System Biology Modeling with Compositional Microbiome Data Reveals Personalized Gut Microbial Dynamics and Keystone Species

Chenhao Li[1,2], Lisa Tucker-Kellogg[3], Niranjan Nagarajan[1,2,*]

[1]*Computational and System Biology, Genome Institute of Singapore, Singapore 138672, Singapore*

[2]*School of Computing, National University of Singapore, Singapore, 117543, Singapore*

[3]*Centre for Computational Biology, Duke–NUS Graduate Medical School, 169857, Singapore*

*Corresponding Author: Niranjan Nagarajan (nagarajann@gis.a-star.edu.sg)

Abstract

A growing body of literature points to the important roles that different microbial communities play in diverse natural environments and the human body. The dynamics of these communities is driven by a range of microbial interactions from symbiosis to predator-prey relationships, the majority of which are poorly understood, making it hard to predict the response of the community to different perturbations. With the increasing availability of high-throughput sequencing based community composition data, it is now conceivable to directly learn models that explicitly define microbial interactions and explain community dynamics. The applicability of these approaches is however affected by several experimental limitations, particularly the compositional nature of sequencing data. We present a new computational approach (BEEM) that addresses this key limitation in the inference of generalised Lotka-Volterra models (gLVMs) by coupling **b**iomass **e**stimation and model inference in an **e**xpectation **m**aximization like algorithm (BEEM). Surprisingly, BEEM outperforms state-of-the-art methods for inferring gLVMs, while simultaneously eliminating the need for additional experimental biomass data as input. BEEM's application to previously inaccessible public datasets (due to the lack of biomass data) allowed us for the first time to analyse microbial communities in the human gut on a per individual basis, revealing personalised dynamics and keystone species.

1. Introduction

A growing body of literature points to the important roles that different microbial communities play in diverse natural environments[1,2] and the human body[3]. This has particularly been aided by advances in next-generation sequencing technology, allowing for rapid, cost-effective taxonomic and functional profiling, combined with computational analysis that has helped associate the state of the microbiome with various environmental conditions[1,4] and human diseases[5–8]. Microbiomes are also constantly evolving and there is now a growing appreciation of the fact that complex interactions between community members[9,10] shape community dynamics[11,12] as well as overall function[13,14]. A systems view of the microbiome is thus essential for understanding and rationally manipulating it[15].

Because of its importance, there have been many approaches proposed to study microbial interactions and dynamics. Experimental approaches have ranged from simple two species co-culture experiments[16–18], all the way to complex, multi-stage reactor models[19]. Analytical approaches[20] frequently use simple correlations between the abundances of various taxa in cross-sectional datasets to infer microbial interactions[21–23]. There are several challenges that need to be addressed in such analysis including the compositionality of sequencing data[21–24], low sensitivity and specificity of such methods[25], and the inability to infer directionality of interactions or dynamics of the system[20].

The most commonly used approach for modeling microbial ecology is based on classical predator-prey systems, also referred to as generalized Lotka-Volterra models (gLVMs). gLVMs are based on ordinary differential equations (ODE) that model the logistic growth of species, naturally capture predator-prey, amensalistic and competitive interactions, and have been applied to study dynamics of microbial ecosystems ranging from simple communities on cheese[26,27] to the human microbiome[15,28–32]. More importantly, from a practical perspective, gLVMs have been used for a range of applications including identifying potential probiotics against pathogens[15,28,30], forecasting changes in microbial density, characterizing important community members (e.g. keystone species[29]) and to analyze community stability[30,32].

Despite this, a key limitation of gLVMs that restricts applicability and wider use is the requirement for microbial abundance data on an absolute scale. Microbiome analysis using high-throughput sequencing naturally provides relative abundance estimates with what is

often referred to as "compositionality bias"[21,22,24], and cannot be directly used to estimate gLVM parameters[31]. Scaling relative abundances to an absolute scale typically requires additional experimental data that is either not readily available (as is true for the vast proportion of publicly available datasets), is technically challenging to directly quantitate for different sample matrices and complex communities (e.g. using flow cytometry[33,34]), or can suffer from significant technical[35–37] and biological noise[38] (e.g. using 16S rRNA qPCR[15,28,30]).

We show that, surprisingly, scaling factors can be directly inferred from microbiome sequencing data, through an algorithm that also simultaneously estimates gLVM parameters (BEEM). This is achieved based on an expectation-maximization-like approach[39] that alternates between learning scaling factors and gLVM parameters, and thus obviates the need for experimental scaling factors which otherwise limit the use of many existing datasets. Based on synthetic data where biomass is precisely known, we show that BEEM estimated gLVM parameters are as accurate as those estimated with true biomass values, and significantly more accurate than what could be expected with commonly used (16S rRNA based) experimentally determined biomass estimates. Using data from a freshwater microbial community with flow cytometry based gold-standard cell counts, we show that biomass estimated using BEEM has good concordance with the gold-standard and improves significantly over existing techniques to normalize data. Leveraging BEEM's unique ability to learn gLVMs from relative abundance data, we analyzed publicly available datasets that represent the longest human gut microbiome time-series data available to-date[40–42]. This analysis revealed, for the first time, the personalized dynamics of gut microbial biomass in different individuals, with communities driven by distinct interaction networks and hub species. Our analysis suggests an emerging model for gut microbial dynamics where relatively low abundance species may play key roles in maintaining gut homeostasis.

2. Results

2.1 Experimentally obtained biomass estimates can lead to inaccurate gLVMs

The gLV equations model the growth rate ($\frac{dx_i(t)}{dt}$) of each microbial species i as a function of absolute densities ($x_i(t)$) of all the p species in a community:

$$\frac{dx_i(t)}{dt} = \mu_i x_i(t) + \sum_{j=1}^{p} \beta_{ij} x_i(t) x_j(t) \quad (1).$$

In the above model, the intrinsic growth rate parameter (μ_i) and self-interaction parameters (β_{ii}) define the logistic growth behavior of species i. In addition, the model also captures the impact of the absolute density of species j on the growth rate of species i through additional parameters (β_{ij}, $i \neq j$), assuming a linear and additive effects model. As high-throughput sequencing based approaches to analyze microbiomes only provide relative abundance estimates, scaling factors related to the total biomass for each sample are then needed to fit gLVMs in practice.

The predominantly used approach to estimate total biomass is to quantify copy number of the 16S rRNA gene using quantitative PCR (qPCR)[15,28,30]. However, 16S qPCR estimates have been reported to have high technical noise, with a coefficient of variation (CV) ranging from 11% to 75%[35–37]. To reconfirm this, we reanalyzed 16S qPCR data from a recent microbiome modeling study on *C. difficile* infections[30] and observed low concordance across technical replicates (Spearman ρ<0.22; **Figure 1A** and **Supplementary Figure 1A**), as well as high coefficient of variation (mean CV=51%). Another critical source of error with 16S qPCR based biomass estimates is biological, and arises due to the fact that bacteria can have widely varying number of copies of the 16S rRNA gene, even within the same ecological niche. For example, the 16S gene copy number of the four major gut bacterial phyla cover a broad spectrum (**Figure 1B**), ranging from a single copy to 15 copies[38]. Correspondingly, 16S qPCR estimated biomass of a community dominated by *Firmicutes* can be twice as much as that of a community dominated by *Bacteriodetes*, even if both communities have exactly the same cell density. Such large relative errors (~100%) can then have a significant impact on the accuracy of gLVMs estimated from the data, as we show below.

To test the impact of biomass estimation errors on model inference, we generated synthetic datasets (10 species community) based on parameters inferred from real datasets, similar to the approach in Bucci *et al*[29] (see **Materials and Methods**). This framework allows us to carefully evaluate the impact of different levels of noise in a setting where model parameters are known. We noted that, given error-free biomass data, a state-of-the-art method (MDSINE[29]) was able to infer model parameters with median relative error <20% and with ~90% median AUC-ROC (area under the sensitive-specificity tradeoff curve) for interaction terms ($\boldsymbol{\beta}$; **Figure 1C**, True). However, as expected[31], directly using relative abundance

estimates without scaling them increased median relative error for parameter estimates to >60% (**Figure 1C**, RA), with AUC-ROC for interaction terms being comparable to randomly generated parameters from the prior model for the simulation (**Figure 1C**, Random). Using biomass estimates with error profile similar to real qPCR data (CV=51%; without systematic errors due to varying copy number of the 16S rRNA gene), surprisingly, did not improve performance substantially when one technical replicate was provided (**Figure 1C**, qPCR_rep1), and even with three technical replicates, growth rate parameter estimates (median relative error >70%) were comparable to random (**Figure 1C**, qPCR_rep3). These results highlight that experimental errors in biomass estimates can significantly impact gLVM parameter estimation even in a relatively well-controlled setting where model assumptions are strictly applied.

2.2 Joint estimation of biomass and model parameters with BEEM

In order to address the challenges of noisy experimental biomass data and, in general, to make gLVM modeling more widely applicable where biomass estimates are not available, we explored the idea of learning gLVM parameters directly from relative abundance data. To achieve this, we first note that model equation 1 can be expressed in terms of relative growth rates by dividing both sides of the equation by $x_i(t)$:

$$\frac{dx_i(t)}{dt}/x_i(t) = \frac{d\ln x_i(t)}{dt} = \mu_i + \sum_{j=1}^{p} \beta_{ij} x_j(t).$$

By explicitly introducing relative abundances ($\tilde{x}_i(t)$) and total biomass ($m(t)$, where $x_i(t) = m(t)\tilde{x}_i(t)$), we get:

$$\frac{d(\ln m(t) + \ln \tilde{x}_i(t))}{dt} = \mu_i + m(t) \sum_{j=1}^{p} \beta_{ij} \tilde{x}_j(t).$$

The biomass terms on the left-hand-side of the equation can be eliminated by subtracting the equation of a selected species r from the equations for all other species, resulting in a new system:

$$\frac{dy_i(t)}{dt} = a_i + m(t) \sum_{j=1}^{p} b_{ij} \tilde{x}_j(t), i \neq r \ ,$$

where $y_i(t) = \ln(\tilde{x}_i(t)/\tilde{x}_r(t))$ and the equations are re-parameterized by a_i and b_{ij}, which are related to the original parameters ($a_i = \mu_i - \mu_r$ and $b_{ij} = \beta_{ij} - \beta_{rj}$). This new system has the advantage that all unknowns are on the right-hand-side of the equation and the

gradient term on the left-hand-side can be estimated directly from relative abundance data through spline smoothing and numerical differentiation[15,28–30].

We then made the observation that the above equations can be re-written as two regression problems across two dimensions of the data matrix ($\tilde{x}_i(t), \forall i, t$):

(1) For each species i, the corresponding parameters a_i and b_{ij} can be solved through gradient matching[15,28–30], given the biomass at each time point t ($m(t)$).
(2) For each time point t, the biomass can be solved for via regression given the model parameters $\boldsymbol{a}$ and $\boldsymbol{b}$ for all the species.

The interlock of the above two problems provides the basis for an expectation-maximization-like algorithm that alternates between estimating model parameters and biomass iteratively and forms the core of BEEM (see **Materials and Methods** for details). Note that estimates provided by BEEM for the biomass act as scaling factors to bring abundances across species and time points to the same scale for learning gLVMs.

