

# Lawrence Berkeley National Laboratory

## LBL Publications

### Title

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

### Permalink

<https://escholarship.org/uc/item/9xm745m8>

### Journal

Nature Communications, 14(1)

### ISSN

2041-1723

### Authors

Lim, Joe J

Diener, Christian

Wilson, James

et al.

### Publication Date

2023-09-01

### DOI

10.1038/s41467-023-41424-1

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

# Growth phase estimation for abundant bacterial populations sampled longitudinally from human stool metagenomes

---

Received: 27 April 2022

---

Accepted: 4 September 2023

---

Published online: 14 September 2023

---

 Check for updates

---

Joe J. Lim<sup>1</sup>, Christian Diener<sup>2</sup>, James Wilson<sup>2</sup>, Jacob J. Valenzuela<sup>2</sup>,  
Nitin S. Baliga<sup>2,3,4,5,6</sup> & Sean M. Gibbons<sup>2,6,7,8,9</sup> ✉

Longitudinal sampling of the stool has yielded important insights into the ecological dynamics of the human gut microbiome. However, human stool samples are available approximately once per day, while commensal population doubling times are likely on the order of minutes-to-hours. Despite this mismatch in timescales, much of the prior work on human gut microbiome time series modeling has 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 stationary system, where population dynamics occur internally and the bacterial population sizes measured in a bolus of stool represent a steady-state endpoint of these dynamics. We formalize this idea as stochastic logistic growth. 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. 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.

The human gut is an anaerobic bioreactor, ecologically distinct to each individual, that transforms dietary and host substrates into bioactive molecules important to host health<sup>1–3</sup>. Disruptions to the ecological composition of the gut have been shown to mediate the progression of various diseases<sup>4–8</sup>. Furthermore, the ecological dynamics of the gut appear to be relevant to both health and disease states<sup>9,10</sup>. 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<sup>11–15</sup>. These models

often assume, either explicitly or implicitly, that day-to-day changes in abundance are proportional to population growth and/or death<sup>16</sup>. However, the underlying data often do not match this assumption<sup>11,16–20</sup>. 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, likely ranging from minutes-to-hours, although precise in vivo estimates are not available (we contend that doubling times of a day or more in the gut would not be sufficient to maintain a stable population size with a daily defecation rate)<sup>21–23</sup>. However, stool sampling frequency is usually limited to,

---

<sup>1</sup>Department of Environmental & Occupational Health Sciences, University of Washington, Seattle, WA 98105, USA. <sup>2</sup>Institute for Systems Biology, Seattle, WA 98109, USA. <sup>3</sup>Departments of Biology and Microbiology, University of Washington, Seattle, WA 98105, USA. <sup>4</sup>Lawrence Berkeley National Laboratory, CA 94720 Berkeley, USA. <sup>5</sup>Molecular and Cellular Biology Program, University of Washington, WA 98105 Seattle, USA. <sup>6</sup>Molecular Engineering Graduate Program, University of Washington, WA 98105 Seattle, USA. <sup>7</sup>Department of Bioengineering, University of Washington, Seattle, WA 98105, USA. <sup>8</sup>Department of Genome Sciences, University of Washington, Seattle, WA 98105, USA. <sup>9</sup>eScience Institute, University of Washington, Seattle, WA 98105, USA.

✉ e-mail: [sgibbons@isbscience.org](mailto:sgibbons@isbscience.org)

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<sup>16</sup>.

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 effective growth rates of bacterial populations in situ is to leverage metagenome-inferred replication rates<sup>21,22</sup>. 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<sup>21–23</sup>. 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.

Early experiments by Jaques Monod<sup>24</sup> identified distinct growth phases for bacterial populations in culture, which can be captured by the stochastic logistic growth equation (sLGE)<sup>25</sup>. The sLGE has been shown to be a good fit for bacterial population growth in vitro and in real-world, steady-state ecosystems<sup>26–32</sup>. 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, and 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 quasi-batch-culture, flow-through, steady-state ecosystem, like the human gut.

To assess our model predictions, we sampled *Escherichia 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 simulations. We also measured population abundance and replication rate trajectories from more than a dozen organisms across four densely sampled human gut metagenomic time series<sup>33</sup>. On average, when controlling for taxonomy, 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. However, we saw more heterogeneity for specific taxa within individual donor time series. 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<sup>8,34,35</sup>, agricultural systems<sup>36,37</sup>, climate change<sup>36,38,39</sup>, and industrial bioreactors<sup>40,41</sup>.

## Results

### Framing the gut as an anaerobic flow-through bioreactor

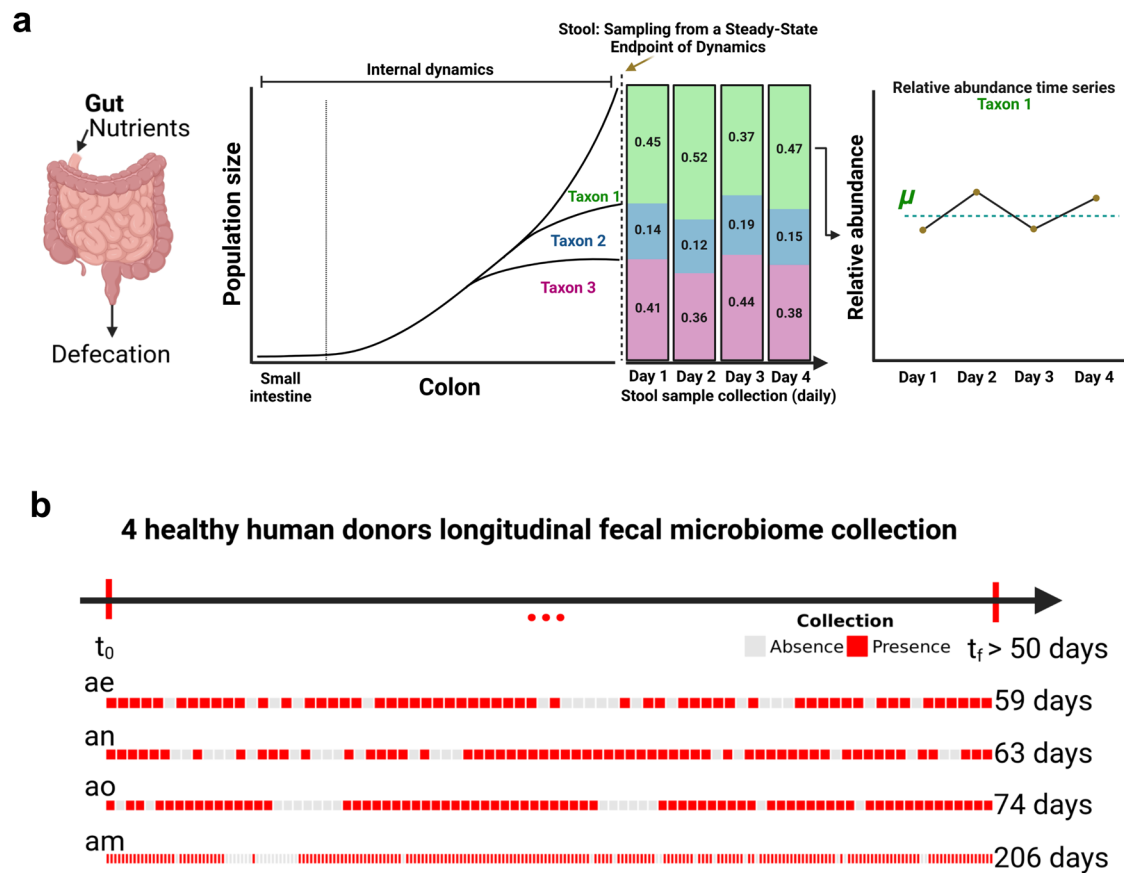
The mammalian gut can be understood as an anaerobic batch culture reactor with a semi-continuous input (i.e., discrete boluses of dietary inputs, mixed with host substrates like mucin and bile acids) and output (i.e., discrete boluses of stool)<sup>42</sup>, and microbial taxa must grow fast enough within 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 ( $\mu$ ) 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)<sup>33</sup>. The BIO-ML cohort contained 84 donors<sup>33</sup>. 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 16 S rRNA gene amplicon sequencing data generated from densely sampled human stool time series<sup>15,33</sup>. Indeed, we found that abundant bacterial populations in the stool of the four BIO-ML donors maintained stable average abundances over time ( $\mu$ ), 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 same negative association (Fig. 2c). Furthermore, similar negative associations were uniformly observed across all taxa analyzed, across all four donors (Fig. 2d). This consistent 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<sup>15,32</sup>.

