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

Authors: Joe J. Lim[1], Christian Diener[2], James Wilson[2], Jacob J. Valenzuela[2], Nitin S. Baliga[2-3], Sean M. Gibbons[2,4-6,*]

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 Microbiology, University of Washington, Seattle, WA 98105; [4] Department of Bioengineering, University of Washington, Seattle, WA 98105; [5] Department of Genome Sciences, University of Washington, Seattle, WA 98105; [6] 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 flow-through ecosystem at a dynamical steady state, where population dynamics occur internally and the bacterial population sizes measured in a bolus of stool represent an endpoint of these internal dynamics. We formalize this idea as stochastic logistic growth of a population in a system held at a semi-constant flow rate. We show how this model provides a path toward estimating the growth phases of gut bacterial populations *in situ*. We validate our model predictions using an *in vitro Escherichia coli* growth experiment. Finally, we show how this method can be applied to densely-sampled human stool metagenomic time series data. Consistent with our model, stool 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 is an anaerobic flow-through 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].

Given these sampling limitations, and in the absence of major perturbations that require multi-day recovery processes in the human gut, it is unclear whether or not meaningful insights into commensal population dynamics can be gleaned from adult human gut microbiome time series. One workaround for inferring growth rates of bacterial populations *in situ* is to leverage metagenome-inferred replication rates [21,22]. 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. 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 [26–32]. We used the sLGE to study statistical relationships between population sizes and growth rates across the various phases of growth (i.e., acceleration, mid-log, deceleration, stationary phases) to see if we could extract *in situ* growth phase information. Overall, the sLGE model yields statistical relationships that may be leveraged to identify the *in situ* growth phase of a bacterial population sampled at a regular period from a flow-through ecosystem, like the human gut.

To assess our model predictions, we sampled *Eschericial coli* populations at different points along the growth curve. We calculated population sizes and replication rates for these samples and observed excellent agreement between this *in vitro* model and our sLGE predictions. We also measured population abundance and replication rate trajectories from more than a dozen organisms across four densely sampled human gut metagenomic time series [33]. On average, gut commensal growth rates and population sizes were positively correlated, both cross-sectionally over 84 stool donors and longitudinally within each of four stool donor time series, which suggests that most abundant taxa in the gut are growing exponentially when sampled in stool. Furthermore, we were able to identify specific growth phase signatures in abundant bacterial populations in the guts of four individuals with long and

dense metagenomic time series by analyzing paired replication rate and abundance trajectories. We describe how our growth phase inference approach can serve to improve statistical inferences derived from microbiome data and 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 human health [8,34,35], agricultural systems [36,37], climate change [36,38,39], and industrial bioreactor production processes [40,41].

RESULTS

Framing the gut as an anaerobic flow-through bioreactor

The mammalian gut can be understood as an anaerobic batch culture reactor with a semi-continuous input (i.e., discrete boluses of dietary inputs, mixed with host substrates like mucin and bile acids) and output (i.e., discrete boluses of stool) [42], and microbial taxa must grow fast enough within the system to avoid dilution-to-extinction (**Fig. 1A**). Thus, stool sampling captures the endpoint of internal gut bacterial population dynamics. For example, in our conceptual 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) [33]. The BIO-ML cohort contained 84 donors [33]. 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 fecal shotgun 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 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 mirrors what one would expect when randomly sampling from a stationary distribution (**Fig. 2B**). 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 uniformly observed across all taxa analyzed, across all four donors (**Fig. 2D**). This negative association between population abundances and changes in abundance between time points is strongly consistent with sampling from a stationary distribution, which is equivalent to ‘regression-to-the-mean’ as an organism fluctuates around a fixed carrying capacity, similar to what we have reported previously [15,32].

One important ecological factor that can impact gut microbial dynamics is host diet [43,44]. Although changes in dietary intake can alter microbial abundance, average dietary choices are

highly conserved within an individual and these choices are notoriously difficult to modify outside of radical changes in geography or lifestyle [45–47]. Prior work demonstrated that macronutrient intake within an individual is largely stable over time and does not show significant autocorrelation or drift [15,48]. Indeed, for donor A from this prior study, we found that longitudinal measurements of macronutrients (i.e. daily intake of calories, carbohydrates, protein, fat, fiber, cholesterol, saturated fat, sugar, sodium, calcium) were stationary over several months, despite day-to-day fluctuations (**Fig. S1**). Combined with the overwhelming stationarity of microbial abundance trajectories within healthy individuals not undergoing major lifestyle changes [15,32,33], these results support our assertion that dietary patterns are largely stable over weeks-to-months and stool samples provide stable, steady-state population abundance estimates of abundant gut commensal bacteria.

Next, we looked at the statistical associations between calculated peak-to-trough ratios (i.e., PTRs; a proxy for growth-rate) for abundant bacterial populations from each metagenomic sample and their respective metagenomic population abundance estimates [22]. 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 $\log_2$PTR and centered log-ratio (CLR) transformed population abundances for the same taxon across the four donors (*Bacteroides ovatus*, **Fig. 3A**). Similarly, we saw a wide range of positive, negative, and null associations between $\log_2$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 [49–51]. Finally, we calculated temporally-averaged (i.e., mean for all collection time within a taxon) PTRs and population sizes for each abundant taxon within each of the four donors. Overall, there was a

significantly positive association (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) between average $\log_2$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 one would expect to observe in exponentially-growing populations. We also looked into whether or not $\log_2$PTR magnitudes were inter-comparable across taxa (**Fig. S2**). We calculated $\log_2$PTRs for all abundant taxa detected across all 84 BIO-ML donors and found that the median $\log_2$PTR was fairly similar across taxonomic classes (~0.45-0.75), with most classes showing a wide range of $\log_2$PTRs (**Fig. S2**). To assess whether or not $\log_2$PTR-CLR associations were robust to controlling for taxonomy, we included either class- or species-level categorizations as covariates in a linear regression model and saw a significant association, independent of taxonomy (class-level □ = 0.0612, $p = 8.359e^{-60}$; species-level □ = 0.0101, $p = 0.0006$).

Stochastic logistic growth equation provides insights into growth phases

In order to better understand and interpret the varying relationships we observe between $\log_2$PTRs and CLR abundance time series, we used a modeling approach. 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 $(1 - x_i(t)/k)$ [52]. 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 growth rate over time, which peaks during mid-log phase (**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 observed abundances of commensal bacterial populations in the gut fluctuate due to myriad factors including interspecies competition, resource fluctuations, technical noise, sampling noise, and stool residence time [27,32,53]. 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. Four growth phases (i.e., acceleration, mid-log, deceleration, and stationary) were defined using the half-maximum and half-minimum, respectively, of the second derivative of the LGE curve (**Fig. S3A**). 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 four major growth phase categories (**Fig. 5A-C**). For example, Pearson's correlations between growth rates and abundances were significantly positive in the acceleration phase and significantly negative in the deceleration phase (**Fig. 5B**). Mid-log phase growth was more variable, but showed little-to-no significant association between growth rates and abundances (**Fig. 5B-C)**. These results were reproduced across a wide range of parameter space and were robust to varying noise levels (**Fig. S3B**).

Even though we expect dietary intake to be stationary within an individual, variation in diet can drive day-to-day fluctuations in the carrying capacities of microbial populations. In order to investigate whether growth-phase specific associations between abundances and growth

rates were influenced by fluctuations in carrying capacity, we added variation to *k* in the sLGE model (**Fig. S4**). Fluctuations in *k* did not alter the sigmoidal shape of the sLGE curve (**Fig. S4A**), and the relationships between abundances and growth rates across growth phases were preserved (**Fig. S4B-C**).

Validating sLGE growth phase inferences in vitro

To validate the relationship between growth rates and abundances across growth phases, we cultured replicate *E. coli* populations *in vitro* and sampled them across their growth curves (**Fig. 6A**). *E. coli* abundances were measured as OD600 values and as the log-ratio of *E. coli* reads to phiX reads (i.e., a fixed amount of the phiX genome was spiked into each DNA extraction) from the shotgun sequencing data (**Fig. 6A-C**). Growth rates were quantified as the $\log_2$PTR for each *E. coli* sample [54]. The relationships between the $\log_2$PTRs and CLR-normalized *E. coli* abundances across growth phases matched the sLGE model predictions (**Fig. 6B-C**). Specifically, growth rates and abundances were significantly positively and negatively correlated in acceleration and deceleration phases, respectively (**Fig. 6B-C**). Furthermore, we saw no significant association between growth rates and abundances in mid-log and stationary phases (**Fig. 6B-C**). Finally, we found that samples in mid-log phase had an average $\log_2$PTR of 1.25±0.167 (± standard deviation), while samples in stationary phase had an average $\log_2$PTR of 0.358±0.059, which clearly distinguished between these phases.

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

Based on the sLGE results and *in vitro* validation work presented above, we assigned putative *in situ* growth phases to abundant gut bacterial populations from the four BIO-ML gut metagenomic time series. The average magnitude of the PTR provides additional information on

whether a population is more likely to be in acceleration/mid-log/deceleration (i.e., $\log_2$PTR >> 0.358) or stationary (i.e., $\log_2$PTR < 0.358) phase (**Fig. 6**). For those taxa with average $\log_2$PTRs above the empirical stationary phase threshold, significantly positive associations (linear regression, adjusted *p*-value < 0.05, with a positive beta-coefficient) between $\log_2$PTRs and CLR abundances likely indicate acceleration-phase and significantly negative associations (linear regression, adjusted *p*-value < 0.05, with negative beta-coefficient) likely indicate deceleration phase. *Bacteroides cellulosilyticus*, *Bacteroides ovatus 1*, and *Megaspaera eldenii* showed significantly positive PTR-abundance associations within donor ae (**Figs. 7A and S5**). *Bacteroides xylanisolvens* had an average $\log_2$PTR less than the stationary threshold in donor am (**Fig. S6**). *Bacteroides ovatus* 1 and *Parabacteroides distasonis* showed positive $\log_2$PTR-CLR abundance associations, while *Alistipes finegoldii*, and *Bacteroides uniformis* showed negative associations in donor am (**Figs. 7A and S6**). *Acidaminococcus intestini*, *Bacteroides xylanisolvens*, and *Odoribacter splanchnicus* showed average $\log_2$PTR below the empirical stationary phase threshold in donor an (**Fig. S7**). *Alistipes shahii*, *Bacteroides intestinalis*, *Bacteroides thetaiotaomicron*, and *Bacteroides uniformis* showed significantly negative $\log_2$PTR-CLR abundance associations in donor an (**Figs. 7A and S7**). Finally, *Favonifractor plautii* showed a positive $\log_2$PTR-CLR abundance association and *Bacteroides fragilis*, *Bacteroides ovatus* 1, *Bacteroides uniformis*, and *Bacteroides xylanisolvens* showed negative associations in donor ao (**Fig. 7A and S8**). In all four donors, many taxa showed average $\log_2$PTRs greater than the stationary threshold but without significant associations between $\log_2$PTR and CLR abundances (**Figs. 7A and S5-8**). The absence of a significant association for these non-stationary taxa likely indicates mid-log phase, but a non-significant association could also represent a false negative (i.e., not powered enough to detect a positive or negative association with the number of time points sampled).

We observed a slight difference in the number of significantly positive and negative PTR-abundance associations between donors ae/am, and an/ao. Donors ae and am tended to have a larger proportion of taxa in acceleration phase, while an and ao tended to have a larger proportion of taxa in deceleration or stationary phases. 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. 7B**). Overall, we were able to at least partially constrain our phase estimates for all taxa with sufficient longitudinal data (**Fig. 7A**). Our approach provides a new path toward providing constraints on *in situ* growth phases for 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 steady-state population fluctuating around a carrying capacity (**Figs. 1-2**). Thus, as prior work has indicated [15,32], 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 do not appear to be related to population growth, PTRs enable direct estimates of *in situ* growth rates from metagenomic samples [21–23,55–57]. 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-B**). While regression-to-the-mean is a plausible mechanism for the consistent negative delta-abundance relationships (**Fig. 2**), the underlying processes driving variable $\log_2$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, and we found clear diagnostic patterns (**Fig. 4**). Simulations showed a wide range of demographic stochasticity (**Fig. 5**) and fluctuations in carrying capacities (**Fig. S4**) could not ablate these patterns, although adding enough noise to these models eventually overrides the signal. We validated these patterns *in vitro* and saw marked correspondence between model predictions and empirical measurements (**Fig. 5-6**). Finally, we applied our sLGE predictions to four human gut metagenomic time series. Consistent with our predictions, we found that individuals with higher defecation rates tended to be enriched for taxa in earlier growth phases (**Fig. 7**). In a recent study, we observed a similar association between PTRs and bowel movement frequency (BMF) in another independent cohort, where PTRs appeared to increase with increasing BMF [58]. Overall, our results reveal a promising approach to inferring *in situ* growth phases for abundant organisms detected in human gut metagenomic time series.

We observed that the average $\log_2$PTR and average CLR abundance of a given taxon over time were positively correlated, which is consistent with exponentially-growing populations (**Fig. 3C**). However, despite this average pattern across taxa, we were also able to identify specific taxa that were abundant in stool that appeared to be in stationary phase (**Fig. 7A**). These results are highly 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 [59,60]. Genome-scale metabolic modeling and flux-balance analysis (FBA) has been effective mechanistic tools for simulating these metabolic exchanges, especially in controlled bioreactor systems [61]. The objective function used to find a solution subspace for these 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 mid-log phases, and to some extent in deceleration phase, but this assumption breaks down completely in stationary phase. Prior work has demonstrated that biomass composition can change depending on the growth phase of a population, which ideally could be taken into account to more accurately model metabolic fluxes within the system [62–64]. Alternatively, organisms that are not actively growing could be omitted from community-scale metabolic models of colonic metabolism [65]. Overall, our work suggests that most abundant organisms in human stool are amenable to FBA, and our growth phase estimation approach allows for the identification of abundant 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. Our results are somewhat reassuring for cross-sectional studies, as they indicate that bacterial abundances in the gut fluctuate around stable carrying capacities within an individual, making inter-individual comparisons fairly robust. Furthermore, this suggests that multi-day averages of abundances will be even more accurate estimates of this carrying capacity, as we have suggested previously

[33]. This work is especially relevant to the design and interpretation of human gut microbiome studies that aim to characterize or investigate ecosystem-scale dynamics. 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

Stationarity testing for daily nutrient intake in a human stool donor

Metadata for daily nutrient intake, excluding the time window when the donor was traveling abroad, was downloaded from David et al. [48]. We tested for stationarity in these nutrient intake time series using the augmented Dickey-Fuller (ADF) test (tseries package in R [66]), with significance threshold for stationarity at $p < 0.1$. ADF tests the null hypothesis that a unit root is present in a time series, with the alternative hypothesis being that the time series is stationary. Thus, significant p-values indicate stationarity of the time series. All analyses throughout the manuscript in R were conducted in R v4.2.2 [67], unless stated otherwise.

E. coli strain information and growth curve analysis with a microplate reader

Escherichia coli strain (MG1655) was streaked from a glycerol stock onto R2A agar plates (Thermo Fisher Scientific: Oxoid CM0906) and incubated overnight at 37°C. A colony was selected using an inoculating loop and transferred to 200 mL of LB-broth (Lennox) and grown at 37°C overnight in a shaking incubator at 225 rpm until the culture reached stationary phase. The overnight culture was then diluted in fresh LB medium to an OD of 0.51 (600 nm). The diluted culture was then chilled for ~25 minutes at ~2°C using an ice bath to synchronize metabolic activity. The chilled culture was then aliquoted (2µL) into a non-treated 96-well flat-bottomed plate (Thomas Scientific Cat No. 1154Q44) containing 198 µL of LB media (Lennox) in each

well. The inoculated plate was then transferred to a BioTek Epoch II plate reader set to 37°C with orbital shaking and programmed to make OD600 readings every minute for the first 60 minutes and every 5 minutes for the remainder of the experiment (~10 hours). The first set of inoculations covered plate rows A and B (n = 24), this was followed by the sequential inoculation of the next 3 sets of rows at 15-minute intervals (i.e., Set 1 = A/B: 0 min; Set 2 = C/D 15 min; Set 3 = E/F: 30 min; Set 4 = G/H 45 min.). This resulted in 4 sets of replicate cultures inoculated 15 minutes apart, allowing sampling every hour for the next 10 hours, spanning 40 time points spaced 15 minutes apart. To ensure there was enough DNA for sequencing at early low OD time-points (first two sample points), we pooled two wells into one sample. All samples were collected in PCR strip tubes (Axygen: PCR-0208-CP-C) and centrifuged at room temperature to pellet the cells. The supernatant was decanted and the remaining cell pellet was immediately frozen in liquid nitrogen for storage at -80°C.

DNA extraction, library preparation, and sequencing

Cell Pellets were resuspended and transferred to 96 deep-well plates for DNA extraction using the IBI Scientific 96-well Genomic DNA Bacteria Kit (IBI Scientific: IB47295) per the manufacturer's protocol. DNA quantification was done using Qubit HS DNA assay, on a Qubit3 device. After DNA quantification, we added PhiX DNA (Thermo Fisher Scientific: SD0031) as an internal standard and run-quality monitor across all samples. A total of 500 fg PhiX DNA was added to each DNA sample before library preparation. DNA libraries were constructed following the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs: E7805L) and indexed using Dual Index Primer Set 2 (New England Biolabs: E7780S). Libraries were quantified again via Qubit 3, and the quality and size of libraries were checked using an Agilent Tapestation, and a D5000 high sensitivity DNA tape assay. Libraries were pooled to 2 nM and sent to NovoGene for sequencing on a NovaSeq 6000 device (Illumina, USA). A partial lane

was used for sequencing, 150 cycles, generating ~64GB (~3.3 million reads per sample) of paired-end reads.

