Title: Growth phase estimation for abundant bacterial populations sampled longitudinally from human stool metagenomes.

Authors: Joe J. Lim[1], Christian Diener[2], Sean M. Gibbons[2-4,*]

Affiliations: [1] Department of Environmental & Occupational Health Sciences, University of Washington, Seattle, WA 98105; [2] Institute for Systems Biology, Seattle, WA 98109; [3] Department of Bioengineering, University of Washington, Seattle, WA 98105; [4] eScience Institute, University of Washington, Seattle, WA 98105

[*] corresponding author: sgibbons@isbscience.org

ABSTRACT

Longitudinal sampling of the stool has yielded important insights into the ecological dynamics of the human gut microbiome. However, due to practical limitations, the most densely sampled time series from the human gut are collected at a frequency of about once per day, while the population doubling times for gut commensals are on the order of minutes-to-hours. Despite this, much of the prior work on human gut microbiome time series modeling has, implicitly or explicitly, assumed that day-to-day fluctuations in taxon abundances are related to population growth or death rates, which is likely not the case. Here, we propose an alternative model of the human gut as a continuous flow ecosystem at a dynamical steady state, where population dynamics occur internally and the bacterial population sizes measured in stool represent an endpoint of these internal dynamics. We formalize this idea as stochastic logistic growth of a population held at a constant dilution rate. We show how this model provides a path toward estimating the growth phases of gut bacterial populations *in situ*. We assess our model predictions against densely-sampled human stool metagenomic time series data. Consistent with our model, donors with slower defecation rates tended to harbor a larger proportion of taxa in later growth phases, while faster defecation rates were associated with more taxa in earlier growth phases. We discuss how these growth phase estimates may be used to better inform metabolic modeling in flow-through ecosystems, like animal guts or industrial bioreactors.

INTRODUCTION

The human gut microbiome is an anaerobic bioreactor, ecologically distinct to each individual, that transforms dietary and host substrates into bioactive molecules important to host health [1–3]. Disruptions to the ecological composition of the gut have been shown to mediate the progression of various complex diseases [4–8]. Furthermore, the ecological dynamics of the gut appear to be relevant to both health and disease states [9, 10]. However, the biological interpretation of densely-sampled adult human fecal microbiome time series is fraught.

Various dynamical models have been applied to gut microbial abundance data collected from adult human donors [11–15]. These models often assume, either explicitly or implicitly, that day-to-day changes in abundance are proportional to population growth and/or death [16]. However, the underlying data often do not match this assumption [11, 16–20]. The gut is a flow-through ecosystem and commensal gut bacteria must grow fast enough to avoid dilution-to-extinction. As such, gut bacterial doubling times tend to be fast, ranging from minutes-to-hours [21–23]. However, stool sampling frequency is usually limited to, at most, about once per day. Consequently, rapid internal population dynamics likely cannot be directly estimated from the day-to-day measurements obtained from stool [16].

In the absence of major perturbations, is it possible to extract meaningful information about commensal population dynamics from adult human gut microbiome time series, despite the fundamental limitations in sampling timescales mentioned above? One work around to inferring growth rates of bacterial populations *in situ* is to leverage metagenome-inferred replication rates [22]. Briefly, instantaneous replication rates can be estimated for abundant bacterial populations in metagenomic samples by taking advantage of the fact that fast-growing taxa show an asymmetry in reads mapping to different genomic loci, with higher read depth near the origin of replication and a lower depth near the terminus due to the initiation of multiple

replication forks [21–23]. However, even when replication rates and population abundances can both be estimated from the same metagenomic samples, it is unclear how these measurements are related to the *in situ* growth phase of a population (e.g. lag, log, or stationary phases). As such, biological interpretations regarding population size and replication rate fluctuations in flow-through ecosystems like the human gut, where internal dynamics are much faster than sampling rates, remain challenging.

Early experiments by Jaques Monod [24] identified distinct growth phases for bacterial populations in culture, which can be captured by the stochastic logistic growth equation (sLGE) [25]. The sLGE has been shown to be a good fit for bacterial population growth *in vitro* and in real-world, steady-state ecosystems [26–31]. We used the sLGE to study statistical relationships between population sizes and growth rates across the various phases of growth (i.e., lag, acceleration, log, deceleration, and stationary phases) to see if we could extract *in situ* growth phase information from longitudinal data from a steady-state, flow-through ecosystem sampled at a consistent frequency. Overall, the sLGE model yields statistical relationships that can be leveraged to identify the *in situ* growth phase of a bacterial population periodically sampled from a continuous-flow ecosystem, like the human gut.

To assess our model predictions, we calculated population abundance and growth/replication rate trajectories from more than a dozen organisms in four densely sampled human gut metagenomic time series [32]. On average, gut commensal growth rates and population sizes were positively correlated within each of the stool donor time series, which suggests that most abundant taxa in the gut are growing exponentially. We were able to identify signatures of specific growth phases in 20-40% of the abundant bacterial populations in the guts of these four individuals. We describe how our growth phase inference approach can serve to inform more accurate mechanistic modeling of flow-through ecosystems (e.g., community-scale metabolic models, which usually assume exponential growth), which could have broad

implications for the gut microbiome and host health [8, 33, 34], flow-through agricultural systems [35, 36], climate change [35, 37, 38], and industrial bioreactor production processes [39, 40].

RESULTS

Framing the gut as an anaerobic flow-through bioreactor

The mammalian gut can be understood as an anaerobic bioreactor with a continuous input (i.e., dietary and host substrates) and output (i.e., stool) [41], and microbial taxa must grow fast enough to avoid dilution to extinction (Fig. 1A). Stool sampling captures the endpoint of internal gut bacterial population dynamics. For example, in our cartoon figure we see that Taxon 1 starts growing higher up in the colon and is in stationary phase by the time a stool sample is collected, while Taxon 3 starts growing lower in the colon and is still growing exponentially at the point of stool sampling (Fig. 1A). Overall, the daily abundances of Taxa 1-3 represent the average (μ) steady-state population size, plus or minus some amount of biological and technical noise, at the time of stool sampling (Fig. 1A). To investigate improved methods for interpreting the dynamics of human gut microbial time series, we downloaded shotgun metagenomic time series data from the BIO-ML cohort (i.e., health-screened stool donors who provided fecal-transplant material to the stool bank OpenBiome) [32]. The BIO-ML cohort contained 74 donors, 70 of which had 1-3 time points collected [32]. To filter for dense longitudinal data, we selected a subset of donors with more than 50 time points. Four donors (i.e. donors ae, am, an, and ao) met this criterion, with 3-5 fecal samples per week for >50 days (Fig. 1B).

Characterizing the relationships between gut commensal population size and growth rate using metagenomic time series data

We first investigated the statistical properties of day-to-day fluctuations in gut bacterial population sizes, estimated from metagenomic time series. Specifically, we looked at the

associations between population abundance estimates (t_n) and the changes in abundance estimates (i.e., deltas) between time points ($t_{n+1} – t_n$). Naïvely, if most bacterial populations in the stool were growing exponentially, we would expect that population abundances and growth rates would be positively correlated. However, prior work has indicated an overall negative correlation between abundances and changes in abundances in stool 16S rRNA gene amplicon sequencing data generated from densely sampled human stool time series [15]. Indeed, we found that abundant bacterial populations in the stool of the four BIO-ML donors maintained stable average abundances over time (μ), with day-to-day fluctuations above and below this average, as pictured in the example of *Bacteroides cellulosilyticus* in donor am (Fig. 2A-B). This kind of pattern fits a regression-to-the-mean model, which one would expect when randomly sampling from a stationary distribution (Fig. 2B). This kind of regression-to-the-mean process will give rise to a negative correlation between population abundances and changes in abundance between time points, consistent with what has been observed previously [15]. We observed that the deltas ($t_{n+1} – t_n$) for the same gut taxon (*Bacteroides uniformis*) measured across each donor time series, when plotted against their respective normalized abundances (t_n), showed the expected negative association (Fig. 2C). Furthermore, similar negative associations were observed across all taxa analyzed, across all donors (Fig. 2D). Overall, these results support our assertion that stool samples provide steady-state population abundance estimates for gut commensal bacteria, which are representative of the endpoint of internal dynamics.

Next, we looked at the statistical associations between calculated peak-to-trough ratios (i.e., PTRs; a proxy for growth-rate) [22] of abundant bacterial populations from each metagenomic sample and their respective metagenomic population abundance estimates. If the deltas, presented above, were truly proportional to growth and/or death rates, we would expect that the statistical relationships between deltas and population size would be similar to those

between PTRs and population size. However, unlike the regression-to-the-mean signature identified for the deltas, we found variable statistical relationships between PTR and centered log-ratio (CLR) transformed population abundances for the same taxon (*Bacteroides ovatus*) across the four donors (Fig. 3A). Similarly, we saw a wide range of positive, negative, and null associations between PTRs and CLR abundances across all measured taxa within each donor (Fig. 3B). These results are inconsistent with a regression-to-the-mean signal, and suggest a more complex relationship between growth rate and population size [42–44]. Finally, we calculated temporally-averaged PTRs and population sizes for each abundant taxon within each of the donors. Overall, there was a significantly positive (linear regression, *p-values* = 0.0318, 0.125, 0.155, 0.031 for donors ae, am, an, and ao, respectively; combined *p-value* using Fisher's method = 0.005), albeit noisy, association between average PTR and average CLR abundance across all four donors (Fig. 3C), indicating that taxa with higher average population sizes tend to have higher average growth rates. This result is consistent with what we would expect to observe in exponentially-growing populations.