On the synthetic datasets used in section 1, we noted that despite not having any biomass data to work with, BEEM was a significant improvement over naïve analysis based on relative abundance data, as well as results based on scaled relative abundances with noisy biomass data (**Figure 1C**, BEEM). In fact, BEEM estimated parameters were nearly as accurate as those obtained using noise-free biomass data, except for a slight decrease in AUC-ROC for interaction terms (primarily due to rounding errors that provide non-zero estimates for zero terms). Additionally, other computational approaches to "normalize" the data could not match BEEM's performance, though they did improve on the experimental approaches described earlier (**Figure 1C**, CSS; see **Materials and Methods**). We noted that BEEM's significant improvement over other experimental and computational approaches, and its ability to closely approximate analysis using true biomass estimates is a robust feature that remains valid even when experimental biomass estimates are significantly better (CV=5%, as expected from flow-cytometry data) and while using different parameter estimation approaches (**Supplementary Figure 1B**).

2.3 BEEM accurately estimates gLVM parameters and biomass in diverse model settings

As in any situation where parameters have to be estimated, a sufficient number of data points (multiple biological replicates) are needed to get accurate gLVM models and this in turn impacts BEEM's biomass estimates. In order to further study BEEM's performance characteristics, we generated synthetic datasets with varying number of species and data points, comparing BEEM's results to those obtained with noise-free biomass data and the same gradient matching algorithm (BLASSO, see **Materials and Methods**) as used internally in BEEM. As expected, when the number of species increases but the number of data points remains constant (60 replicates with 30 timepoints), gLVM parameter estimation becomes harder (**Figure 2A**). However, despite the quadratic increase in the number of parameters, performance for both BLASSO (with true biomass) and BEEM seems to only degrade linearly (**Figure 2A**). In addition, even when the model has 25 species (650 model parameters) and can thus capture over 90% of the human gut microbiome[43] (**Supplementary Figure 2**), interaction parameters estimated by BEEM were nearly as accurate as those with true biomass (**Figure 2A**), though growth rate parameters were more affected. We also noted that median relative error for biomass estimates from BEEM was generally well controlled (<10%; **Figure 2B**).

Increasing the number of data points available for model fitting for a fixed number of species (10) improved performance for both BLASSO with true biomass and BEEM, as expected. Performance improvements were most notable when going from 10 to 20 replicates and plateaued out after that (30 timepoints; **Figure 2C**). In general, after 20 replicates, differences between BLASSO and BEEM were small, especially in terms of estimating interaction parameters. Similarly, biomass estimates from BEEM had median relative error <5% when 20 replicates were available (**Figure 2D**). In general, our analysis suggests that inherent limitations in gradient matching based on estimated gradients from data were a greater source of error for gLVM parameter estimation in many of our experiments than errors in BEEM estimated biomass values.

To assess BEEM's performance for biomass inference in real-world datasets we analyzed data from a recently published study on freshwater microbial communities[33,34], which to our knowledge is the only one to have sufficient longitudinal microbiome sequencing data as well as flow-cytometry based gold-standard biomass estimation. Notably, the flow cytometry data

in this study was reported to have high reproducibility (CV<5%)[33], and therefore was suitable for use as the ground truth for total biomass. Surprisingly, with only 57 time points in total across two replicate experiments, BEEM was able to infer the total biomass for a 26-species community accurately solely based on 16S sequencing based relative abundances. Compared to CSS normalization[44], a state-of-the-art approach to scale metagenome taxonomic profiles, BEEM estimated biomass had both significantly lower relative error rates (median 29% *vs* 40% and 34% *vs* 41%, respectively, for the two replicates; Wilcoxon test *p*-value=0.03; **Figure 3A**), as well as much better concordance with ground truth (Spearman's ρ=0.72 *vs* 0.36; **Figure 3B**).

2.4 Personalized gut microbial dynamics and keystone species

The development of BEEM allows us to analyze previously generated datasets in a gLVM framework, even when biomass measurements were not made in the original study. To showcase this capability, we applied BEEM to the longest (over one year) and most densely (almost daily) sampled human gut microbiome time-series datasets available to date (four individuals: DA, DB from David *et al*[41] and M3, F4 from Caporaso *et al*[40]). BEEM estimated models exhibited a good fit to the data, with predicted relative abundances for a day based on numerical integration from the previous day being in high concordance with observed data (median Spearman's ρ = 0.83). As BEEM directly infers daily biomass values, we plotted these and observed distinct individual-specific patterns: while subject DA's biomass was found to vary relatively smoothly, following an approximately cyclic pattern with a period of about three months (**Figure 4A**), subject M3's biomass fluctuated to a greater extent on a day to day basis with no clear trend (**Figure 4B**). Similar patterns were observed in parts for subjects DB and F4, which had a greater resemblance to DA overall (**Supplementary Figure 3A, B**). The fluctuations predicted in M3's biomass were also found to be accompanied by frequent blooms of rare taxa that were otherwise not detected at other time points[42] and may be a consequence of this instability in the community. In contrast, the smoother progression of DA's biomass may be a reflection of the relative stability of the gut community in this individual, though the source of the observed cyclic patterns deserves to be explored further. As an initial hint, we noted that the strongest association between DA's biomass and reported metadata was a negative correlation with calcium intake (**Supplementary Figure 4**).

Concordant with their distinct biomass dynamics, DA and M3 also exhibited microbial interaction networks that were unique to them (**Figure 4C, D**). DA's network was defined by

hub nodes for *Feacalibacterium prausnitzii* and *Bacteroides uniformis*, two species with many beneficial roles and frequent associations with a healthy gut[45,46]. The hubs were found to negatively affect the growth of *Enterobacteriaceae* species, consistent with previous reports for *B. uniformis*[47] and *F. prausnitzii*[48–50]. In comparison, the major hub nodes in M3's network were a *Blautia* and an *Oscillospira* species that were connected by a positive feed-forward loop. Additionally, we found that abundances of the *Blautia* and *Oscillospira* species were significantly negatively correlated with total biomass in M3's gut microbiome (**Supplementary Figure 5**). Feed-forward loops have been implicated in destabilizing effects on ecosystems[32] and so these observations may explain the unstable behavior of M3's biomass as well as the corresponding susceptibility to invasive blooms of rare taxa[42]. *Oscillospira*'s protective role in M3's gut flora is further indicated by its parasitic relationship (negative-positive loop) with another hub species *B. fragilis*, an opportunistic pathogen that has been associated with diarrhea[51]. Interestingly, several of the transient species in M3's gut microbiome were observed to be at the periphery of the network, with a single incoming edge indicating that their abundances were being influenced by a hub species. For example, this was observed for several *Streptococcus* species that are primarily oral commensals and could be transient colonizers of the gut[52,53].

Despite differences in the identity of species in their interaction networks, the various individual-specific networks shared some common features, including the presence of a few hub nodes that negatively influenced many other species, and were generally not the most abundant species in the community (**Figure 4C**, **D** and **Supplementary Figure 3C**, **D**). Overall, we also found that the ratio between out- and in- degree of species in the networks were negatively correlated with their median relative abundances (**Supplementary Figure 6**), suggesting that the hub species in the interaction network, that are often considered as keystone species for the community[29,54], are typically not the abundant species in the gut microbiome. We further confirmed this observation by analyzing a large collection (840 healthy individuals) of gut microbiome datasets[43], to find that the core species in the gut microbiome were also frequently not the most abundant species (**Supplementary Figure 7**). Together, these observations suggest a model for the gut microbiome where relatively less abundant species in the community are more stable colonizers of the host, and by virtue of their impact on the growth of other species in the community, play an important role in defining its dynamics in different individuals.

3. Discussion

A major limitation of most microbiome profiling datasets available to date is the restriction to relative abundances and the 'compositionality' of this data has led to significant challenges even when performing common statistical tests for correlated abundances[55]. These issues are amplified when considering systems models such as gLVMs, and our analysis here confirms that model parameter estimates can be severely distorted if relative abundances are not properly scaled. Experimental approaches to measure scaling factors are generally seen as a laborious and sometimes feasible way to address this issue, but as we show here, this may not be the case if care is not taken to ensure that experimental noise is minimized and sufficient number of replicates are analyzed. By eliminating the need for additional experimental data, BEEM greatly expands the applicability of gLVMs to microbiome datasets, and its robustness could simultaneously improve the quality of models and scaling factor estimates, as observed in our synthetic and real datasets. Explicitly modelling microbial interactions through gLVMs has proven to be a powerful framework for studying microbial community dynamics[15,26–32], and the approach used in BEEM could also be extended (with minimal modifications) to time-series with external perturbations (e.g. antibiotics usage)[15,28,30], as well as systems models for gene expression regulation based on RNA-seq data[56].

Due to limited availability of absolute abundance data, gLVM models have generally been constructed by aggregating information across experiments and individuals[15,28,30]. We exploited the availability of year-long time series datasets and BEEM's facility with relative abundances to construct individual specific gut microbiome models for the first time. Intriguingly, we observed that our inferred scaling factors suggest that gut microbial biomass has distinct dynamics across different individuals. Consistent with a recent study on 20 individuals where human gut microbial biomass (measured via flow-cytometry) was found to have high variation (CV≈ 53% within a week)[55], we also noted high variability over time across the four individuals we analyzed (CV ranging from 49% to 76% over a year). Additionally, we observed cyclic behavior of biomass trajectories in multiple individuals, similar to the seasonal patterns reported in hunter-gatherers of western Tanzania[57], and the evolutionarily conserved patterns observed in other mammals[58]. Similar patterns have not been reported before for western city dwellers, perhaps due to the confounding effects of aggregate analysis across individuals and the impact of highly diverse diets. BEEM analysis,

however, suggests that the underlying patterns may still be conserved in urban subjects (without overt seasonal shifts in diet), and may be more general than previously believed.

Our inference of gLVM models for each individual allows us to identify specific microbial species and the kinds of interactions that they have, to account for the distinct dynamics that were observed. For example, the positive feed forward loop observed between the hubs in M3's gut microbiome provides a specific plausible and testable hypothesis to explain the instability observed there. While we were unable to explore this further in this study, this capability should be very useful in future studies and in the design of targeted perturbations. Despite differences in the microbial interaction networks observed for different individuals, a shared feature seems to be the presence of relatively lowly abundant species that act as hub nodes in the network. A similar pattern was seen in cross-sectional data as well where frequently shared "core" gut microbiome species tend to not be the most abundant species in the community. These observations point to a model where species at low relative abundances stably colonize the gut (perhaps because they are mucosa-associated) compared to abundant but transient (lumen-associated) bacteria, and play an important role in defining gut microbiome dynamics. In particular, hub species were frequently found to negatively regulate more transient species in the community, in agreement with the known role of mucosa-associated species in providing colonization resistance against foreign, potentially pathogenic species[59].

An important point that we noted in the gut microbiome datasets that were analyzed here is the limited number of core species (prevalent in most time points for an individual) that are shared across individuals. This feature makes it infeasible to learn gLVM models by merging short time-series datasets across different individuals. Similar constraints might be present in other microbial communities as well, including specific challenges in measuring total biomass in complex matrices[55], and thus the development of BEEM makes it more feasible to generate the long and densely sampled datasets that are needed for such models. We also note that the analysis in BEEM can be directly extended to cross-sectional datasets if the corresponding communities are believed to be at equilibrium (i.e. $\frac{dx_i(t)}{dt} = 0$, for all species). This extension would significantly expand the amount of data that could be used and thus allow us to learn even more complex models in the future. As is the case for any modelling approach, no model is expected to be perfect, but as they capture more and more features of real systems, we can expect that their predictions become increasingly useful. BEEM's

development therefore serves as an important step in expanding the use of modelling approaches to study microbial community dynamics and rationally identify appropriate perturbations.