Shotgun metagenomics data processing and analysis

Longitudinal shotgun metagenomics sequencing data from healthy human stool samples (BIO-ML) was downloaded from NCBI BioProject accession PRJNA544527, and the associated metadata was downloaded from the associated article [33]. Raw FASTQ files from the BIO-ML cohort and from the *in vitro E. coli* experiment were filtered and trimmed using FASTP [68], 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 [69]. Abundances on the kingdom, phylum, genus, and species ranks were then obtained using Bracken [70]. Trimmed and filtered reads were then aligned to 2,935 representative bacterial reference genomes taken from the IGG database (version 1.01) using Bowtie2 [71,72]. Coverage profiles and $\log_2$ estimates of peak-to-trough ratios (PTRs) were estimated using COPTR v1.1.2 at the species-level within each sample for taxa that passed our abundance threshold [54]. PTR estimates were then merged with Bracken abundance estimates, retaining only those species identified by both methods (Kraken2 and Bowtie2 alignment to IGGdb). For the *in vitro E. coli* experiment, reads were aligned to a custom database containing the *E. coli K12* strain genome (NCBI accession NC_000913.3) and the phiX174 genome (NCBI accession NC_001422.1). CLR abundances were then calculated from the read counts for the *E. coli* genome and the phiX174 genome.

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 [67]. All plots were generated using ggplot2 [73], unless indicated otherwise. BIO-ML donors were selected by 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 [74]. Taxa that had matched $\log_2$PTR and CLR abundance 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. Regression coefficients with FDR-adjusted p-values < 0.05 were considered significant. Taxa with average $\log_2$PTRs < 0.358 (experimentally-determined stationary threshold) were designated as being in stationary phase. For those taxa not designated as being in stationary phase, significantly positive or negative associations between $\log_2$PTRs and abundances were considered to be in acceleration or deceleration phase, respectively. Those with no correlation and an average $\log_2$PTR above the stationary threshold were constrained to be in mid-log phase or in acceleration/deceleration phase (i.e., if there was a false negative due to lack of statistical power in detecting a positive or negative slope). 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 [75].

Stochastic logistic growth model simulation

The stochastic logistic growth equation (sLGE) 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 [76], 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 sLGE 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. S3**). 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 phases were combined, as these phases display similar delta-abundance relationships along the logistic growth curve. Conceptual diagrams were created using BioRender.

Death or dilution terms were not explicitly added to the simulated sLGE models. Here, we discuss how death or dilution rates are equivalent to changing the carrying capacity term, which has no impact on our growth phase inferences. Analytically, a decrease in abundance at a given time can be represented as a fraction of the current abundance subtracted from the LGE: $\frac{dx_i}{dt} = rx_i(t)\left(1 - \frac{x_i(t)}{K}\right) - Hx_i(t)$. Here, H is the "harvest rate", which determines the proportional decrease in each timepoint in the equation. At steady state, $rx*_i\ (t)\left(1 - \frac{x*_i\ (t)}{K}\right) - Hx*_i\ (t) = 0$, where $x*_i\ (t)$ represents the fixed point. Two equilibria exists in this equation: $x*_i\ (t) = 0$ and $x*_i\ (t) = K(1 - \frac{H}{r})$, with the latter being asymptotically stable. As H increases, the stable population size $x*_i\ (t)$ decreases due to the proportional decrease in K. As long as H does not exceed the intrinsic growth rate of gut microbes, which is expected for highly abundant and stably colonized taxa, the resulting K becomes the new stable K. To show that variation in k does not impact the relationship between growth rate and abundance, we simulated the LGE with stochastically varying K by adding the stochastic term, i.e. $\sigma k_i(t)\omega(t)$, to $k_i(t)$ (**Fig. S4**). In base R, simulation was performed for 100 iterations with the same noise levels (σ = 0.1) as the representative sLGE simulations with stochastic x. Major growth phases were defined the same way as sLGE simulations with stochastic x.

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). These DNA datasets are

publicly available from the National Center for Biotechnology (NCBI) Sequence Read Archive (SRA), accession code PRJNA942341. Scripts 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. Thanks to Pamela Troisch for help with DNA library preparation for the *in vitro* experiment. 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). Research reported in this publication was supported by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health (NIH) under award number R01DK133468 (to SMG).

Conflict of Interest Statement

The authors report no conflicts of interest.

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* **4**, 118–124 (2013).
2. Martin, A. M., Sun, E. W., Rogers, G. B. & Keating, D. J. The Influence of the Gut Microbiome on Host Metabolism Through the Regulation of Gut Hormone Release. *Front.*

Physiol. **10**, 428 (2019).

3. Zhang, P. *et al.* Commensal Homeostasis of Gut Microbiota-Host for the Impact of Obesity. *Front. Physiol.* **8**, 1122 (2017).
4. Stefan, K. L., Kim, M. V., Iwasaki, A. & Kasper, D. L. Commensal Microbiota Modulation of Natural Resistance to Virus Infection. *Cell* **183**, 1312–1324.e10 (2020).
5. Fischbach, M. A. Microbiome: Focus on Causation and Mechanism. *Cell* **174**, 785–790 (2018).
6. Zmora, N. *et al.* Personalized Gut Mucosal Colonization Resistance to Empiric Probiotics Is Associated with Unique Host and Microbiome Features. *Cell* **174**, 1388–1405.e21 (2018).
7. Suez, J. *et al.* Post-Antibiotic Gut Mucosal Microbiome Reconstitution Is Impaired by Probiotics and Improved by Autologous FMT. *Cell* **174**, 1406–1423.e16 (2018).
8. Bullman, S. *et al.* Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer. *Science* **358**, 1443–1448 (2017).
9. Halfvarson, J. *et al.* Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat Microbiol* **2**, 17004 (2017).
10. Cuna, A., Morowitz, M. J., Ahmed, I., Umar, S. & Sampath, V. Dynamics of the preterm gut microbiome in health and disease. *Am. J. Physiol. Gastrointest. Liver Physiol.* **320**, G411–G419 (2021).
11. Fisher, C. K. & Mehta, P. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PLoS One* **9**, e102451 (2014).
12. Bashan, A. *et al.* Universality of human microbial dynamics. *Nature* **534**, 259–262 (2016).
13. Stein, R. R. *et al.* Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLoS Comput. Biol.* **9**, e1003388 (2013).
14. Faust, K., Lahti, L., Gonze, D., de Vos, W. M. & Raes, J. Metagenomics meets time series analysis: unraveling microbial community dynamics. *Curr. Opin. Microbiol.* **25**, 56–66 (2015).

15. Gibbons, S. M., Kearney, S. M., Smillie, C. S. & Alm, E. J. Two dynamic regimes in the human gut microbiome. *PLoS Comput. Biol.* **13**, e1005364 (2017).
16. Carr, A., Diener, C., Baliga, N. S. & Gibbons, S. M. Use and abuse of correlation analyses in microbial ecology. *ISME J.* **13**, 2647–2655 (2019).
17. Momeni, B., Xie, L. & Shou, W. Lotka-Volterra pairwise modeling fails to capture diverse pairwise microbial interactions. *Elife* **6**, (2017).
18. Berry, D. & Widder, S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* **5**, 219 (2014).
19. Freilich, M. A., Wieters, E., Broitman, B. R., Marquet, P. A. & Navarrete, S. A. Species co-occurrence networks: Can they reveal trophic and non-trophic interactions in ecological communities? *Ecology* **99**, 690–699 (2018).
20. Coenen, A. R. & Weitz, J. S. Limitations of Correlation-Based Inference in Complex Virus-Microbe Communities. *mSystems* **3**, (2018).
21. Korem, T. *et al.* Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* **349**, 1101–1106 (2015).
22. Brown, C. T., Olm, M. R., Thomas, B. C. & Banfield, J. F. Measurement of bacterial replication rates in microbial communities. *Nat. Biotechnol.* **34**, 1256–1263 (2016).
23. Gibson, B., Wilson, D. J., Feil, E. & Eyre-Walker, A. The distribution of bacterial doubling times in the wild. *Proc. Biol. Sci.* **285**, (2018).
24. Monod, J. THE GROWTH OF BACTERIAL CULTURES. *Annu. Rev. Microbiol.* **3**, 371–394 (1949).
25. Logistic stochastic growth models and applications. in *Handbook of the Logistic Distribution* 419–419 (CRC Press, 1991).
26. Descheemaeker, L. & de Buyl, S. Stochastic logistic models reproduce experimental time series of microbial communities. *Elife* **9**, (2020).
27. Grilli, J. Macroecological laws describe variation and diversity in microbial communities.

Nat. Commun. **11**, 4743 (2020).

28. Lobry, J. R., Flandrois, J. P., Carret, G. & Pave, A. Monod's bacterial growth model revisited. *Bull. Math. Biol.* **54**, 117–122 (1992).
29. Fujikawa, H., Kai, A. & Morozumi, S. A new logistic model for bacterial growth. *Shokuhin Eiseigaku Zasshi* **44**, 155–160 (2003).
30. Baranyi, J., McClure, P. J., Sutherland, J. P. & Roberts, T. A. Modeling bacterial growth responses. *J. Ind. Microbiol.* **12**, 190–194 (1993).
31. Ho, P.-Y., Good, B. H. & Huang, K. C. Competition for fluctuating resources reproduces statistics of species abundance over time across wide-ranging microbiotas. *Elife* **11**, (2022).
32. Wolff, R., Shoemaker, W. & Garud, N. Ecological Stability Emerges at the Level of Strains in the Human Gut Microbiome. *mSystems* (2023) doi:10.1128/mbio.02502-22.
33. Poyet, M. *et al.* A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. *Nat. Med.* **25**, 1442–1452 (2019).
34. Kadosh, E. *et al.* The gut microbiome switches mutant p53 from tumour-suppressive to oncogenic. *Nature* **586**, 133–138 (2020).
35. Kolodziejczyk, A. A., Zheng, D., Shibolet, O. & Elinav, E. The role of the microbiome in NAFLD and NASH. *EMBO Mol. Med.* **11**, (2019).
36. Day, J. A. *et al.* Lettuce (Lactuca sativa) productivity influenced by microbial inocula under nitrogen-limited conditions in aquaponics. *PLoS One* **16**, e0247534 (2021).
37. 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.* **110**, 12804–12809 (2013).
38. Otwell, A. E. *et al.* Sulfur Metabolites Play Key System-Level Roles in Modulating Denitrification. *mSystems* **6**, (2021).
39. Wallace, R. J., Snelling, T. J., McCartney, C. A., Tapio, I. & Strozzi, F. Application of meta-omics techniques to understand greenhouse gas emissions originating from ruminal

metabolism. *Genet. Sel. Evol.* **49**, 9 (2017).

40. Roy, P., Dutta, A. & Deen, B. Greenhouse gas emissions and production cost of ethanol produced from biosyngas fermentation process. *Bioresour. Technol.* **192**, 185–191 (2015).
41. Zhong, C. Industrial-Scale Production and Applications of Bacterial Cellulose. *Front Bioeng Biotechnol* **8**, 605374 (2020).
42. Guzman-Rodriguez, M. *et al.* Using bioreactors to study the effects of drugs on the human microbiota. *Methods* **149**, 31–41 (2018).
43. Wilson, A. S. *et al.* Diet and the Human Gut Microbiome: An International Review. *Dig. Dis. Sci.* **65**, 723–740 (2020).
44. Dahl, W. J., Rivero Mendoza, D. & Lambert, J. M. Diet, nutrients and the microbiome. *Prog. Mol. Biol. Transl. Sci.* **171**, 237–263 (2020).
45. Baird, J. *et al.* The effect of a behaviour change intervention on the diets and physical activity levels of women attending Sure Start Children's Centres: results from a complex public health intervention. *BMJ Open* **4**, e005290–e005290 (2014).
46. Vangay, P. *et al.* US Immigration Westernizes the Human Gut Microbiome. *Cell* **175**, 962–972.e10 (2018).
47. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
48. David, L. A. *et al.* Host lifestyle affects human microbiota on daily timescales. *Genome Biol.* **15**, R89 (2014).
49. Blätke, M.-A. & Bräutigam, A. Evolution of C4 photosynthesis predicted by constraint-based modelling. *Elife* **8**, (2019).
50. Atolia, E. *et al.* Environmental and Physiological Factors Affecting High-Throughput Measurements of Bacterial Growth. *MBio* **11**, (2020).
51. Nguyen, J., Lara-Gutiérrez, J. & Stocker, R. Environmental fluctuations and their effects on microbial communities, populations and individuals. *FEMS Microbiol. Rev.* **45**, (2021).

52. Tsoularis, A. & Wallace, J. Analysis of logistic growth models. *Math. Biosci.* **179**, 21–55 (2002).

53. O'Sullivan, O. *et al.* Exercise and the microbiota. *Gut Microbes* **6**, 131–136 (2015).

54. Joseph, T. A., 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.* **32**, 558–568 (2022).

55. Long, A. M., Hou, S., Ignacio-Espinoza, J. C. & Fuhrman, J. A. Benchmarking microbial growth rate predictions from metagenomes. *ISME J.* **15**, 183–195 (2021).

56. Szafrańska, A. K., Junker, V., Steglich, M. & Nübel, U. Rapid cell division of Staphylococcus aureus during colonization of the human nose. *BMC Genomics* **20**, 229 (2019).

57. Gao, Y. & Li, H. Quantifying and comparing bacterial growth dynamics in multiple metagenomic samples. *Nat. Methods* **15**, 1041–1044 (2018).

58. Johnson, J. P. *et al.* Generally-healthy individuals with aberrant bowel movement frequencies show enrichment for microbially-derived blood metabolites associated with impaired kidney function. *bioRxiv* 2023.03.04.531100 (2023) doi:10.1101/2023.03.04.531100.

59. Braga, R. M., Dourado, M. N. & Araújo, W. L. Microbial interactions: ecology in a molecular perspective. *Braz. J. Microbiol.* **47 Suppl 1**, 86–98 (2016).

60. Zelezniak, A. *et al.* Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6449–6454 (2015).

61. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).

62. Pramanik, J. & Keasling, J. D. Stoichiometric model of Escherichia coli metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnol. Bioeng.* **56**, 398–421 (1997).

63. Dikicioglu, D., Kırdar, B. & Oliver, S. G. Biomass composition: the ‘elephant in the room’ of metabolic modelling. *Metabolomics* **11**, 1690–1701 (2015).
64. Schulz, C., Kumelj, T., Karlsen, E. & Almaas, E. Genome-scale metabolic modelling when changes in environmental conditions affect biomass composition. *PLoS Comput. Biol.* **17**, e1008528 (2021).
65. Diener, C., Gibbons, S. M. & Resendis-Antonio, O. MICOM: Metagenome-Scale Modeling To Infer Metabolic Interactions in the Gut Microbiota. *mSystems* **5**, (2020).
66. Adrian Trapletti, K. H. tseries: Time Series Analysis and Computational Finance. Preprint at https://CRAN.R-project.org/package=tseries (2022).
67. Team, R. C. R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna Austria URL (2022).
68. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
69. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
70. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* **3**, e104 (2017).
71. Nayfach, S., Shi, Z. J., Seshadri, R., Pollard, K. S. & Kyrpides, N. C. New insights from uncultivated genomes of the global human gut microbiome. *Nature* **568**, 505–510 (2019).
72. Langmead, B., Wilks, C., Antonescu, V. & Charles, R. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**, 421–432 (2019).
73. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer, 2016).
74. Lin, H. & Peddada, S. D. Analysis of microbial compositions: a review of normalization and differential abundance analysis. *NPJ Biofilms Microbiomes* **6**, 60 (2020).
75. Fisher, S. R. A. *Statistical Methods for Research Workers*. (Oliver and Boyd, 1925).
76. Iacus, S. M. SDE : simulation and inference for stochastic differential equations. (2007).

FIGURE LEGENDS

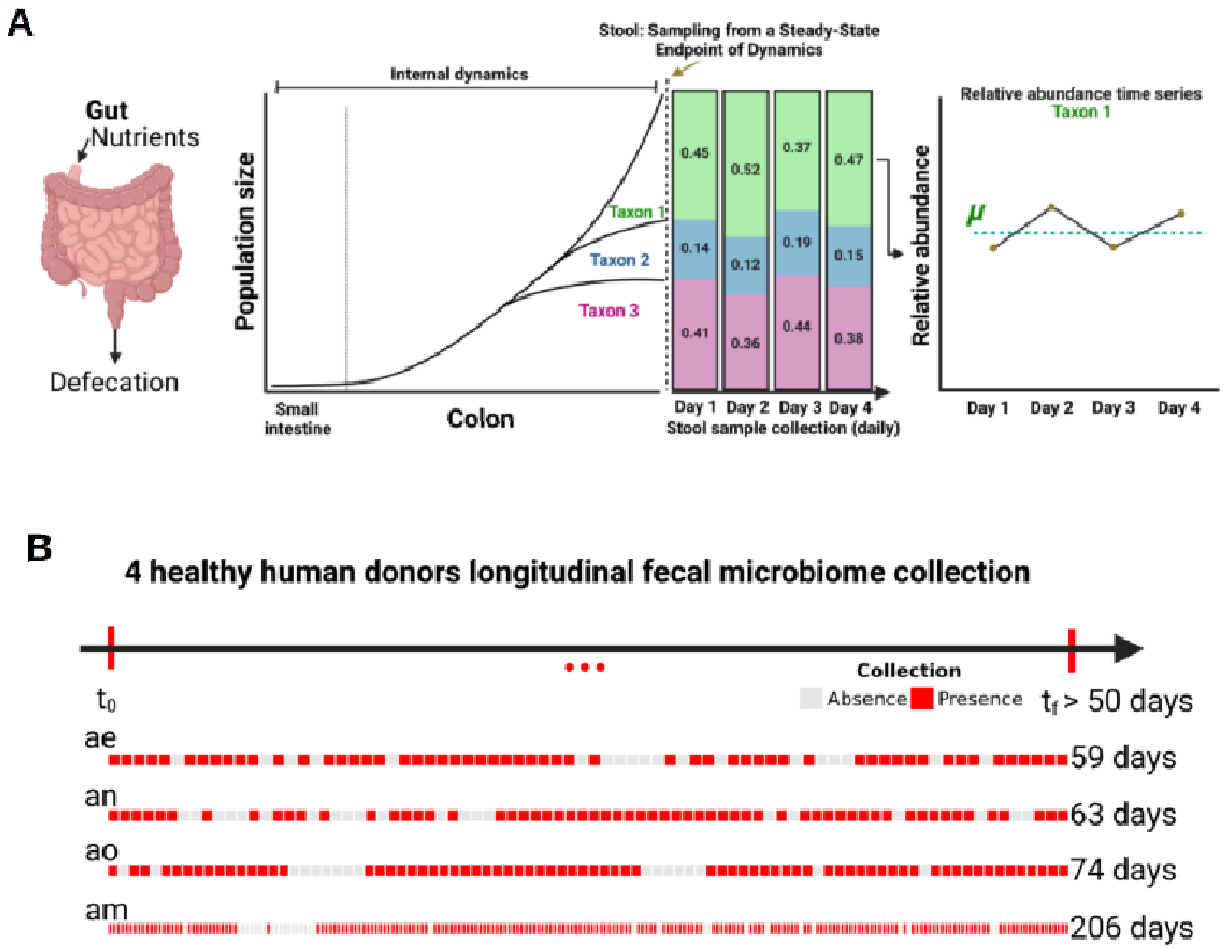


Figure 1. Conceptual figure showing two flow-through microbial ecosystems: a bioreactor and a human gut. **A.** Both bioreactors and guts are continuous flow-through systems. Prior to reaching the measured abundances in stool, taxa grow in the large intestine with varying growth rates, carrying capacities, 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.

