Inferring Resource Competition in Microbial Communities from Time Series

Xiaowen Chen [1], Kyle Crocker [2,3,4] Seppe Kuehn [2,3,4,5,*,†] Aleksandra M. Walczak [1,4,*,‡] and Thierry Mora [1,4,*,§]

[1]*Laboratoire de Physique de l'École Normale Supérieure*, ENS, Université PSL, CNRS, Sorbonne Université,
Université Paris Cité, 75005 Paris, France
[2]*Department of Ecology and Evolution, The University of Chicago, Chicago, Illinois 60637, USA*
[3]*Center for the Physics of Evolving Systems, The University of Chicago, Chicago, Illinois 60637, USA*
[4]*Center for Living Systems, The University of Chicago, Chicago, Illinois 60637, USA*
[5]*National Institute for Theory and Mathematics in Biology, The University of Chicago and Northwestern University,
Chicago, Illinois 60611, USA*

The competition for resources is a defining feature of microbial communities. In many contexts, from soils to host-associated communities, highly diverse microbes are organized into metabolic groups or guilds with similar resource preferences. The resource preferences of individual taxa that give rise to these guilds are critical for understanding fluxes of resources through the community and the structure of diversity in the system. However, inferring the metabolic capabilities of individual taxa, as well as their competition with other taxa, within a community is challenging and unresolved. Here we address this gap in knowledge by leveraging dynamic measurements of abundances in communities. We show that simple correlations are often misleading in predicting resource competition. We show that spectral methods such as the cross-power spectral density and coherence that account for time-delayed effects are superior metrics for inferring the structure of resource competition in communities. We demonstrate this fact on synthetic data generated from consumer-resource models with time-dependent resource availability, where taxa are organized into groups or guilds with similar resource preferences. By applying spectral methods to oceanic plankton time-series data, we demonstrate that these methods detect interaction structures among species with similar genomic sequences. Our results indicate that analyzing temporal data across multiple timescales can reveal the underlying structure of resource competition within communities.

I. INTRODUCTION

The collective biological activity of ecosystems is defined by fluxes of resources. From plant [1] to microbial [2] communities, the availability of resources such as sunlight or reduced carbon and nitrogen enables the generation of energy and the production of biomass. Competition for resources therefore drives ecological interactions between members of the system. As a result, resource-mediated interactions such as competition have long been recognized as central structuring properties of ecosystems across scales.

Perhaps nowhere are the structuring forces of resources more clear than in microbial consortia. Communities of microbes utilize resources in almost every niche on the planet, from carbon remineralization in the photic zone of the ocean [3] to fiber utilization in the rumen of ungulates [4] and the collective transformation of oxidized nitrogen compounds in the soil [5,6]. In each of these cases, complex communities of hundreds or thousands of species compete for and utilize resources to generate energy and biomass.

In these contexts, the metabolic traits of microbes utilizing each resource become a key property of interest because they dictate which species compete for which resources. In part due to the conserved structure of the biochemical pathways that enable the utilization of resources, microbial communities are frequently observed to be comprised of guilds, or groups of species, that utilize a similar set of resources [7,8]. These guilds, often comprised of tens (or more) of taxa, compete for diverse pools of resources. For example, marine communities utilize complex mixtures of carbon sources [9] and many members of the human gut microbiome utilize mixtures of carbon sources [10]. Thus the mapping of species to the resources they utilize typically has a blocklike structure, with groups of species utilizing groups of resources [11].

A second key feature of resource utilization in microbial ecosystems is temporal fluctuations. Temporal variability in resource availability is the norm across many if not most habitats. For example, diurnal fluctuations drive the availability of carbon in photosynthetically driven ecosystems [12], and feeding temporally structures the availability of resources in host-associated microbial communities [13]. Similarly, in

---

*These authors contributed equally to this work.
†Contact author: seppekuehn@uchicago.edu
‡Contact author: aleksandra.walczak@phys.ens.fr
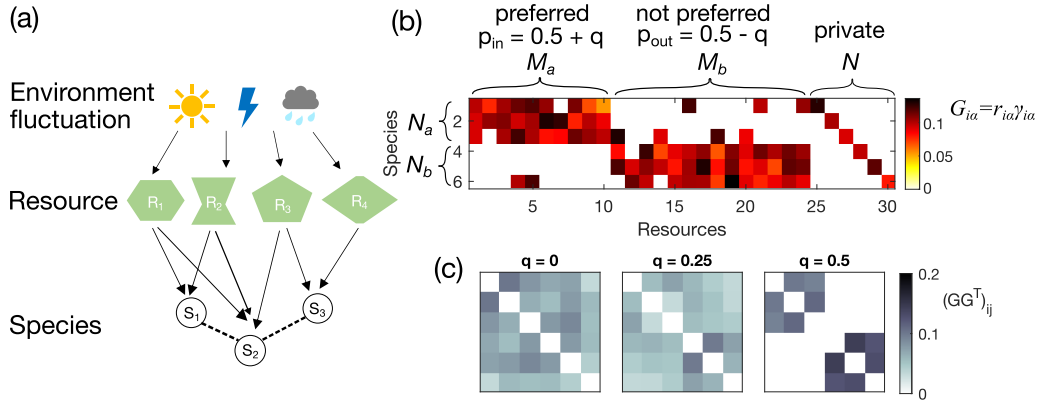§Contact author: thierry.mora@phys.ens.fr

FIG. 1. Environment-mediated ecological interactions between resources and consumers. (a) Cartoon of interactions among environment-mediated resources and consumer species. (b) Example of the gain matrix for consumer-resource interaction **G** with elements $G_{i\alpha} = r_{i\alpha}\gamma_{i\alpha}$ (color bar). The probability of nonzero consumer-resource interactions depends on whether the resource is preferred by the species or not, with an adjustable parameter q that encodes guild-structure bias ($q = 0.4$ for this example). All resources also have private resources to prevent extinction. (c) Effective interactions among the species are captured by the effective resource-utilization overlap matrix **GG**$^\mathsf{T}$ (color bar), shown here for three different values of the guild-structure bias q.

soils, transient dynamics in moisture drive large fluctuations in nutrient availability [14].

Given the importance of resources, the central role of microbial traits and temporal fluctuations in mediating the dynamic utilization of resources by communities, a central challenge in the field, is understanding resource-mediated interactions in communities. This challenge is highlighted by the fact that, in the context of a complex community, we cannot easily access the resources each and every species is utilizing. In most cases, isolating individual strains or taxa and assaying their resource preferences is infeasible. While genome-scale models [15] show potential for predicting resource utilization from sequencing data (shotgun metagenomics), these methods remain highly error prone for genomes of nonmodel organisms [16]. In light of these considerations, it is important to develop empirically grounded methods to infer the resource-mediated interactions in a complex consortium.