Stochastic logistic growth equation provides insights into growth phases

In order to better understand and interpret the varying relationships we observe between PTRs and CLR abundances, we turn to modeling. The basic properties of growth curves of microbial taxa can be captured using the logistic growth equation (Fig. 4). This model is defined such that the change in abundance for each taxon *i* (dx_i/dt) is captured by the current abundance at time *t*, $x_i(t)$, multiplied by the maximal growth rate, *r*, and the carrying capacity (*k*) term $(1-x_i(t)/k)$ [45]. In this model, population size over time shows a sigmoidal curve, with the abundance asymptotically approaching *k* (Fig. 4A, top panel). The derivative of this curve with respect to time yields the change in growth rate over time, which peaks during log-phase growth (Fig. 4A, middle panel). The second derivative of abundance with respect to time, which is the

instantaneous change in growth with respect to time and is often referred to as the acceleration rate, shows a peak during the acceleration phase and a trough during the deceleration phase (Fig. 4A, bottom panel). Based on this second-derivative curve, we show the expected relationships between growth rate and abundance as you move across the logistic growth curve, along the time axis (Fig. 4B). These expected relationships provide a potential path forward for inferring the *in situ* growth phase of a bacterial population sampled at a consistent frequency from a flow-through ecosystem.

The logistic growth model is a deterministic equation. However, the abundances of commensal bacterial populations in the gut fluctuate due to myriad factors including interspecies competition, resource fluctuations, and stool residence time [46]. In order to approximate these fluctuations in our modeling, we introduced a stochastic term to the logistic growth model (Fig. 5A). Herein, σ denotes the noise magnitude and $\omega(t)$ represents a white noise term [27]. Five growth phases (i.e., lag, acceleration, log, deceleration, and stationary phases) were defined using the half-maximum and half-minimum, respectively, of the second derivative LGE curve (Fig. S1A). We grouped these phases into three major categories: lag-acceleration phase, log phase, and deceleration-stationary phase. We simulated 100 iterations of the stochastic logistic growth equation (sLGE) for each of a range of parameterizations (see Methods), which recapitulated the expected statistical relationships between growth rates and abundances for populations consistently sampled within our three major growth phase categories (Fig. 5A-C). For example, the Pearson's R values between growth rates and abundances were significantly positive in lag-acceleration phase and significantly negative in deceleration-stationary phase (Fig. 5B). Log phase growth was more variable, but showed little-to-no significant association between growth rates and abundances. These results were reproduced across a wide range of parameter space and were robust to varying the noise term (Fig. S1B). Overall, the relationships

between growth rate and abundance across growth phases were highly consistent with our expectations (Fig. 5C).

Inferring in situ growth phases for abundant gut commensal populations sampled in metagenomic time series

Based on these sLGE results, we assigned putative *in situ* growth phases to abundant gut bacterial populations from the four BIO-ML gut metagenomic time series. Specifically, we suggest that significantly positive associations (linear regression, adjusted *p*-value < 0.05, with a positive beta-coefficient) between PTRs and CLR abundances indicate early-phase exponential growth (i.e. acceleration phase; we can likely exclude lag phase due to the fact that we would be unlikely to detect taxa with very low levels of biomass), significantly negative associations (linear regression, adjusted *p*-value < 0.05, with negative beta-coefficient) indicate deceleration or stationary phases, and the absence of a significant association could indicate either log-phase growth or a false negative (i.e., not powered enough to detect a positive or negative association with the number of time points sampled). *Bacteroides cellulosilyticus*, *Bacteroides ovatus 1*, and *Megaspaera eldenii* showed significantly positive PTR-abundance associations within donor ae (Figs. 6A and S2). *Bacteroides ovatus* 1 and *Parabacteroides distasonis* showed positive PTR-abundance associations, while *Alistipes finegoldii*, *Bacteroides uniformis*, and *Bacteroides xylanisolvens* showed negative associations in donor am (Figs. 6A and S3). *Alistipes shahii*, *Bacteroides intestinalis*, *Bacteroides thetaiotaomicron*, *Bacteroides uniformis*, *Bacteroides xylanisolvens*, and *Odoribacter splanchnicus* showed significantly negative PTR-abundance associations in donor an (Fig. 6A and S4). Finally, *Favonifractor plautii* showed a positive PTR-abundance association and *Bacteroides fragilis*, *Bacteroides ovatus* 1, *Bacteroides uniformis*, and *Bacteroides xylanisolvens* showed negative associations in donor ao (Fig. 6A and S5).

We observed a slight difference in the number of significantly positive and negative PTR-abundance associations between donors ae/am, and an/ao, with donors an and ao tending to have a larger proportion of negative associations when compared to donors ae and am. Interestingly, donors an and ao had a lower average defecation frequency (≤ 1 per day) than donors ae and am (> 1 per day). Concordantly, based on our flow-through model of the gut ecosystem (Fig. 1A), we would expect that bacterial populations would be pushed towards earlier growth phases at faster flow rates (Fig. 6B). Over half of the taxa with PTR and abundance time series data did not show significant associations (Fig. 6A and Fig. S2-5). This suggests that either these taxa are in the log growth phase or we were not powered enough to detect significant positive or negative associations for these taxa given the effect sizes and the number of samples. We suggest that many of these taxa may well be in log phase, due to the significant association observed between average PTRs and average CLR abundances across donors (Fig. 3C). Overall, our approach provides a new path toward identifying the *in situ* growth phase of microbial populations in flow-through ecosystems.

DISCUSSION

Many prior studies assumed, either implicitly or explicitly, that the growth and death rates of gut bacterial populations were proportional to day-to-day changes in abundances, as measured from human stool samples. However, we outline how this assumption is likely invalid due to the

fact that human gut bacterial population growth/death processes inside the intestinal tract are known to be faster (minutes-to-hours) than our sampling timescales (days). In support of this assertion, we show how the statistical relationships between changes in abundance ($t_{n+1} - t_n$) and abundances (t_n), estimated from stool metagenomic time series, indicate a regression-to-the-mean effect that one would expect when sampling from a stationary distribution (Figs. 1-2). Thus, as prior work has indicated [15], bacterial taxa in the gut have stable average population sizes, which likely represent steady-state endpoints of internal dynamics (Figs. 1-2). Despite the fundamental mismatch between gut bacterial population dynamics and sampling timescales, we attempt to identify statistical signatures within these daily-sampled human gut time series that might provide accurate insights into *in situ* population dynamics.

While changes in abundance between time points (i.e., deltas) do not appear to be related to population growth, PTRs enable direct estimates of *in situ* growth rates from metagenomic samples [22, 23, 47–49]. Unlike the relationships between deltas and abundances, which were always negative (Fig. 2C-D), the relationships between PTRs and abundances were quite variable (Fig. 3A-3B). While regression-to-the-mean is a plausible mechanism for the consistent negative delta-abundance relationships (Fig. 2), the underlying processes driving variable PTR-abundance relationships appear to be more nuanced (Fig. 3).

We turned to the sLGE to explore relationships between growth rate and abundance across different phases of growth (Fig. 4). The sLGE recapitulated the key relationships observed in the metagenomic time series between PTRs and abundances and provided predictions for *in situ* growth phases (Figs. 5-6). Consistent with our sLGE predictions, we found that individuals with higher defecation rates tended to be enriched for taxa in earlier growth phases (Fig. 6). Thus, our results reveal a promising approach to inferring *in situ* growth phases for abundant organisms detected in human gut metagenomic time series. A major limitation of this approach is our lack of knowledge about the effect size distributions for these relationships

in real-world data and an understanding of the statistical power needed for detecting these associations from metagenomic time series. The absence of a PTR-abundance relationship could represent a false negative or it could indicate log-phase growth. Future *in vitro* experimental work (e.g., in chemostats) should focus on better quantifying these PTR-abundance relationships across parameter space to build a more quantitative understanding of these phenomena (e.g., through varying maximal growth rates, the carrying capacities, the flow rates, and volumes).

We observed that the average PTR and average abundance of a given taxon over time were positively correlated, which is consistent with exponentially-growing populations (Fig. 3C). This result is especially relevant to the metabolic modeling community. Ecological interactions within free-living and host-associated microbial communities are largely governed by exchanges of small-molecule metabolites [50, 51]. Genome-scale metabolic modeling and flux-balance analysis (FBA) has been effective mechanistic tools for simulating these metabolic exchanges, especially in controlled bioreactor systems [52]. The objective function used to find a unique solution to bacterial FBA models is often biomass maximization, which assumes that these organisms are growing exponentially at steady state. Exponential growth is a valid assumption for organisms in acceleration or log growth phases, but this assumption is violated for organisms in deceleration or stationary phases. Prior work has demonstrated that biomass composition can change depending on the growth phase of a population, which ideally would be taken into account to more accurately model metabolic fluxes within the system [53–55]. Overall, our work suggests that most organisms in the human gut are amenable to FBA, and our growth phase estimation approach allows for the identification of populations that may not fit classical FBA assumptions.

In conclusion, we provide a new path forward for the biological interpretation of metagenomic time series data generated from adult human stool samples. We hope that *in situ*

growth phase estimation will be applied more broadly to other kinds of flow-through environments to improve our understanding of internal dynamics in these systems and provide improved constraints for mechanistic modeling of microbial communities.

METHODS

Shotgun metagenomics data processing and analysis

Longitudinal shotgun metagenomics sequencing data from healthy human stool samples was downloaded from NCBI BioProject accession PRJNA544527, and the associated metadata was downloaded from the associated article [32]. Raw FASTQ files were filtered and trimmed using FASTP [56], removing the first 5 nucleotides of the read 5' end to avoid leftover primer and adapter sequencing not removed during demultiplexing and an adaptive sliding window filter on the 3' end of the read with a required minimum quality score of 20. Reads containing ambiguous base calls, having a mean quality score less than 20, or with a length smaller than 50nt after trimming were removed from the analysis. Taxonomic assignment on the read level was performed with Kraken2 using the Kraken2 default database [57]. Abundances on the kingdom, phylum, genus, and species ranks were then obtained using Bracken [58]. Trimmed and filtered reads were then aligned to 2,935 representative bacterial reference genomes taken from the IGG database (version 1.01) using Bowtie2 [59, 60]. Coverage profiles and log2 estimates of peak-to-trough ratios were estimated using COPTR v1.1.2 on the species level within each sample [61]. PTR estimates were then merged with Bracken abundance estimates, retaining only those species identified by both methods (Kraken2 and Bowtie2 alignment to IGGdb).

The processed data containing the raw reads and $\log_2$ peak-to-trough ratios ($\log_2$PTRs) were read into R version 4.1.3 for analysis ([62]). All plots were generated using ggplot2 [63], unless indicated otherwise. Donor time series were selected by only retaining individuals with over 50 metagenomic time points, resulting in four time series (i.e., donors ae, am, an, and ao).