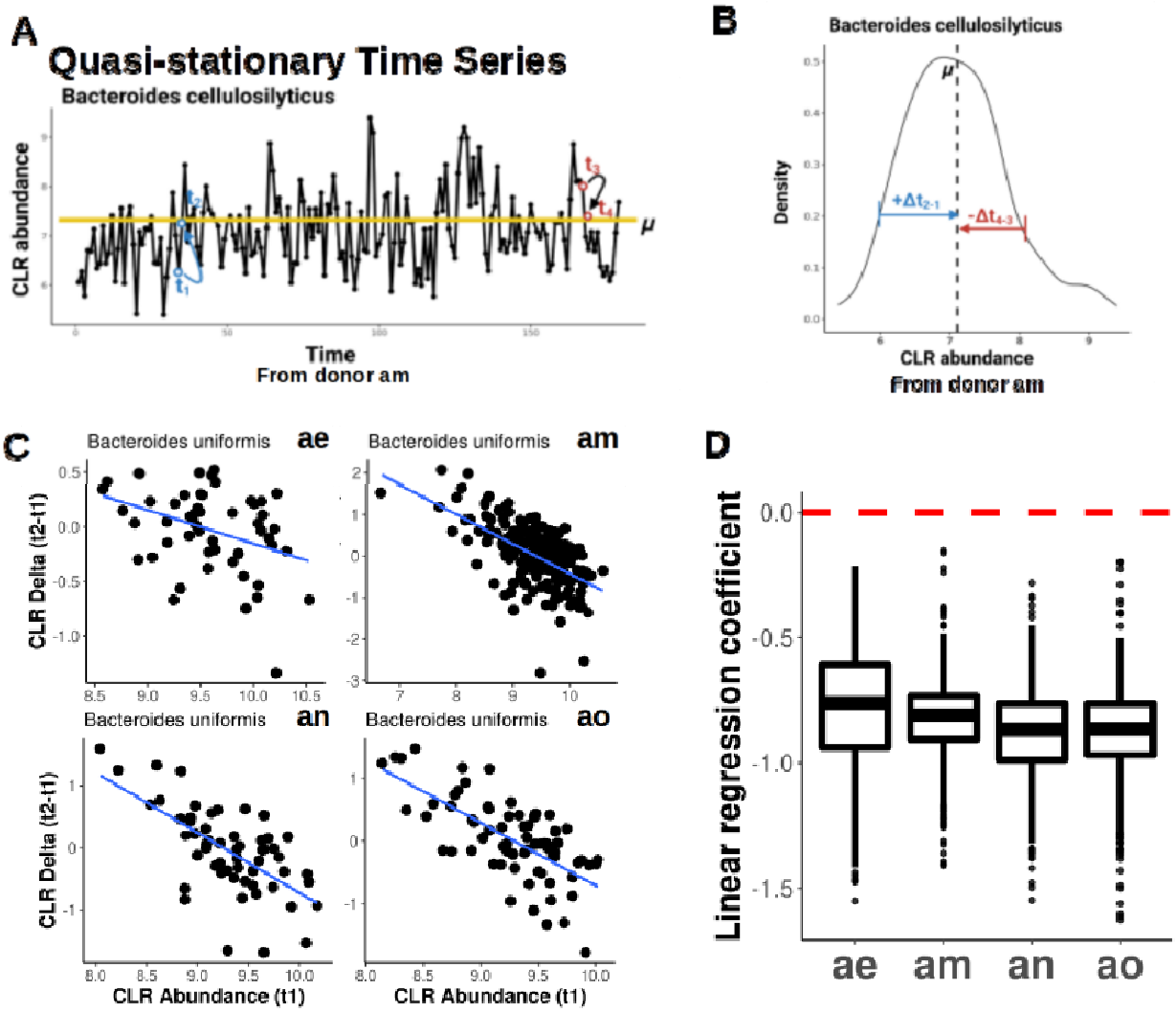


Figure 2. Regression-to-the-mean effect in human microbial time series data. A. Yellow line represents the mean abundance (μ) of *Bacteroides cellulosilyticus* over time in donor am. Time points t_1 and t_3 indicate fluctuations below and above the mean abundance, and t_2 and t_4 show the return to the 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.** Delta vs. abundance 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.

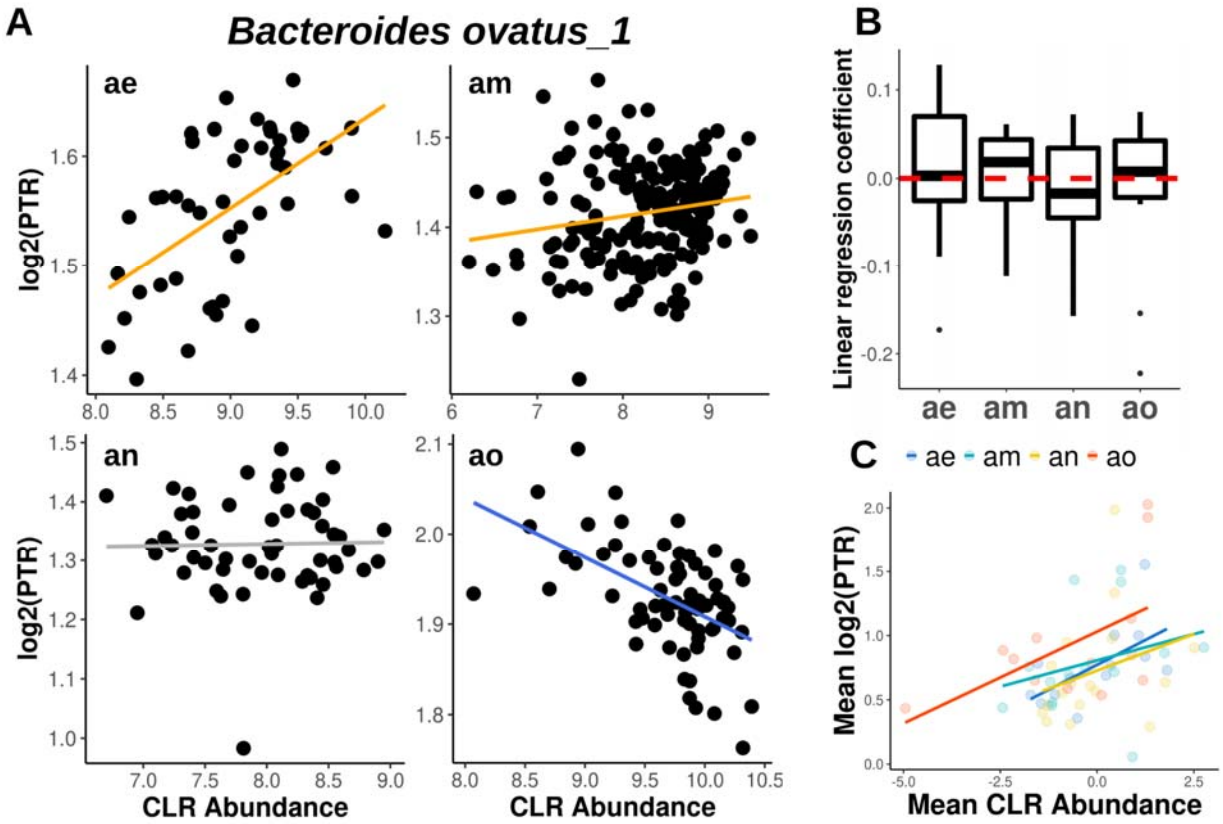


Figure 3. Variable relationships between PTRs and CLR-normalized abundances across human gut microbial time series. The ratio of sequencing coverage near the replication origin to the replication terminus for each species (i.e. peak-to-trough ratio, or PTR), was calculated using COPTR. **A.** Log_2(PTR) 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). PTR was calculated.

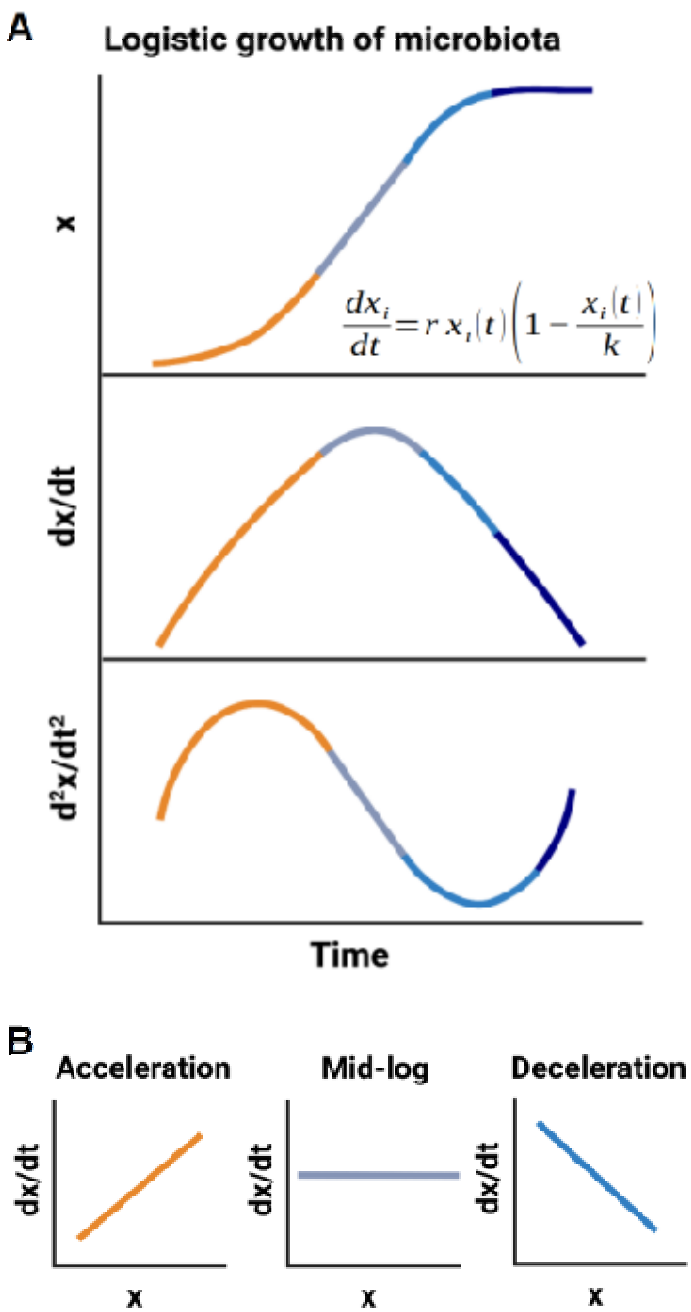


Figure 4. Diagram of the logistic growth equation. A. The logistic growth curve models abundance (*x*) with respect to time (top panel). Orange, grey, blue, and navy describes acceleration, mid-log, deceleration, and stationary phases, respectively. The first derivative of the logistic growth curve models 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.

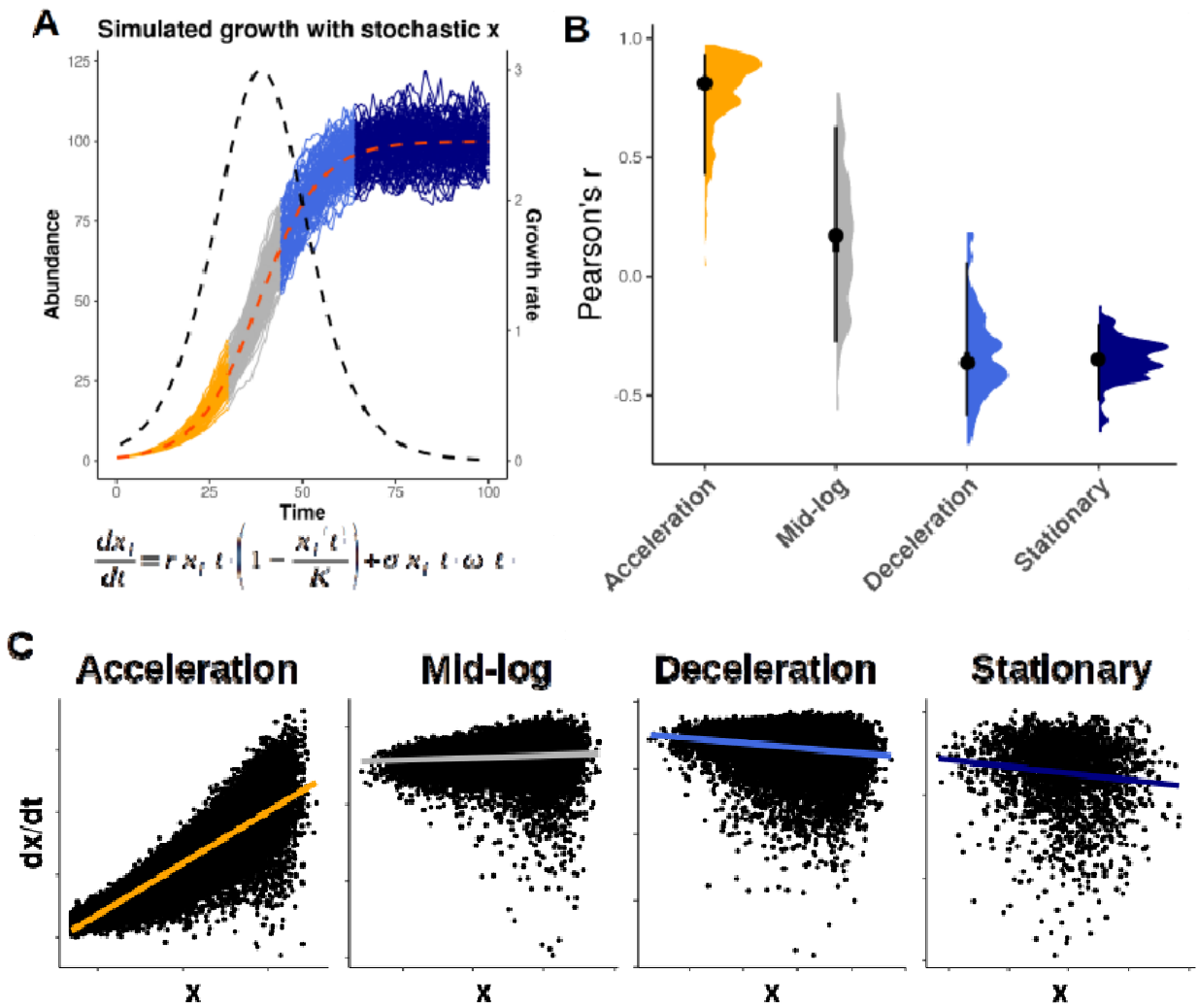


Figure 5. Distinguishing growth phases using the stochastic logistic growth model. A. Stochastic logistic growth curves with growth rate (r) = 1.2, carrying capacity (k) = 100, and noise level (n) = 0.1 across 100 iterations. Major growth phase groups in orange (acceleration), gray (mid-log), blue (deceleration),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