One important ecological factor that can impact gut microbial dynamics is host diet<sup>43,44</sup>. 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<sup>45–47</sup>. Prior work demonstrated that macronutrient intake within an individual is largely stable over time and does not show significant autocorrelation or drift<sup>15,48</sup>. 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<sup>15,32,33</sup>, 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. Furthermore, we have shown in prior studies on these same BIO-ML



**Fig. 1 | Conceptual figure showing internal bacterial population dynamics within a human gut and the observed steady-state abundance derived from stool sampling.** **a** Mammalian guts are continuous flow-through systems. Taxa grow in the large intestine with varying growth rates, carrying capacities, and steady-state population sizes, and may be in different growth phases at the time of stool sampling. 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.

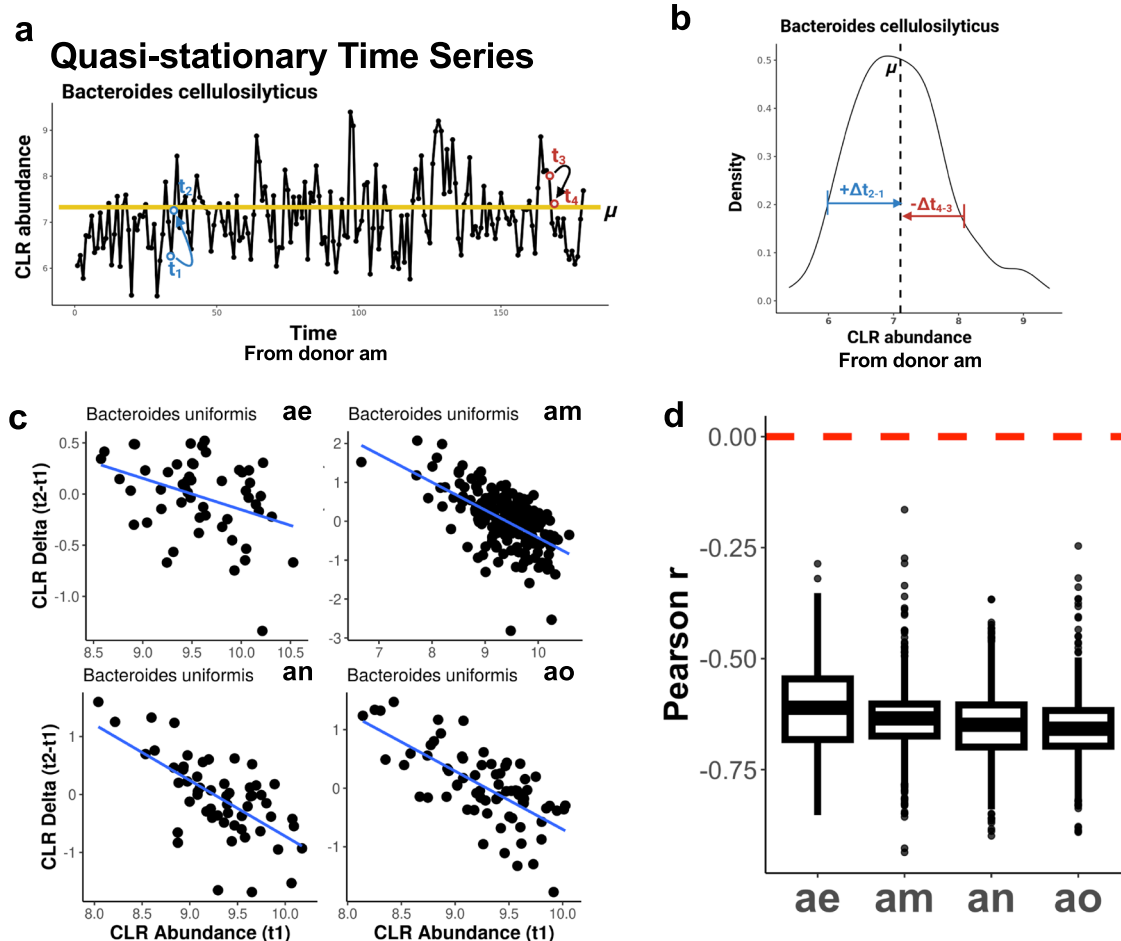
participants that average within-person taxon abundances are highly correlated across people, which suggests that these carrying capacities, likely representing dietary and host substrate niches, are fairly conserved across humans<sup>15,33</sup>.

Next, we looked at the statistical associations between calculated peak-to-trough ratios (i.e., PTRs; a proxy for the effective growth-rate) for abundant bacterial populations from each metagenomic sample and their respective metagenomic population abundance estimates<sup>22</sup>. 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 putative regression-to-the-mean signature identified for the deltas, we found variable statistical relationships between  $\log_2(\text{PTR})$  and centered log-ratio (CLR) transformed population abundances for the same taxon across the four donors (*Bacteroides ovatus\_1*, Fig. 3a). Similarly, we saw a wide range of positive, negative, and null associations between  $\log_2(\text{PTRs})$  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<sup>49–51</sup>. Finally, we calculated temporally-averaged 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  $\log_2(\text{PTR})$  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. We also looked into whether or not  $\log_2(\text{PTR})$  magnitudes were inter-comparable across taxa (Fig. S2). We calculated  $\log_2(\text{PTRs})$  for all abundant taxa detected across all 84 BIO-ML donors and found that the median  $\log_2(\text{PTR})$  was fairly similar across taxonomic classes (–0.45–0.75), with most classes showing a wide range (Fig. S2). To assess whether or not  $\log_2(\text{PTR})$ -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 positive association, independent of taxonomy (class-level  $\beta = 0.0612$ ,  $p = 8.359e^{-60}$ ; species-level  $\beta = 0.0101$ ,  $p = 0.0006$ ). While we saw a significantly positive cross-sectional association between PTRs and abundances when controlling for species identity, we found that the vast majority of individual species showed null associations between abundance-PTR relationships, indicating that we were perhaps underpowered to detect this weak effect at the level of individual taxa (Figs. 3 and S3). For three species, we actually saw significant negative associations: *Alistipes shaii*, *Alistipes finegoldii*, and *Odoribacter splanchnicus* (FDR-adjusted  $p < 0.1$ ; Fig. S3).