4. Materials and Methods

4.1 BEEM's core algorithm

As introduced in **Section 2.2**, the gLVM model in equation 1 can be first simplified by dividing $x_i(t)$ on each side, and then re-written in terms of total biomass $m(t)$ (i.e. $m(t) = \sum_{i=1}^{p} x_i(t)$) and relative abundances $\tilde{x}_i(t)$ (i.e. $\tilde{x}_i(t) = x_i(t)/m(t)$) as shown below:

$$\frac{d \ln m(t) + \ln \tilde{x}_i(t)}{dt} = \mu_i + m(t) \sum_{j=1}^{p} \beta_{ij} \tilde{x}_j(t) \ . \tag{2}$$

To eliminate the biomass related term in the left-hand-side of the equation, we subtract the corresponding equation for a reference species r (species with lowest CV, by default) from both sides of the system, resulting in additive log ratio (ALR) transformed[60] relative abundances ($y_i(t) = \ln(\tilde{x}_i(t)/\tilde{x}_r(t))$) on the left-hand-side and a re-parameterized right-hand-side:

$$\frac{d y_i(t)}{dt} = a_i + m(t) \sum_{j=1}^{p} b_{ij} \tilde{x}_j(t) \, , i \neq r \ \ ,$$

where $a_i = \mu_i - \mu_r$ and $b_{ij} = \beta_{ij} - \beta_{rj}$.

An estimate for $d y_i(t)/dt$, denoted as Y_{it}, can be calculated as the derivative of a piece-wise polynomial spline fitted to the ALR transformed relative abundances ($y_i(t)$, see **Section 4.2** for details). BEEM then estimates the model parameters $\boldsymbol{a}, \boldsymbol{b}$ and the biomass $\boldsymbol{m}$ using an EM-like algorithm with the following sum of squared error objective function:

$$\Theta(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{m}) = \sum_{i,t} \left(Y_{it} - \left(a_i + m_t \sum_{j=1}^{p} b_{ij} \tilde{X}_{jt} \right) \right)^2 ,$$

where $\tilde{X}_{it} = \tilde{x}_i(t)$ and $m_t = m(t)$ are the variables written in their matrix representations.

The EM-like algorithm in BEEM works by iterating two steps, an E-step and an M-step, to convergence as detailed below:

Model parameter estimation with Bayesian lasso (E-step): In iteration T, with estimated biomass from the previous iteration $\hat{m}_t^{(T-1)}$, BEEM estimates $\hat{a}_i^{(T)}$ and $\hat{b}_{ij}^{(T)}$ for each i ($i \neq r$) based on the following regression problem (also known as gradient matching):

$$Y_{it} \sim a_i^{(T)} + \widehat{m}_t^{(T-1)} \sum_{j=1}^{p} b_{ij}^{(T)} \tilde{X}_{jt}.$$

Solving the above system is often limited by the amount of data available in practice. For microbial communities, it is usually assumed that the interaction vector (β_{ij}) is sparse (i.e. a species is only directly affected by a small number of other species). Consequently, the transformed matrix b_{ij} is also sparse and BEEM estimates it using a sparse regression technique based on a Bayesian approach (Bayesian lasso - BLASSO[30]; R package "monomvn" version 1.9-7; default parameters)[61].

Biomass estimation with linear regression (M-step): With $\hat{a}_i^{(T)}$ and $\hat{b}_{ij}^{(T)}$ from the E-step, the biomass $\widehat{m}_t^{(T)}$ for each T can be computed as the coefficient of the following linear regression:

$$U_{ti}^{(T)} \sim m_t^{(T)} V_{ti}^{(T)}, i \neq r \ ,$$

where $U_{ti}^{(T)} = Y_{it} - \hat{a}_i^{(T)}$ and $V_{ti}^{(T)} = \sum_{j=1}^{p} \hat{b}_{ij}^{(T)} \tilde{X}_{jt}$.

Initialization: For the initialization step in its EM-like algorithm, BEEM assumes that scaling factors inferred from a commonly used normalization approach for metagenomic data (Cumulative Sum Scaling - CSS[44]) provides a reasonable starting point for the algorithm to then learn better scaling factors. Note that, as expected, scaling factors from CSS normalization and BEEM cannot recapitulate the absolute scale corresponding to experimental measurements (e.g. by qPCR or flow cytometry), and so their estimates were scaled to the same median value across the time series as experimental measurements for subsequent comparisons.

Termination and parameter estimation: The E- and M-step in BEEM are run until convergence or a user specified maximal number of iterations. The search was assumed to have reached convergence (to a local optimum) when the mean squared error (MSE) for the M-step starts to increase by more than 10% compared to the minimal MSE observed. In practice, on the datasets analyzed in this study, convergence takes a few hours using 4 CPUs. Estimates for $\hat{a}_i$, $\hat{b}_{ij}$ and $\widehat{m}_t$ were calculated as the median of the values from all iterations whose MSE was within 10% of the minimal MSE.

4.2 Robust parameter estimation with BEEM

In our experiments with synthetic and real data, we noted that gLVM modelling can be sensitive to noise and outliers in the data, and this in turn could affect estimation of scaling factors with BEEM. To address this, we refined the core algorithm in BEEM with additional pre-processing steps that further enable robust parameter estimation.

Outliers in relative abundance data: We observed in our numerical analysis that outliers in the abundance data could notably affect the spline fitting procedure and lead to spurious gradient estimates. To obtain more robust spline fitting, an over-smoothed spline was first fitted to $y_i(t)$ (function "smooth.Pspline" from R package "pspline"[62] with maximal degree of five and a large smoothing parameter "spar=1e10") to calculate the absolute error in fitted values ($e_{it} = |y_i(t) - y_i(t)^{\text{smoothed}}|$), and points with absolute error larger than expected ($(e_{it} - \underset{j}{\text{median}}(e_{ij}))/\underset{j}{\text{MAD}}(e_{ij}) > \tau$, $\tau = 5$ by default) were then filtered out. The final smoothing spline was fitted (degree of five and smoothing parameter selected using cross validation) to the remaining data to calculate the estimated gradients Y_{it}. In addition, outliers in biomass estimated from the previous iteration ($\hat{m}_t^{(T-1)}$) were identified in the same way and replaced with interpolated values from the spline.

Outliers in estimated gradients: In practice, gradient matching based methods (including the various algorithms implemented in MDSINE) were found to be sensitive to outliers in the estimated gradients (i.e. Y_{it}). To identify outliers in a time series (Y_{it}, for all t) a local regression (LOESS) smoother was fitted to de-trend Y_{it}, and the outliers were filtered out as described above.

Estimating constrained biomass values: For each time point, biomass was estimated as the slope of a linear regression ($U_{tk}^{(T)}$ against $V_{tk}^{(T)}$) where outliers in both $U_{tk}^{(T)}$ and $V_{tk}^{(T)}$ were identified and removed following a standard boxplot approach i.e. as deviations from the median by more than 1.5× inter-quartile range. In addition, the biomass was constrained to be positive by removing points where $U_{ti}^{(T)}$ and $V_{ti}^{(T)}$ had different signs.

4.3 Recovering gLVM parameters

Based on the previously stated assumption that the interaction matrix $\boldsymbol{\beta}$ is sparse, most entries in each column are expected to be zero and thus the median value for the jth column in

$\boldsymbol{b}$ would be expected to be $-\beta_{rj}$, allowing us to infer back all the other rows of $\boldsymbol{\beta}$ ($\beta_{ij} = b_{ij} + \beta_{rj}$). BEEM then assigns a Z-score like confidence value (s_{ij}) to each entry of $\boldsymbol{\beta}$, by dividing the estimated interaction strength by the column standard deviation ($s_{ij} = |\hat{\beta}_{ij}/\sigma_j|$). The growth rate vector $\boldsymbol{\mu}$ is not expected to be sparse but can be recovered by directly solving the original gLVM system (equation 2), using the already derived estimates for scaling factors and $\boldsymbol{\beta}$. For robustness, BEEM estimates the growth rate for each species as the median of positive estimates across all time points.

4.4 Datasets and evaluation metrics

Simulated datasets: MDSINE's Bayesian variable selection (BVS) algorithm (with spline smoothing option and minor bug fixes: https://bitbucket.org/chenhao_li/mdsine) was used to estimate parameters from the *C. difficile* infection dataset provided with the package[30]. Simulated datasets were then generated based on these estimated parameters following the procedure described in Bucci *et al*[30] (excluding perturbations). Noisy abundances were obtained by sampling from Poisson distributions[63] with means based on scaled abundances at each time point (sum = 5×10^4). Simulated qPCR and flow cytometry based values for total biomass were generated from log normal distributions with coefficients of variation (CV) that matched those seen in real datasets (qPCR=51%[30], flow cytometry=5%[33,34]).

Dataset from Props et al: The original OTU table was obtained from the authors[34]. Samples for the "operation" stage, where the environment had roughly constant temperature were selected for BEEM analysis. OTUs with low mean relative abundances (<0.1%) were excluded, resulting in 26 OTUs across 58 time points from two replicates.

Dataset from Gibbons et al: This dataset included four long time series collected by David *et al*[40] and Caporaso *et al*[41]. The original OTU tables[42] were filtered to keep only top OTUs based on prevalence (>10 reads in most of the samples). In total, 26 and 22 OTUs were left for samples from David *et al* and Caporaso *et al*, respectively. In order to assess the robustness of the inferred network, BEEM was run with 30 different seeds and edges with confidence score $s_{ij} \leq 1$ in more than 50% of the networks were kept. The final biomass was obtained by taking the geometric mean across all 30 runs (**Supplementary File 1**).

Metrics for evaluation: The following metrics were used for evaluating inference algorithms:

1. Median relative error (MRE) for estimates $\widehat{\boldsymbol{\theta}}$ when the true values are $\boldsymbol{\theta}$: $\underset{\theta_i \neq 0}{\text{median}} \left| \frac{\widehat{\theta}_i - \theta_i}{\theta_i} \right|$.
2. Area under receiver operating characteristic curve (AUC-ROC) for the inferred microbial interactions: Absolute values of parameters was used to rank predicted edges for BLASSO, while confidence scores were used for Bayesian Variable Selection (BVS) in MDSINE and for BEEM.

4.5 Software and reproducibility of results

BEEM is available under the MIT license and can be downloaded from https://github.com/csb5/BEEM. Scripts to reproduce the results presented in this work are also available at this website.

Figure Legends

Figure 1. Noise in experimentally determined biomass severely distorts gLVM parameter estimation. (A) Scatter plot with fitted linear regression line for two 16S qPCR technical replicates from Bucci *et al*. (B) Copy number variation for 16S rRNA genes in members of four major phyla of human gut bacteria. (C) Relative impact of different experimental (qPCR_rep1 – 1 qPCR replicate, qPCR_rep3 – mean of 3 qPCR replicates) and computational (RA – relative abundance, CSS – CSS normalization) data scaling approaches on gLVM parameter estimation (BVS algorithm for MDSINE), in comparison to using true biomass or using BEEM. Boxplots represent the summary of 15 simulations (10 species, 30 replicates with 30 time points each) and three different metrics are shown here including median relative error for growth rate ($\boldsymbol{\mu}$) and interaction ($\boldsymbol{\beta}$) parameters, and AUC-ROC for the interaction network. Dashed horizontal lines represent the performance of randomly generated parameters from the simulation model.