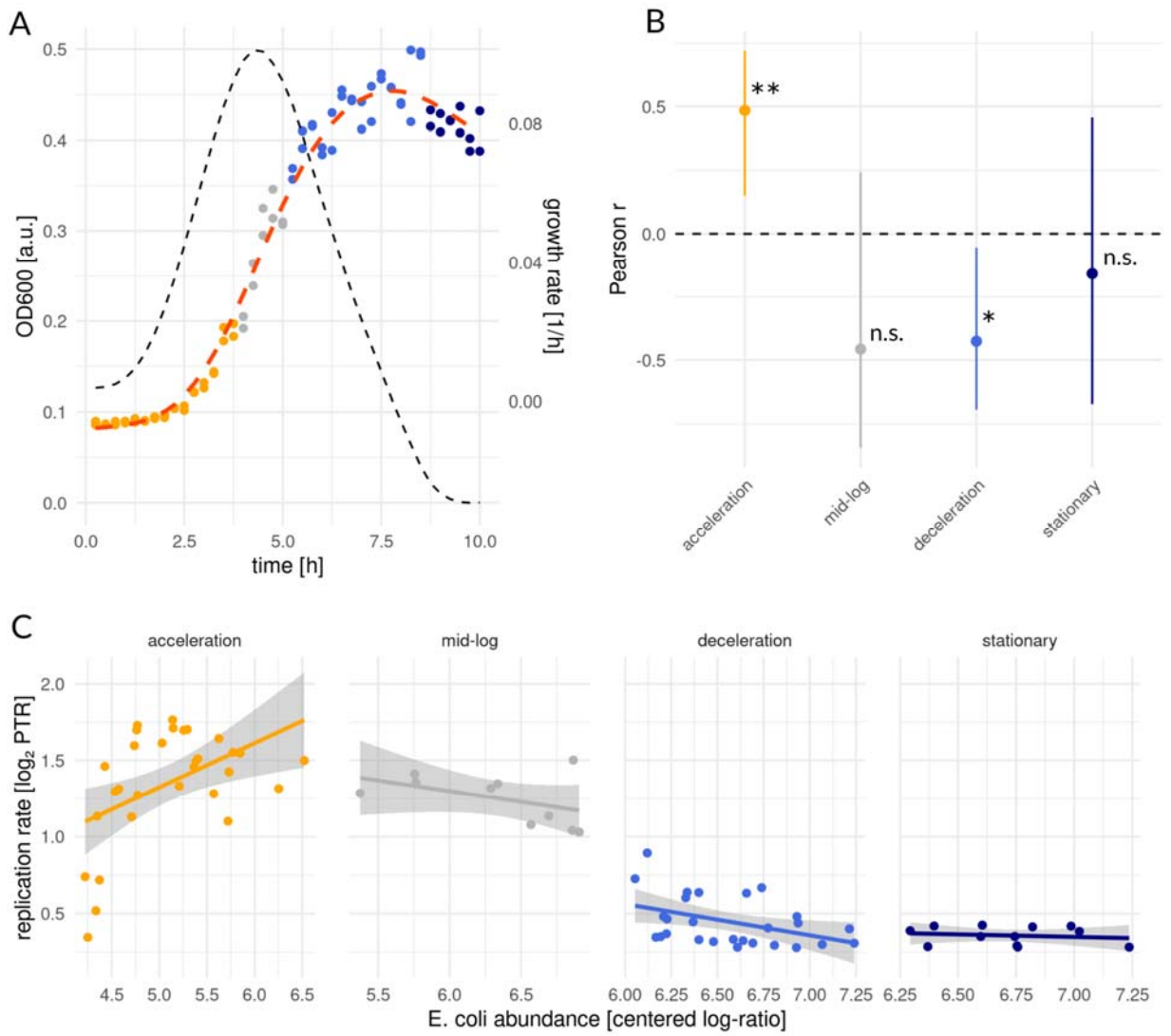


Figure 6. Relationship between growth rate and abundance in major growth phases in *E. coli* populations. A. Growth curve of *E. coli* (MG1655) using OD measurements. Colors describe major growth phases. Dotted black and red lines show the growth rate derived from OD measurements and mean growth trajectory, respectively. **B.** Pearson r values between abundance and growth rate in each of the four growth phase windows. Asterisks show statistical significance. **: $p < 0.01$, *: $p < 0.05$, n.s.: not significant. **C.** Scatter plots in log scale showing relationships between abundance and replication rate ($\log_2$PTR) across the four growth phase regions defined in panel A.

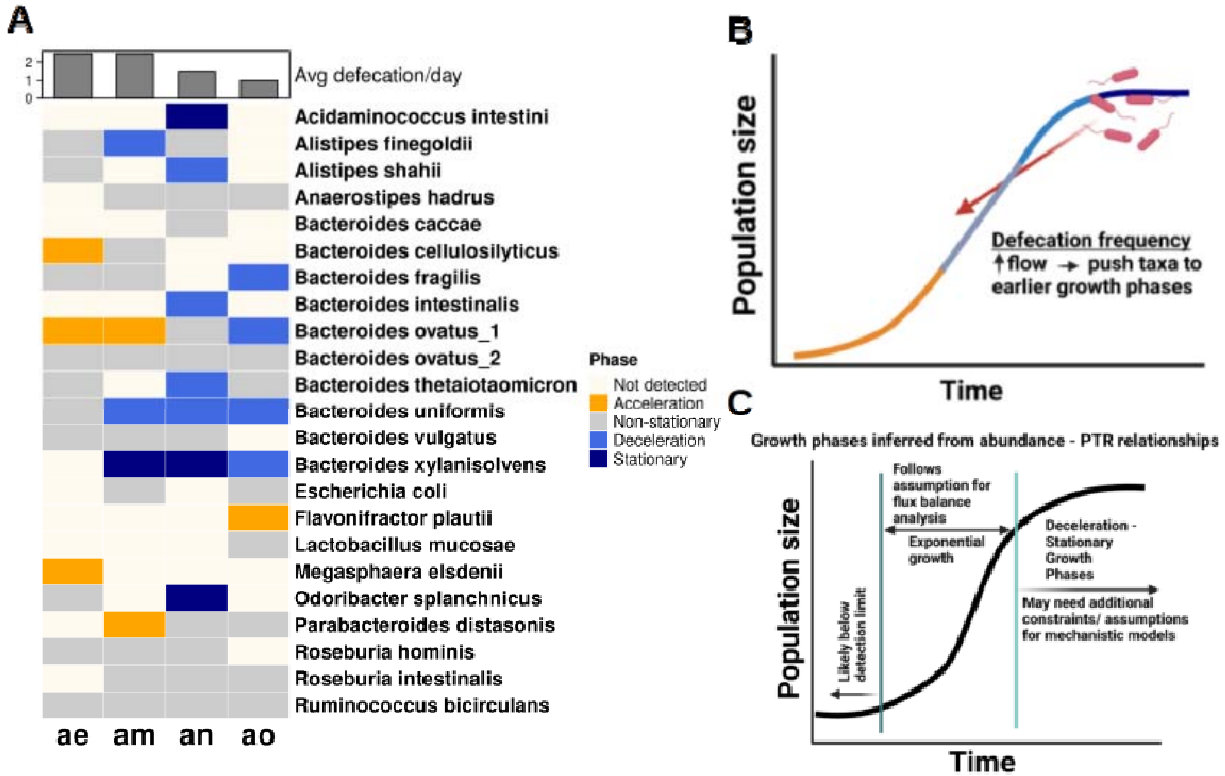


Figure 7. *in vivo* growth phase estimation. A. We find variable relationships between $\log_2$PTRs and population abundances across taxa in each of the four donors, consistent with the growth phase patterns observed in sLGE simulations. Donors with higher defecation rates tended to have a larger fraction taxa with positive $\log_2$PTR-abundance associations and fewer with negative associations, indicating acceleration and deceleration-stationary phases, respectively. Taxa in stationary phase were classified using an empirical threshold (average $\log_2$PTR < 0.358). Non-stationary taxa (i.e., above the stationary phase threshold, but lacking a significant correlation between $\log_2$PTRs and abundances) are likely in mid-log phase, but these taxa could also be in acceleration/deceleration phases (i.e., underpowered to detect the correlation). **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, such as Flux Balance Analysis, which assumes exponential growth.

Supplemental Figures

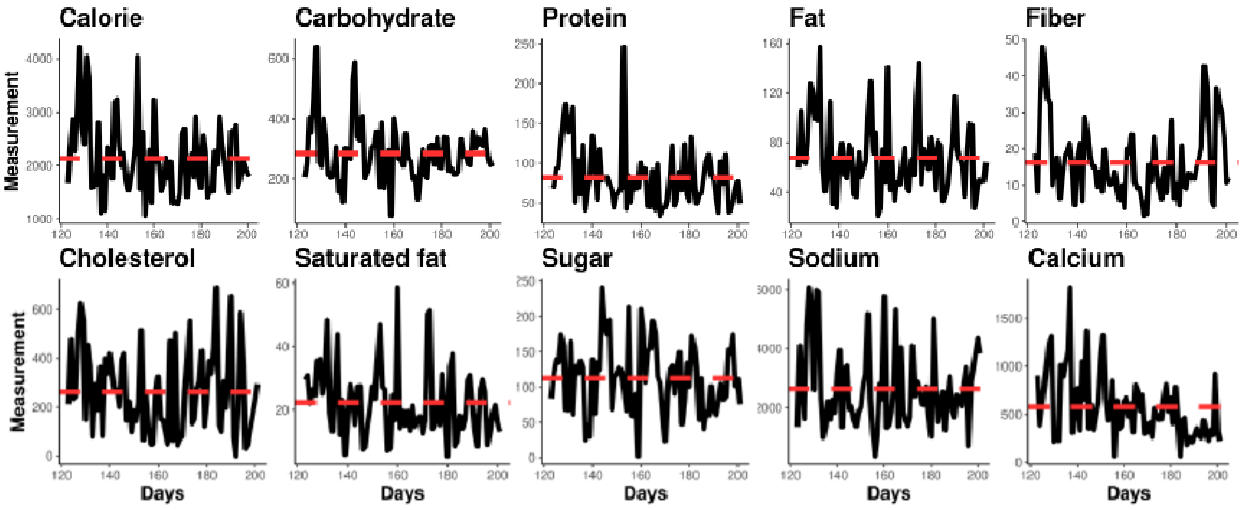


Figure S1. Lack of autocorrelation detected in most daily macronutrient intake. Daily measurements of nutrient intake were downloaded from David et al. [48]. Post-travel time points are shown. Autocorrelation was tested using the augmented Dickey-Fuller test (i.e., $p < 0.1$ indicates significant stationarity of a dietary variable). For each nutrient, p-values are reported for the recordings post-travel period of the subject. Calorie ($p = 0.0341$), carbohydrate ($p = 0.0144$), protein ($p = 0.0314$), fat ($p = 0.0172$), fiber ($p = 0.0369$), cholesterol ($p = 0.0534$), saturated fat ($p < 0.01$), sugar ($p = 0.0123$), sodium ($p = 0.0341$), calcium ($p < 0.01$). Dotted red lines show the mean measurement values.

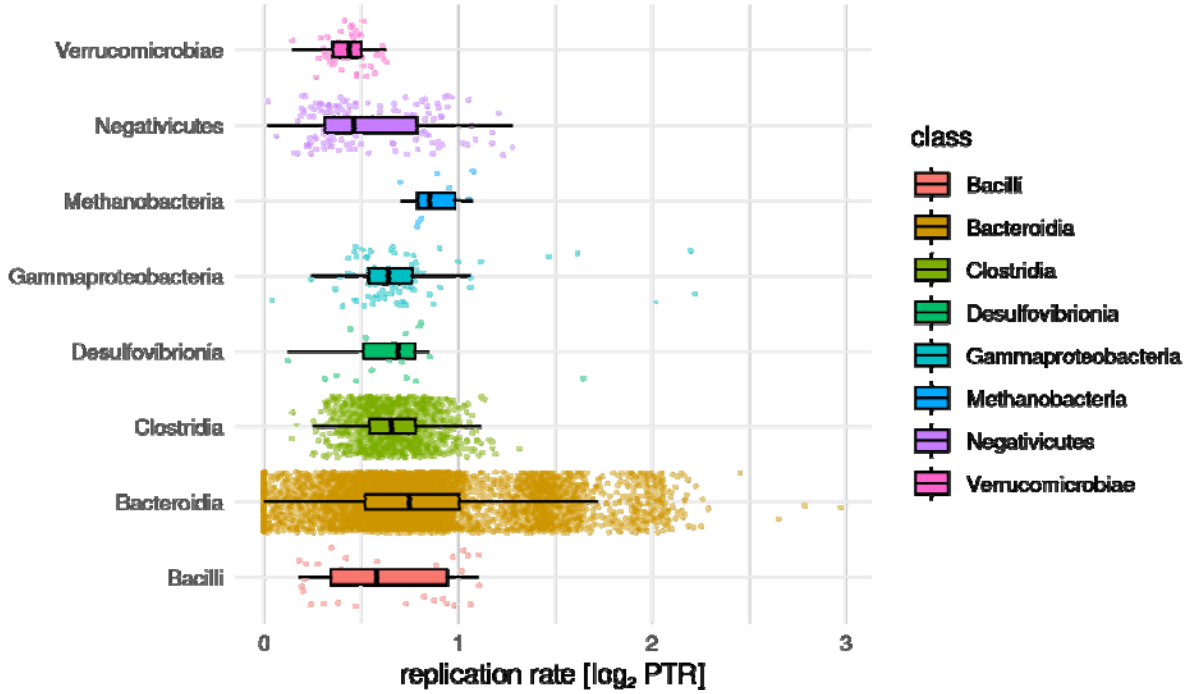


Figure S2. Distributions of log_2PTR values across 84 BIO-ML donors, broken down by phylogenetic class. We see a fairly wide range of log_2PTRs within each taxonomic class. The median log_2PTR across classes varies between ~0.45 and ~0.75. In a linear regression model, controlling from taxonomic group as a covariate, we see a significant positive association between log_2PTRs and CLR abundances at the class-level (β = 0.0612, p = 8.359e^{-60}). This positive taxonomy-controlled association is preserved at the species-level (β = 0.0101, p = 0.0006).

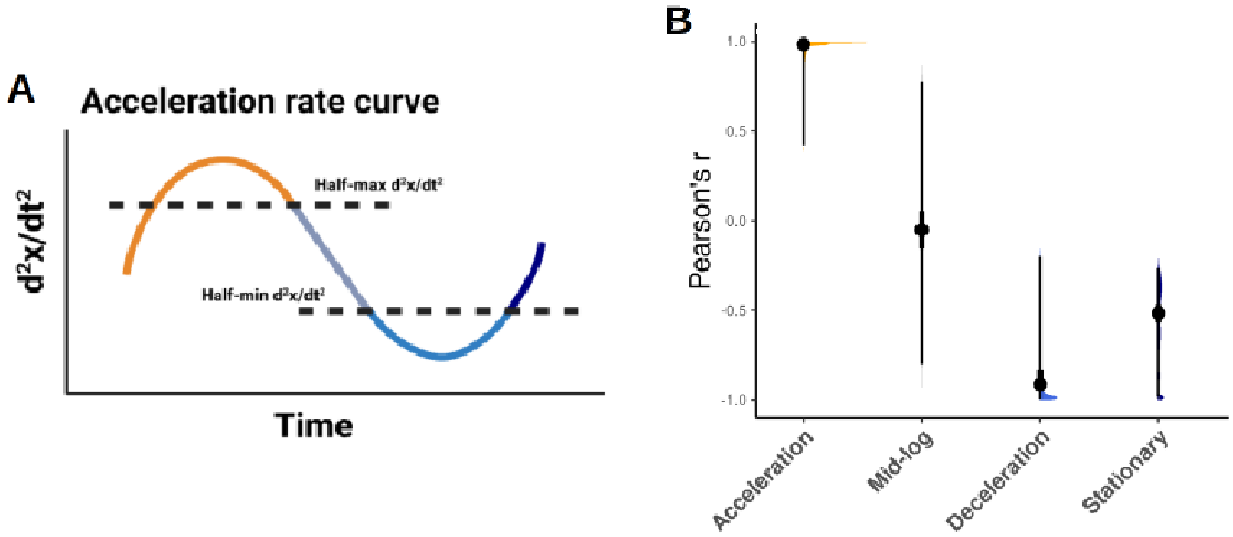


Figure S3. Definition of major growth phases using the stochastic logistic growth model. A. The half-maximum of the peak and half-minimum of the trough of the second derivative of abundance were used to define growth phases across model parameterizations. **B.** Pearson r values between abundances and 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).

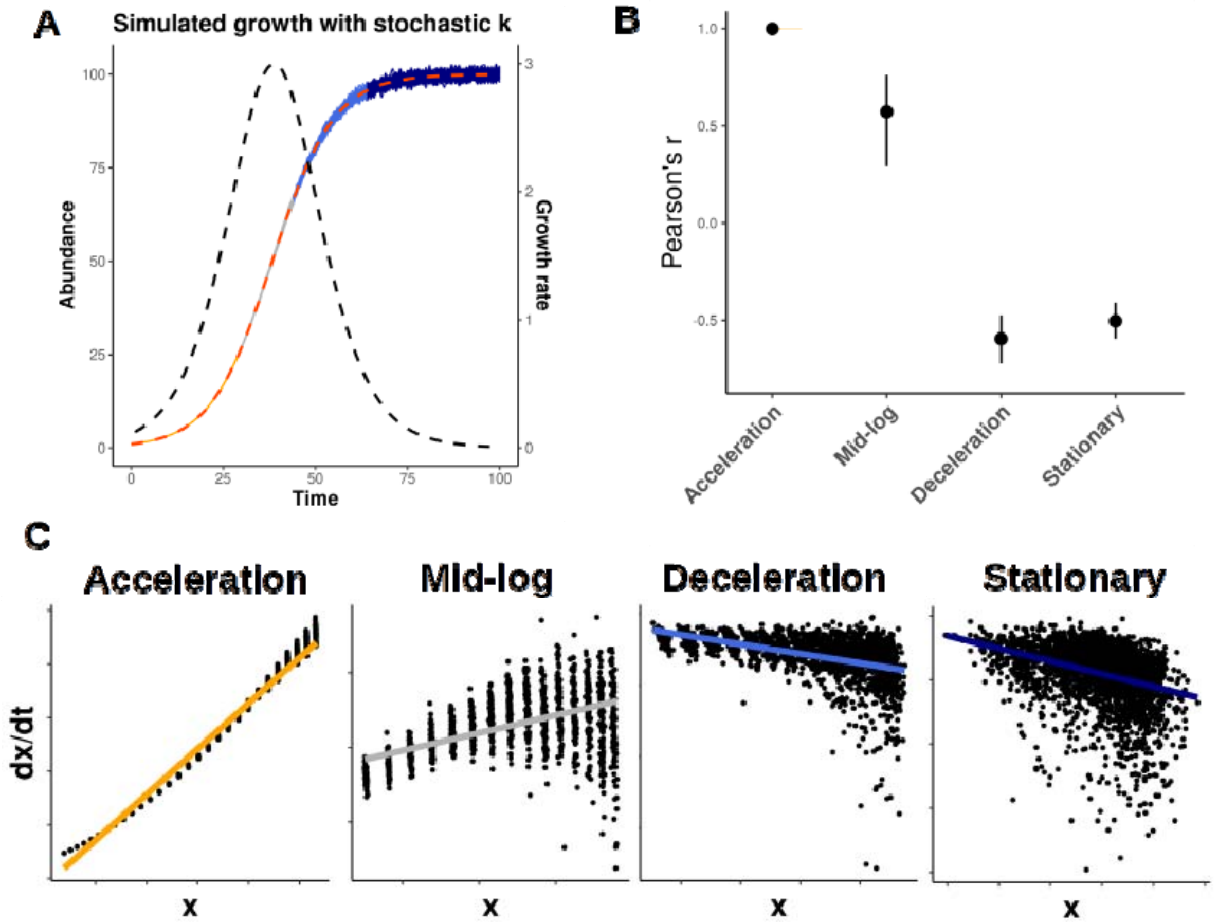


Figure S4. Relationship between growth rate and abundance across the major growth phases, simulated using the logistic growth model with stochastically varying carrying capacities. A. Stochastic logistic growth curves with growth rate (r) = 1.2, carrying capacity (k) = 100, and noise level (n) = 0.1 applied to k across 100 iterations. Major growth phase groups in orange (acceleration), gray (mid-log), blue (deceleration), and navy (stationary). **B.** Pearson r values between abundances and growth rates in each of the four growth phase windows across variable model parameterizations (r = 1-3, k = 10-1000) and a fixed noise level (□ = 0.1). **C.** Scatter plots in log scale showing relationships between abundance and growth rate across the four growth phase regions defined in panel A.