Here we ask what metrics on the observed time series of species abundances can give us insight into the structure of resource-mediated interactions in the community. In particular, if there are metabolic guilds in the community, how can we reliably infer them from time-series measurements? We approach this problem in the context of consumer-resource models [17]. We focus on consumer-resource formalisms because recent work has shown these models to be powerful quantitative formalisms for understanding abundance and resource dynamics in real communities [6,18], suggesting that this formalism has some predictive power in the microbial context.

While a host of powerful methods have been developed to infer interactions in ecological communities, many of these methods rely on statistical analysis of correlations in species abundances and make simplifying assumptions such as sparsity [19,20]. Leveraging the widespread availability of time-series data in microbial communities [21–23], we present temporal pairwise measurements as novel predictors of resource-mediated species-species interactions and apply them to existing datasets.

II. RESULTS

A. Environment-mediated consumer-resource model

In real experiments, the underlying consumer-resource interactions and resource dynamics are exceedingly challenging to measure. To develop a method that can be applied to analyze experimental time traces, we first introduce a model with known ground truth to develop an appropriate metric, which can reveal the effective resource overlap between species, using only the observed abundances of consumer species.

We consider an ecological system with N species with abundances x_i ($i = 1, \ldots, N$) and M resources with concentrations R_α ($\alpha = 1, \ldots, M$). Interactions are given by an externally supplied resource model. The dynamical effect of environmental fluctuations is explicitly modeled by the time dependence of the external supply, $K_\alpha(t)$, that mediates resource availability [Fig. 1(a)] [17,24–26]. Mathematically,

$$\dot{x}_i = \left(\sum_{\alpha=1}^{M} r_{i\alpha}\gamma_{i\alpha}R_\alpha(t) - d_x \right) x_i(t), \tag{1}$$

$$\dot{R}_\alpha = K_\alpha(t) - \left(d_R + \sum_{i=1}^{N} r_{i\alpha}x_i(t) \right) R_\alpha(t), \tag{2}$$

where $r_{i\alpha}$ is the uptake rate of resource α for species i (and simultaneously the depletion rate of resource α given species i), $\gamma_{i\alpha}$ is the yield of species i given resource α, i.e., how much species i grows by consuming resource α, and d_x and d_R are the death rates of the species and the resources, respectively. For simplicity, d_x is set to be equal for all species and d_R across all resources. We call this model the environment-mediated consumer-resource model (ECRM).

Because resources impact species abundances through exponential growth, it will be useful to log-transform species abundances before analysis, defining

$$\tilde{s}_i(t) \equiv \log x_i(t), \tag{3}$$

which we further normalize as the z score,

$$s_i(t) = \frac{\tilde{s}_i(t) - \langle \tilde{s}_i \rangle}{\sigma_s}. \qquad (4)$$

Intuitively, the z score describes the deviation from the mean in terms of expected standard deviation.

To model the impact of environment, we assume that its fluctuations directly influence the resource abundance, which then, by means of the consumer-resource relation, is transmitted through the ecological system. We consider two types of environmental fluctuations $K_\alpha(t)$.

First, sinusoidal drives are given by

$$K_\alpha(t) = K_\alpha^0 + A_\alpha \sin(\omega_\alpha t - \phi_\alpha), \qquad (5)$$

where K_α^0 is the base supply for resource α, A_α the amplitude of the sinusoidal drive, ω_α the frequency, and ϕ_α the phase. This form models the periodic drives omnipresent in nature, for example, diurnal cycles in aquatic systems, or regular food intakes in host-associated microbiomes.

Second, Ornstein-Uhlenbeck (OU) drives evolve according to

$$K_\alpha(t) = K_\alpha^0 + Q_\alpha(t), \qquad (6)$$

with K_α^0 the base supply for resource α as in the sinusoidal case and Q_α a stochastic OU process defined by

$$\frac{dQ_\alpha}{dt} = -\omega_\alpha Q_\alpha + A_\alpha \sqrt{\omega_\alpha} \eta_\alpha(t). \qquad (7)$$

Here ω_α and A_α are defined such that both the power and the timescales in the sinusoidal and the OU processes match, and the white noise $\{\eta_\alpha(t)\}$ satisfies

$$\langle \eta_\alpha(t) \rangle = 0, \qquad \langle \eta_\alpha(t) \eta_\beta(t') \rangle = \delta_{\alpha\beta} \delta(t - t'). \qquad (8)$$

Compared to sinusoidal drives, OU drives are stochastic, which are useful in modeling environmental fluctuations with unknown strength and regularity such as redox fluctuations in soils [27].

The gain matrix for the consumer-resource interaction **G**, where each entry is $G_{i\alpha} = r_{i\alpha} \gamma_{i\alpha}$, describes the intake of resource α by species i [Fig. 1(b)]. Summing over the resources, we obtain the resource-utilization overlap matrix $\mathbf{GG}^\mathsf{T}$ [Fig. 1(c)]. Large off-diagonal entries of this matrix $(\mathbf{GG}^\mathsf{T})_{i,j}$ indicate that species i and j have high overlap in the resources they utilize. Species with high resource overlap are expected to compete and those without are expected not to compete. One can consider the resource-utilization matrix as a proxy for effective interactions between species, which is informative about the community structure.

Specifically, we are interested in detecting guilds among the interacting species. To introduce a guild structure in resource overlap, we write the uptake rate as $r_{i\alpha} = r g_{i\alpha}$, where $g_{i\alpha}$ is an adjacency matrix that takes the value of 0 if species i does not uptake resource α and 1 if it does consume this resource. In order to generate ensembles of random interaction strengths such that the resulting $\mathbf{GG}^\mathsf{T}$ matrix has a block (guild) structure, we construct a bipartite graph between N species and M resources. We then separate both the species and the resources into k groups, with each group of species preferring to consume resources from one group of resources. The elements of the adjacency matrix that specifies

these preferences, $g_{i\alpha}$, are drawn from a Bernoulli distribution parametrized by a probability p_{in}. For the other $k - 1$ groups of nonpreferred resources, the elements of the adjacency matrix $g_{i\alpha}$ are drawn from a Bernoulli distribution parametrized by a probability $p_{\text{out}} < p_{\text{in}}$. In addition, each species is assigned its private resource to prevent extinction [Fig. 1(b)]. The yield $\gamma_{i\alpha}$ is drawn from a Gaussian distribution with a positive mean and small variance to introduce randomness. All yields are non-negative (see Sec. IV for specific parameter values).