### Stochastic logistic growth equation provides insights into growth phases

In order to better understand and interpret the varying relationships we observe between  $\log_2(\text{PTRs})$  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 (LGE) (Fig. 4). Although simple models like the LGE do not capture the



**Fig. 2 | Regression-to-the-mean effect in human microbial time series data.**  
**a** Yellow line represents the mean abundance ( $\mu$ ) 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

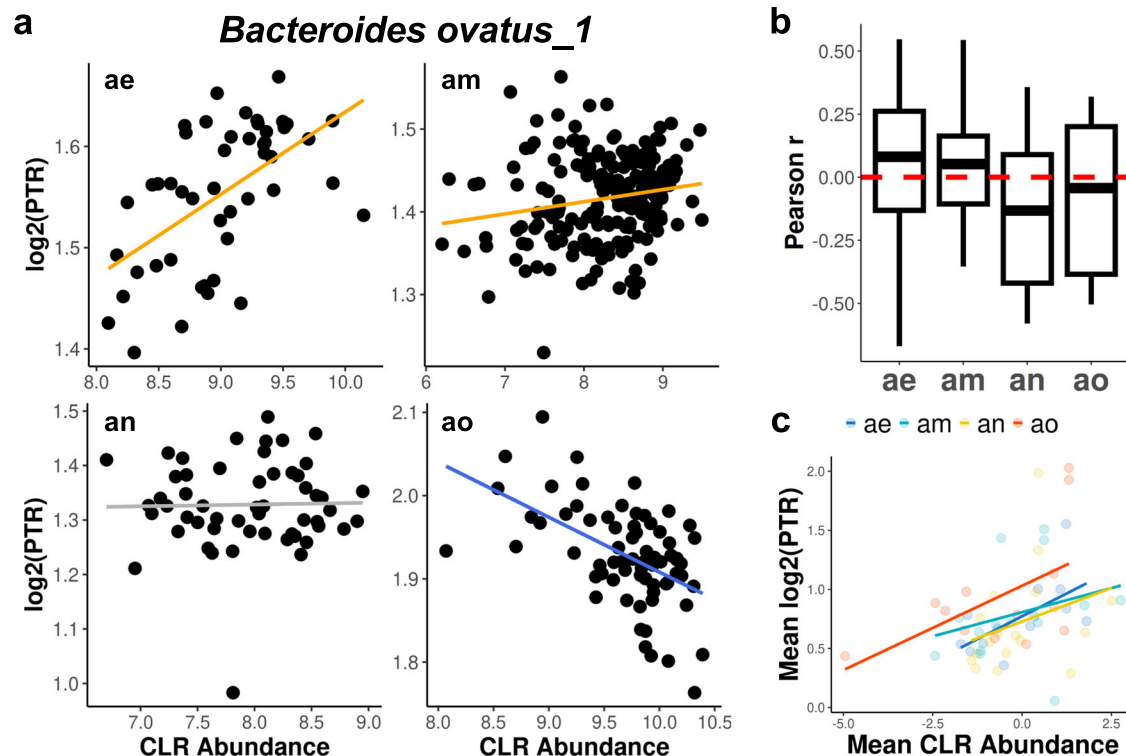
donor am, which is approximately normally distributed. **c** Delta vs. abundance for *Bacteroides uniformis* time series from donors ae, am, an, and ao. **d** Box plots (showing minima, 25th percentile, median, 75th percentile, and maxima) of Pearson  $r$  values for deltas vs. abundances across all taxa time series in all four donors. Red line indicates a Pearson correlation coefficient of 0.

full complexity of the gut, such as spatial structure and specific resource usage, the LGE and its variations have been widely applied to population and community dynamics<sup>26,27,31,32,52</sup>. 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  $(1 - x_i(t)/k)$ <sup>53</sup>. 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 abundance, which reflects the effective rate of population growth at a given point along the curve (i.e., effective growth rate over time), and peaks during mid-log phase (Fig. 4a, middle panel). The second derivative of abundance with respect to time, which is the instantaneous change in effective 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 the effective growth rate (i.e.,  $dx/dt$ ) and abundance moving 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 semi-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<sup>27,32,54</sup>. In order to approximate these fluctuations in our modeling, we introduced a stochastic term to the logistic growth model (Fig. 5a). Herein,  $\sigma$  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. S4A). 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 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 when combining simulation results across a wide range of parameter space and varying noise levels (Fig. S4B).

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. S5). Fluctuations in  $k$  did



**Fig. 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<sub>2</sub>(PTR) and CLR-normalized abundance relationships for *Bacteroides ovatus\_1* in 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. Donor ae:  $p = 0.0005$ , donor am:  $p = 0.0317$ , donor an:  $p = 0.925$ , donor ao:  $p = 0.00005$ . **b** Box plots (showing minima, 25th percentile, median, 75th percentile, and maxima) of Pearson  $r$  values combined for all filtered taxa for each donor. **c** Mean log<sub>2</sub>(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).