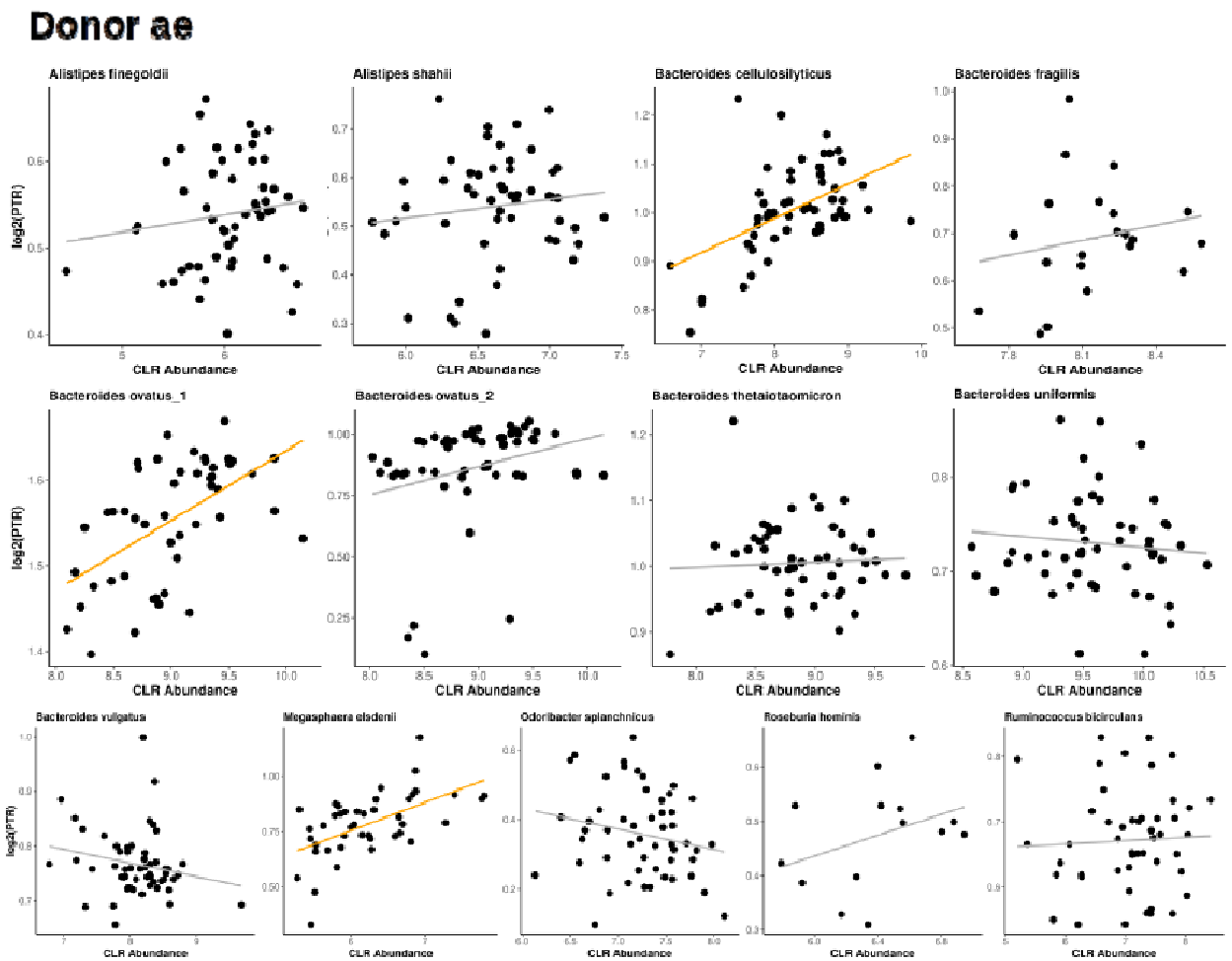


Figure S5. Relationships between abundance and log2(PTR) for abundant taxa in donor ae. Abundant taxa with relatively dense longitudinal PTR and abundance data (at least 5 matched data points; time differences between adjacent samples less than three days) were selected for analysis. Gray trend lines 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).

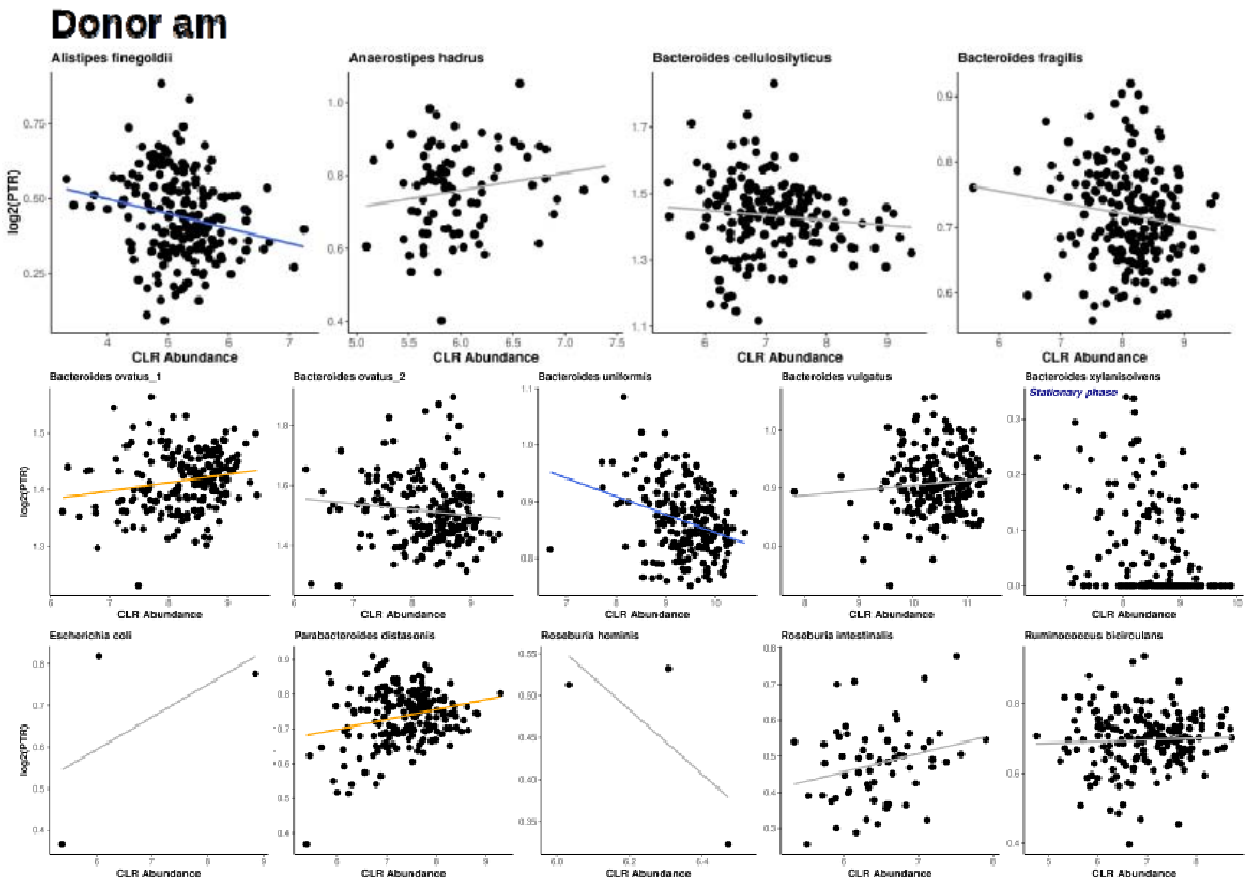


Figure S6. Relationships between abundance and log2(PTR) for abundant taxa in donor am. Abundant taxa with relatively dense longitudinal PTR and abundance data (at least 5 matched data points; time differences between adjacent samples less than three days) were selected for analysis. Gray trend lines 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).

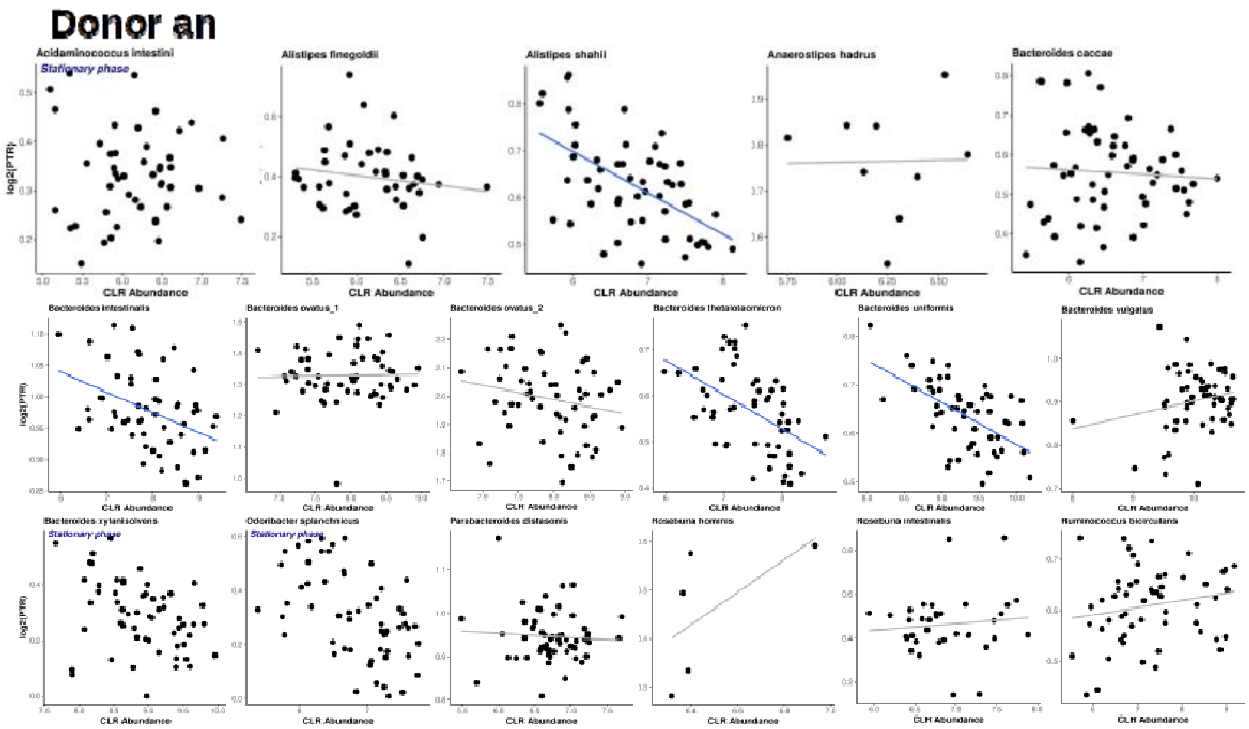


Figure S7. Relationships between abundance and log2(PTR) for abundant taxa in donor an. Abundant taxa with relatively dense longitudinal PTR and abundance data (at least 5 matched data points; time differences between adjacent samples less than three days) were selected for analysis. Gray trend lines 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).

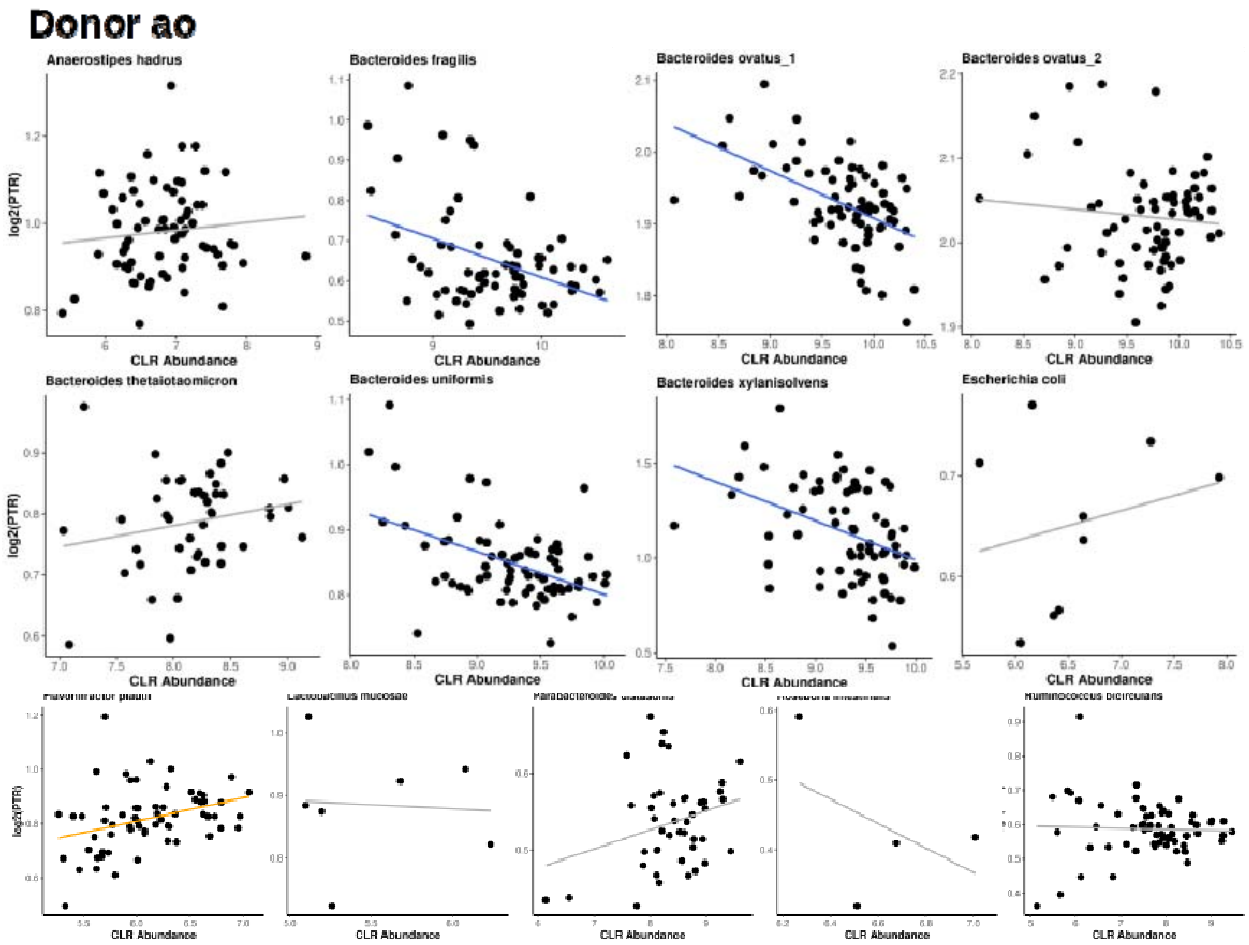
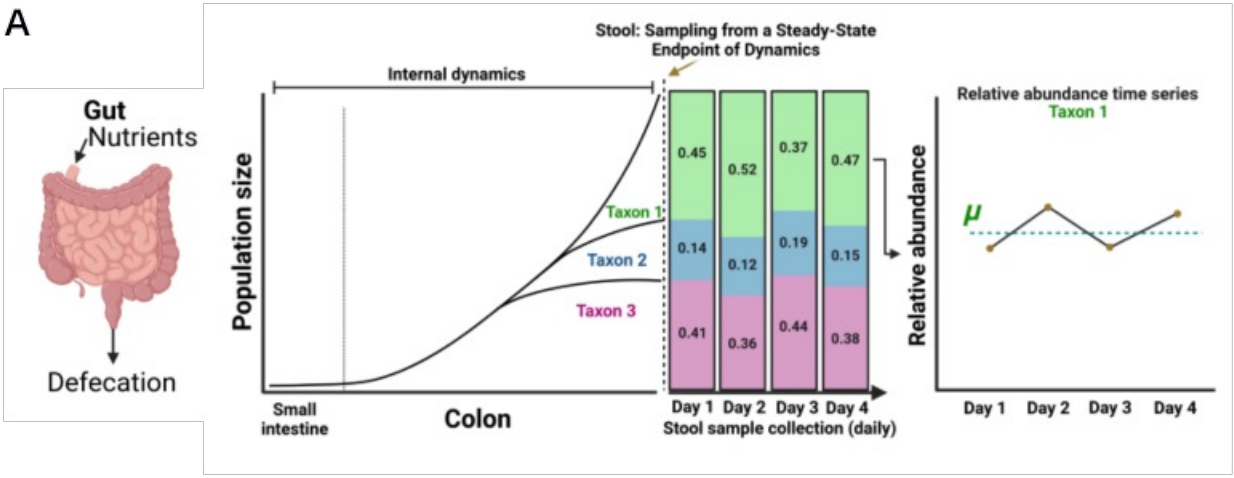
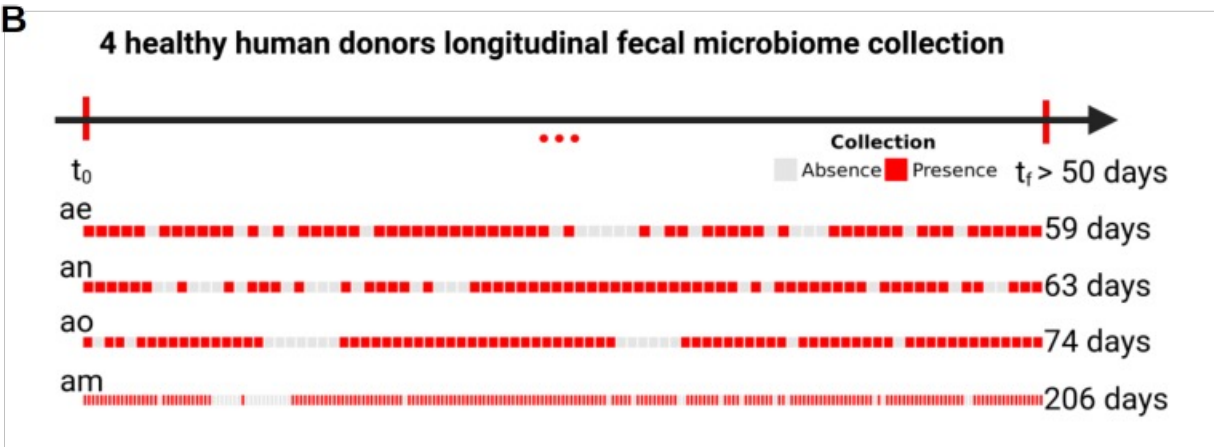