Distinct *Bacteroides ovatus* strains across all four donors contained duplicated taxon names with unique taxonomic identifiers, and were renamed to "*Bacteroides ovatus_1*" and "*Bacteroides ovatus_2.*" Raw read counts for a given taxon within a sample were centered log-ratio (CLR) transformed [64]. Taxa that had matched $\log_2$PTR information available across more than 5 time points within an individual, with time differences between samples less than three days, were used in subsequent analyses. Changes in normalized abundance were calculated as $Abundance\ changes(delta) = x(t+1) - x(t)$,where $\Delta t < 3\ days$. To assess the regression-to-the-mean effect, CLR-normalized abundances were plotted against deltas for each taxon, and the regression coefficients, aggregating all microbial taxa, were plotted as boxplots (showing median and interquartile range), summarized by donor.

For each donor, to estimate the growth phase of each individual taxon, we used linear regression of CLR-normalized abundances vs. $\log_2$PTRs, followed by a Benjamini-Hochberg p-value correction to control for the false discovery rate (FDR) in base R. FDR-adjusted p-values < 0.05 were considered significant. Taxa with significantly positive or negative associations were considered to be in lag-acceleration or deceleration-stationary phase, respectively. Those with no correlation were not assigned a growth phase, as this result could either be a false negative or indicative of log-phase growth. Linear regression was also used to test whether or not average CLR-normalized abundances and average $\log_2$PTRs were significantly associated within each donor, and p-values from individual tests were combined using Fisher's method [65].

Stochastic logistic growth model simulation

The stochastic logistic growth equation (SLE) was implemented as: $\frac{dx_i}{dt} = rx_i(t)\left(1 - \frac{x_i(t)}{K}\right) + \sigma x_i(t)\omega(t)$, where *t* is time, *r* is the growth rate, x_i is the abundance of taxon *i*, *K* is the carrying capacity, σ is the noise magnitude term, and $\omega(t)$ is the noise distribution term. Using the R

package sde [66], taxonomic growth was simulated with $x_{i,0} = 1$, $t_0 = 1$ to $t_{final} = 100$, for 100 iterations. The other parameters were varied as described in the results and below. To investigate the impact of noise on sLGE trajectories, noise levels were set from 0.001 to 1, with *r* and *K* ranging from 1 to 3 and 10 to 1000, respectively. To investigate the statistical relationships between deltas and abundances across growth phases and across model parameterizations, Pearson's R coefficients and p-values were calculated for each of the three growth phase categories. The growth phases for each model parameterization were defined using the non-stochastic logistic growth equation (LGE): $\frac{dx_i}{dt} = rx_i(t)\left(1 - \frac{x_i(t)}{K}\right)$, the solution for which can be written as $x_i = \frac{x_{i,0}Ke^{rt}}{(K - x_{i,0}) + x_{i,0}e^{rt}}$.

The x_i values for each simulated time point from solving the LGE were used to calculate the first derivative (i.e., the growth rate), which is exactly equal to the LGE. The second derivative (i.e., growth acceleration), $\frac{d^2x_i}{dt^2} = K^2x_i\left(1 - \frac{x_i}{K}\right)\left(1 - \left(\frac{2x_i}{K}\right)\right)$, was calculated using solved x_i values. Growth phases from the SLM were defined using the second derivative curves. First, the intersections of the acceleration curve and the half-max, a_1 and a_2, and the half-min, a_3 and a_4, were calculated (Fig. S1). The corresponding simulated time points of a_j, denoted as s_j, where *j = 1 - 4*, were then used to define growth phases as follows: lag phase: $t < s_1$; acceleration phase: $s_1 < t < s_2$; log phase: $s_2 < t < s_3$; deceleration phase: $s_3 < t < s_4$; and stationary phase: $t > s_4$. Here, lag and acceleration, and deceleration and stationary phases were combined, as these phases display similar delta-abundance relationships along the logistic growth curve. Conceptual diagrams were created using BioRender.

Data and code availability

Nextflow pipelines implementing the processing of metagenomic shotgun sequencing data from raw reads to taxonomic abundance matrices and PTR estimates can be found at

https://github.com/Gibbons-Lab/pipelines/ (metagenomics pipelines). R scripts and code used to analyze the data, run the sLGE simulations, and produce the figures in the manuscript have been deposited at https://github.com/Gibbons-Lab/human-microbiome-time-series-growth-phase-estimation.

Acknowledgements

We would like to thank Shijie Zhao for suggesting that we investigate PTR-abundance relationships in the BIO-ML data set. We would also like to thank Nitin Baliga, Amy Willis, Julia Cui, and the members of the Gibbons Lab for helpful discussions of this work. SMG and CD were supported by a Washington Research Foundation Distinguished Investigator Award and by startup funds from the Institute for Systems Biology. JL was supported by the Environmental Pathology/Toxicology training grant (ES007032).

References

1. El Aidy S, Hooiveld G, Tremaroli V, Bäckhed F, Kleerebezem M. The gut microbiota and mucosal homeostasis: colonized at birth or at adulthood, does it matter? *Gut Microbes* 2013; **4**: 118–124.
2. Martin AM, Sun EW, Rogers GB, Keating DJ. The Influence of the Gut Microbiome on Host Metabolism Through the Regulation of Gut Hormone Release. *Front Physiol* 2019; **10**: 428.
3. Zhang P, Meng X, Li D, Calderone R, Mao D, Sui B. Commensal Homeostasis of Gut Microbiota-Host for the Impact of Obesity. *Front Physiol* 2017; **8**: 1122.
4. Stefan KL, Kim MV, Iwasaki A, Kasper DL. Commensal Microbiota Modulation of Natural Resistance to Virus Infection. *Cell* 2020; **183**: 1312–1324.e10.
5. Fischbach MA. Microbiome: Focus on Causation and Mechanism. *Cell* 2018; **174**: 785–790.

6. Zmora N, Zilberman-Schapira G, Suez J, Mor U, Dori-Bachash M, Bashiardes S, et al. Personalized Gut Mucosal Colonization Resistance to Empiric Probiotics Is Associated with Unique Host and Microbiome Features. *Cell* 2018; **174**: 1388–1405.e21.
7. Suez J, Zmora N, Zilberman-Schapira G, Mor U, Dori-Bachash M, Bashiardes S, et al. Post-Antibiotic Gut Mucosal Microbiome Reconstitution Is Impaired by Probiotics and Improved by Autologous FMT. *Cell* 2018; **174**: 1406–1423.e16.
8. Bullman S, Pedamallu CS, Sicinska E, Clancy TE, Zhang X, Cai D, et al. Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer. *Science* 2017; **358**: 1443–1448.
9. Halfvarson J, Brislawn CJ, Lamendella R, Vázquez-Baeza Y, Walters WA, Bramer LM, et al. Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat Microbiol* 2017; **2**: 17004.
10. Cuna A, Morowitz MJ, Ahmed I, Umar S, Sampath V. Dynamics of the preterm gut microbiome in health and disease. *Am J Physiol Gastrointest Liver Physiol* 2021; **320**: G411–G419.
11. Fisher CK, Mehta P. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PLoS One* 2014; **9**: e102451.
12. Bashan A, Gibson TE, Friedman J, Carey VJ, Weiss ST, Hohmann EL, et al. Universality of human microbial dynamics. *Nature* 2016; **534**: 259–262.
13. Stein RR, Bucci V, Toussaint NC, Buffie CG, Rätsch G, Pamer EG, et al. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLoS Comput Biol* 2013; **9**: e1003388.
14. 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.
15. Gibbons SM, Kearney SM, Smillie CS, Alm EJ. Two dynamic regimes in the human gut microbiome. *PLoS Comput Biol* 2017; **13**: e1005364.

16. Carr A, Diener C, Baliga NS, Gibbons SM. Use and abuse of correlation analyses in microbial ecology. *ISME J* 2019; **13**: 2647–2655.
17. Momeni B, Xie L, Shou W. Lotka-Volterra pairwise modeling fails to capture diverse pairwise microbial interactions. *Elife* 2017; **6**.
18. Berry D, Widder S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front Microbiol* 2014; **5**: 219.
19. Freilich MA, Wieters E, Broitman BR, Marquet PA, Navarrete SA. Species co-occurrence networks: Can they reveal trophic and non-trophic interactions in ecological communities? *Ecology* 2018; **99**: 690–699.
20. Coenen AR, Weitz JS. Limitations of Correlation-Based Inference in Complex Virus-Microbe Communities. *mSystems* 2018; **3**.
21. Korem T, Zeevi D, Suez J, Weinberger A, Avnit-Sagi T, Pompan-Lotan M, et al. Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* 2015; **349**: 1101–1106.
22. Brown CT, Olm MR, Thomas BC, Banfield JF. Measurement of bacterial replication rates in microbial communities. *Nat Biotechnol* 2016; **34**: 1256–1263.
23. Gibson B, Wilson DJ, Feil E, Eyre-Walker A. The distribution of bacterial doubling times in the wild. *Proc Biol Sci* 2018; **285**.
24. Monod J. THE GROWTH OF BACTERIAL CULTURES. *Annu Rev Microbiol* 1949; **3**: 371–394.
25. Logistic stochastic growth models and applications. *Handbook of the Logistic Distribution*. 1991. CRC Press, pp 419–419.
26. Descheemaeker L, de Buyl S. Stochastic logistic models reproduce experimental time series of microbial communities. *Elife* 2020; **9**.
27. Grilli J. Macroecological laws describe variation and diversity in microbial communities. *Nat Commun* 2020; **11**: 4743.