Figure 2. Robustness of parameter estimation with BEEM. (A) Results with increasing number of species but fixed number of replicates (50). As expected, parameter estimation gets harder but BEEM's performance tracks the ideal case using BLASSO with true biomass values, especially for interaction parameters. (B) Median relative error in biomass estimates remains less than 10%. (C) Results with increasing number of replicates and fixed number of species (15). BEEM's performance converges to that of BLASSO with true biomass as the number of replicates increases. (D) Median relative error in biomass estimates reduces noticeably as the number of replicates increases.

Figure 3. Concordance of BEEM estimated biomass with gold standard experimental measurements. (A) Trajectories and relative error for BEEM estimated biomass (blue) and CSS normalization (orange), compared to gold standard measurements using flow cytometry (black). (B) Scatter plots with fitted linear regression line highlighting that BEEM's biomass estimates are notably more concordant with flow cytometry based values compared to CSS normalization based estimates.

Figure 4. BEEM analysis of year long gut microbial time-series datasets. (A, B) BEEM estimated biomass values for two individuals (DA and M3) with daily sampled, year long gut microbial time-series datasets from David *et al*[40] and Caporaso *et al*[41]. Interestingly, while M3's biomass fluctuates rapidly, DA's biomass seems to vary in a more defined fashion with a periodicity of around 3 months. (C, D) Graphs representing non-zero interaction terms in gLVM models learnt individually for DA and M3 using BEEM. Dashed and solid edges represent positive and negative interactions respectively. Edge widths are proportional to the interaction strength and node sizes are proportional to the log-transformed mean relative abundance of the corresponding species. Nodes are labeled with GreenGenes IDs and colored according to order level taxonomic annotations.

References

1. Sunagawa, S. *et al.* Ocean plankton. Structure and function of the global ocean microbiome. *Science* **348,** 1261359 (2015).

2. Gilbert, J. A., Jansson, J. K. & Knight, R. The Earth Microbiome project: successes and aspirations. *BMC Biol.* **12,** 69 (2014).

3. Turnbaugh, P. J. *et al.* The human microbiome project. *Nature* **449,** 804–10 (2007).

4. Hayat, R., Ali, S., Amara, U., Khalid, R. & Ahmed, I. Soil beneficial bacteria and their role in plant growth promotion: a review. *Ann. Microbiol.* **60,** 579–598 (2010).

5. Halfvarson, J. *et al.* Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat. Microbiol.* **2,** 17004 (2017).

6. Qin, N. *et al.* Alterations of the human gut microbiome in liver cirrhosis. *Nature* **513,** 59–64 (2014).

7. Li, Q., Han, Y., Dy, A. B. C. & Hagerman, R. J. The Gut Microbiota and Autism Spectrum Disorders. *Front. Cell. Neurosci.* **11,** 120 (2017).

8. Chng, K. R. *et al.* Whole metagenome profiling reveals skin microbiome-dependent susceptibility to atopic dermatitis flare. *Nat. Microbiol.* **1,** 16106 (2016).

9. Hol, F. J. H., Rotem, O., Jurkevitch, E., Dekker, C. & Koster, D. A. Bacterial predator-prey dynamics in microscale patchy landscapes. *Proceedings. Biol. Sci.* **283,** 20152154 (2016).

10. Miller, M. B. & Bassler, B. L. Quorum Sensing in Bacteria. *Annu. Rev. Microbiol.* **55,** 165–199 (2001).

11. Martin, M., Hölscher, T., Dragoš, A., Cooper, V. S. & Kovács, Á. T. Laboratory Evolution of Microbial Interactions in Bacterial Biofilms. *J. Bacteriol.* **198,** 2564–71 (2016).

12. Embree, M., Liu, J. K., Al-Bassam, M. M. & Zengler, K. Networks of energetic and metabolic interactions define dynamics in microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 15450–5 (2015).

13. Cordero, O. X. & Datta, M. S. Microbial interactions and community assembly at microscales. *Curr. Opin. Microbiol.* **31,** 227–234 (2016).

14. Fraune, S. *et al.* Bacteria–bacteria interactions within the microbiota of the ancestral metazoan Hydra contribute to fungal resistance. *ISME J.* **9,** 1543–1556 (2015).

15. Buffie, C. G. *et al.* Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. *Nature* **517,** 205–208 (2014).

16. Freilich, S. *et al.* Competitive and cooperative metabolic interactions in bacterial communities. *Nat. Commun.* **2,** 589 (2011).

17. Lim, K. M. K., Li, C., Chng, K. R. & Nagarajan, N. @MInter: Automated Text-mining of Microbial Interactions. *Bioinformatics* btw357 (2016). doi:10.1093/bioinformatics/btw357

18. Friedman, J., Higgins, L. M. & Gore, J. Community structure follows simple assembly rules in microbial microcosms. *Nat. Ecol. Evol.* **1,** 109 (2017).

19. Blasche, S., Kim, Y., Oliveira, A. P. & Patil, K. R. Model microbial communities for ecosystems biology. *Curr. Opin. Syst. Biol.* **6,** 51–57 (2017).

20. Li, C., Kenneth, L. K. M., Chng, K. R. & Nagarajan, N. Predicting Microbial Interactions through Computational Approaches. *Methods* (2016). doi:10.1016/j.ymeth.2016.02.019

21. Faust, K. *et al.* Microbial Co-occurrence Relationships in the Human Microbiome. *PLoS Comput. Biol.* **8,** e1002606 (2012).

22. Friedman, J. & Alm, E. J. Inferring Correlation Networks from Genomic Survey Data. *PLoS Comput. Biol.* **8,** e1002687 (2012).

23. Kurtz, Z. D. *et al.* Sparse and Compositionally Robust Inference of Microbial Ecological Networks. *PLoS Comput. Biol.* **11,** e1004226 (2015).

24. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* **35,** 833–844 (2017).

25. Berry, D. & Widder, S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* **5,** 219 (2014).

26. Maoz, A., Mayr, R. & Scherer, S. Temporal stability and biodiversity of two complex antilisterial cheese-ripening microbial consortia. *Appl. Environ. Microbiol.* **69,** 4012–8 (2003).

27. Mounier, J. *et al.* Microbial interactions within a cheese microbial community. *Appl. Environ. Microbiol.* **74,** 172–81 (2008).

28. 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).

29. 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).

30. Bucci, V. *et al.* MDSINE: Microbial Dynamical Systems INference Engine for microbiome time-series analyses. *Genome Biol.* **17,** 121 (2016).

31. Cao, H.-T., Gibson, T. E., Bashan, A. & Liu, Y.-Y. Inferring human microbial dynamics from temporal metagenomics data: Pitfalls and lessons. *BioEssays* **39,** 1600188 (2017).

32. Coyte, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: Networks,

competition, and stability. *Science* **350,** 663–6 (2015).

33. Props, R., Monsieurs, P., Mysara, M., Clement, L. & Boon, N. Measuring the biodiversity of microbial communities by flow cytometry. *Methods Ecol. Evol.* **7,** 1376–1385 (2016).

34. Props, R. *et al.* Absolute quantification of microbial taxon abundances. *ISME J.* **11,** 584–587 (2017).

35. Smith, C. J., Nedwell, D. B., Dong, L. F. & Osborn, A. M. Evaluation of quantitative polymerase chain reaction-based approaches for determining gene copy and gene transcript numbers in environmental samples. *Environ. Microbiol.* **8,** 804–815 (2006).

36. White, R. A., Blainey, P. C., Fan, H. C. & Quake, S. R. Digital PCR provides sensitive and absolute calibration for high throughput sequencing. *BMC Genomics* **10,** 116 (2009).

37. Sze, M. A., Abbasi, M., Hogg, J. C. & Sin, D. D. A Comparison between Droplet Digital and Quantitative PCR in the Analysis of Bacterial 16S Load in Lung Tissue Samples from Control and COPD GOLD 2. *PLoS One* **9,** e110351 (2014).

38. Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. K. & Schmidt, T. M. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res.* **43,** D593–D598 (2015).

39. Do, C. B. & Batzoglou, S. What is the expectation maximization algorithm? *Nat. Biotechnol.* **26,** 897–899 (2008).

40. Caporaso, J. G. *et al.* Moving pictures of the human microbiome. *Genome Biol.* **12,** R50 (2011).

41. David, L. A. *et al.* Host lifestyle affects human microbiota on daily timescales. *Genome Biol.* **15,** R89 (2014).

42. 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).

43. Pasolli, E. *et al.* Accessible, curated metagenomic data through ExperimentHub. *Nat. Methods* **14,** 1023–1024 (2017).

44. Paulson, J. N., Stine, O. C., Bravo, H. C. & Pop, M. Differential abundance analysis for microbial marker-gene surveys. *Nat. Methods* **10,** 1200–1202 (2013).

45. Clemente, J. C., Ursell, L. K., Parfrey, L. W. & Knight, R. The impact of the gut microbiota on human health: An integrative view. *Cell* **148,** 1258–1270 (2012).

46. Munukka, E. *et al.* Faecalibacterium prausnitzii treatment improves hepatic health and reduces adipose tissue inflammation in high-fat fed mice. *ISME J.* **11,** 1667–1679 (2017).

47. Gauffin Cano, P., Santacruz, A., Moya, Á. & Sanz, Y. Bacteroides uniformis CECT 7771 Ameliorates Metabolic and Immunological Dysfunction in Mice with High-Fat-

Diet Induced Obesity. *PLoS One* **7,** e41079 (2012).

48. De Filippo, C. *et al.* Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 14691–6 (2010).

49. Azcarate-Peril, M. A. *et al.* Impact of short-chain galactooligosaccharides on the gut microbiome of lactose-intolerant individuals. *Proc. Natl. Acad. Sci.* **114,** E367–E375 (2017).

50. Candela, M. *et al.* Unbalance of intestinal microbiota in atopic children. *BMC Microbiol.* **12,** 95 (2012).

51. Wexler, H. M. Bacteroides: the good, the bad, and the nitty-gritty. *Clin. Microbiol. Rev.* **20,** 593–621 (2007).

52. Derrien, M. & van Hylckama Vlieg, J. E. T. Fate, activity, and impact of ingested bacteria within the human gut microbiota. *Trends Microbiol.* **23,** 354–366 (2015).

53. Zhang, C. *et al.* Ecological robustness of the gut microbiota in response to ingestion of transient food-borne microbes. *ISME J.* **10,** 2235–45 (2016).

54. Agler, M. T. *et al.* Microbial Hub Taxa Link Host and Abiotic Factors to Plant Microbiome Variation. *PLOS Biol.* **14,** e1002352 (2016).

55. Vandeputte, D. *et al.* Quantitative microbiome profiling links gut community variation to microbial load. *Nature* **551,** 507 (2017).

56. Ay, A. & Arnosti, D. N. Mathematical modeling of gene expression: a guide for the perplexed biologist. *Crit. Rev. Biochem. Mol. Biol.* **46,** 137–51 (2011).

57. Smits, S. A. *et al.* Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science* **357,** 802–806 (2017).

58. Ren, T. *et al.* Seasonal, spatial, and maternal effects on gut microbiome in wild red squirrels. *Microbiome* **5,** 163 (2017).

59. Ashida, H., Ogawa, M., Kim, M., Mimuro, H. & Sasakawa, C. Bacteria and host interactions in the gut epithelial barrier. *Nat. Chem. Biol.* **8,** 36–45 (2012).

60. Aitchison, J. The statistical analysis of compositional data. (1986).

61. Gramacy, R. B. monomvn: Estimation for Multivariate Normal and Student-t Data with Monotone Missingness. (2017).

62. Ramsey, J. & Ripley, B. pspline: Penalized Smoothing Splines. (2017).

63. Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18,** 1509–17 (2008).