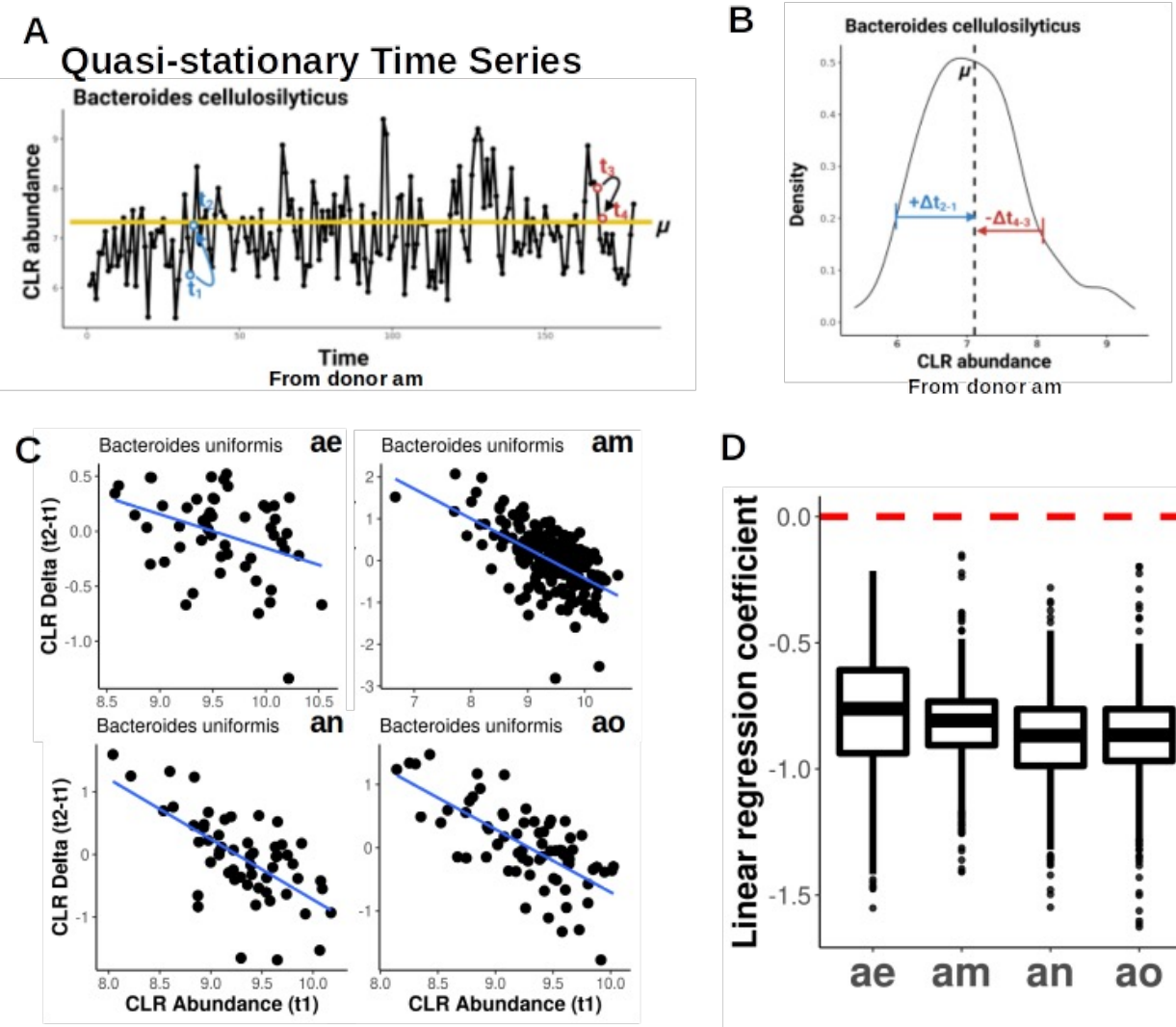
Figure S8. Relationships between abundance and log2(PTR) for individual taxon in donor ao. Abundant taxa with relatively dense longitudinal PTR and abundance data (at least 5 matched data points; time differences between adjacent samples less than three days) were selected for analysis. Gray trend lines 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).

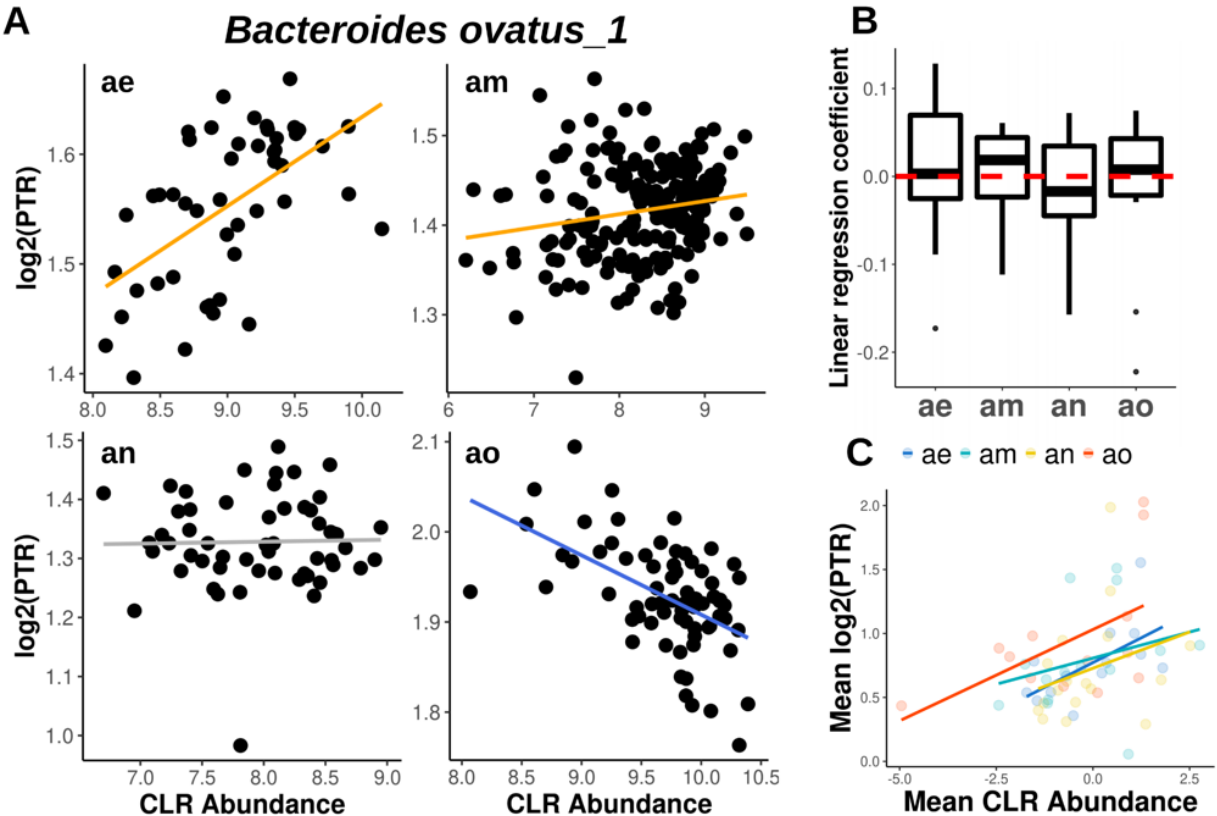
A

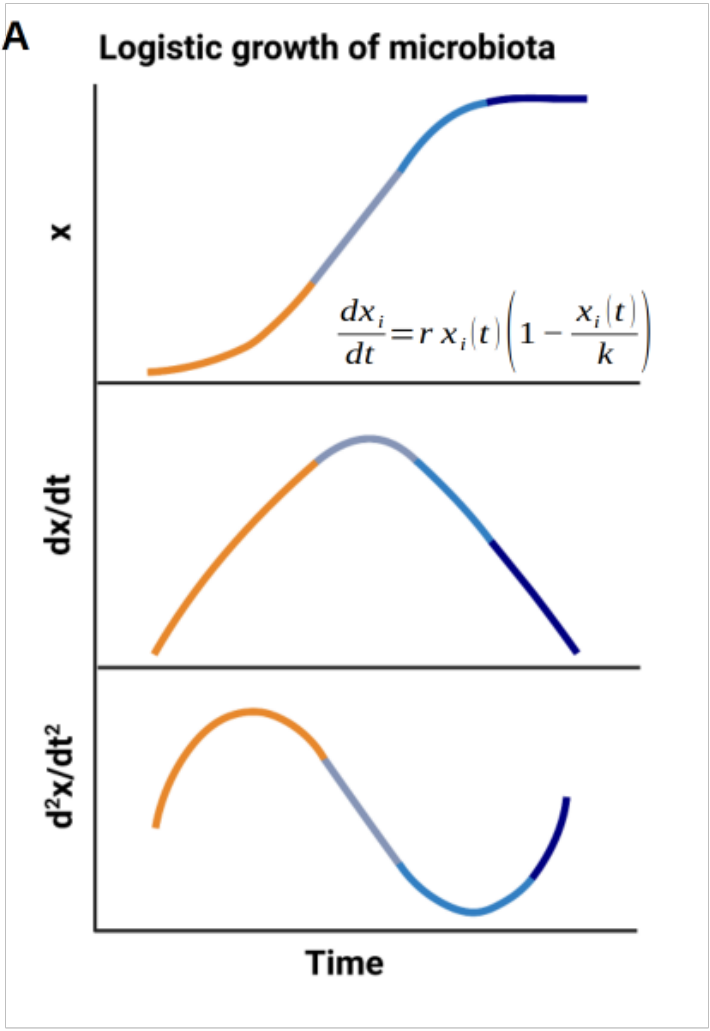

Gut
Nutrients
Defecation
Stool: Sampling from a Steady-State Endpoint of Dynamics
Internal dynamics
Population size
Taxon 1
Taxon 2
Taxon 3
Small intestine
Colon
0.45
0.14
0.41
0.52
0.12
0.36
0.37
0.19
0.44
0.47
0.15
0.38
Day 1 Day 2 Day 3 Day 4
Stool sample collection (daily)
Relative abundance
Relative abundance time series
Taxon 1
μ
Day 1 Day 2 Day 3 Day 4

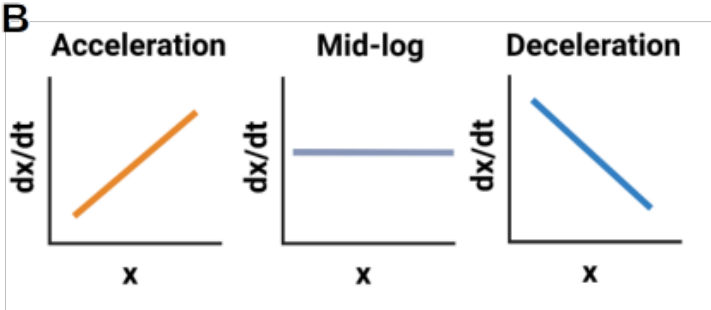
B

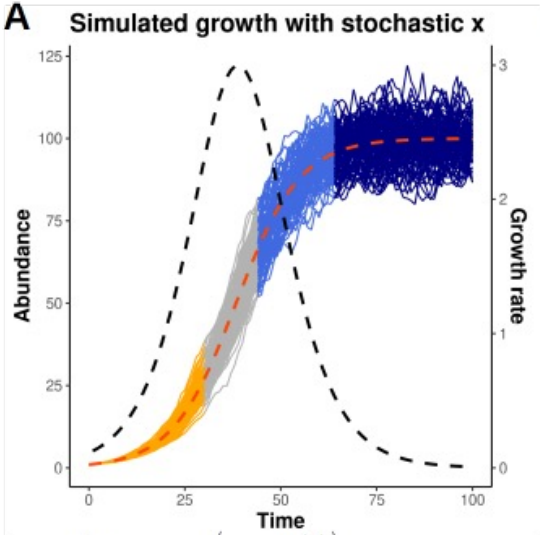

4 healthy human donors longitudinal fecal microbiome collection
Collection
Absence
Presence
t_0
t_f > 50 days
ae
59 days
an
63 days
ao
74 days
am
206 days


A
Quasi-stationary Time Series
Bacteroides cellulosilyticus
CLR abundance
Time
From donor am
B
Bacteroides cellulosilyticus
Density
CLR abundance
From donor am
C
Bacteroides uniformis
ae
am
an
ao
CLR Delta (t2-t1)
CLR Abundance (t1)
D
Linear regression coefficient
ae
am
an
ao

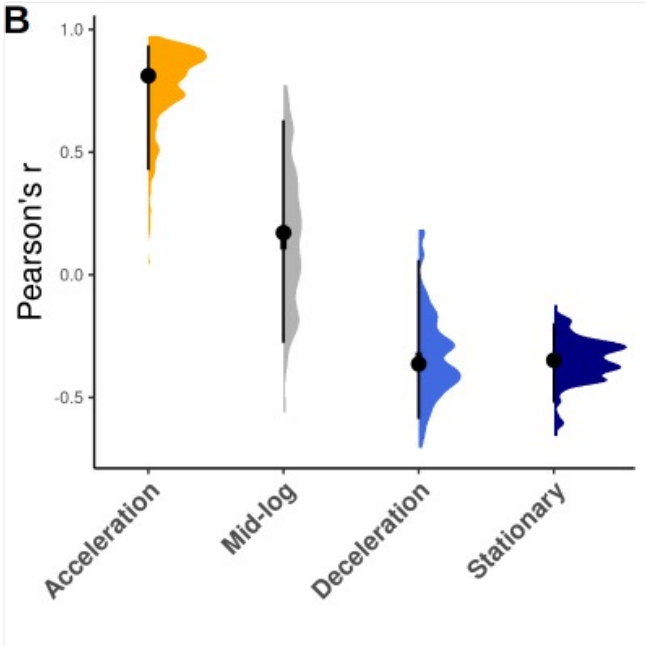

A
Bacteroides ovatus_1
ae
am
an
ao
log2(PTR)
CLR Abundance
B
Linear regression coefficient
ae
am
an
ao
C
ae
am
an
ao
Mean log2(PTR)
Mean CLR Abundance

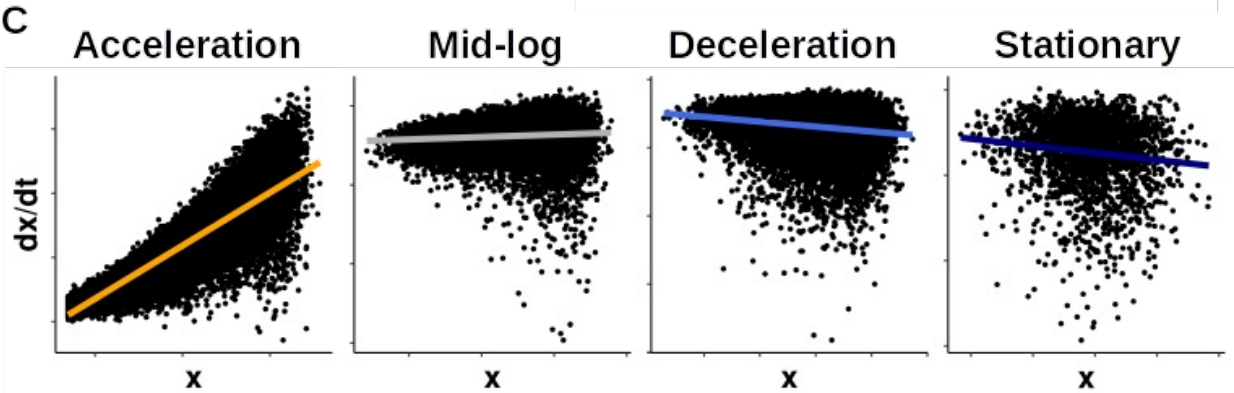

A
Logistic growth of microbiota
x
$\frac{dx_i}{dt} = r\, x_i(t)\left(1 - \frac{x_i(t)}{k}\right)$
dx/dt
d²x/dt²
Time