When there are only two guilds a and b, we can reduce the number of parameters by defining a guild-structure bias q and set $p_{\text{in}} = 0.5 + q$ and $p_{\text{out}} = 0.5 - q$. For the first set of N_a species we set the probability that they consume the M_a resources to $p_{\text{in}} = 0.5 + q$, while their probability to consume the remaining M_b resources is $p_{\text{out}} = 0.5 - q$. The N_b species from the second group have opposite preferences: With probability $p_{\text{out}} = 0.5 - q$ they consume the first M_a resources and with probability $p_{\text{in}} = 0.5 + q$ the remaining M_b resources. Here q acts as a tuning variable for block structures. If $q = 0.5$, there are zero interactions between species that belong to different guilds. If $q = 0$, there is no bias of any species towards any resource [see Figs. 1(b) and 1(c)].

The consumer-resource model [Eqs. (1) and (2)], environmental drive [Eqs. (5)–(8)], and resource guild structure $r_{i\alpha}$ among consumers completely define the model. To illustrate its behavior and gain intuition, we first simulate a small network with $N = 6$ species and $M = 30$ resources, with a guild-structure bias of $q = 0.4$. The resource-utilization matrix $\mathbf{GG}^\mathsf{T}$ is given by Fig. 2(a). The specific set of parameters is given in Sec. IV. The first $N_a = 3$ species form one guild, while the remaining $N_b = 3$ species form the other. To study the impact of environmental drive on species abundances, we simulate the ECRM under sinusoidal and OU environmental drives, each sampled at three different ranges of timescales: slow, fast, and with a mixture of timescales [Fig. 2(b)]. Example traces of pairs of species within the same guild show complex coupled dynamics depending on the environmental timescales [Fig. 2(d)].

B. Pairwise measures of species couplings

To explore how the coupled dynamics of species abundances can be informative about their resource overlap $\mathbf{GG}^\mathsf{T}$, we consider observables of increasing complexity that measure the deviation from independence between the time traces of two species.

1. Equal-time correlation **C(0)**

We first consider the equal-time correlation coefficient, defined as

$$C_{ij}(0) = \lim_{T \to \infty} \frac{1}{T} \int_0^T s_i(t) s_j(t) dt, \qquad (9)$$

which for finite data time-series size sampled at a finite sampling rate is

$$C_{ij}(0) \approx \frac{1}{T} \sum_{t=1}^{T} s_i(t) s_j(t), \qquad (10)$$
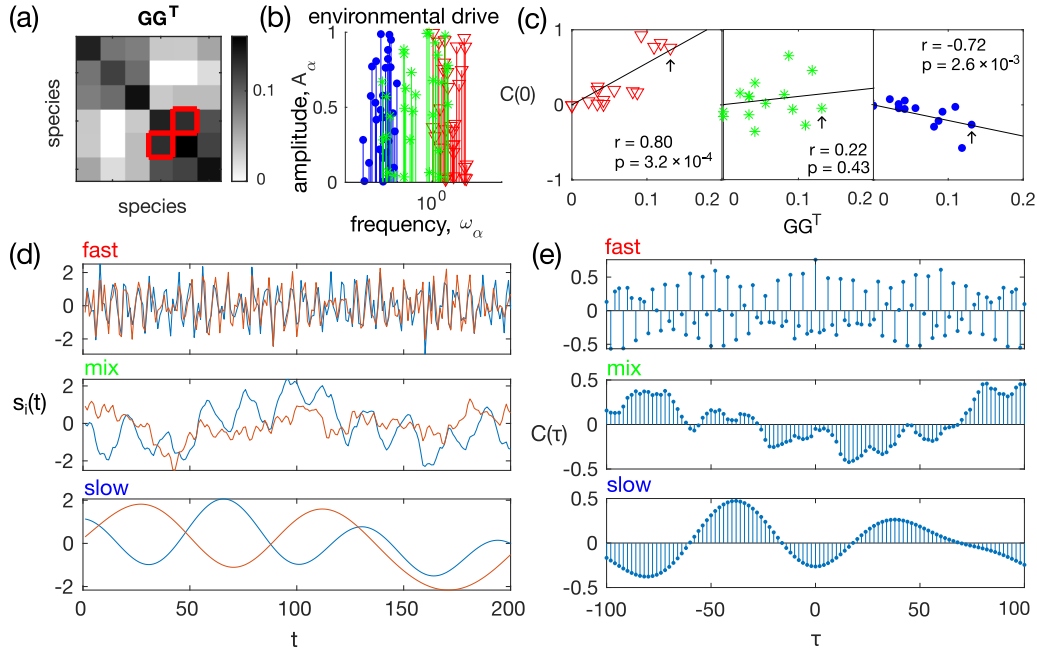
FIG. 2. Simulated guild-based ecosystem: an example environment-mediated consumer-resource model (number of species $N = 6$, number of resources $M = 30$, and guild-structure bias $q = 0.4$) demonstrating that the relation between the equal-time species abundance correlation $\mathbf{C}(0)$ and the underlying resource-utilization-overlap matrix $\mathbf{GG}^\mathsf{T}$ depend on the timescale of the environmental drive. (a) Resource-utilization-overlap matrix $\mathbf{GG}^\mathsf{T}$. The red squares indicate an example pair of interacting species. (b) Environmental drives are chosen to be sinusoidal, with fast (red, $\omega_\alpha \in [1, 10]$), slow (blue, $\omega_\alpha \in [0.01, 0.1]$), and mixed timescales (green, $\omega_\alpha \in [0.03, 3]$). (c) The correlation between the elements of the equal-time correlation function $\mathbf{C}(0)$ and the elements of the resource-utilization matrix $\mathbf{GG}^\mathsf{T}$ is positive when the environmental drive is fast (red), negative when the drive is slow (blue), and shows no correlation when the drive is composed of a mixture of slow and fast timescales (green). In each panel, r is Spearman's rank correlation coefficient and p is the p value for the null hypothesis that there is no correlation between $\mathbf{C}(0)$ and $\mathbf{GG}^\mathsf{T}$. Black arrows indicate the example pair of species in the red squares of (a). (d) Example segments of time series of the log-transformed species abundances for the red-square pair of (a), under the three types of environmental drive given in (b). (e) Cross-correlation function between the species abundances of the example red-square pair of (a). The equal-time correlation is the value of $\mathbf{C}(\tau)$ at $\tau = 0$.

where $t = 1, \ldots, T$ now denotes the time frame. As the equal-time correlation describes how much the concentrations of pairs of species fluctuate together, it is a measure commonly used in detecting species-species interactions. However, it is known that equal-time correlations cannot accurately reproduce the interactions [20].