Figure 1

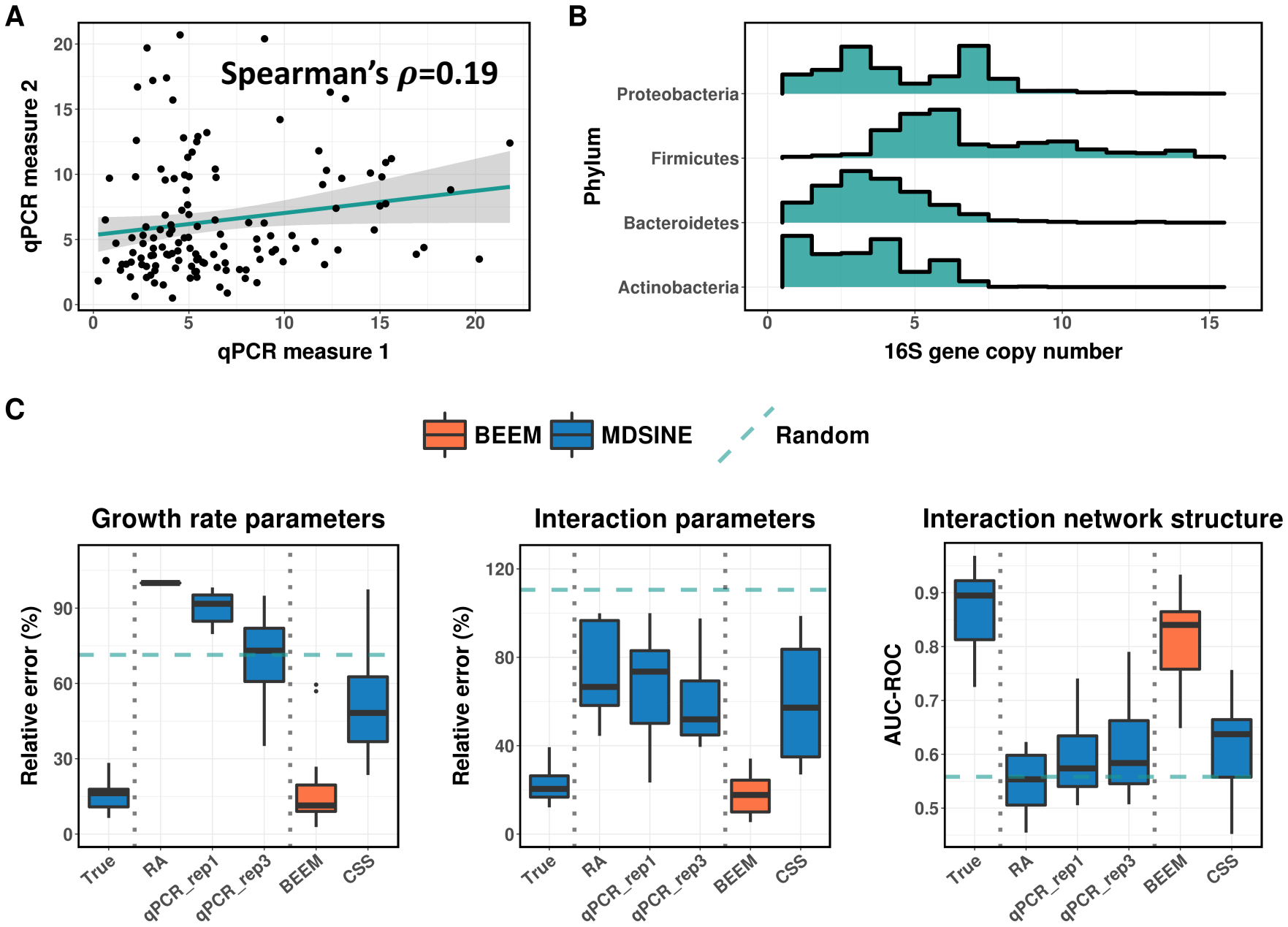

A
Spearman's ρ=0.19
qPCR measure 2
qPCR measure 1
0
5
10
15
20
B
Phylum
Proteobacteria
Firmicutes
Bacteroidetes
Actinobacteria
16S gene copy number
C
BEEM
MDSINE
Random
Growth rate parameters
Interaction parameters
Interaction network structure
Relative error (%)
AUC-ROC
True
RA
qPCR_rep1
qPCR_rep3
BEEM
CSS

Figure 2

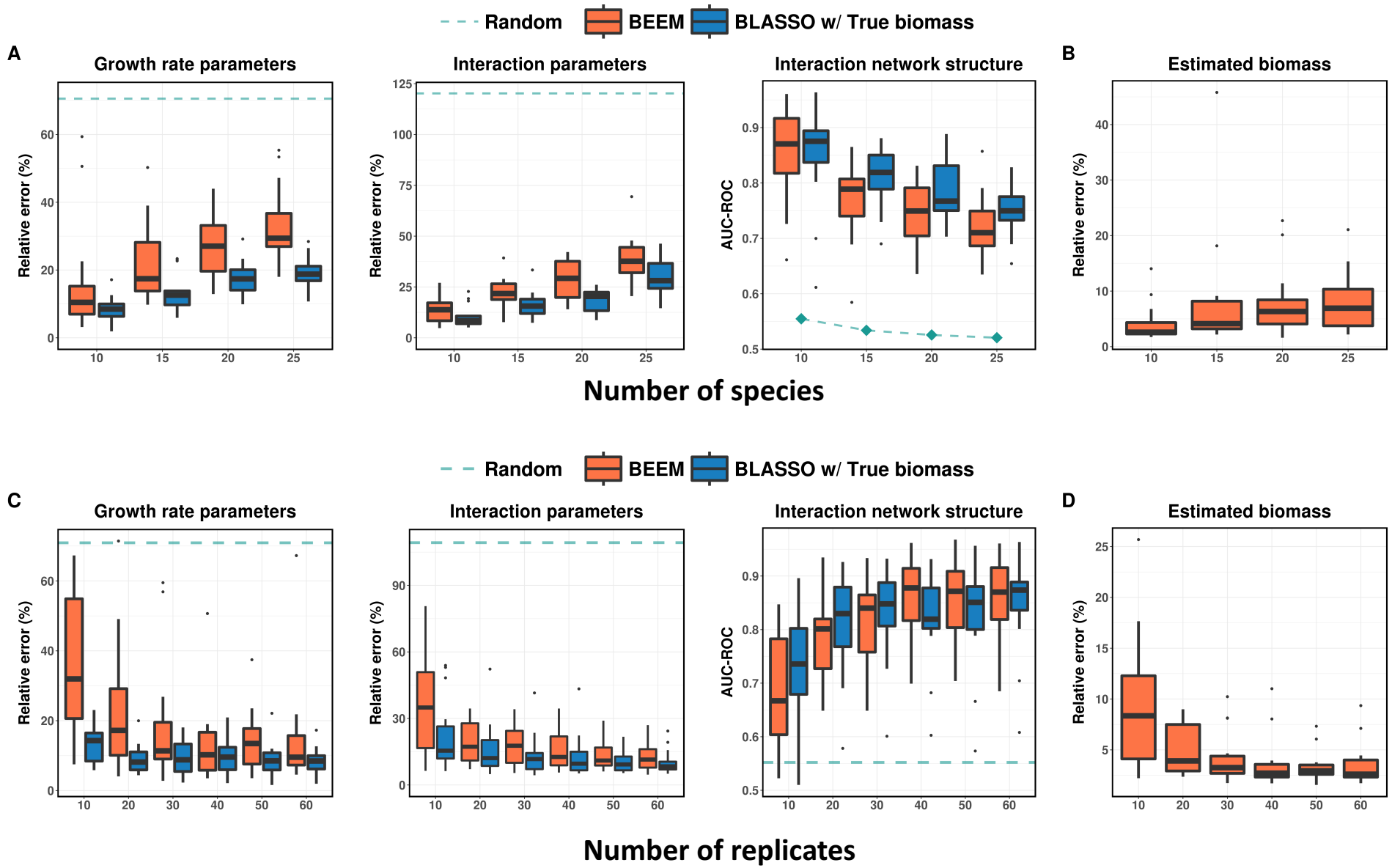

Random
BEEM
BLASSO w/ True biomass
A
Growth rate parameters
Interaction parameters
Interaction network structure
B
Estimated biomass
Relative error (%)
AUC-ROC
Number of species
C
D
Number of replicates