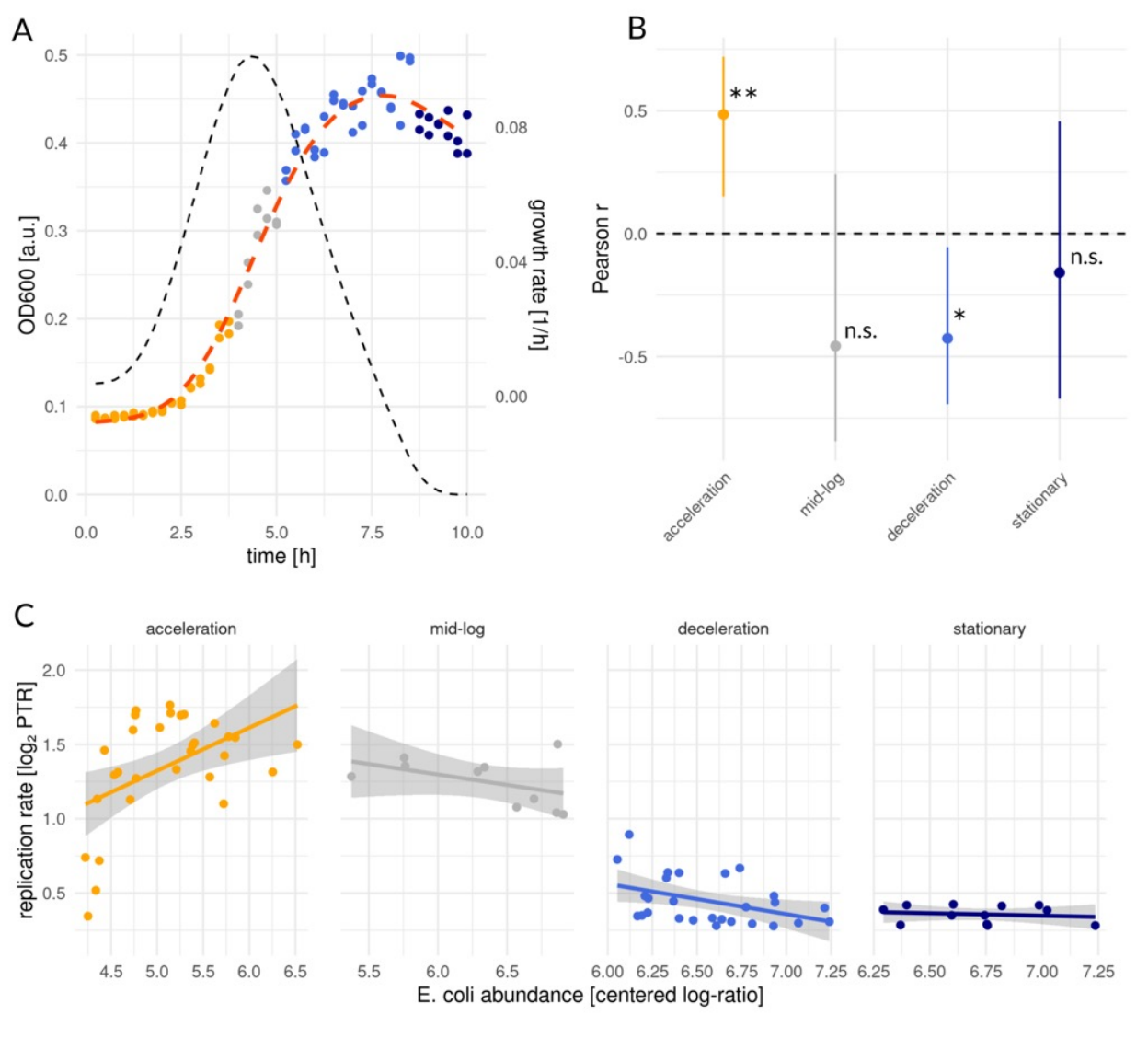

B
Acceleration
Mid-log
Deceleration
dx/dt
x
dx/dt
x
dx/dt
x

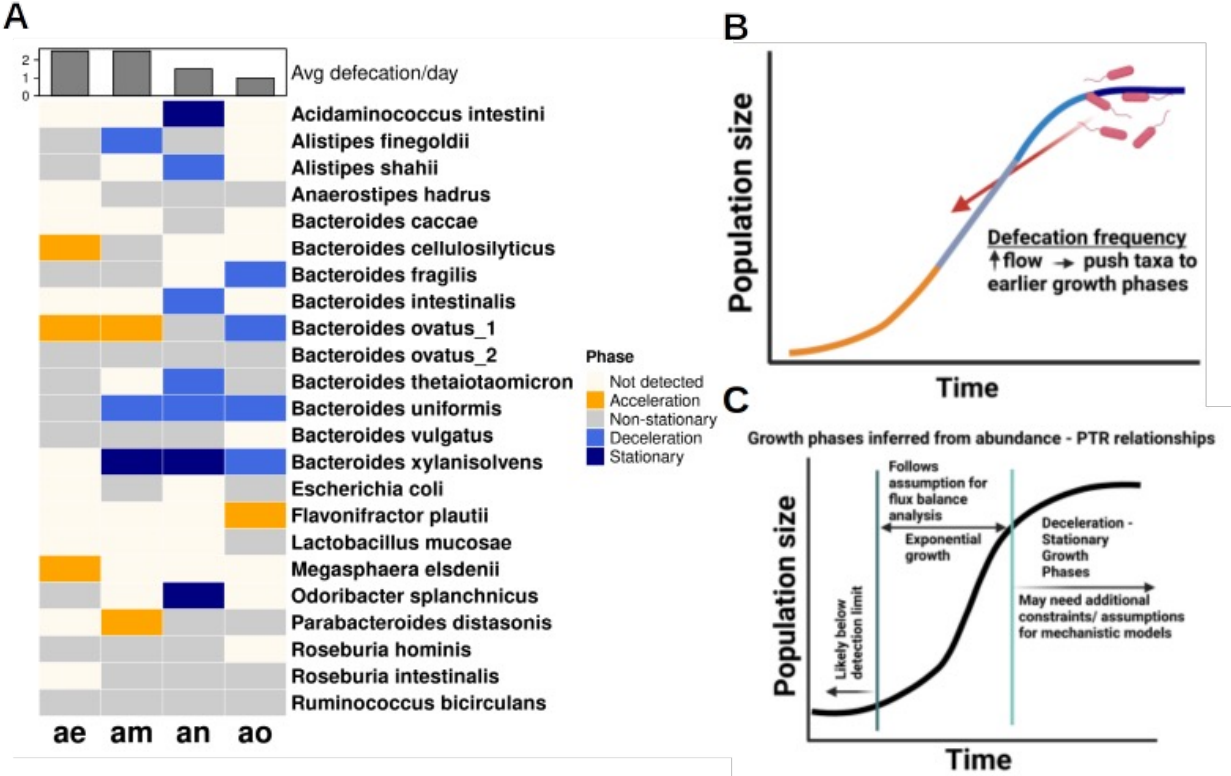


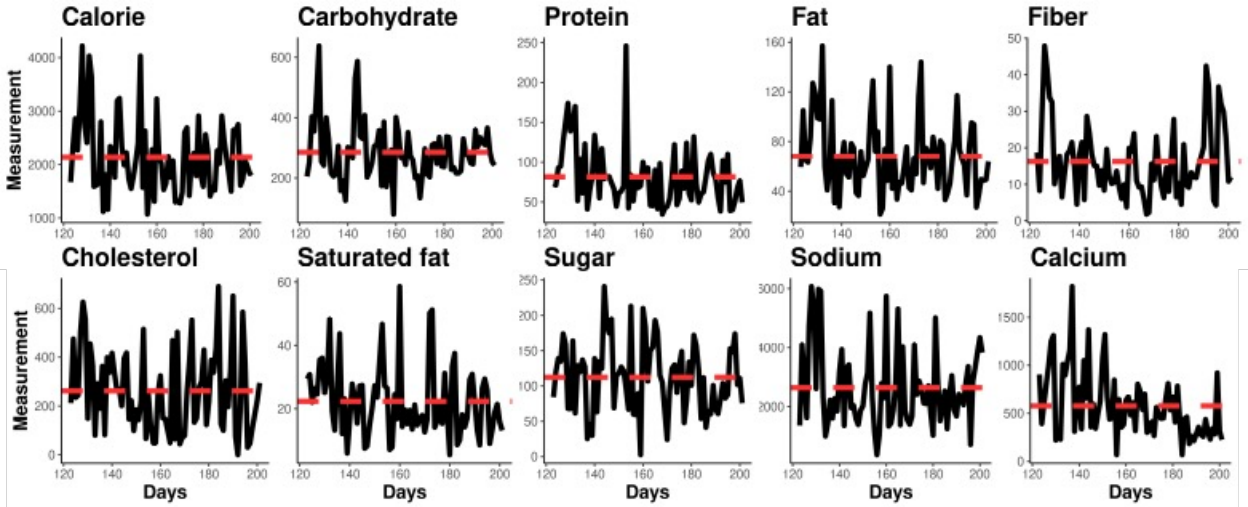
$$\frac{dx_i}{dt} = r\,x_i(t)\left(1 - \frac{x_i(t)}{K}\right) + \sigma\,x_i(t)\,\omega(t)$$

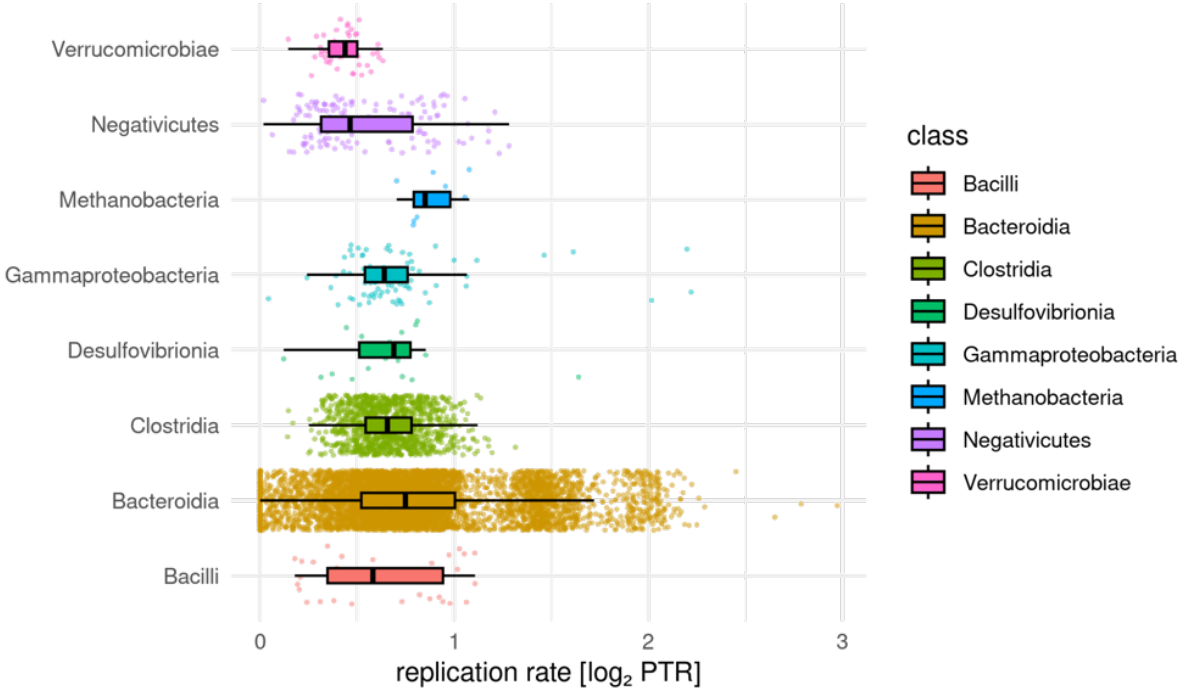


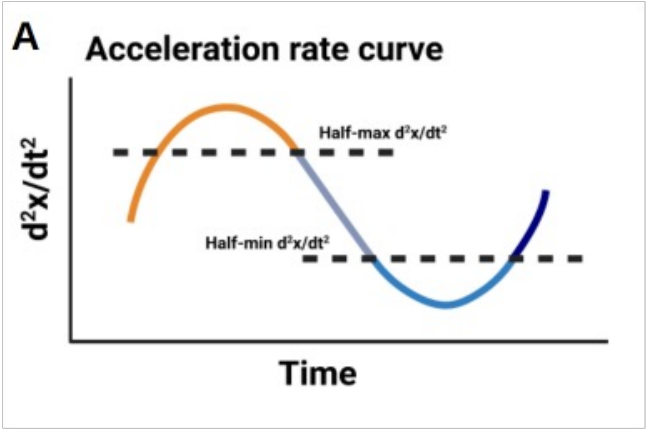


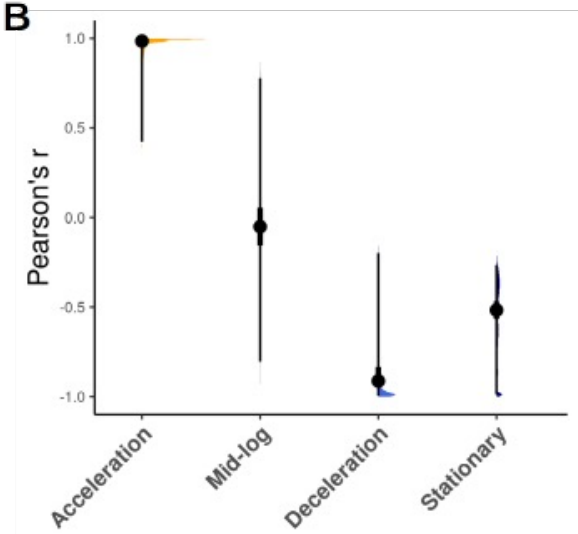

A
OD600 [a.u.]
growth rate [1/h]
time [h]
B
Pearson r
**
*
n.s.
n.s.
acceleration
mid-log
deceleration
stationary
C
acceleration
mid-log
deceleration
stationary
replication rate [log2 PTR]
E. coli abundance [centered log-ratio]

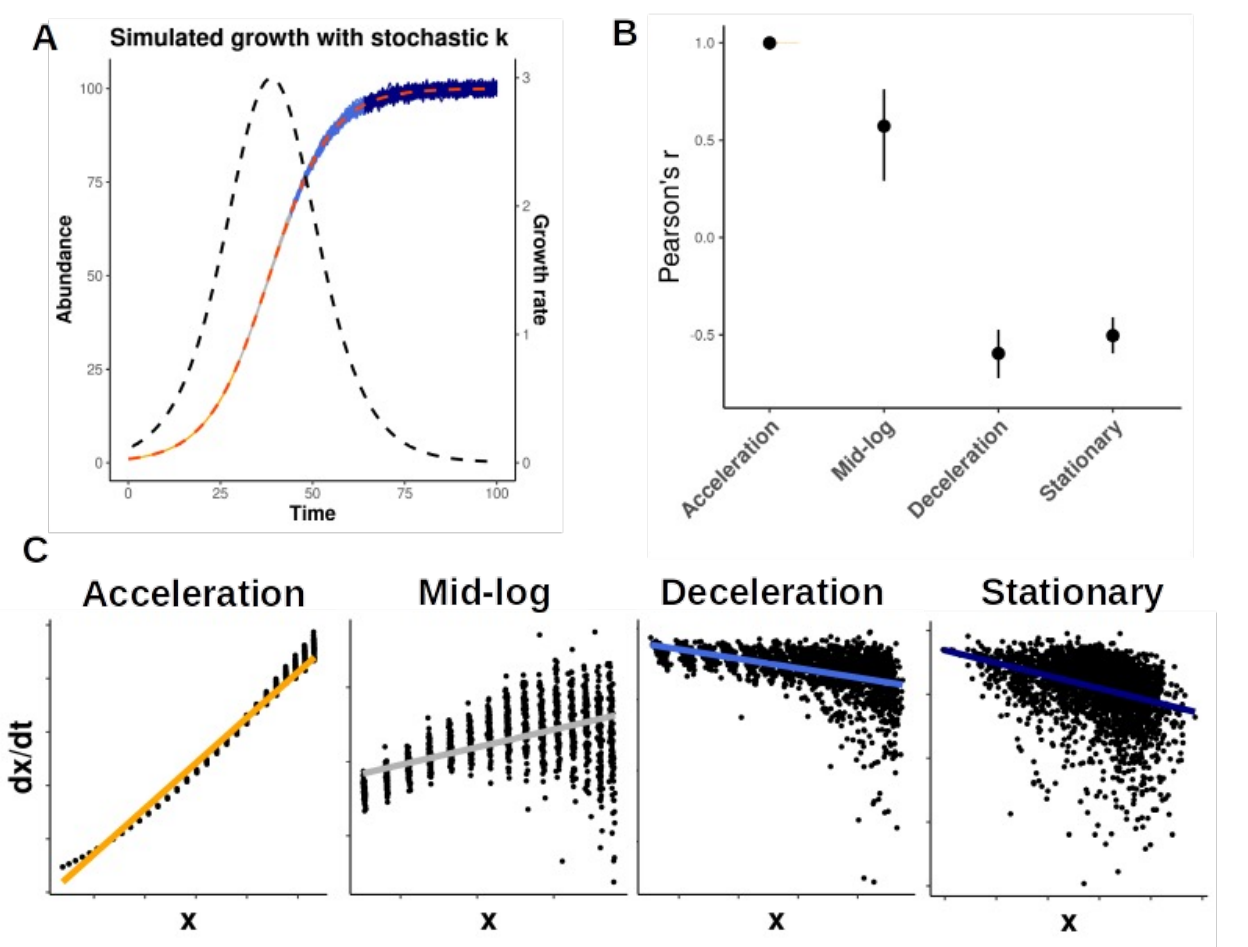

A
Avg defecation/day
2
1
0
Acidaminococcus intestini
Alistipes finegoldii
Alistipes shahii
Anaerostipes hadrus
Bacteroides caccae
Bacteroides cellulosilyticus
Bacteroides fragilis
Bacteroides intestinalis
Bacteroides ovatus_1
Bacteroides ovatus_2
Bacteroides thetaiotaomicron
Bacteroides uniformis
Bacteroides vulgatus
Bacteroides xylanisolvens
Escherichia coli
Flavonifractor plautii
Lactobacillus mucosae
Megasphaera elsdenii
Odoribacter splanchnicus
Parabacteroides distasonis
Roseburia hominis
Roseburia intestinalis
Ruminococcus bicirculans
ae
am
an
ao
Phase
Not detected
Acceleration
Non-stationary
Deceleration
Stationary
B
Population size
Time
Defecation frequency
↑flow → push taxa to
earlier growth phases
C
Growth phases inferred from abundance - PTR relationships
Population size
Time
Follows
assumption for
flux balance
analysis
Exponential
growth
Deceleration -
Stationary
Growth
Phases
May need additional
constraints/ assumptions
for mechanistic models
Likely below
detection limit