28. Lobry JR, Flandrois JP, Carret G, Pave A. Monod's bacterial growth model revisited. *Bull Math Biol* 1992; **54**: 117–122.
29. Fujikawa H, Kai A, Morozumi S. A new logistic model for bacterial growth. *Shokuhin Eiseigaku Zasshi* 2003; **44**: 155–160.
30. Baranyi J, McClure PJ, Sutherland JP, Roberts TA. Modeling bacterial growth responses. *J Ind Microbiol* 1993; **12**: 190–194.
31. Ho P-Y, Good BH, Huang KC. Competition for fluctuating resources reproduces statistics of species abundance over time across wide-ranging microbiotas. *Elife* 2022; **11**.
32. Poyet M, Groussin M, Gibbons SM, Avila-Pacheco J, Jiang X, Kearney SM, et al. A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. *Nat Med* 2019; **25**: 1442–1452.
33. Kadosh E, Snir-Alkalay I, Venkatachalam A, May S, Lasry A, Elyada E, et al. The gut microbiome switches mutant p53 from tumour-suppressive to oncogenic. *Nature* 2020; **586**: 133–138.
34. Kolodziejczyk AA, Zheng D, Shibolet O, Elinav E. The role of the microbiome in NAFLD and NASH. *EMBO Mol Med* 2019; **11**.
35. Day JA, Diener C, Otwell AE, Tams KE, Bebout B, Detweiler AM, et al. Lettuce (Lactuca sativa) productivity influenced by microbial inocula under nitrogen-limited conditions in aquaponics. *PLoS One* 2021; **16**: e0247534.
36. Levy R, Borenstein E. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proc Natl Acad Sci U S A* 2013; **110**: 12804–12809.
37. Otwell AE, Carr AV, Majumder ELW, Ruiz MK, Wilpiszeski RL, Hoang LT, et al. Sulfur Metabolites Play Key System-Level Roles in Modulating Denitrification. *mSystems* 2021; **6**.
38. Wallace RJ, Snelling TJ, McCartney CA, Tapio I, Strozzi F. Application of meta-omics techniques to understand greenhouse gas emissions originating from ruminal metabolism.

Genet Sel Evol 2017; **49**: 9.

39. Roy P, Dutta A, Deen B. Greenhouse gas emissions and production cost of ethanol produced from biosyngas fermentation process. *Bioresour Technol* 2015; **192**: 185–191.
40. Zhong C. Industrial-Scale Production and Applications of Bacterial Cellulose. *Front Bioeng Biotechnol* 2020; **8**: 605374.
41. Guzman-Rodriguez M, McDonald JAK, Hyde R, Allen-Vercoe E, Claud EC, Sheth PM, et al. Using bioreactors to study the effects of drugs on the human microbiota. *Methods* 2018; **149**: 31–41.
42. Blätke M-A, Bräutigam A. Evolution of C4 photosynthesis predicted by constraint-based modelling. *Elife* 2019; **8**.
43. Atolia E, Cesar S, Arjes HA, Rajendram M, Shi H, Knapp BD, et al. Environmental and Physiological Factors Affecting High-Throughput Measurements of Bacterial Growth. *MBio* 2020; **11**.
44. Nguyen J, Lara-Gutiérrez J, Stocker R. Environmental fluctuations and their effects on microbial communities, populations and individuals. *FEMS Microbiol Rev* 2021; **45**.
45. Tsoularis A, Wallace J. Analysis of logistic growth models. *Math Biosci* 2002; **179**: 21–55.
46. O'Sullivan O, Cronin O, Clarke SF, Murphy EF, Molloy MG, Shanahan F, et al. Exercise and the microbiota. *Gut Microbes* 2015; **6**: 131–136.
47. Long AM, Hou S, Ignacio-Espinoza JC, Fuhrman JA. Benchmarking microbial growth rate predictions from metagenomes. *ISME J* 2021; **15**: 183–195.
48. Szafrańska AK, Junker V, Steglich M, Nübel U. Rapid cell division of Staphylococcus aureus during colonization of the human nose. *BMC Genomics* 2019; **20**: 229.
49. Gao Y, Li H. Quantifying and comparing bacterial growth dynamics in multiple metagenomic samples. *Nat Methods* 2018; **15**: 1041–1044.
50. Braga RM, Dourado MN, Araújo WL. Microbial interactions: ecology in a molecular perspective. *Braz J Microbiol* 2016; **47 Suppl 1**: 86–98.

51. Zelezniak A, Andrejev S, Ponomarova O, Mende DR, Bork P, Patil KR. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc Natl Acad Sci U S A* 2015; **112**: 6449–6454.
52. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol* 2010; **28**: 245–248.
53. Pramanik J, Keasling JD. Stoichiometric model of Escherichia coli metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnol Bioeng* 1997; **56**: 398–421.
54. Dikicioglu D, Kırdar B, Oliver SG. Biomass composition: the 'elephant in the room' of metabolic modelling. *Metabolomics* 2015; **11**: 1690–1701.
55. Schulz C, Kumelj T, Karlsen E, Almaas E. Genome-scale metabolic modelling when changes in environmental conditions affect biomass composition. *PLoS Comput Biol* 2021; **17**: e1008528.
56. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018; **34**: i884–i890.
57. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019; **20**: 257.
58. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput Sci* 2017; **3**: e104.
59. Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC. New insights from uncultivated genomes of the global human gut microbiome. *Nature* 2019; **568**: 505–510.
60. Langmead B, Wilks C, Antonescu V, Charles R. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* 2019; **35**: 421–432.
61. Joseph TA, Chlenski P, Litman A, Korem T, Pe'er I. Accurate and robust inference of microbial growth dynamics from metagenomic sequencing reveals personalized growth rates. *Genome Res* 2022; **32**: 558–568.

62. Team RC. RA language and environment for statistical computing. R Foundation for Statistical Computing Vienna Austria URL (2022).
63. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2016. Springer.
64. Lin H, Peddada SD. Analysis of microbial compositions: a review of normalization and differential abundance analysis. *NPJ Biofilms Microbiomes* 2020; **6**: 60.
65. Fisher SRA. Statistical Methods for Research Workers. 1925. Oliver and Boyd.
66. Iacus SM. SDE : simulation and inference for stochastic differential equations. 2007.

FIGURES

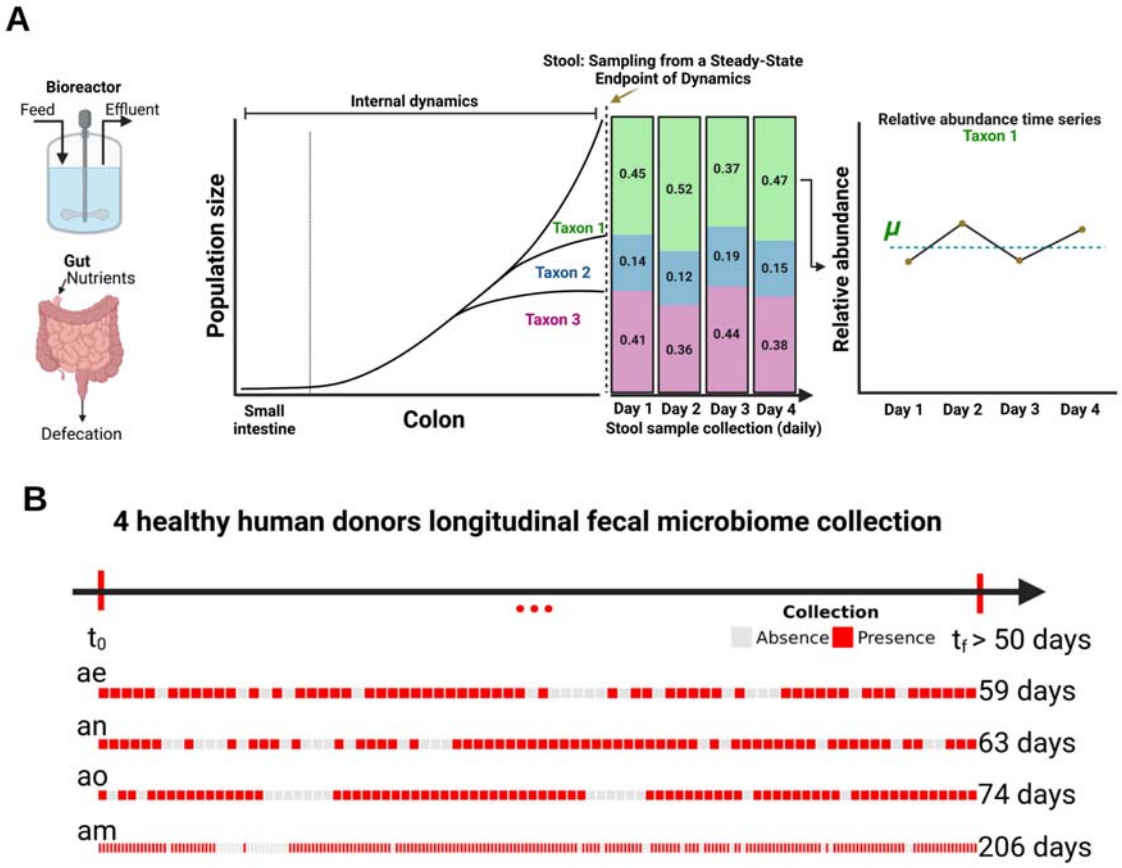

Figure 1. Conceptual figure showing two flow-through microbial ecosystems: a bioreactor and a human gut. **A.** Both bioreactors and guts are continuous flow-through systems. Prior to reaching the measured abundances in stool, taxa grow in the large intestine with varying growth rates, carrying capacities, and steady-state population sizes, which may be in different growth phases at the time of measurement. For example, see dynamics for Taxa 1-3. Daily stool collections show variation in abundances, but this variation likely does not reflect internal growth dynamics in the gut. **B.** Healthy BIO-ML stool donors (subject IDs: ae, am, an, and ao) with samples collected 3-5 days per week for a total of >50 time points. Red indicates presence of shotgun metagenomic sequencing data and gray represents absence of metagenomic data from consecutive daily time points.