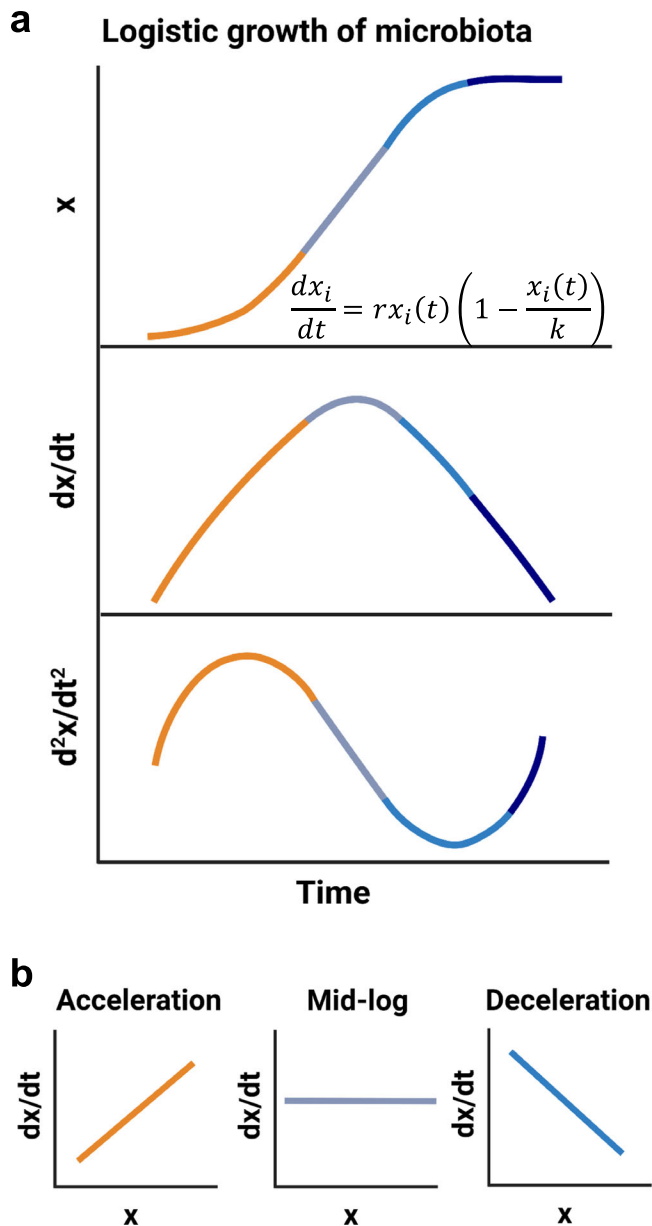
not alter the sigmoidal shape of the sLGE curve (Fig. S5A), and the relationships between abundances and growth rates across growth phases were preserved (Fig. S5B–C). Simulations with the noise level set at 0.2 recapitulated the correlation coefficient distributions seen in the human time series data, while maintaining the distinct growth-abundance relationships previously associated with each growth phase (Fig. S6). However, these distinct relationships could be destroyed at high enough noise levels (Fig. S6).

### Validating sLGE growth phase inferences in vitro

To validate the relationship between growth/replication 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). Effective growth rates were quantified as the log<sub>2</sub>(PTR) for each *E. coli* sample<sup>55</sup>. The relationships between the log<sub>2</sub>(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<sub>2</sub>(PTR) of  $1.25 \pm 0.167$  ( $\pm$  standard deviation), while samples in stationary phase had an average log<sub>2</sub>(PTR) of  $0.358 \pm 0.059$ , which clearly distinguished between these phases that lacked a differentiable correlation signal.

### 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<sub>2</sub>(PTR) > 0.358) or stationary (i.e., log<sub>2</sub>(PTR) < 0.358) phase (Fig. 6). This log<sub>2</sub>(PTR) threshold of 0.358 is somewhat arbitrary, based on experimental data from a single organism (Fig. 6), but it was able to classify several gut species (*Escherichia coli*, *Citrobacter rodentium*, *Lactobacillus gasseri*, *Enterococcus faecalis*, and several more) grown into stationary phase from prior in vitro experiments<sup>21</sup>. For those taxa with average log<sub>2</sub>(PTRs) above the empirical stationary phase threshold, significantly positive associations (linear regression, FDR-adjusted  $p$  value < 0.05, with a positive beta-coefficient) between log<sub>2</sub>(PTRs) and CLR abundances likely indicate acceleration phase and significantly negative associations (linear regression, FDR-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 S7). *Bacteroides xylanisolvens* had an average log<sub>2</sub>(PTR) that was lower than the stationary threshold in donor am (Fig. S8). *Bacteroides ovatus\_1* and *Parabacteroides distasonis* showed positive log<sub>2</sub>(PTR)-CLR abundance associations, while *Alistipes finegoldii* and *Bacteroides uniformis* showed negative associations in donor am (Figs. 7a and S8). *Acidaminococcus intestini*, *Bacteroides xylanisolvens*, and *Odoribacter splanchnicus* showed average log<sub>2</sub>(PTR) below the empirical stationary phase threshold in donor an (Fig. S9). *Alistipes shahii*, *Bacteroides*



**Fig. 4 | Diagram of the logistic growth equation.** **a** The logistic growth curve models abundance ( $x$ ) with respect to time (top panel). Orange, gray, blue, and navy indicates acceleration, mid-log, deceleration, and stationary phases, respectively. The first derivative of the logistic growth curve shows the growth rate with respect to time (middle panel). The second derivative of the logistic growth curve shows growth acceleration with respect to time (bottom panel). **b** Expected relationships between abundance and growth rate at different locations along the logistic growth curve.

*intestinalis*, *Bacteroides thetaiotaomicron*, and *Bacteroides uniformis* showed significantly negative  $\log_2(\text{PTR})$ -CLR abundance associations in donor an (Figs. 7a and S9). Finally, *Favonifractor plautii* showed a positive  $\log_2(\text{PTR})$ -CLR abundance association and *Bacteroides fragilis*, *Bacteroides ovatus\_1*, *Bacteroides uniformis*, and *Bacteroides xylosovens* showed negative associations in donor ao (Fig. 7a and S10). In all four donors, many taxa showed average  $\log_2(\text{PTRs})$  greater than the stationary threshold but without significant associations between  $\log_2(\text{PTR})$  and CLR abundances (Figs. 7a and S7–10). The absence of a significant association for these putatively non-stationary taxa could indicate 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).