Calorie
Carbohydrate
Protein
Fat
Fiber
Cholesterol
Saturated fat
Sugar
Sodium
Calcium
Measurement
Days

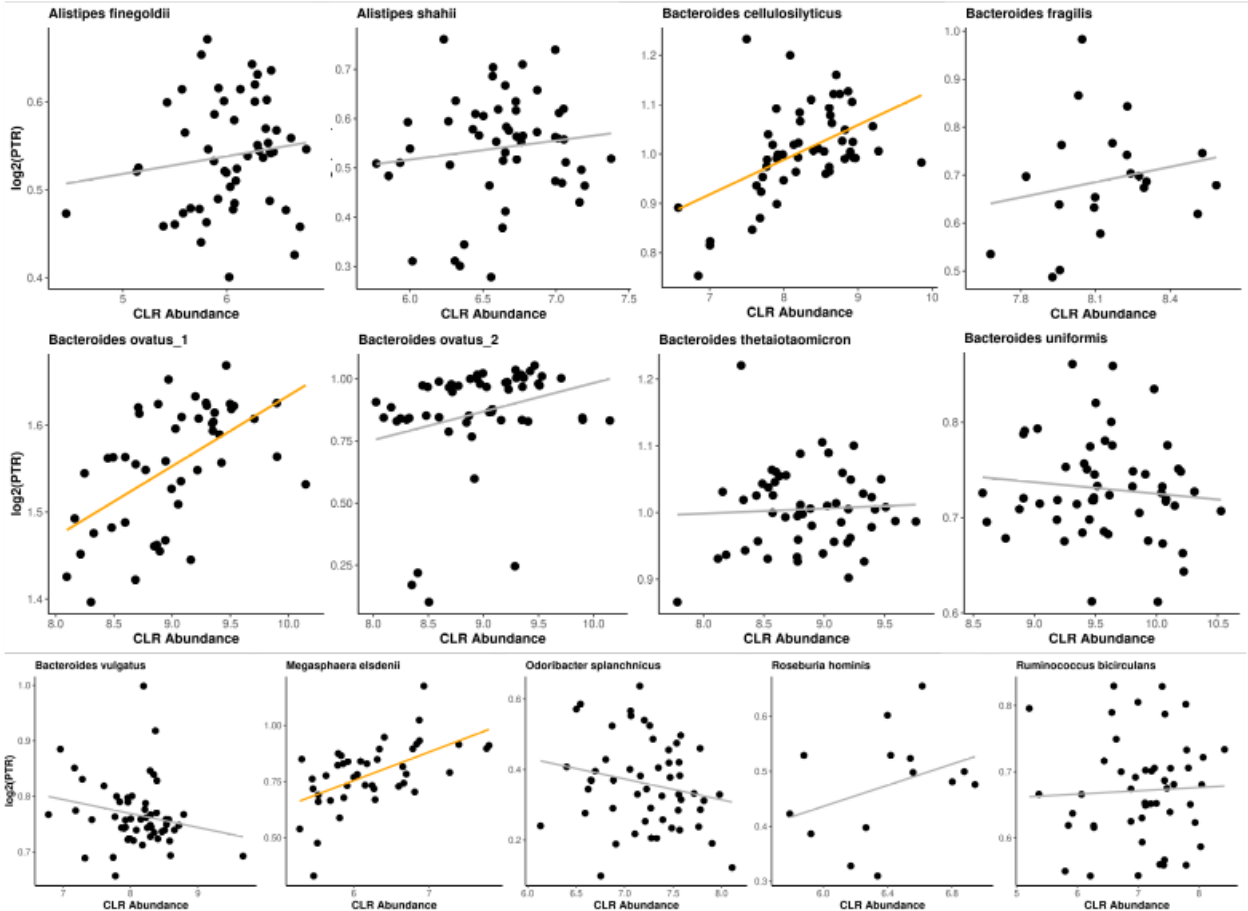

Verrucomicrobiae
Negativicutes
Methanobacteria
Gammaproteobacteria
Desulfovibrionia
Clostridia
Bacteroidia
Bacilli
0
1
2
3
replication rate [log2 PTR]
class
Bacilli
Bacteroidia
Clostridia
Desulfovibrionia
Gammaproteobacteria
Methanobacteria
Negativicutes
Verrucomicrobiae


A
Acceleration rate curve
Half-max d²x/dt²
Half-min d²x/dt²
d²x/dt²
Time

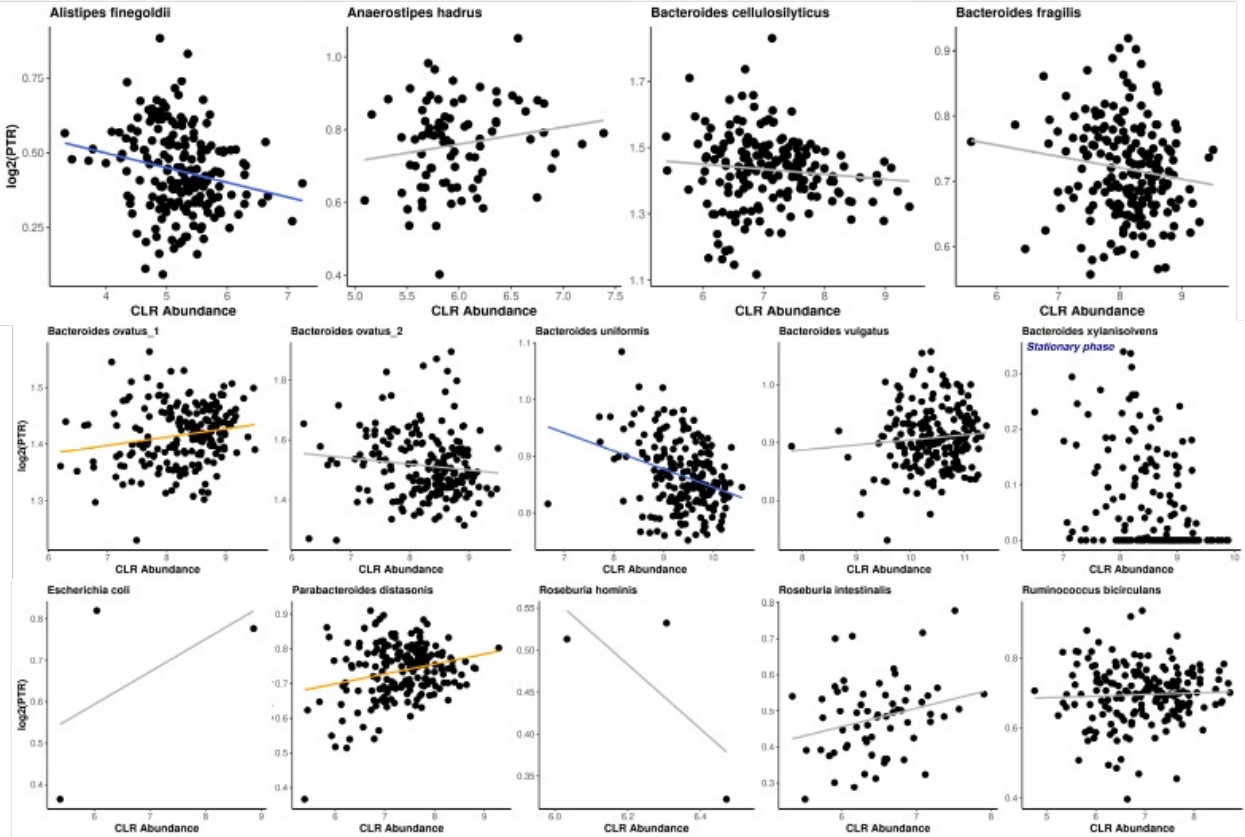

B
Pearson's r
1.0
0.5
0.0
-0.5
-1.0
Acceleration
Mid-log
Deceleration
Stationary


A
Simulated growth with stochastic k
Abundance
Growth rate
Time
B
Pearson's r
Acceleration
Mid-log
Deceleration
Stationary
C
Acceleration
Mid-log
Deceleration
Stationary
dx/dt
x

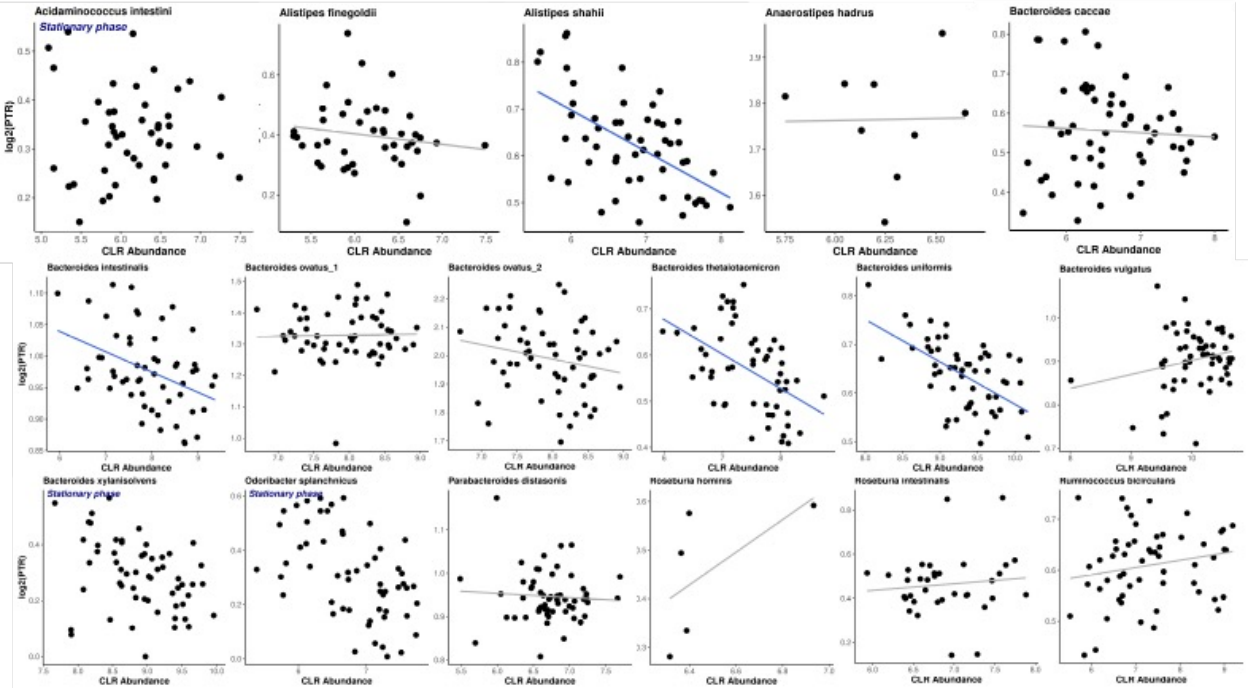
Donor ae


Alistipes finegoldii
Alistipes shahii
Bacteroides cellulosilyticus
Bacteroides fragilis
Bacteroides ovatus_1
Bacteroides ovatus_2
Bacteroides thetaiotaomicron
Bacteroides uniformis
Bacteroides vulgatus
Megasphaera elsdenii
Odoribacter splanchnicus
Roseburia hominis
Ruminococcus bicirculans
log2(PTR)
CLR Abundance

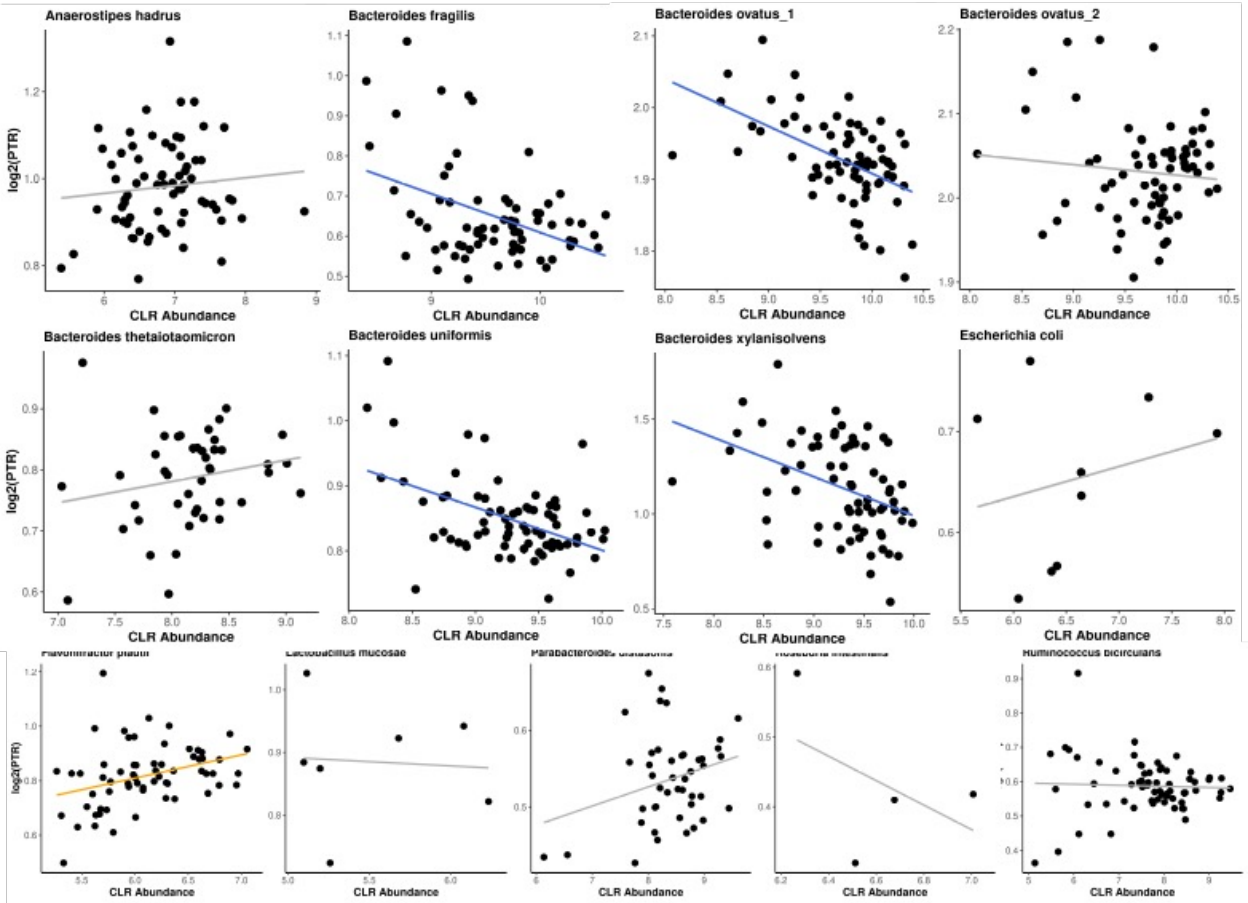
Donor am


Alistipes finegoldii
Anaerostipes hadrus
Bacteroides cellulosilyticus
Bacteroides fragilis
Bacteroides ovatus_1
Bacteroides ovatus_2
Bacteroides uniformis
Bacteroides vulgatus
Bacteroides xylanisolvens
Stationary phase
Escherichia coli
Parabacteroides distasonis
Roseburia hominis
Roseburia intestinalis
Ruminococcus bicirculans
log2(PTR)
CLR Abundance

Donor an


Acidaminococcus intestini
Stationary phase
Alistipes finegoldii
Alistipes shahii
Anaerostipes hadrus
Bacteroides caccae
Bacteroides intestinalis
Bacteroides ovatus_1
Bacteroides ovatus_2
Bacteroides thetaiotaomicron
Bacteroides uniformis
Bacteroides vulgatus
Bacteroides xylanisolvens
Stationary phase
Odoribacter splanchnicus
Stationary phase
Parabacteroides distasonis
Roseburia hominis
Roseburia intestinalis
Ruminococcus bicirculans
log2(PTR)
CLR Abundance

Donor ao


Anaerostipes hadrus
Bacteroides fragilis
Bacteroides ovatus_1
Bacteroides ovatus_2
Bacteroides thetaiotaomicron
Bacteroides uniformis
Bacteroides xylanisolvens
Escherichia coli
Lactobacillus mucosae
Ruminococcus bicirculans
log2(PTR)
CLR Abundance