Figure 3

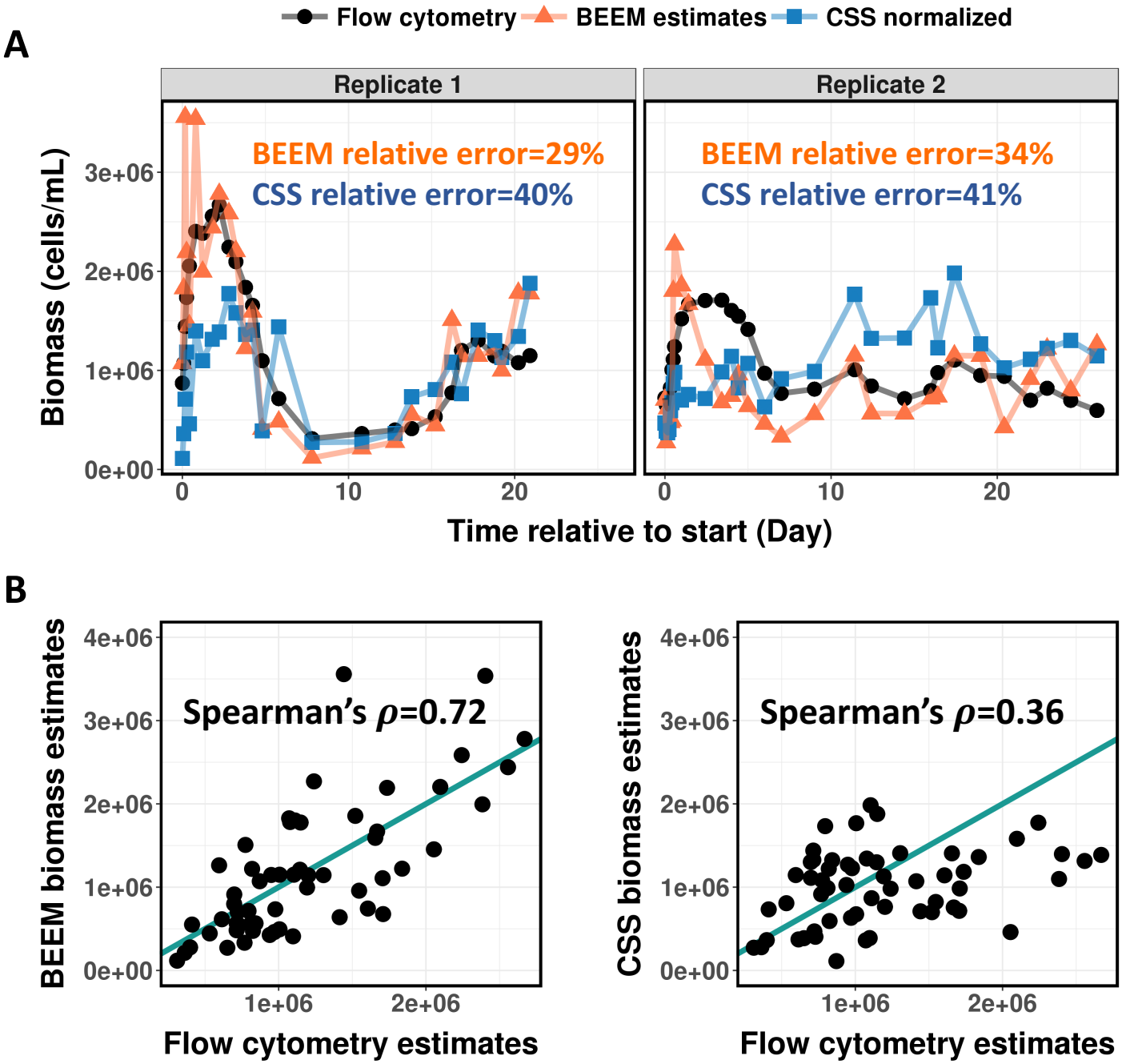
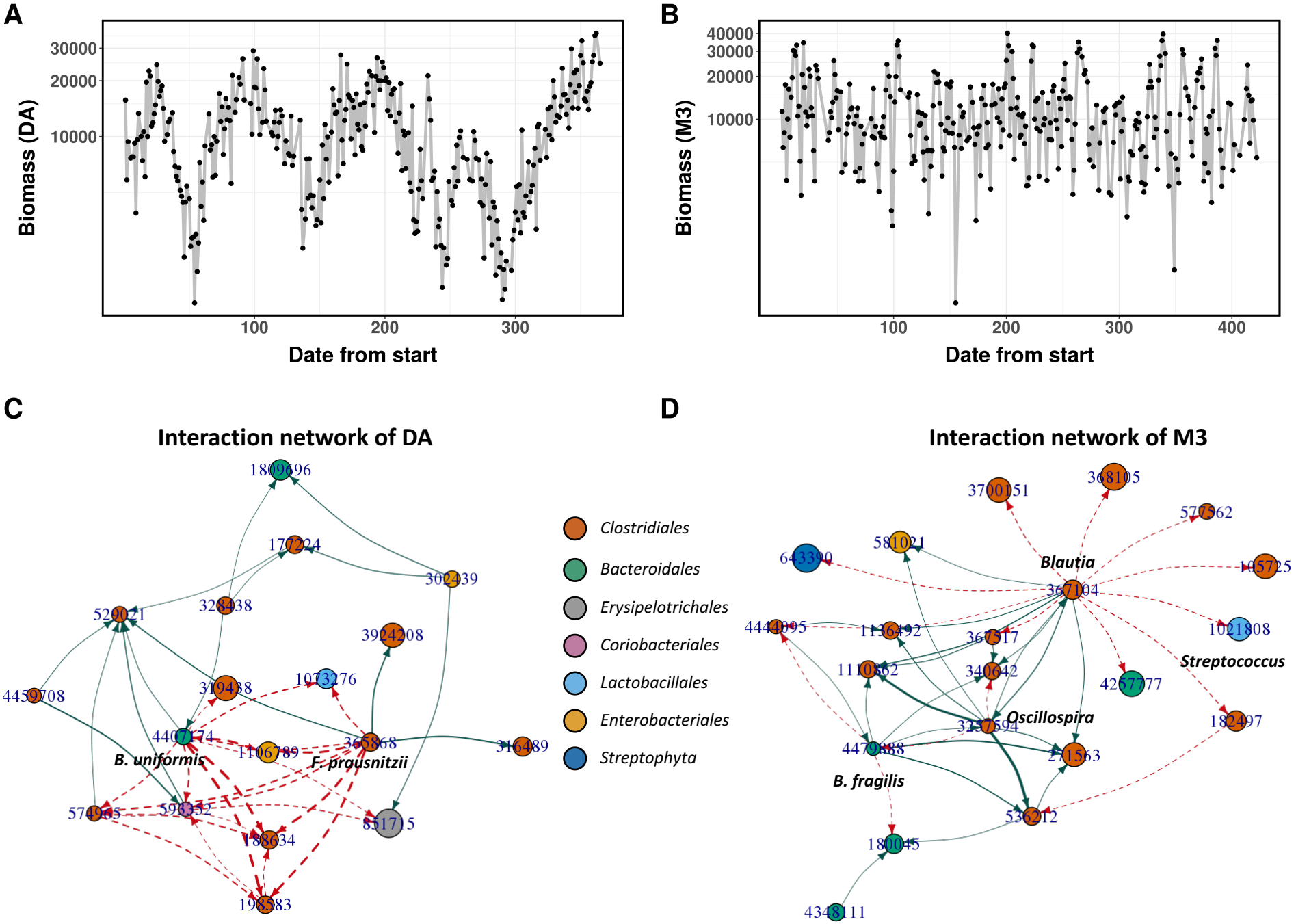
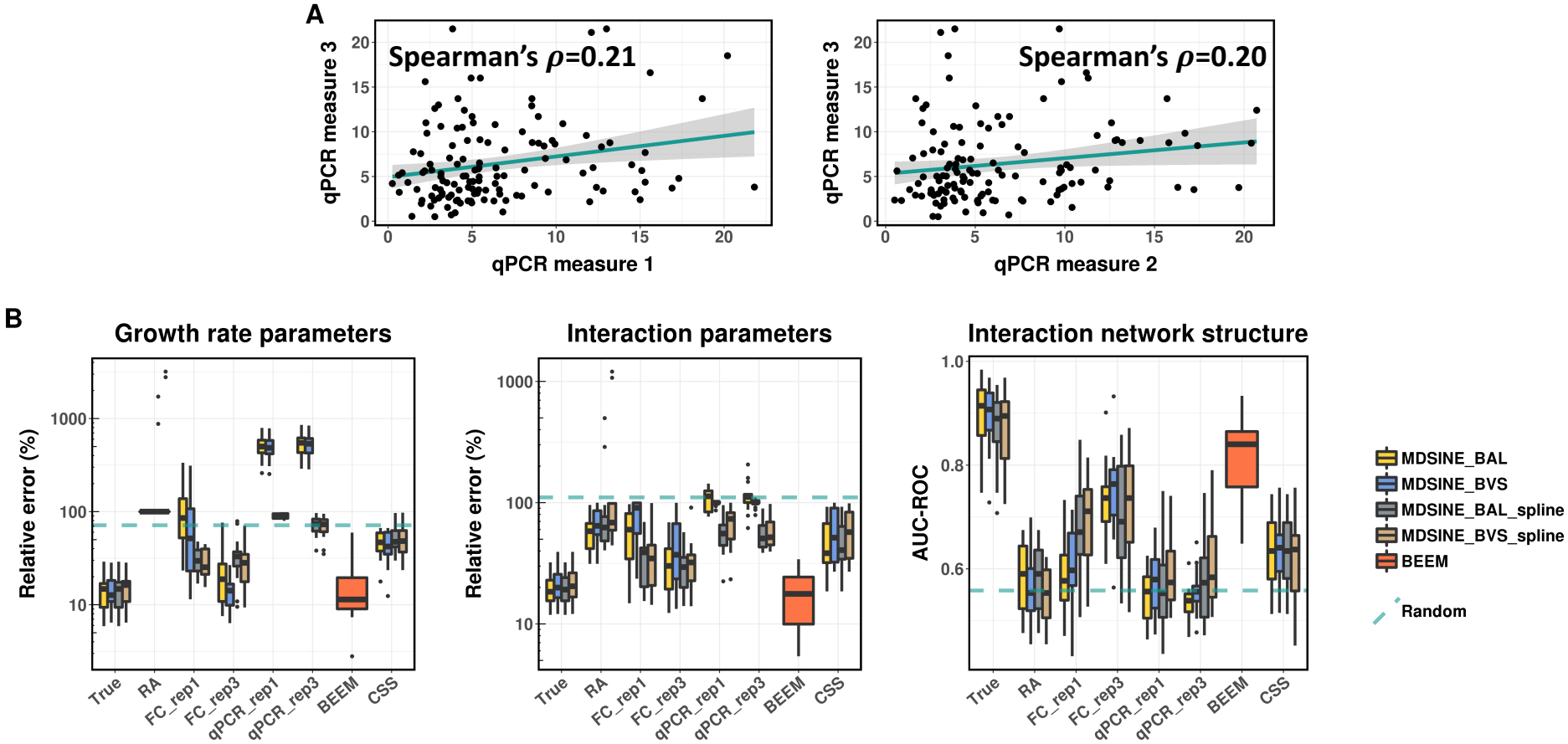
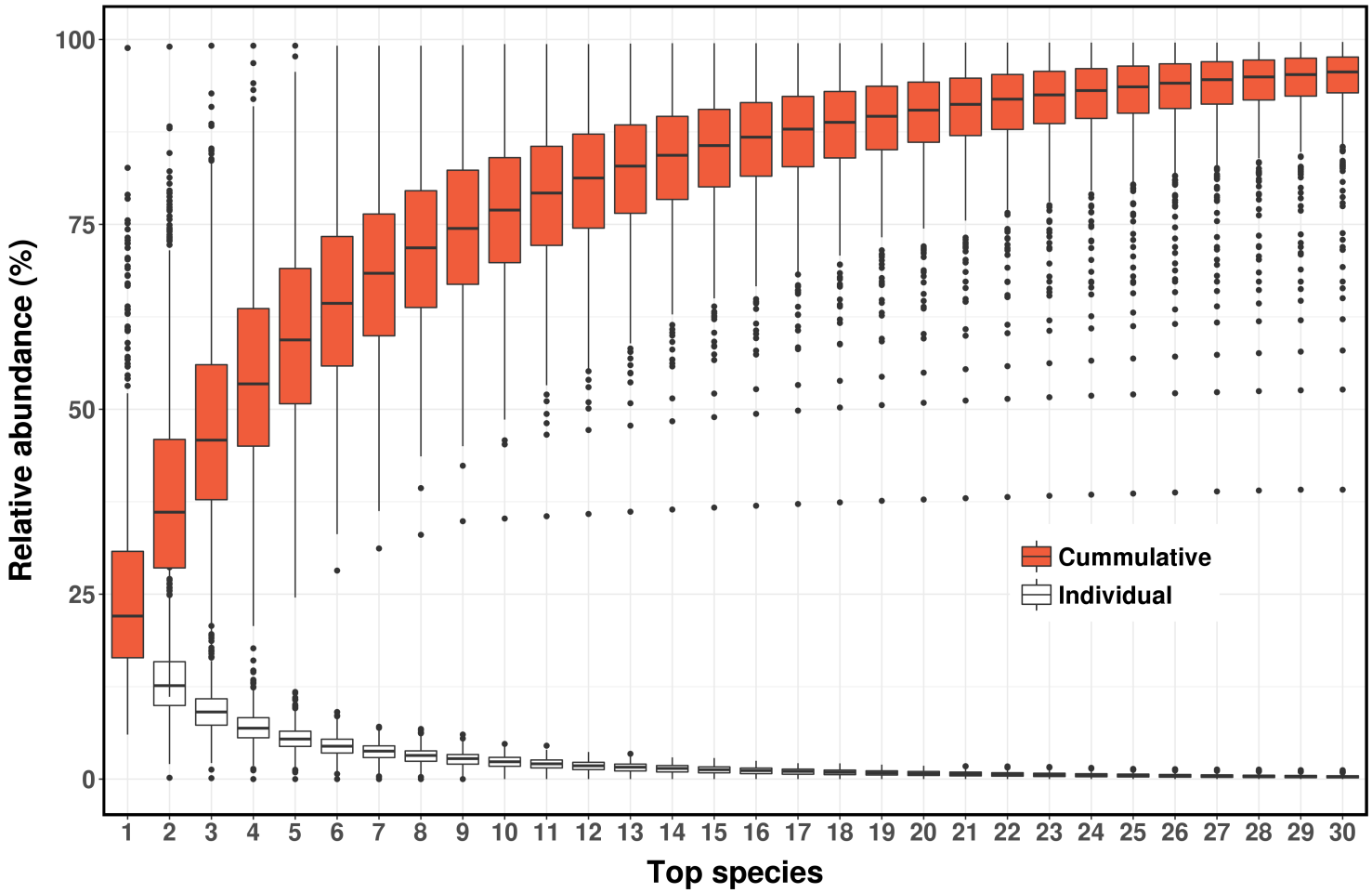