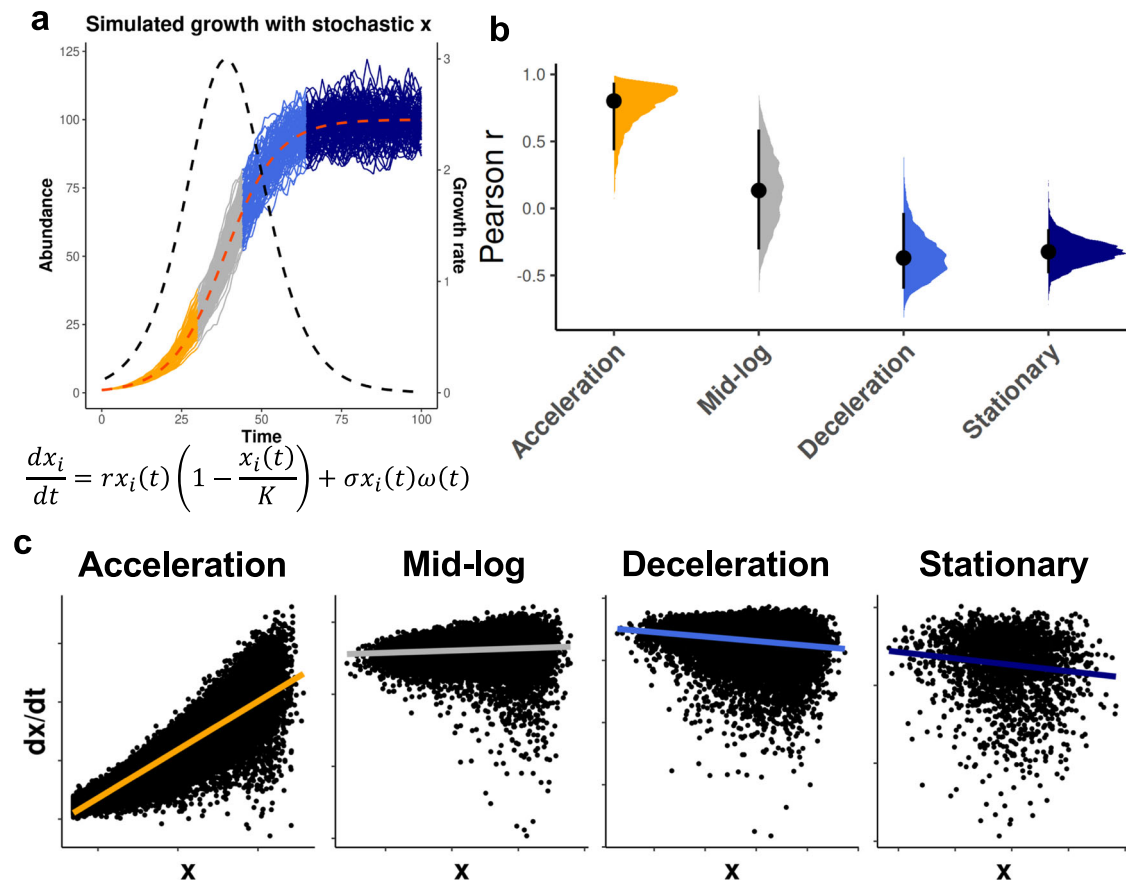
We observed a slight difference in the number of significantly positive and negative PTR-abundance associations between donors ae/am, and an/ao. Donors ae and am tended to have a larger proportion of taxa in acceleration phase, while an and ao tended to have a larger proportion of taxa in deceleration or stationary phases. Interestingly, donors an and ao had a lower average defecation frequency ( $\leq 1$  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 putative phase estimates for all taxa with sufficient longitudinal data (Fig. 7a). Our approach provides a new potential 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 likely faster (minutes-to-hours) than our sampling timescales (days). 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, metagenome-derived PTRs have been shown to be proportional to effective growth rate measures in vitro<sup>21–23,56–58</sup>. However, normalizing PTR values across taxa remains a challenge. For example, one might assume that the time required to replicate a particular bacterial chromosome (i.e., C-period)<sup>21,59</sup> would be an important conversion factor between PTRs and effective growth rates. However, prior work has shown that genome size and growth rates are uncorrelated in bacteria and archaea, and that ribosomal gene copy number explains much more variance in bacterial growth rates<sup>60</sup>. Ribosomal gene copy number, on average, has been shown to be higher for microbes living in the human gut, as compared to taxa in non-host-associated environments<sup>61–64</sup>, which suggests that gut microbial taxa are optimized for higher maximal growth rates. Even if we cannot precisely convert PTRs to effective growth rates that are intercomparable across taxa, prior work has shown similar ranges of minimum and maximum PTRs across diverse gut-associated taxa in vitro<sup>21</sup>, so we have made the coarse assumption that these values can be roughly compared across organisms (Fig. 3c).

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  $\log_2(\text{PTR})$ -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). Prior work has shown that the sLGE is an optimal model for capturing within-person strain- and species-level dynamics for the vast majority of human gut commensals<sup>32</sup>. Simulations showed how these diagnostic patterns held across a wide range of external stochastic noise levels (Fig. 5). Furthermore, fluctuations in carrying capacities, which could be interpreted as fluctuations in niche-size due to dietary variation, could not ablate these patterns (Fig. S5). However, as one might expect, adding



**Fig. 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 ( $\sigma=0.1$ ). Black circles represent the median and black bars show 95% confidence interval. **c** Scatter plots in log scale showing relationships between abundance and growth rate across the four growth phase regions defined in (a).

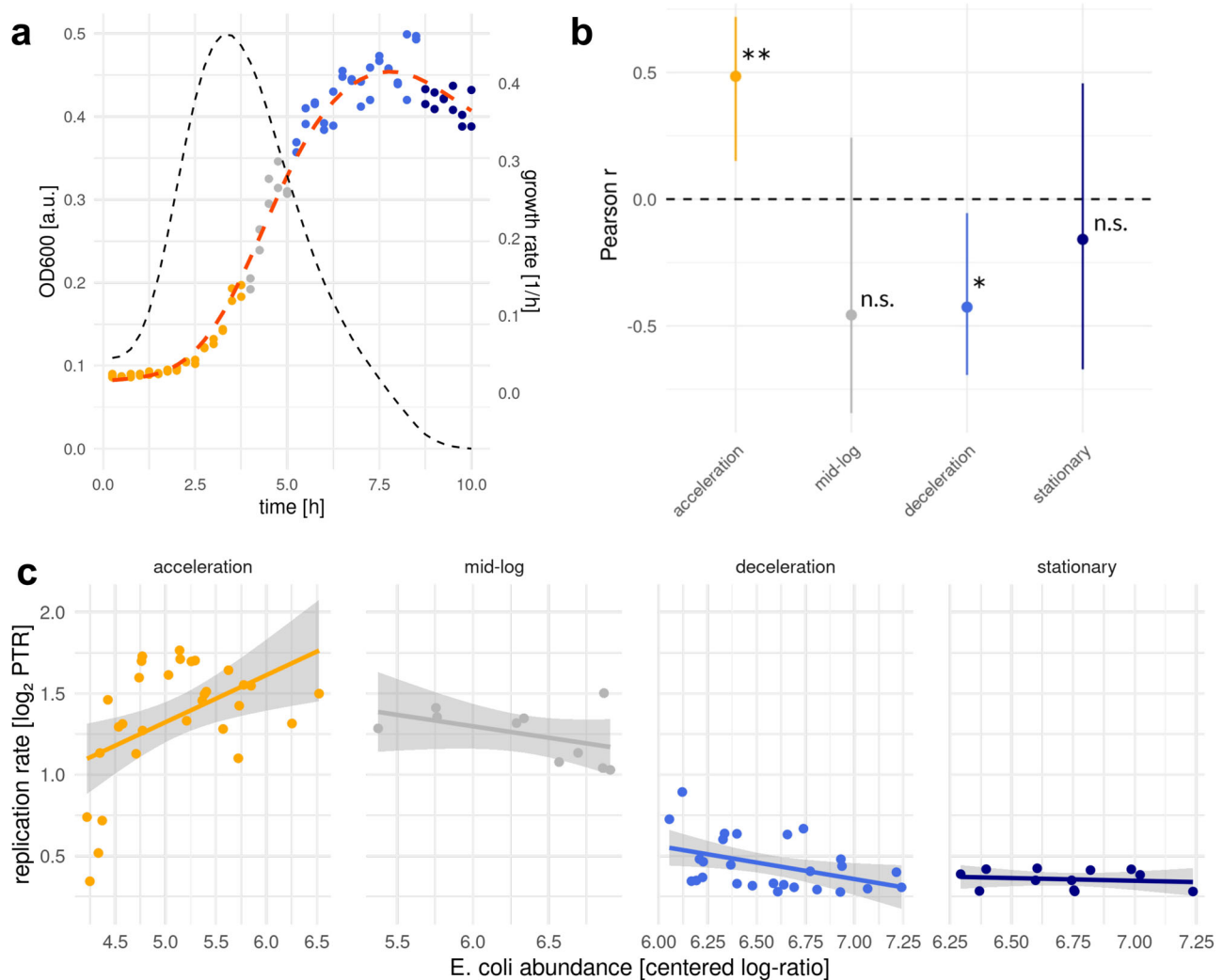
enough noise to these models eventually did override the signal (Fig. S6).

