbaugh PJ. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. Cell. 2013;152:39–50.
74. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3.
75. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
76. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7:335–6.
77. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73:5261–7.
78. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41:D590–6.
79. Rajilic-Stojanovic M, de Vos WM. The first 1000 cultured species of the human gastrointestinal microbiota. FEMS Microbiol Rev. 2014;38:996–1047.
80. Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barcelo-Vidal C. Isometric logratio transformations for compositional data analysis. Math Geol. 2003;35: 279–300.
81. Aitchison J, Shen SM. Logistic-normal distributions—some properties and uses. Biometrika. 1980;67:261–72.
82. Barceló-Vidal C, Martín-Fernández JA, Pawlowsky-Glahn V. Mathematical foundations of compositional data analysis. In: Proceedings of IAMG; 2001. p. 1–20.
83. Gelman A, Lee D, Guo JQ. Stan: a probabilistic programming language for Bayesian inference and optimization. J Educ Behav Stat. 2015;40:530–43.
84. Hoffman MD, Gelman A. The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J Mach Learn Res. 2014;15:1593–623.
85. Lewandowski D, Kurowicka D, Joe H. Generating random correlation matrices based on vines and extended onion method. J Multivar Anal. 2009; 100:1989–2001.
86. Kucukelbir A, Ranganath R, Gelman A, Blei D. Automatic variational inference in Stan. In: Advances in neural information processing systems; 2015. p. 568–76.
87. Stan Development Team: Stan modeling language: user's gude and reference manual. 2.17.0 (pg. 371) edition; 2017.
88. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB: Bayesian data analysis, third edition. Boca Ratton: Chapman and Hall/CRC; 2013. p. 285.
89. Dryden IL, Koloydenko A, Zhou DW. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. Ann Appl Stat. 2009; 3:1102–23.

90. Dryden IL: Shapes: statistical shape analysis. 1.2.3 edition. CRAN; 2017.
91. Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R. Principal Balances. In: Egozcue JJ, Tolosana-Delgado R, Ortego M, editors. 4th International Workshop on Compositional Data Analysis. Girona; 2011. p. 1–10.
92. Egozcue JJ, Pawlowsky-Glahn V, Gloor GB. Linear association in compositional data analysis. Austrian J Stat. 2018;37:3–31.
93. Martín-Fernández J-A, Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R. Advances in principal balances for compositional data. Math Geosci. 2018; 50:273–98.
94. Ward JH. Hierarchical grouping to optimize an objective function. J Am Stat Assoc. 1963;58:236.
95. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 2004;20:289–90.
96. Duong T: ks: Kernel smoothing. R package version 1.10.6. 2017.
97. Silverman JD: philr: Phylogenetic partitioning based ILR transform for metagenomics data. R package version 1.0.0 edition. http://bioconductor.org/packages/philr/: Bioconductor; 2016.
98. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer; 2016.
99. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2016;8(1):28–36.
100. van den Boogaart KG, Tolosana-Delgado R, Bren M. Compositions: compositional data analysis. R package version 1; 2014. p. 40–1.