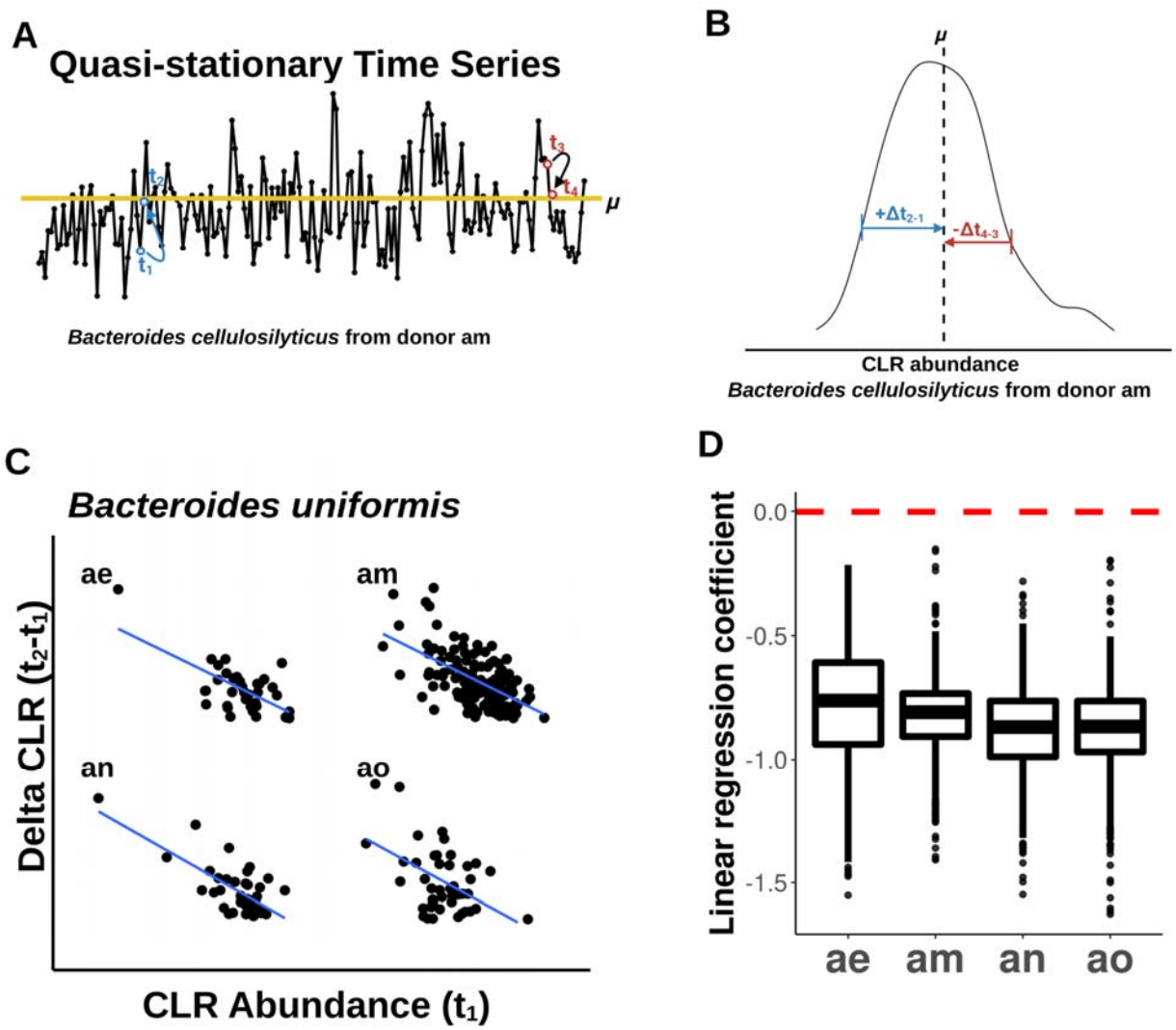

Figure 2. Regression-to-the-mean effect in human microbial time series data. A. Yellow line represents the mean abundance (μ) of *Bacteroides cellulosilyticus* over time in donor am. Time points t_1 and t_3 indicate fluctuations below and above the mean abundance, and t_2 and t_4 show the return to the mean abundance. **B.** Distribution of time series delta values (e.g., t_2-t_1) for *Bacteroides cellulosilyticus* in donor am, which is approximately normally distributed. **C.** Deltas vs. abundances for *Bacteroides uniformis* time series from donors ae, am, an, and ao. **D.** Boxplots (showing median and interquartile range) of linear regression coefficients for deltas vs. abundances across all taxa time series in all four donors. Red line indicates a regression coefficient of 0.

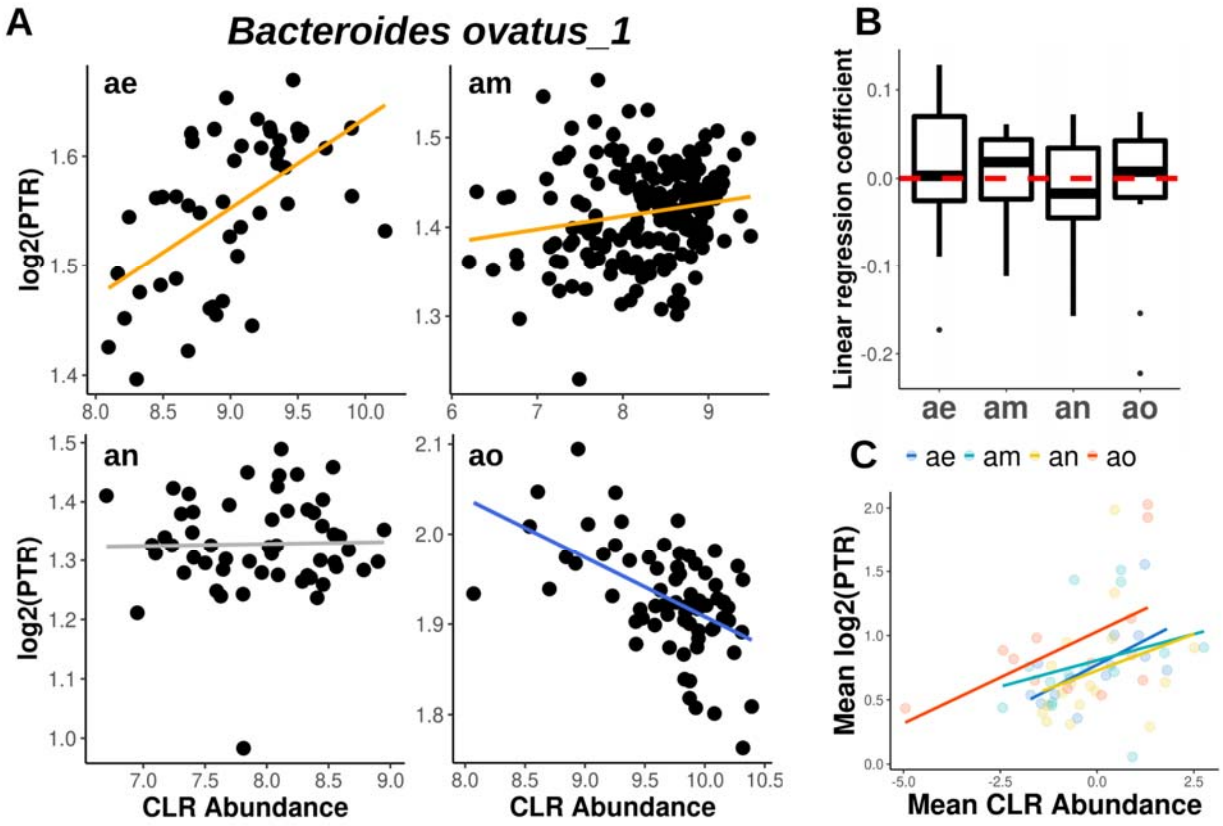

Figure 3. Variable relationships between PTRs and CLR-normalized abundances across human gut microbial time series. A. Log_2(PTR) and CLR-normalized abundance relationships for donors ae, am, an, and ao. Orange and blue lines show significantly positive and negative linear regression coefficients (linear regression, FDR adjusted *p*-value < 0.05), respectively. Gray lines indicate no statistically significant association. **B.** Boxplots (showing median and interquartile range) of linear regression coefficient combined for all filtered taxa for each donor. **C.** Mean log_2(PTR) and mean CLR-normalized abundance for all abundant taxa in each donor (*p-values* for regressions run within each donor were combined using Fisher's method; combined *p-value* = 0.005).

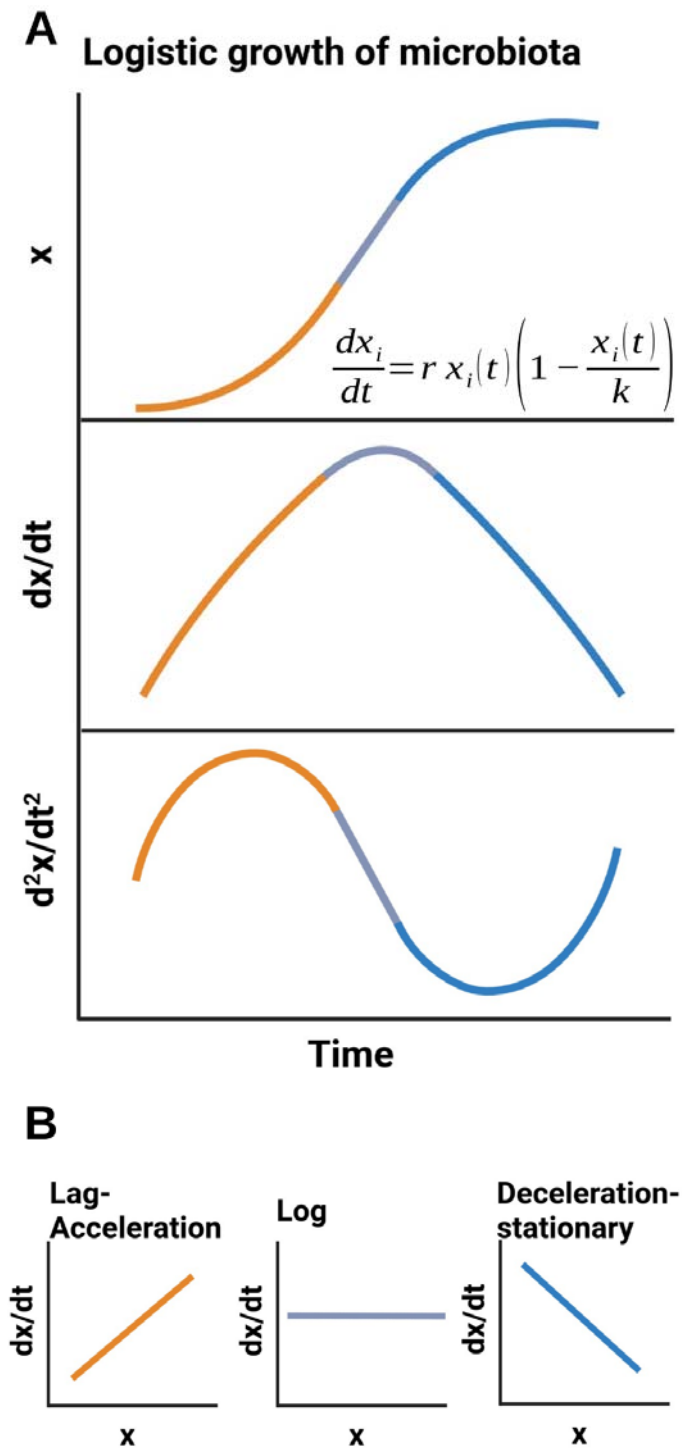

Figure 4. Logistic growth model. A. The logistic growth curve models abundance (x) with respect to time (top panel). The first derivative of the logistic growth curve models the growth rate with respect to time (middle panel). The second derivative of the logistic growth curve models growth rate acceleration with respect to time (bottom panel). **B.** Expected relationships between abundance and growth rate at different locations along the logistic growth curve.