We validated the putative growth phase diagnostic patterns from our sLGE model in vitro and saw marked correspondence (Figs. 5 and 6). The in vitro data indicated that the average  $\log_2$ (PTR) across the growth curve could easily distinguish between stationary phase (i.e., no effective growth) and the other phases (Fig. 6). This is particularly useful for distinguishing stationary phase from log-phase growth, which are both expected to show a null association between  $\log_2$ (PTRs) and abundances (Figs. 4 and 5). We saw that *E. coli* grown in vitro showed a maximum  $\log_2$ (PTR) of -1.75 and a minimum of -0.3, which is quite consistent with  $\log_2$ (PTR) ranges observed for other organisms growing in vitro and in vivo<sup>21</sup>. While  $\log_2$ (PTR) ranges can vary across taxa, we tentatively set a  $\log_2$ (PTR) threshold of  $\leq 0.358$  to assign taxa measured in vivo to putative stationary phase, which is on the lower end of what has been observed in vitro and in vivo<sup>21</sup>.

We applied our sLGE predictions to four human gut metagenomic time series. Consistent with our sLGE 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 average community PTRs appeared to increase with increasing BMF<sup>65</sup>. 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, albeit weakly, correlated, which is consistent with exponentially-growing populations (Fig. 3c). We were also able to identify specific taxa that were abundant in stool that appeared to be in stationary phase, based on our in vitro stationarity threshold (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<sup>66,67</sup>. Genome-scale metabolic modeling and flux-balance analysis (FBA) has been effective mechanistic tools for simulating these metabolic exchanges, especially in controlled bioreactor systems<sup>68</sup>. 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<sup>69-71</sup>. Alternatively, organisms that are not actively growing in the distal colon could be omitted from community-scale metabolic models of colonic metabolism<sup>72</sup>. 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.





**Fig. 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 from two-sided correlation tests without adjustment for multiple comparisons. \*\*:  $p < 0.01$

(acceleration:  $p = 0.007$ ), \*:  $p < 0.05$  (deceleration:  $p = 0.003$ ), n.s.: not significant (mid-log: 0.184, stationary: 0.622). Black circles represent the median and black bars show 95% confidence interval. Pooled duplicate samples (4 sets of replicate cultures) for 40 time points in total were used (see Methods). **c** Scatter plots in log scale showing relationships between abundance and replication rate ( $\log_2$  PTR) across the four growth phase regions defined in (a). Gray regions represent 95% confidence intervals.

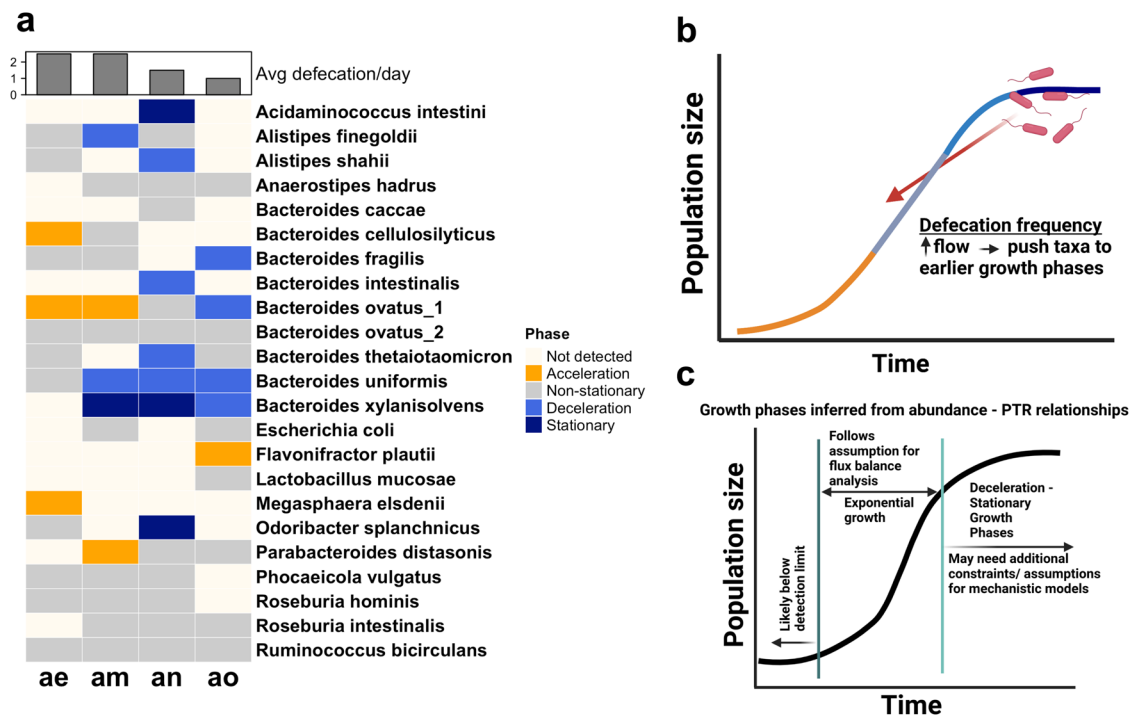
In conclusion, we provide a new path forward for the biological interpretation of metagenomic time series data generated from adult human stool samples. Our results are somewhat reassuring for cross-sectional studies, as they indicate that bacterial abundances in the gut fluctuate around stable carrying 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<sup>33</sup>. 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.

#### Study limitations

One critique of our approach is that dense longitudinal sampling of human stool metagenomes is necessary for growth phase inference.

Currently, these data sets are rare, which limits the immediate use of our approach. However, interest is growing for long-term monitoring of gut microbial variation<sup>73–75</sup>. Through technological advances, dense metagenomic time series will become more and more common over time<sup>76,77</sup>.

Another limitation is in our ability to relate PTRs to effective growth rates. Earlier research compared PTR curves and measured growth curves (OD600)<sup>21</sup>, which allows for a continuous mapping between  $dx/dt$  (i.e., the effective population growth rate) and PTRs. Although our findings lack thorough quantitative estimates of in vivo effective population growth rates, we make the assumption here that in vivo PTRs are indeed related to growth. For some of our analyses, we also assume that PTRs are roughly comparable across taxa. For example, we chose a simple threshold ( $\log_2(\text{PTR}) \leq 0.358$ ) for determining which taxa in the gut are in stationary phase, based on our in vitro *E. coli* experiment. However, as mentioned above, we do find that this threshold also successfully classifies many other gut species into stationary phase from prior in vitro experiments<sup>21</sup>. Nonetheless, future work is needed in order to develop better



**Fig. 7 | In vivo growth phase estimation.** **a** We find variable relationships between  $\log_2(\text{PTRs})$  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  $\log_2(\text{PTR})$ -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  $\log_2\text{PTR} < 0.358$ ). Non-stationary taxa (i.e., above the stationary phase threshold, but lacking a significant

correlation between  $\log_2(\text{PTRs})$  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 (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.

conversion factors that enable direct comparisons of PTRs across taxa.