Flow cytometry
BEEM estimates
CSS normalized
A
Replicate 1
Replicate 2
BEEM relative error=29%
CSS relative error=40%
BEEM relative error=34%
CSS relative error=41%
Biomass (cells/mL)
0e+00
1e+06
2e+06
3e+06
0
10
20
Time relative to start (Day)
B
BEEM biomass estimates
Spearman's ρ=0.72
CSS biomass estimates
Spearman's ρ=0.36
4e+06
Flow cytometry estimates

Figure 4

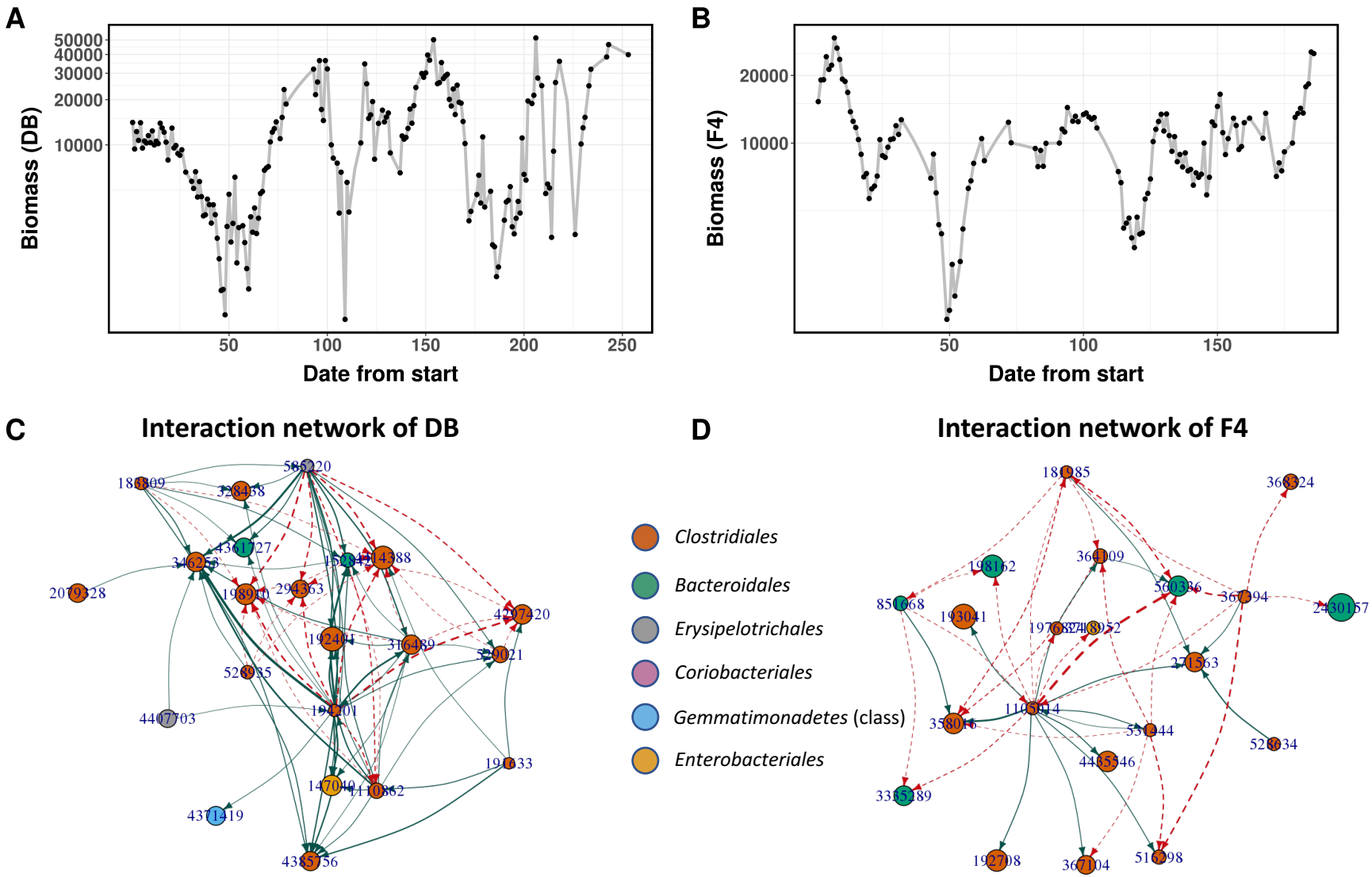

A
Biomass (DA)
30000
20000
10000
100
200
300
Date from start
B
Biomass (M3)
40000
30000
20000
10000
100
200
300
400
Date from start
C
Interaction network of DA
1809696
177224
302439
328438
529021
3924208
4459708
319438
1073276
4407174
1106789
365868
316489
B. uniformis
F. prausnitzii
574965
596352
851715
188634
198583
Clostridiales
Bacteroidales
Erysipelotrichales
Coriobacteriales
Lactobacillales
Enterobacteriales
Streptophyta
D
Interaction network of M3
3700151
368105
577562
581021
643390
Blautia
105725
367104
4444095
1136492
367517
1021808
1110862
340642
Streptococcus
4257777
Oscillospira
3357594
182497
4479888
271563
B. fragilis
536212
180045
4348111



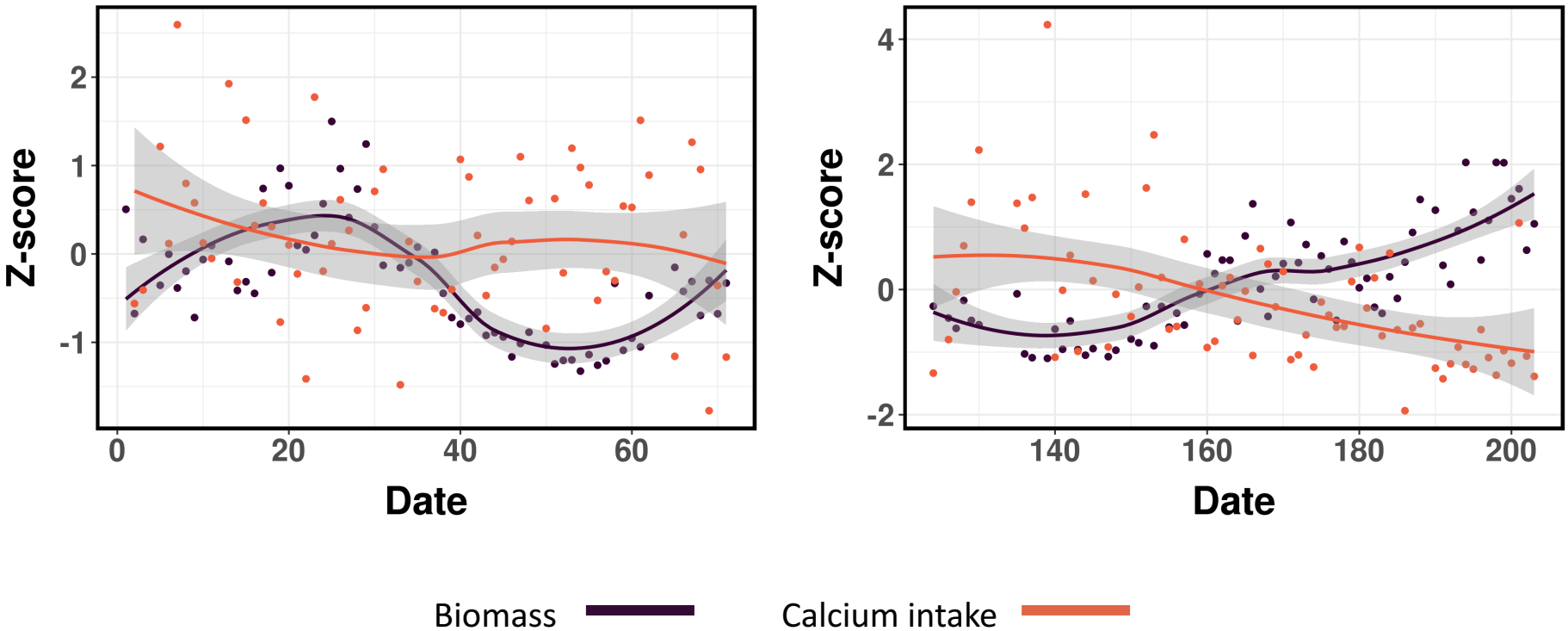
Supplementary Figure 1. Noise in experimentally determined biomass severely distorts gLVM parameter estimation. (A) Scatter plots with fitted linear regression lines for three 16S qPCR technical replicates from Bucci *et al*. (B) Relative impact of different gLVM parameter estimation algorithms (BAL, BVS, BAL_spline and BVS_spline as implemented in MDSINE) and data scaling approaches. Boxplots represent the summary of 15 simulations (10 species, 30 replicates with 30 time points each). Note that, in general, different data scaling approaches were found to impact performance more than the different estimation algorithms. Dashed horizontal lines represent the performance of randomly generated parameters from the simulation model. In general, scaling with true factors and using BEEM provided notably good results, and among competing experimental (FC_rep1, FC_rep3, qPCR_rep1, qPCR_rep3) and computational approaches (RA – relative abundance, CSS – CSS normalization), having three replicates from flow cytometry was the closest (FC_rep3).