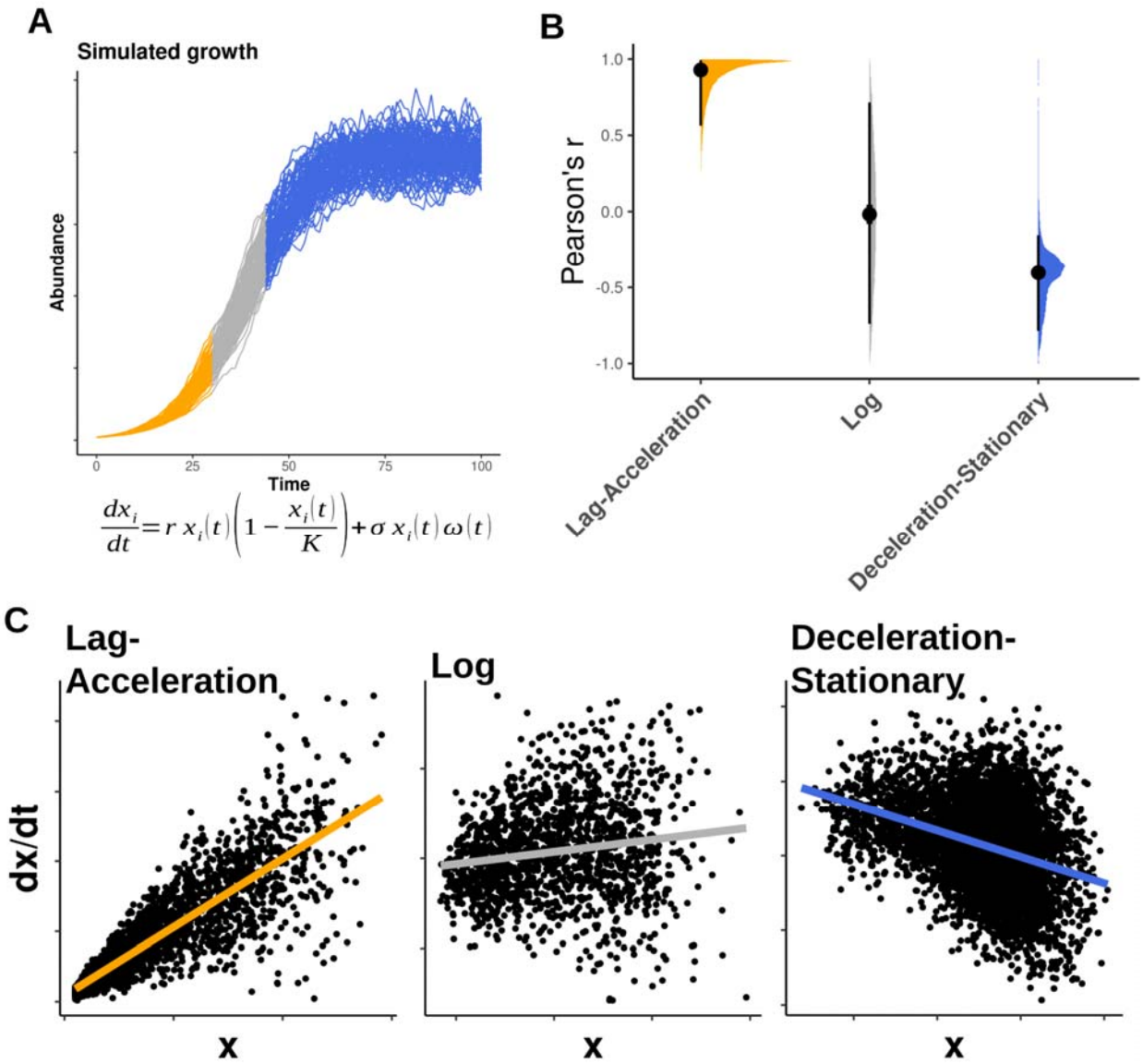

Figure 5. Distinguishing growth phases using the stochastic logistic growth model. A. Stochastic logistic growth curves with growth rate (r) = 1.2, carrying capacity (k) = 100, and noise level (n) = 0.1 across 100 iterations. Major growth phase groups in orange (lag-acceleration), gray (log), and blue (deceleration-stationary). **B.** Pearson's R values between abundances and growth rates in each of our three growth phase windows across variable model parameterizations (r = 1-3, k = 10-1000) and a fixed noise level (□ = 0.1). **C.** Scatter plots showing relationships between abundances and deltas across the three growth phase regions defined in panel A.

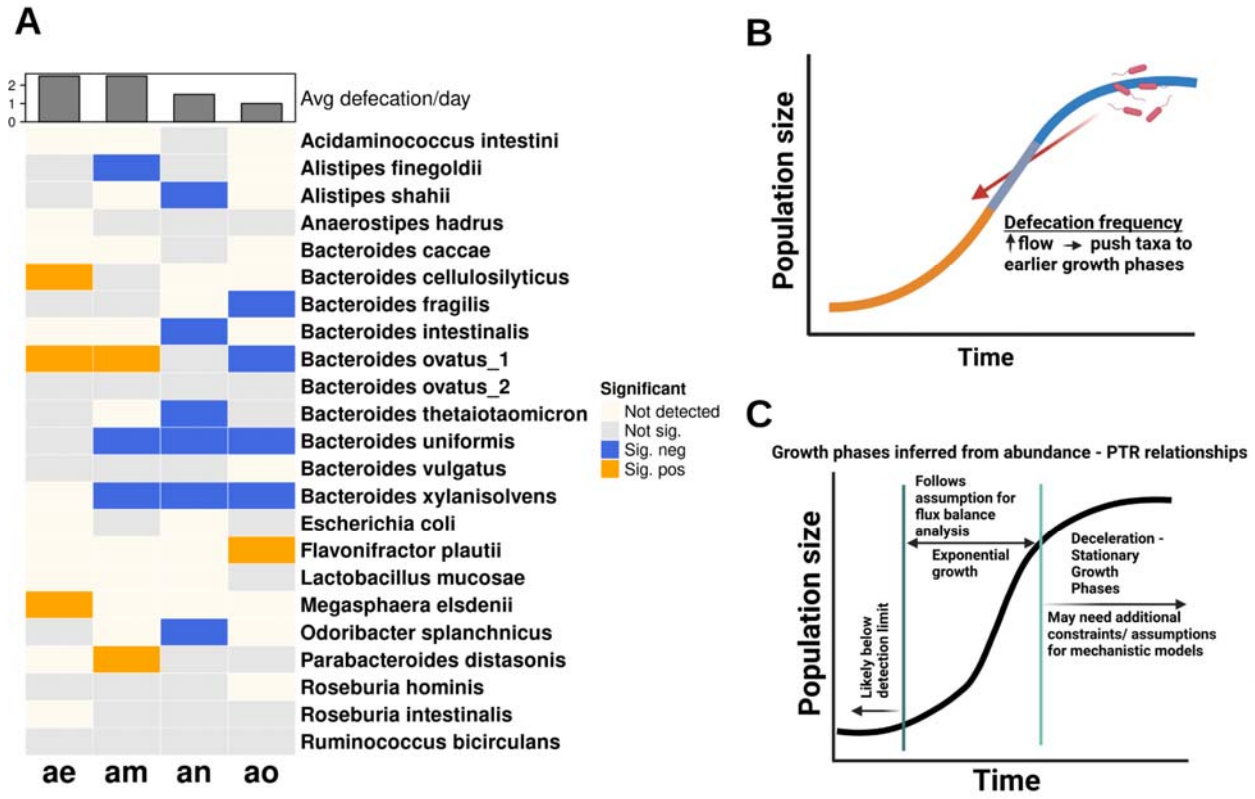

Figure 6. *in vivo* growth phase estimation. A. We find variable relationships between PTRs and population abundances across taxa in each of the four donors, consistent with the growth phase patterns observed in sLGE simulations. Donors with higher defecation rates tended to have a larger fraction of taxa with positive PTR-abundance associations and fewer with negative associations, indicating lag-acceleration and deceleration-stationary phases, respectively. **B.** We suggest that higher defecation rates (i.e., higher dilution rates) push bacterial populations towards earlier growth phases, which is consistent with our results in panel A. **C.** Growth phase estimates can be leveraged to identify taxa that are more-or-less amenable to metabolic modeling techniques, like Flux Balance Analysis, which assumes exponential growth.