Specifically, as shown in Fig. 2(c), in our example of six species distributed into two guilds, when the environmental drive is fast, $C_{ij}(0) \propto (\mathbf{GG}^\mathsf{T})_{ij}$, while when the environmental drive is slow, $C_{ij}(0) \propto -(\mathbf{GG}^\mathsf{T})_{ij}$. Intuitively this happens because fast resource dynamics drives coherent responses of members within a guild, but slow resource dynamics leads to competition within guilds (the underlying mechanism was first explored and explained in detail in [28]). This problem is more pronounced when environmental fluctuations are driven by a mixture of timescales [middle panel in Fig. 2(c)]. There is no correlation between the equal-time correlation $\mathbf{C}(0)$ and the resource-utilization overlap $\mathbf{GG}^\mathsf{T}$ (p value equals 0.4, Spearman correlation test). In this example, the pair of species with the strongest resource-utilization overlap [indicated by red squares in Fig. 2(a)] has an equal-time correlation close to 0, as indicated by the black arrow in Fig. 2(c).

Without knowing *a priori* the timescale of the environmental fluctuation, the equal-time correlation $\mathbf{C}(0)$ does not return a good enough estimator of the resource-utilization overlap.

2. Cross correlation with time delay reveals species-interaction structure

Since equal-time correlations do not reveal the structure of effective species-species interactions, we take a closer look at time-series data. Let us focus again on the pair of species with the strongest resource overlap [red squares in Fig. 2(a)]. As shown by Fig. 2(d), the time evolution of species abundances depends heavily on the environmental drive. Under a fast environmental drive, the two species abundances track each other closely, while under a slow drive, they exhibit a phase shift. When the environmental drive is mixed with fast and slow modes, the dynamics of each species also exhibits a mixture of modes and the relation between the two species is less clear.

Given a complicated time series, we need to consider dynamical pairwise measures, such as the time-delayed cross-correlation function

$$C_{ij}(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_0^T s_i(t) s_j(t + \tau) dt. \tag{11}$$

As shown in Fig. 2(e), under all three types of environmental drive, the absolute value of the cross correlation at certain time lags is large, which suggests that it carries information about the species-species interaction structure. However, if one only
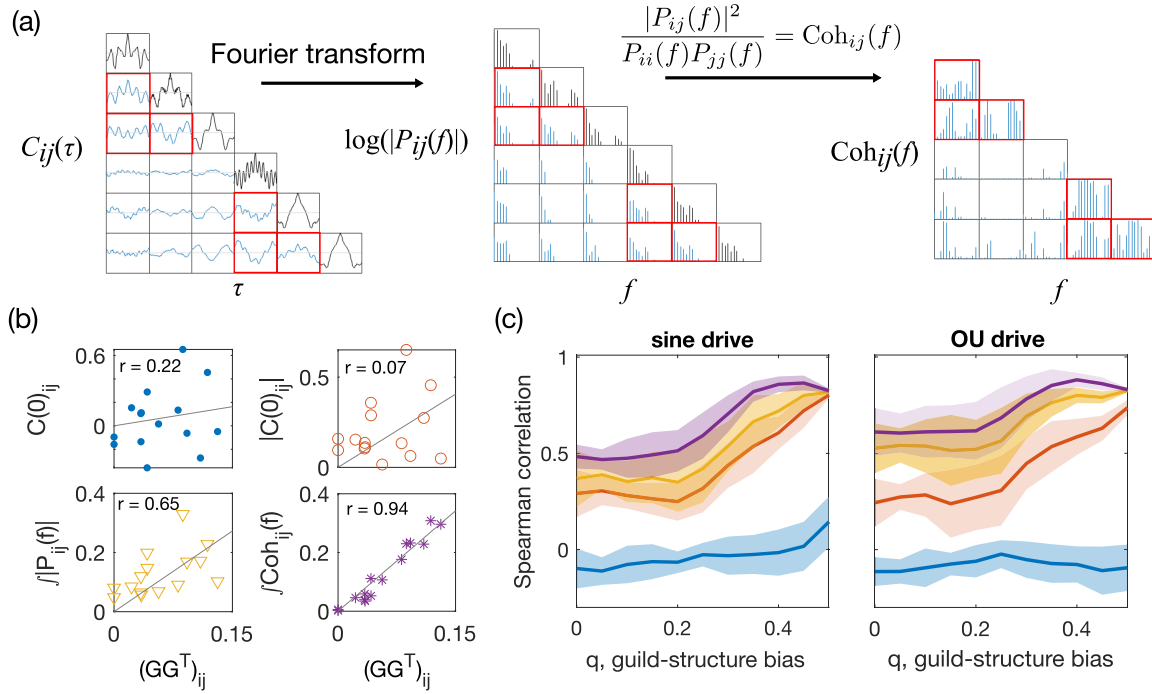
FIG. 3. Dynamical pairwise observables reveal more about the resource-utilization-overlap structure than equal-time correlations. (a) Schematic representation for calculating dynamical observables from the cross-correlation function. The pairwise cross-correlation function (left) is Fourier transformed into the CPSD (middle) and squared and normalized to obtained the squared coherence (right). All plots show results for the ECRM example in Fig. 2 under an environmental drive with mixed timescales. Pairs of species within the same guild are indicated by red squares. The guild-structure bias is $q = 0.4$. (b) Scatter plots between pairwise observables for each pair of species and the elements of the effective resource-utilization matrix $\mathbf{G}\mathbf{G}^\mathsf{T}$ for same data as in (a). Pairwise observables include the equal-time correlation $\mathbf{C}(0)$ [reproduced from the middle panel of Fig. 2(b)], its absolute value $|\mathbf{C}(0)|$, the total magnitude of the CPSD [$\mathcal{P}_{ij} = \int df |P_{ij}(f)|$], and the total coherence [$\mathcal{TC}_{ij} = \int df \, \mathrm{Coh}_{ij}(f)$]. The r is Spearman's rank correlation coefficient. (c) Prediction performance of the pairwise measures of species abundance for resource-utilization overlap, given by Spearman's correlation coefficient between the measure and the resource-utilization matrix value, as a function of the guild-structure bias q [same color code as in (b)]. Shaded areas represent standard deviation across ten realizations of the network, each with ten random realizations of environmental drive given by a sinusoidal process with ω_α sampled log-uniformly $\in [0.03, 3]$ (left) or by an OU process with $\omega_\alpha = 1$ (right). The total duration of the simulated trajectory is $t_f = 20\,000$. The sampling time step is $\Delta t_{\mathrm{sampling}} = 1$.

considers a single time delay, e.g., $\tau = 0$, the value can be very close to 0.