The jury is still out on whether sLGE, in all its simplicity, is an appropriate coarse-grained model for human gut commensal dynamics. However, recent work has found that 86.6% of species time series from the same BIO-ML stool donors studied here could be optimally fit by the sLGE, using a fairly stringent significance cutoff<sup>32</sup>. *Phocaeicola vulgatus* was one of the few taxa that overlapped with our analysis that previously showed a poor fit to the sLGE across stool donors<sup>32</sup>. In our analysis *P. vulgatus* never showed a significant association between PTRs and abundances across any of the stool donors, but its average PTR was higher than the stationary phase threshold (Fig. 7). Thus, we only applied a stationary phase cutoff to this organism. We look forward to other independent research groups testing, and hopefully improving, our in situ growth phase inference approach as additional data sets become available.

## 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.<sup>48</sup>. We tested for stationarity in these nutrient intake time series using the augmented Dickey-Fuller (ADF) test (tseries package in R<sup>78</sup>), 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<sup>79</sup>, 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 (2  $\mu\text{L}$ ) into a non-treated 96-well flat-bottomed plate (Thomas Scientific Cat No. 1154Q44) containing 198  $\mu\text{L}$  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<sup>33</sup>. Raw FASTQ files from the BIO-ML cohort and from the in vitro *E. coli* experiment were filtered and trimmed using FASTP<sup>80</sup>, 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<sup>81</sup>. Abundances on the kingdom, phylum, genus, and species ranks were then obtained using Bracken<sup>82</sup>. Trimmed and filtered reads were then aligned to 2,935 representative bacterial reference genomes taken from the IGG database (version 1.01) using Bowtie2<sup>83,84</sup>. Coverage profiles and log<sub>2</sub> 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<sup>85</sup>. 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<sub>2</sub> peak-to-trough ratios (i.e., log<sub>2</sub>(PTRs)) were read into R version 4.1.3 for analysis<sup>79</sup>. All plots were generated using ggplot2<sup>85</sup>, unless indicated otherwise. Raw read counts for a given taxon within a sample were centered log-ratio (CLR) transformed<sup>86</sup>. Taxa that had matched log<sub>2</sub>(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  $Abundance\ changes(\Delta t) = x(t+1) - x(t)$ , where  $\Delta t < 3\text{days}$ . 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<sub>2</sub>(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<sub>2</sub>(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<sub>2</sub>(PTRs) and abundances were considered to be in acceleration or deceleration phase, respectively. Those with no correlation and an average log<sub>2</sub>(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<sub>2</sub>(PTRs) were significantly associated within each donor, and *p*-values from individual tests were combined using Fisher's method<sup>87</sup>.

### Stochastic logistic growth model simulation

The stochastic logistic growth equation (sLGE) was implemented as:  $\frac{dx_i}{dt} = rx_i(t)(1 - \frac{x_i(t)}{K}) + \sigma x_i(t)\omega(t)$ , where *t* is time, *r* is the growth rate, *x<sub>i</sub>* is the abundance of taxon *i*, *K* is the carrying capacity,  $\sigma$  is the noise magnitude term, and  $\omega(t)$  is the noise distribution term. Using the R package sde<sup>88</sup>, 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):  $\frac{dx_i}{dt} = rx_i(t)(1 - \frac{x_i(t)}{K})$ , the solution for which can be written as  $x_i = \frac{x_{i,0}K e^{rt}}{(K - x_{i,0}) + x_{i,0}e^{rt}}$ .

The *x<sub>i</sub>* 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),  $\frac{d^2x_i}{dt^2} = K^2 x_i(1 - \frac{x_i}{K})(1 - \frac{2x_i}{K})$ , was calculated using solved *x<sub>i</sub>* 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<sub>1</sub>* and *a<sub>2</sub>*, and the half-min, *a<sub>3</sub>* and *a<sub>4</sub>*, were calculated (Fig. S4A). The corresponding simulated time points of *a<sub>j</sub>*, denoted as *s<sub>j</sub>*, 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:  $\frac{dx_i}{dt} = rx_i(t)(1 - \frac{x_i(t)}{K}) - Hx_i(t)$ . Here, *H* is the "harvest rate", which determines the proportional decrease in each timepoint in the equation. At steady state,  $rx_i^*(t)(1 - \frac{x_i^*(t)}{K}) - Hx_i^*(t) = 0$ , where  $x_i^*(t)$  represents the fixed point. Two equilibria exist 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. S5).

In base R, simulation was performed for 100 iterations with the same noise levels ( $\sigma = 0.1$ ) as the representative sLGE simulations with stochastic  $x$ . Major growth phases were defined the same way as sLGE simulations with stochastic  $x$ .

### Statistics and reproducibility

Data for longitudinal daily nutrient intake<sup>48</sup> and human stool metagenomic sequencing<sup>33</sup> were secondary analyses using the profiling data obtained from the original studies. Sample sizes were not calculated, experiments were not randomized, and investigators were not blinded to allocation during outcome assessments. 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*”.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The raw shotgun sequencing data from the in vitro experiment are publicly available from the National Center for Biotechnology (NCBI) Sequence Read Archive (SRA), BioProject accession number [PRJNA942341](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA942341). The raw BIO-ML metagenomic data are available under the SRA BioProject accession number [PRJNA544527](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA544527).

### 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). 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><sup>89</sup>.