SUPPLEMENTAL FIGURES

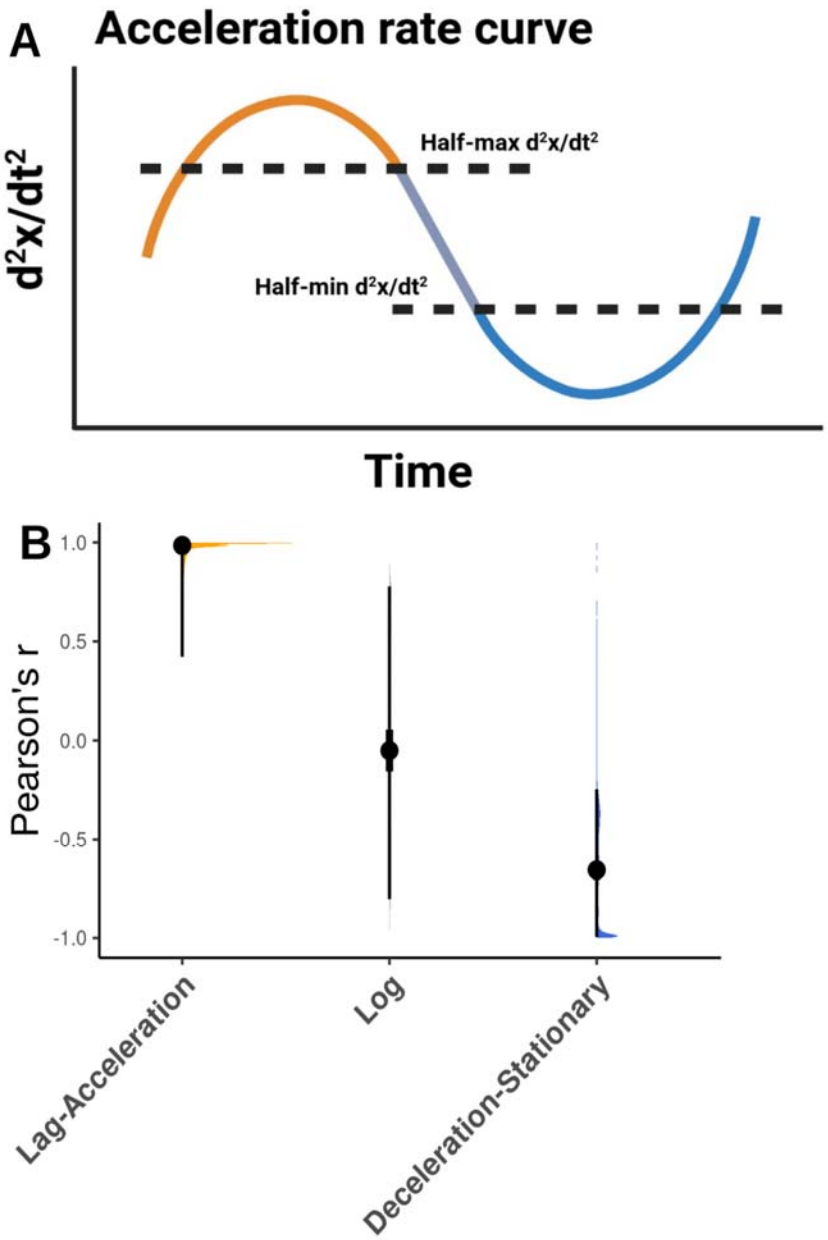

Figure S1. Definition of major growth phases using the stochastic logistic growth model. A. The half-maximum of the peak and half-minimum of the trough of the second derivative of abundance were used to define growth phases across model parameterizations. **B.** Pearson's R values between abundances and growth rates in the three growth phase categories obtained from combined sLGE simulation results across a range of growth rates (r = 1-3), carrying capacities (k = 10-1000), and noise levels (n = 0.001-1).

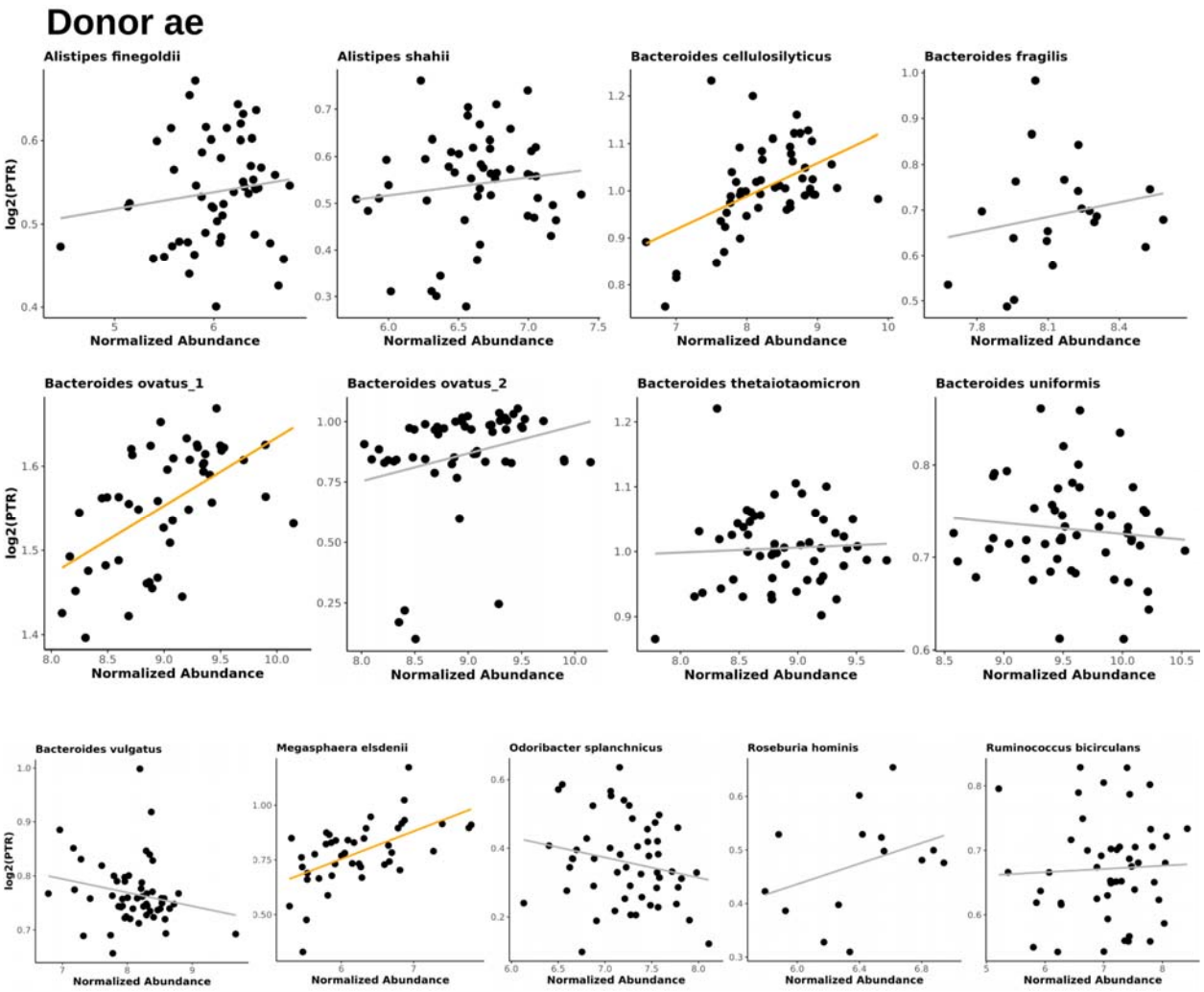

Figure S2. Relationships between abundance and log2(PTR) for abundant taxa in donor ae. Abundant taxa with relatively dense longitudinal PTR and abundance data (at least 5 matched data points; time differences between adjacent samples less than three days) were selected for analysis. Gray trend lines show no significant correlations, orange trend lines indicate significant positive correlations, and blue trend lines represent significant negative correlations (linear regression, BH-FDR < 0.05).