Comparing across different pairs of species, we observe that the cross-correlation function between species of the same guild varies with a much larger amplitude than between species from a different guild [Fig. 3(a), left panel; pairs of distinct species belonging to the same guild are outlined in red]. However, different pairs of interacting species show signatures of strong cross correlations at different time delays τ, implying that there is no *a priori* choice of any single time delay that would reveal the interaction structure between all pairs of species.

3. Spectral method for time-series analysis: Cross power spectral density and coherence

In order to focus on the interplay of timescales, we reformulated the dynamical pairwise measures in the frequency domain. Spectral analysis has been used to probe many aspects of ecological systems [29]. Here we measure the cross power spectral density (CPSD) by taking the Fourier trans-

form of the cross-correlation function,

$$P_{ij}(f) = \int_{-\infty}^{\infty} C_{ij}(\tau) e^{-i2\pi f \tau} d\tau. \tag{12}$$

We divide its magnitude squared by the power spectral density of each of the two time series to define the (magnitude-squared) coherence

$$\mathrm{Coh}_{ij}(f) = \frac{|P_{ij}(f)|^2}{P_{ii}(f)P_{jj}(f)}. \tag{13}$$

Coherence ranges between 0, when the two signals are unrelated, and 1, when they are perfectly correlated up to a phase shift. It is used in signal processing to detect the relation between two signals [30].

Figure 3(a) shows a schematic of how to compute the CPSD and coherence from the cross-correlation function for the example network in Fig. 2(a) with mixed timescales of the environmental drive. Recalling that the underlying resource utilization overlap network has a block structure (the first three species belong to the same guild and the last three species belong to another guild), it becomes clear from

Fig. 3(a) that both the magnitude of the CPSD and the coherence are larger between intraguild pairs of species than between interguild species. As coherence further reduces the impact of individual species (due to normalization), the difference between intra- and interguild pairs is larger in the coherence than in the CPSD magnitude [Fig. 2(a)].

Both the CPSD and coherence are frequency dependent and contain information about the relation between the species abundance z scores $s_i(t)$ and $s_j(t)$ at each frequency. To obtain an aggregate measure of the coupling between species abundances over many timescales, we consider the integral of the CPSD magnitude and coherence across all frequencies. Specifically,

$$\mathcal{P} \equiv \int_0^{f_{\max}} |P_{ij}(f)| df \approx \sum_{k=0}^{k_{\max}} |P_{ij}(f_k)| \Delta f, \qquad (14)$$

where Δf is the frequency resolution we choose when performing the discrete Fourier transform. Note that if we integrated the CPSD without taking its magnitude, we would recover the equal-time correlation by Parseval's theorem.

Likewise, we define the total coherence as the integral of the magnitude-squared coherence across all frequencies,

$$\mathcal{TC}_{ij} \equiv \int_0^{f_{\max}} \text{Coh}_{ij}(f) df \approx \sum_{k=0}^{k_{\max}} \text{Coh}_{ij}(f_k) \Delta f. \qquad (15)$$

The maximum frequency is given by the Nyquist frequency, $f_{\max} = 1/2\Delta t_{\text{sampling}}$. The frequency resolution Δf is chosen heuristically such that the power spectral density estimate is relatively stable when Δf is doubled or halved ($\Delta f = f_{\max}/15$ for all the results presented in this paper). Using the simulated ECRM data, we estimate the power spectral density of the species abundance time series using the MVGC Toolbox [31] and the MATLAB built-in function with Welch's method to reduce noise from using finite data. As shown in Fig. 3(b), compared to the equal-time correlation $\mathbf{C}(0)$ and its absolute value, both the total magnitude of the CPSD $\mathcal{P}$ and the total coherence $\mathcal{TC}$ are much more informative about the underlying resource-utilization-overlap matrix $\mathbf{GG}^{\mathsf{T}}$. Total coherence is the best of the four predictors. This advantage of the total coherence persists for ECRMs under both sinusoidal and OU drives and over the range of guild-structure biases q [Fig. 3(c)]. Thus, the two proposed measures $\mathcal{P}$ and $\mathcal{TC}$ offer a proxy for the resource-overlap interactions between species.

C. Binary classifier for guild-structure detection

In many cases, one is interested in recovering the guild structure of the community from time series [32]. Identification of the guild structure can enable a massive dimensional reduction in the dynamic description of the system and provide insight into key metabolic properties [33].

Motivated by this, we ask whether total coherence and CPSD magnitude could be used to learn the large-scale guild structure of $\mathbf{GG}^{\mathsf{T}}$. We use the four pairwise measures $\mathbf{C}(0)$, $|\mathbf{C}(0)|$, $\mathcal{P}$, and $\mathcal{TC}$ as scores to perform a binary classification between pairs of species belonging to the same guild and pairs belonging to different guilds. We apply this approach to our small $N = 6$ simulated ECRM system with two guilds [Fig. 2(a)]. To obtain a consistent partition into guilds, we
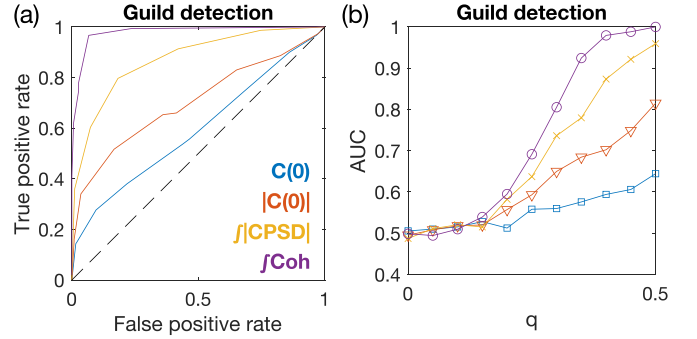

FIG. 4. Guild detection of resource-utilization overlap using the pairwise measures of species abundances with single-linkage clustering. (a) Receiver operator curve for successfully distinguishing same-guild from different-guild pairs of species using the four pairwise measures, from 50 random realizations of the simulated model in Fig. 2. The guild-structure bias is set to $q = 0.4$. (b) Area under the ROC for guild detection. The color code is the same as in (a). Environmental drives are given by OU processes with $\omega_\alpha = 1$. The total duration is $t_f = 20\,000$. The sampling time step $\Delta t_{\text{sampling}} = 1$.

perform single-linkage clustering initiated from all pairs of species whose pairwise measure is larger than a threshold within the range of the pairwise measure. For example, if after thresholding species i and j are determined to belong to the same guild, as well as species j and k, then all three species i, j, and k are grouped into to the same guild.