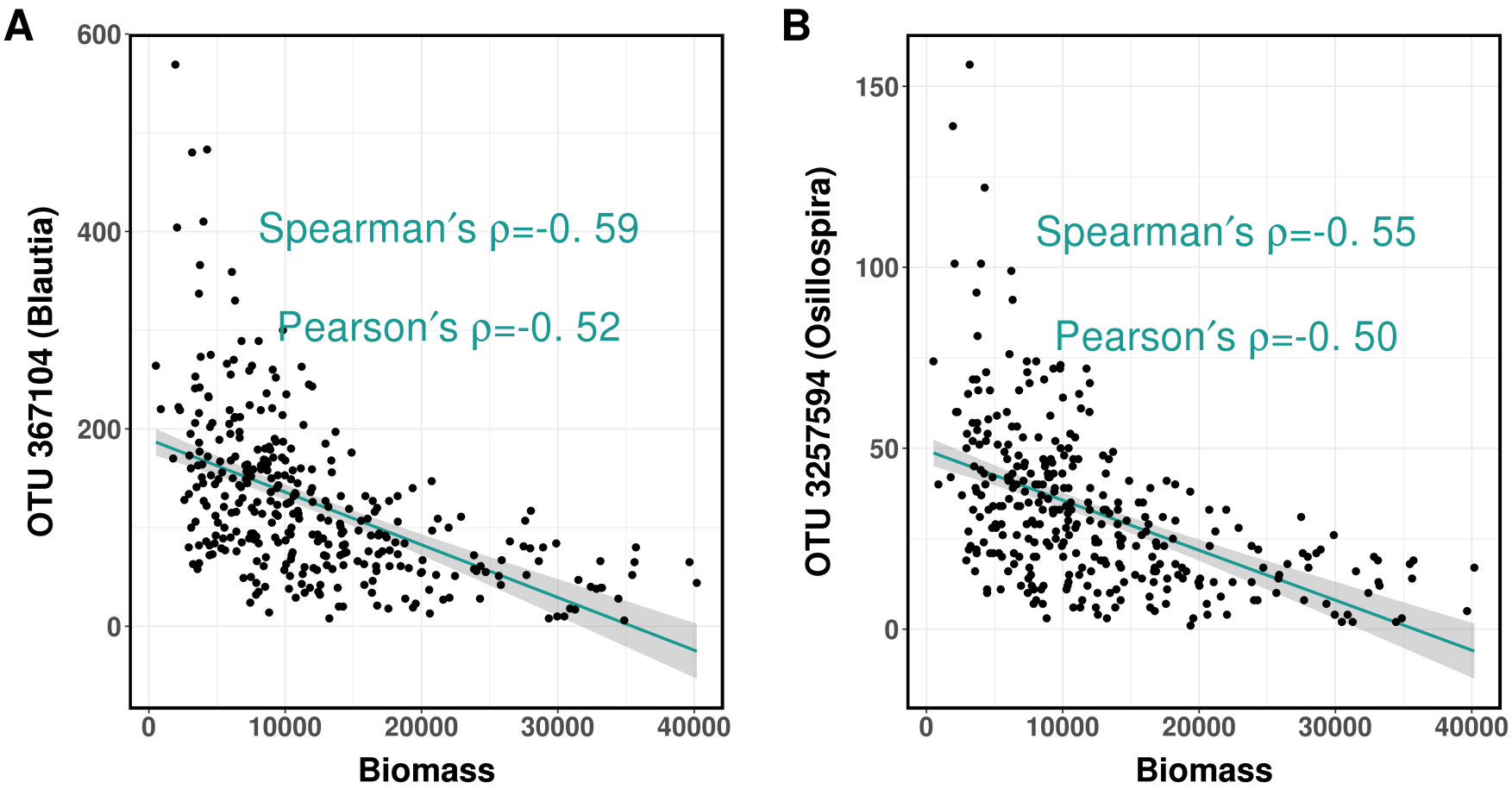
Supplementary Figure 2. Relative abundances observed for the most abundant species in 840 normal stool metagenomic samples from Pasolli *et al*. Filled boxplots show cumulative values.



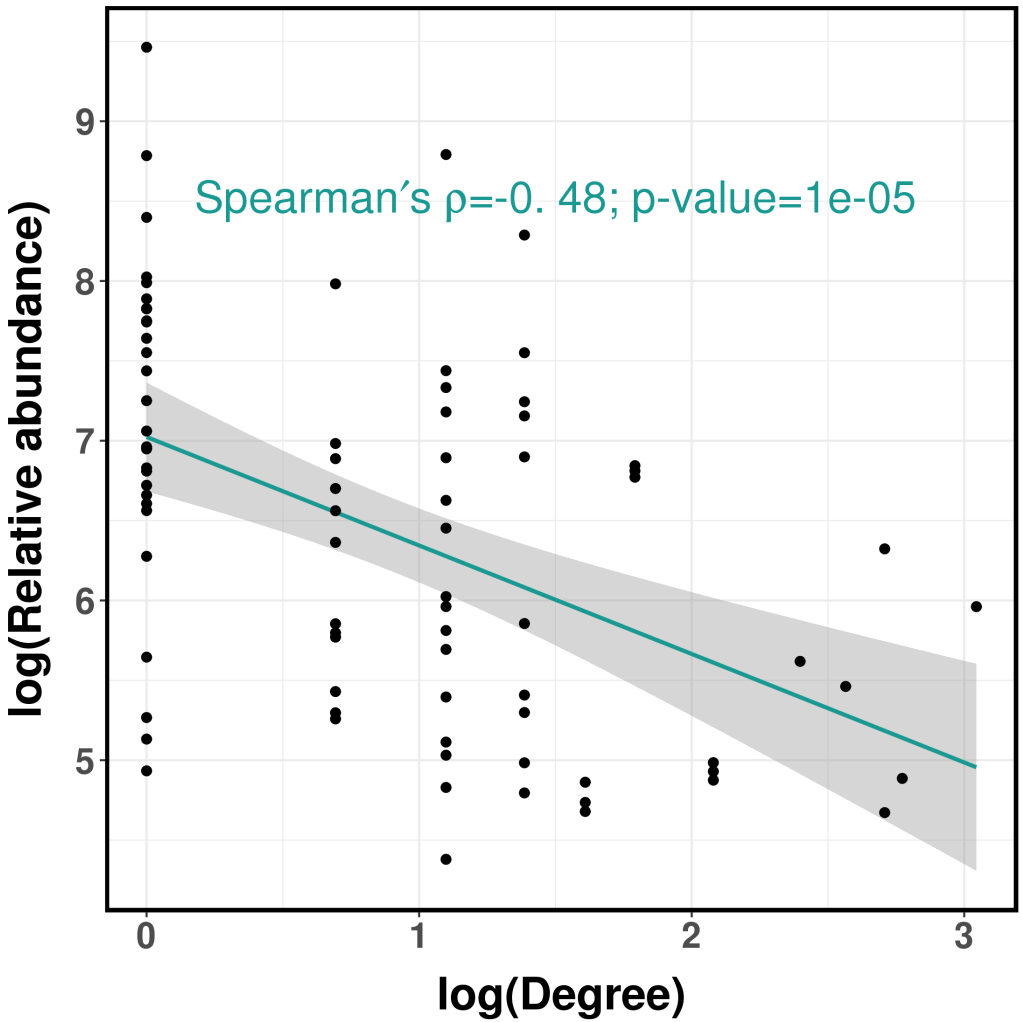
Supplementary Figure 3. BEEM estimated biomass (A and B; scaled to median of 10^4) and interaction networks (C and D) from the two shorter gut microbial longitudinal profiles from David *et al* and Caporaso *et al*. Dashed and solid edges represent positive and negative interactions, respectively, in the networks. Edge widths are proportional to the interaction strength and node sizes are proportional to log-transformed mean relative abundance of the corresponding species (OTU). Nodes are labeled with GreenGenes IDs and colored according to order level of taxonomic annotation.



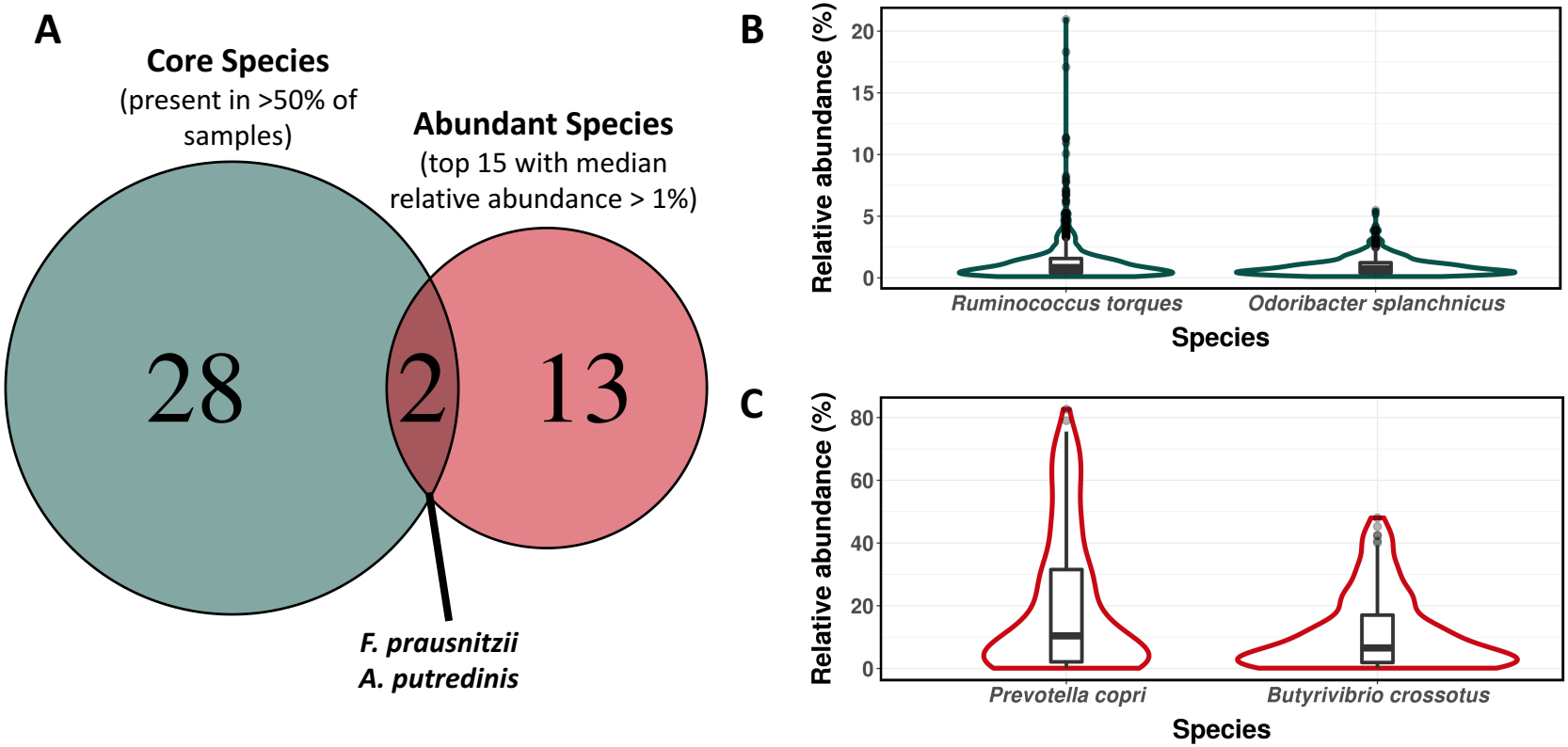
Supplementary Figure 4. Changes in calcium intake (z-score normalized) for the preceding day (orange) in relation to BEEM-estimated biomass (z-score normalized) for subject DA's gut microbiome (purple; only time points with calcium intake data are shown). Lines represent loess smoothers and shaded regions depict 95% confidence intervals. Overall, the two variables were found to be significantly correlated (Spearman's ρ=-0.40, p-value=1.7×10^{-6}).



Supplementary Figure 5. Scatter plots with fitted linear regression lines between the two hub OTUs and the estimated biomass of M3's gut microbiome. All correlations are significant (*p*-value<2.2×10^{-16}).



Supplementary Figure 6. Scatter plot with fitted linear regression line between the out- and in-degree of the OTU versus its mean relative abundance on log scale (based on networks for all 4 subjects).



Supplementary Figure 7. (A) Venn diagram for core species (present in >50% of samples) and abundant species (top 15 with median relative abundance > 1%) in healthy human gut microbiomes from Pasolli *et al* (N=840). (B) Examples of core gut species with low relative abundances. *R. torques* and *O. splanchnicus* were present in 95% and 69% of samples but both of them rarely have relative abundance >1%. (C) Examples of more abundant species that are only found in a fraction of individuals. *P. copri* and *B. crossotus* are frequently present at abundances >5%, but are only present in 34% and 15% of all samples, respectively.