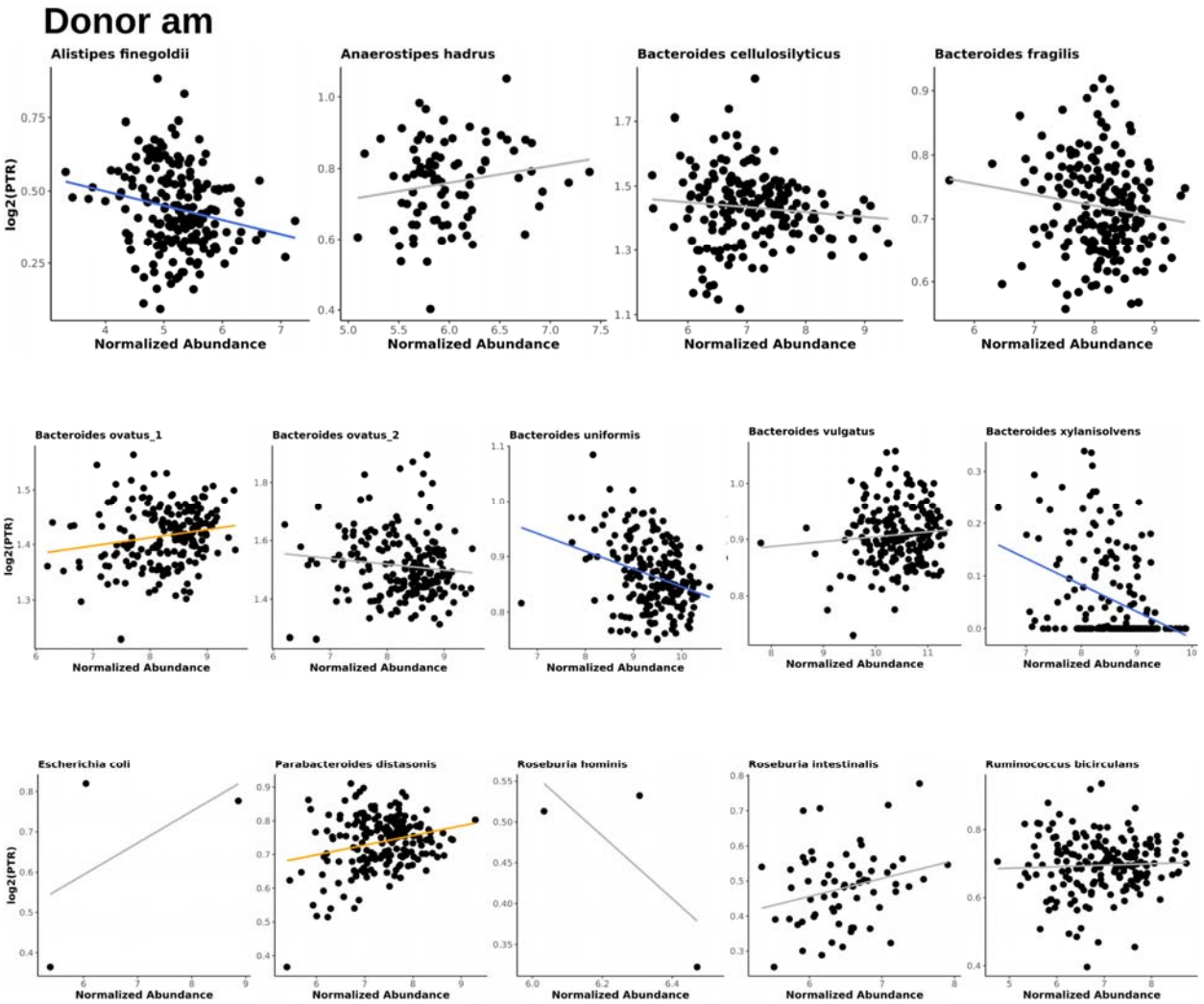

Figure S3. Relationships between abundance and log2(PTR) for abundant taxa in donor am. Abundant taxa with relatively dense longitudinal PTR and abundance data (at least 5 matched data points; time differences between adjacent samples less than three days) were selected for analysis. Gray trend lines show no significant correlations, orange trend lines indicate significant positive correlations, and blue trend lines represent significant negative correlations (linear regression, BH-FDR < 0.05).

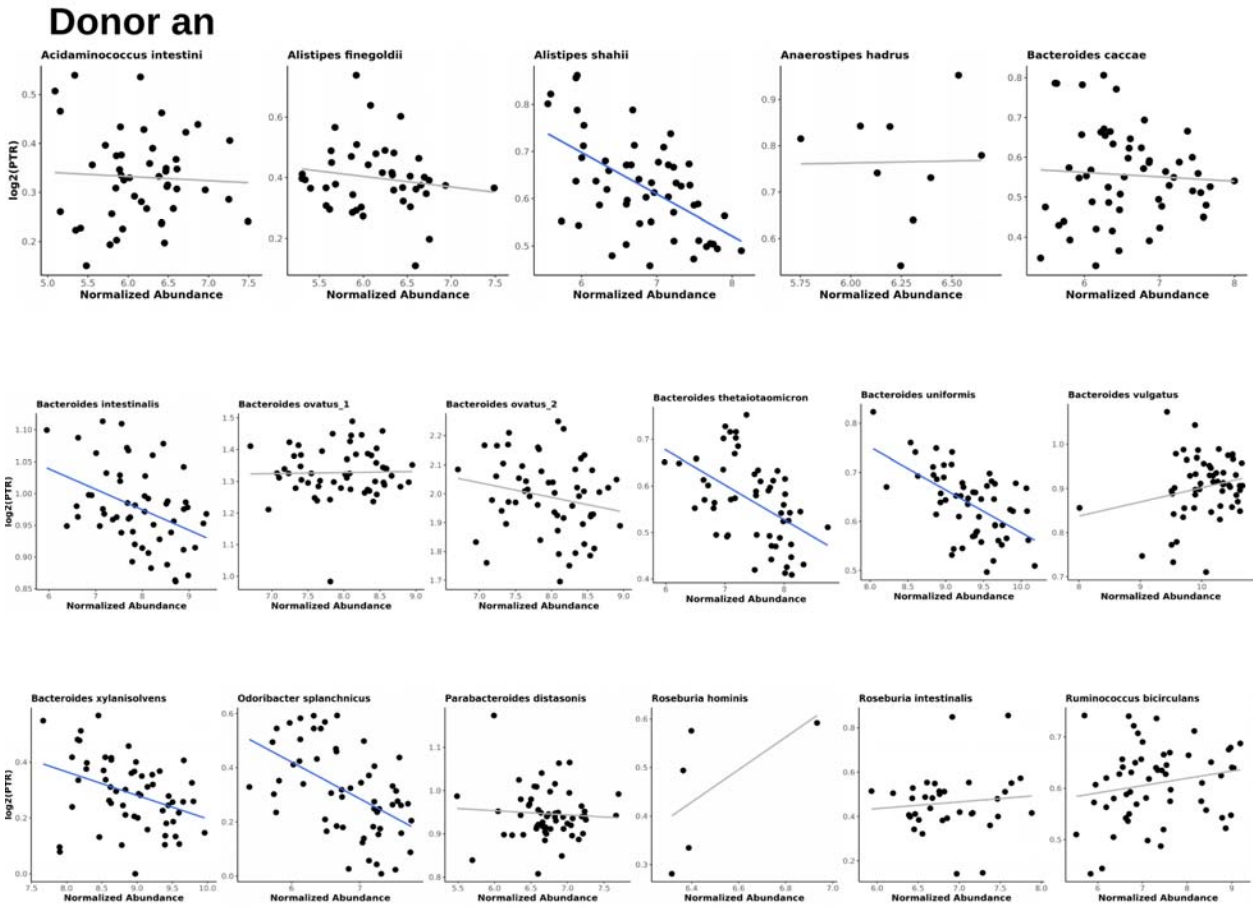

Figure S4. Relationships between abundance and log2(PTR) for abundant taxa in donor an. Abundant taxa with relatively dense longitudinal PTR and abundance data (at least 5 matched data points; time differences between adjacent samples less than three days) were selected for analysis. Gray trend lines show no significant correlations, orange trend lines indicate significant positive correlations, and blue trend lines represent significant negative correlations (linear regression, BH-FDR < 0.05).

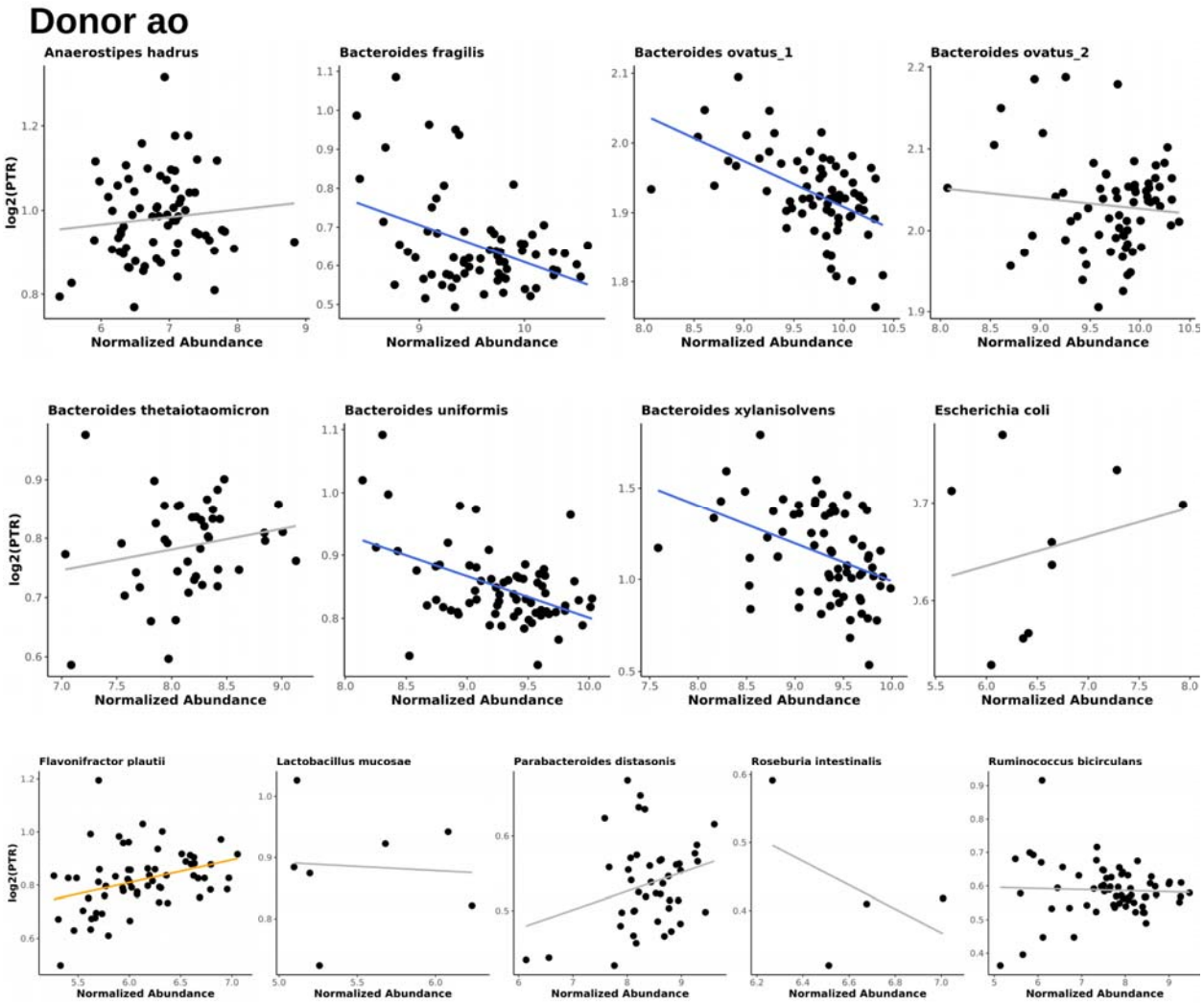

Figure S5. Relationships between abundance and log2(PTR) for individual taxon in donor ao. Abundant taxa with relatively dense longitudinal PTR and abundance data (at least 5 matched data points; time differences between adjacent samples less than three days) were selected for analysis. Gray trend lines show no significant correlations, orange trend lines indicate significant positive correlations, and blue trend lines represent significant negative correlations (linear regression, BH-FDR < 0.05).