To visualize the performance of the binary classifiers, we plot the receiver operator curve (ROC) across different thresholds [Fig. 4(a)]. The total coherence outperforms all other measures in correctly partitioning species into guilds. The area under the ROC (AUC) shows the advantage of coherence persists for all guild biases q [Fig. 4(b)]. For $q = 0$ there is no guild structure to detect and all measures fail at the task, as they should. The ROC and AUC of the equal-time correlation predictor $C_{ij}(0)$ show that it performs only slightly better than random guessing (with the AUC equal to 0.5), for all values of q. Taking its absolute value improves prediction beyond random guessing, but is still worse than the spectral measures $\mathcal{P}$ and $\mathcal{TC}$.

D. Spectral methods work for general ECRM systems with larger sizes

In natural ecological systems, both the number of guilds and the number of species in each guild can be large. How well can the dynamical pairwise observables predict the resource-utilization overlap for different system sizes? We now consider general ECRMs with a larger number of guilds. To explore finite-size scaling, we consider three families of ECRMs, with the numbers of species, resources, and guilds given in Table I.

As we increase the size of the ECRMs by increasing the species size N, we also increase the number of resources in proportion, $M \propto N$, and the number of guilds as $k \propto \sqrt{N}$. We take $p_{\text{in}} = (1/2 + q)(N/6)^{-1/2}$ and $p_{\text{out}} = (1/2 - q)(N/6)^{-1}$. The scalings with N are chosen such that the average connectivity degree for each species in the effective resource-utilization matrix $\mathbf{GG}^{\mathsf{T}}$ becomes constant at large

TABLE I. Parameters for the ECRM with different system sizes.

N	$M \propto N$	$k \propto \sqrt{N}$
6	30	2
24	120	4
96	480	8

N (see Sec. IV B for details). The particular prefactors are chosen so that the numbers are consistent with the $N = 6$ case considered earlier.

We monitor the predictive performance of the same four pairwise measures as before [$\mathbf{C}(0)$, $|\mathbf{C}(0)|$, $\mathcal{P}$, and $\mathcal{TC}$] as we increase the system size from $N = 6$ and $k = 2$ to $N = 24$ and $k = 4$ and then to $N = 96$ and $k = 8$ [Figs. 5(a), 5(e), and

5(i)]. We compute Spearman's correlation coefficient between the pairwise observables and the elements of the $\mathbf{GG}^{\mathsf{T}}$ matrix [Figs. 5(b), 5(f), and 5(j)] and the ROC of guild detection after single-linkage clustering [Figs. 5(d), 5(h), and 5(i)].

By the choice of scaling for p_{in} and p_{out}, the $\mathbf{GG}^{\mathsf{T}}$ matrix is sparse for all system sizes. We can set a threshold on our measures to perform link detection, i.e., identify whether there exists an effective interaction between species due to competition for the same resource. As shown in Fig. 5 for $q = 0.5$ (and Fig. S1 in the Supplemental Material [34] for $q = 0.4$), for all three system sizes tested, the total coherence remains the best among the four predictors. The summary statistics in Fig. S2 in Ref. [34] show that the total coherence always performs the best among the four predictors, although all get worse as the system size increases. For $q = 0.5$, guild detection [Fig. S2(c) in Ref. [34]] has a better AUC than link
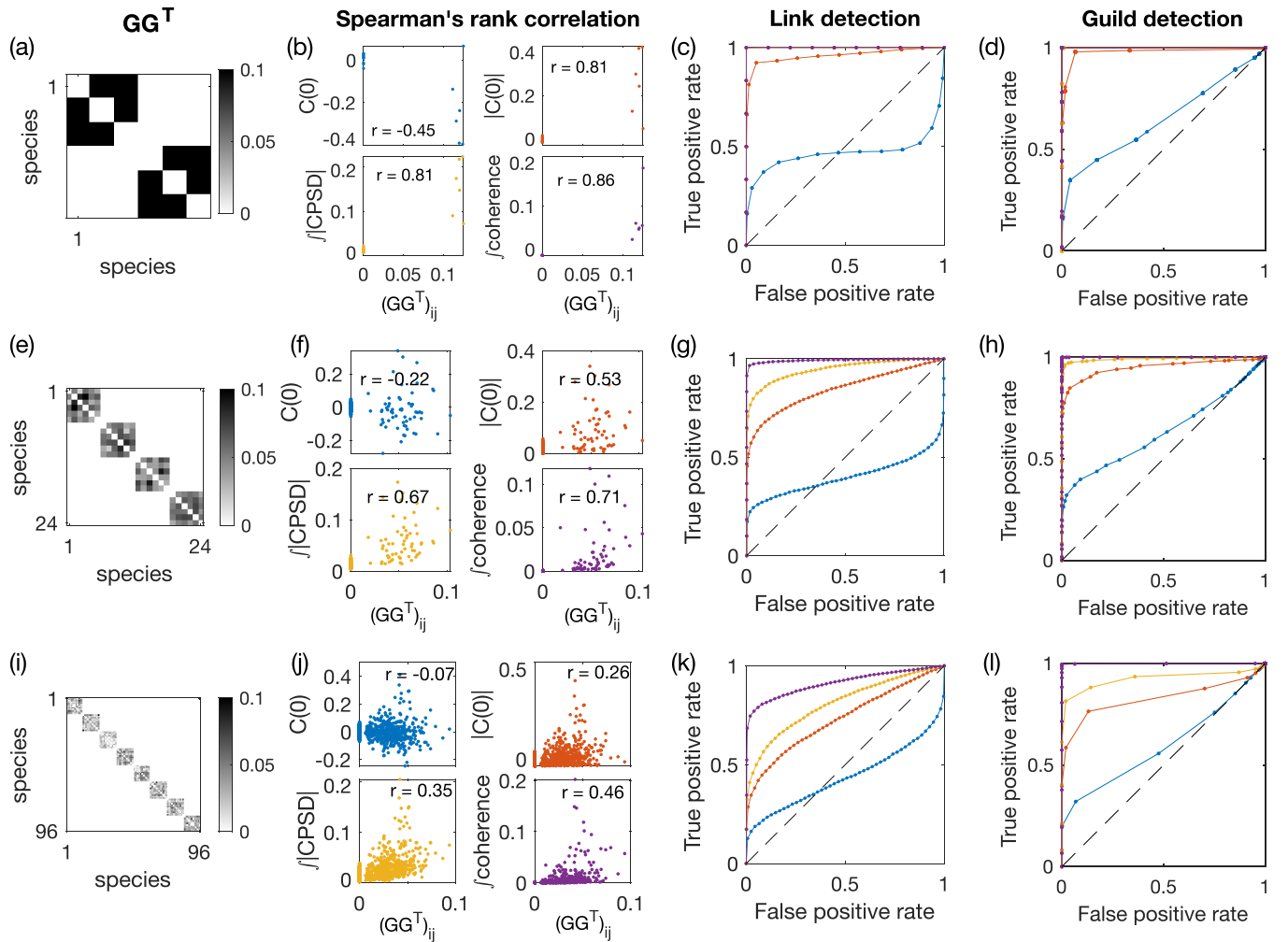