### References

- 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).
- 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).
- Zhang, P. et al. Commensal homeostasis of gut microbiota-host for the impact of obesity. *Front. Physiol.* **8**, 1122 (2017).
- 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).
- Fischbach, M. A. Microbiome: focus on causation and mechanism. *Cell* **174**, 785–790 (2018).
- 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).
- 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).
- Bullman, S. et al. Analysis of *Fusobacterium* persistence and antibiotic response in colorectal cancer. *Science* **358**, 1443–1448 (2017).
- Halfvarson, J. et al. Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat. Microbiol.* **2**, 17004 (2017).
- 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).
- 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).
- Bashan, A. et al. Universality of human microbial dynamics. *Nature* **534**, 259–262 (2016).
- 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).
- 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).
- 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).
- 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).
- Momeni, B., Xie, L. & Shou, W. Lotka-Volterra pairwise modeling fails to capture diverse pairwise microbial interactions. *Elife* **6**, e25051 (2017).
- Berry, D. & Widder, S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* **5**, 219 (2014).
- 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).
- Coenen, A. R. & Weitz, J. S. Limitations of correlation-based inference in complex virus-microbe communities. *mSystems* **3**, e00084–18 (2018).
- Korem, T. et al. Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* **349**, 1101–1106 (2015).
- 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).
- Gibson, B., Wilson, D. J., Feil, E. & Eyre-Walker, A. The distribution of bacterial doubling times in the wild. *Proc. Biol. Sci.* **285**, 20180789 (2018).
- Monod, J. THE GROWTH OF BACTERIAL CULTURES. *Annu. Rev. Microbiol.* **3**, 371–394 (1949).
- Logistic stochastic growth models and applications. in *Handbook of the Logistic Distribution* 419–419 (CRC Press, 1991).
- Descheemaeker, L. & de Buyl, S. Stochastic logistic models reproduce experimental time series of microbial communities. *Elife* **9**, (2020).
- Grilli, J. Macroecological laws describe variation and diversity in microbial communities. *Nat. Commun.* **11**, 4743 (2020).
- Lobry, J. R., Flandrois, J. P., Carret, G. & Pave, A. Monod’s bacterial growth model revisited. *Bull. Math. Biol.* **54**, 117–122 (1992).
- Fujikawa, H., Kai, A. & Morozumi, S. A new logistic model for bacterial growth. *Shokuhin Eiseigaku Zasshi* **44**, 155–160 (2003).
- Baranyi, J., McClure, P. J., Sutherland, J. P. & Roberts, T. A. Modeling bacterial growth responses. *J. Ind. Microbiol.* **12**, 190–194 (1993).
- 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**, e75168 (2022).
- Wolff, R., Shoemaker, W. & Garud, N. Ecological Stability Emerges at the Level of Strains in the Human Gut Microbiome. *mSystems* (2023) <https://doi.org/10.1128/mbio.02502-22>.
- 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).
- 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**, e9302 (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. USA* **110**, 12804–12809 (2013).
38. Otwell, A. E. et al. Sulfur metabolites play key system-level roles in modulating denitrification. *mSystems* **6**, e01025–20 (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**, e49305 (2019).
50. Atolia, E. et al. Environmental and physiological factors affecting high-throughput measurements of bacterial growth. *MBio* **11**, e01378–20 (2020).
51. Nguyen, J., Lara-Gutiérrez, J. & Stocker, R. Environmental fluctuations and their effects on microbial communities, populations and individuals. *FEMS Microbiol. Rev.* **45**, fuaa068 (2021).
52. Zaoli, S. & Grilli, J. A macroecological description of alternative stable states reproduces intra- and inter-host variability of gut microbiome. *Sci. Adv.* **7**, eabj2882 (2021).
53. Tsoularis, A. & Wallace, J. Analysis of logistic growth models. *Math. Biosci.* **179**, 21–55 (2002).
54. O'Sullivan, O. et al. Exercise and the microbiota. *Gut Microbes* **6**, 131–136 (2015).
55. 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).
56. 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).
57. Szafrńska, 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).
58. Gao, Y. & Li, H. Quantifying and comparing bacterial growth dynamics in multiple metagenomic samples. *Nat. Methods* **15**, 1041–1044 (2018).
59. Bremer, H. & Churchward, G. An examination of the Cooper-Helmstetter theory of DNA replication in bacteria and its underlying assumptions. *J. Theor. Biol.* **69**, 645–654 (1977).
60. Westoby, M. et al. Cell size, genome size, and maximum growth rate are near-independent dimensions of ecological variation across bacteria and archaea. *Ecol. Evol.* **11**, 3956–3976 (2021).
61. Klappenbach, J. A., Dunbar, J. M. & Schmidt, T. M. rRNA operon copy number reflects ecological strategies of bacteria. *Appl. Environ. Microbiol.* **66**, 1328–1333 (2000).
62. Lin, Q., De Vrieze, J., Fang, X., Li, L. & Li, X. Microbial life strategy with high rRNA operon copy number facilitates the energy and nutrient flux in anaerobic digestion. *Water Res.* **226**, 119307 (2022).
63. Li, J. et al. Predictive genomic traits for bacterial growth in culture versus actual growth in soil. *ISME J.* **13**, 2162–2172 (2019).
64. Thompson, L. R. et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).
65. 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) <https://doi.org/10.1101/2023.03.04.531100>.
66. Braga, R. M., Dourado, M. N. & Araújo, W. L. Microbial interactions: ecology in a molecular perspective. *Braz. J. Microbiol.* **47**, 86–98 (2016).
67. Zelezniak, A. et al. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc. Natl Acad. Sci. USA* **112**, 6449–6454 (2015).
68. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
69. 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).
70. Dikicioglu, D., Kırdar, B. & Oliver, S. G. Biomass composition: the 'elephant in the room' of metabolic modelling. *Metabolomics* **11**, 1690–1701 (2015).
71. 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).
72. Diener, C., Gibbons, S. M. & Resendis-Antonio, O. MICOM: Metagenome-Scale Modeling To Infer Metabolic Interactions in the Gut Microbiota. *mSystems* **5**, e00606–e00619 (2020).
73. Gambhir, S. S., Ge, T. J., Vermesh, O. & Spitler, R. Toward achieving precision health. *Sci. Transl. Med.* (2018).
74. Gambhir, S. S., Ge, T. J., Vermesh, O., Spitler, R. & Gold, G. E. Continuous health monitoring: An opportunity for precision health. *Sci. Transl. Med.* **13**, eabe5383 (2021).
75. Wang, X. J. & Camilleri, M. A smart toilet for personalized health monitoring. *Nat. Rev. Gastroenterol. Hepatol.* **17**, 453–454 (2020).
76. Ge, T. J. et al. Passive monitoring by smart toilets for precision health. *Sci. Transl. Med.* **15**, eabk3489 (2023).
77. Park, S.-M. et al. A mountable toilet system for personalized health monitoring via the analysis of excreta. *Nat. Biomed. Eng.* **4**, 624–635 (2020).
78. Trapletti, A. & Hornik, K. tseries: Time Series Analysis and Computational Finance. R Package v. 0.1-25 8 (2020) <https://CRAN.R-project.org/package=tseries>.
79. Team, R. C. R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna Austria URL (2022).
80. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

81. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
82. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* **3**, e104 (2017).
83. 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).
84. Langmead, B., Wilks, C., Antonescu, V. & Charles, R. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**, 421–432 (2019).
85. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer, 2016).
86. Lin, H. & Peddada, S. D. Analysis of microbial compositions: a review of normalization and differential abundance analysis. *NPJ Biofilms Microbiomes* **6**, 60 (2020).
87. Fisher, S. R. A. *Statistical Methods for Research Workers*. (Oliver and Boyd, 1925).
88. Iacus, S. M. SDE: simulation and inference for stochastic differential equations. (2007).
89. Lim, J. J. et al. *Growth phase estimation for abundant bacterial populations sampled longitudinally from human stool metagenomes*. (Github, 2023).

## 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 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) and Environmental Health and Microbiome Research Center (EHMBRACE). 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). Work conducted by JJV, JW, and NSB, as part of ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies (<http://enigma.lbl.gov>), a Science Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231 (to NSB).

## Author contributions

S.M.G., J.J.V. and J.L. designed and conceived the study. J.L. performed the analyzes, ran simulations, made the figures, and wrote the first draft of the manuscript. J.W. performed the in vitro experiment. CD processed metagenomic and genomic data and contributed to data analysis and figure construction. S.M.G. and N.S.B. contributed resources and mentorship. All authors contributed to writing and editing the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-41424-1>.

**Correspondence** and requests for materials should be addressed to Sean M. Gibbons.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023