FIG. 5. Pairwise observables for the simulated ECRM ecosystem of different sizes (Table I), with $q = 0.5$. (a) Effective resource-utilization matrix $\mathbf{GG}^{\mathsf{T}}$ with $N = 6$ species, $M = 30$ resources, $k = 2$ guilds, and $q = 0.5$. (b) Correlation between pairwise measures of the species abundances and the resource-utilization overlap. Measured pairwise observables include the equal-time correlation $\mathbf{C}(0)$ and its absolute value $|\mathbf{C}(0)|$, the total absolute value of the CPSD, and the total coherence. (c) ROC of link detection for the four pairwise observables. The color code is the same as in (b); note that the total absolute value of the CPSD overlaps with the total coherence, as both measures return perfect predictions. (d) Same as (c) but for guild detection, after performing single-linkage clustering on the pairwise metrics with a tunable threshold. (e)–(h) Same as (a)–(d) but for $N = 24$, $M = 120$, and $k = 4$. (i)–(l) Same as (a)–(d) but for $N = 96$, $M = 480$, and $k = 8$. For all ECRMs, the environment is driven by OU processes with the intrinsic timescale set to $\omega_\alpha = 1$. For all system sizes, the total duration of the simulated trajectory is set to $t_f = 20\,000$ and the sampling time step is set to $\Delta t_{sampling} = 1$.
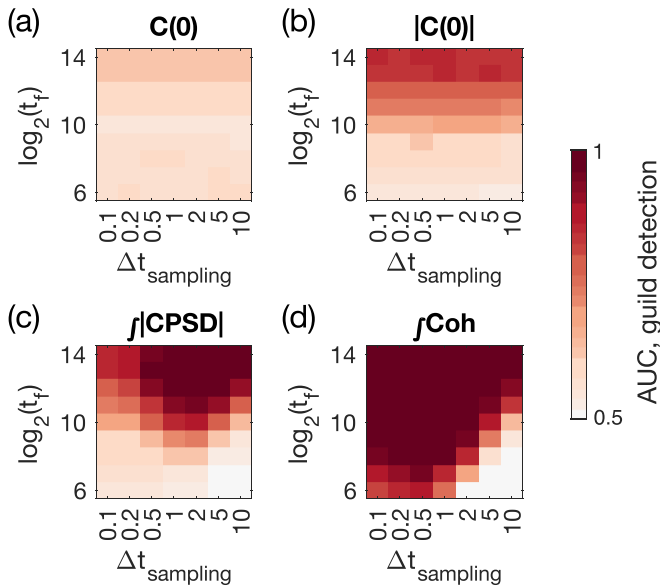
FIG. 6. Influence of finite data duration and sampling rate on guild detection using pairwise measures of species abundance. Shown is the AUC of the guild detection task using (a) equal-time correlation $\mathbf{C}(0)$, (b) its absolute value $|\mathbf{C}(0)|$, (c) the total absolute value of the CPSD, and (d) the total coherence, as a function of sampling duration t_f and interval $\Delta t_{\text{sampling}}$. For reference the species turnover timescale is $d_x^{-1} = 3.33$. The results are averaged over 50 random realizations of the ECRMs with $N = 6$, with OU environmental drive at $\omega_\alpha = 1$ and guild-structure bias $q = 0.4$.

detection [Fig. S2(b) in Ref. [34]], especially for large system sizes, due to error correction afforded by single-linkage clustering. For $q = 0.4$, however, guild detection performance is worse at larger system size, as single-linkage clustering overcorrects given the presence of nonzero resource competition across different guilds [Figs. S1(d) and S1(i) in Ref. [34]]. Alternatively, one can perform single-linkage clustering but only for one iteration, such that immediately adjacent neighbors are now linked. While suggesting more sophisticated community-detection methods, this middle ground provides a better recovery of the guild structure compared to link detection and guild detection with full single-linkage clustering [Figs. S1(e) and S1(j) in Ref. [34]].

E. Reconstruction of the $\mathbf{GG}^{\mathsf{T}}$ matrix depends on sampling rate and total time-series length

The success of the CPSD and coherence in detecting guild structures depends on the interplay of multiple timescales. These timescales include the doubling timescale $(\gamma r)^{-1}$ of the species, their typical lifetime d_x^{-1}, the environmental fluctuation timescales ω_α^{-1}, and their relaxation time d_R^{-1}. Additionally, there are experimental timescales: the total time t_f of data acquisition and sampling interval $\Delta t_{\text{sampling}}$. Here we discuss the effects of the finite time series and sampling rate on recovering the elements and structure of the $\mathbf{GG}^{\mathsf{T}}$ matrix.

As we increase the total time of acquisition t_f, the accuracy of guild detection improves for all four observables as measured by the AUC, as expected [Figs. 6(a)–6(d)]. For equal-time measurements $\mathbf{C}(0)$ and $|\mathbf{C}(0)|$ [Figs. 6(a) and

6(b)], the sampling interval $\Delta t_{\text{sampling}}$ does not change the prediction significantly. However, for the total absolute CPSD and the total coherence [Figs. 6(c) and 6(d)], the AUC for guild detection shows an optimal value near $\Delta t_{\text{sampling}} = 2$ and 0.5, respectively, for which the AUC for guild detection is largest for the biggest range of total data acquisition timescale. For comparison, in our simulation the species death rate is $d_x = 0.3$, corresponding to a turnover timescale of 3. In the limit of small t_f and large $\Delta t_{\text{sampling}}$, the number of time points is less than the number of time points required by the spectral method given the frequency resolution we specified (for the fast Fourier transform the number of data points needs to be at least double the number of the frequency resolution), which means the methods cannot work [bottom right corner of Fig. 6(d)]. Similar results hold for Spearman's correlation between the pairwise metrics and the $\mathbf{GG}^{\mathsf{T}}$ matrix element (Fig. S3 in Ref. [34]).

In summary, predictions always benefit from longer time traces, but there exists an optimal, finite sampling frequency for the two spectral methods which perform best.

F. Reconstructing guild structure from relative species abundances

When dealing with genomic data, often only relative abundances of the species rather than absolute abundances are available. We tested whether the CPSD and coherence observables remain useful for species structure inference from relative abundances, for systems of different sizes.

Figure S4 in Ref. [34] shows that, for small numbers of species ($N = 6$ and 24), total coherence has poor predictive power when using relative abundances, while the total magnitude of the CPSD still performs well. This result is surprising since coherence as a function of frequency has more structure than the CPSD. However, total coherence outperforms other measures for a large number of species ($N = 96$). This result is consistent with previous work [20], where the impact of relative abundance was reported to be less severe when the number of species is large.

G. Reconstruction of guild structure with heterogeneity in guild size and evenness

So far we have examined ECRMs where guild size is uniform across a community. Here we test whether the dynamical observables still perform well when there is heterogeneity in guild size or unevenness in species abundances.

Figure S5 in Ref. [34] shows that the ability to successfully detect links remains high for communities comprised of guilds of variable size. Coherence remains the best predictor among the four pairwise measures. Guild detection, however, becomes worse when the ECRM allows for the sharing of resources across different guilds ($q = 0.4$). Detectability depends on guild size: While link detectability is unaffected by guild structure, smaller guilds are harder to detect after single-linkage clustering.

Another source of heterogeneity is species evenness, i.e., certain species can be more abundant in the community. To allow variability in species abundances, especially across different guilds, we adjust the number of resources associated

with each guild. Figure S6 in Ref. [34] shows that the local guild detection is improved by species unevenness, especially if species in small guilds are present at high relative abundance. While local detectability is not monotonic with species evenness, higher relative abundance improves the AUC. Thus, a small guild (e.g., two species) in a community with larger guilds may be misidentified, unless its members are highly abundant.

H. Applying temporal metrics to marine data

To illustrate the usefulness of the CPSD and coherence on real data, we tested our four pairwise metrics on a public dataset consisting of time series of relative abundance of a total of 49 637 bacterial and eukaryotic operational taxonomic units (OTUs) from coastal plankton collected over 93 consecutive days [22]. For each of the 93 days we have three samples for each OTU, and we use the averaged relative abundance across the three samples as the relative abundance for each OTU. As the spectral methods have so far been developed considering consecutive time points sampled at regular intervals, we choose to analyze only the eukaryotic OTUs, because the bacterial data contain missing days across all OTUs. Furthermore, since our methods have been developed and tested on synthetic data without zero abundances, we limit our analysis to the $N = 31$ eukaryotic OTUs which are present on each of the 93 days. As the total number of eukaryotic OTUs is much larger than the considered subset, we expect that the total coherence and the total magnitude of the CPSD should not be adversely effected by the use of relative abundances.

We use the time series of the relative abundances of the $N = 31$ eukaryotic OTUs to compute the four pairwise metrics as described in the previous sections: the equal-time correlation $\mathbf{C}(0)$, its absolute value $|\mathbf{C}(0)|$, the total magnitude of the CPSD $\mathcal{P}$, and the total coherence $\mathcal{TC}$. Ideally, we would like to compare the pairwise metrics with the elements of the resource-utilization-overlap matrix $\mathbf{GG}^\mathsf{T}$. However, since there is no *a priori* measurement or knowledge of the guild structure in natural communities, we use phylogenetic distance as a proxy, following previous work [35]. The intuition is that the closer the species are genetically, the more likely they are to share the resources [36].

Comparing the pairwise metrics of the species abundances and the pairwise phylogenetic distance [Fig. 7(a)], all four pairwise metrics exhibit a weak negative correlation with the phylogenetic distance, indicating that more closely related taxa are more likely to share resource-utilization capabilities, which is consistent with the literature [35,37]. After binning the phylogenetic distances using regular intervals and computing the mean and average of the pairwise metrics in each bin, the trend in negative correlations becomes more visible [Fig. 7(b), colored dots]. To examine whether this negative correlation is statistically significant, we create a null model by destroying any phylogenetic correlation by randomly shuffling the indices of the OTUs before computing the pairwise metrics. The randomization is performed 10^5 times to collect statistics. As shown by the gray error bars in Fig. 7(c), the null model results in flat pairwise metrics for all metrics, as a function of the phylogenetic distance. At large phylogenetic distances, the data and the index-shuffled null model return

the same mean and variance, as indicated by the overlaying error bars in Fig. 7(c).

Among the four metrics, total coherence exhibits the biggest difference between the data and the null model for small phylogenetic distances, with a p value of 5×10^{-5} for the test of the mean in the first bin against the index-shuffled null model (see Sec. IV F for details). These results indicate that even for data collected experimentally in nature, total coherence works well in distinguishing structures in OTUs sharing common resources.

With the pairwise metrics, now we can infer guild structures among the $N = 31$ eukaryotic OTUs using single-linkage clustering with a tunable threshold on the metrics [Fig. S7(a) in Ref. [34]]. As there is no *a priori* knowledge of how those eukaryotic OTUs share resources, i.e., no ground truth to compare the guild structure to, we take the strategy of comparing the inferred structures across the four pairwise metrics. The threshold for each metric is selected such that the resulting number of clusters, including singletons, is the same for all four pairwise metrics ($k = 13$). The thresholded connections can be found in Fig. S7(b) in Ref. [34].

As shown by Fig. 7(c), different metrics yield different predicted guild structures. Clustering based on coherence gives two nontrivial clusters that exclusively contain OTUs of the same class, which does not occur for other metrics. The equal-time correlation $\mathbf{C}(0)$ and its absolute value $|\mathbf{C}(0)|$ return similar networks. Meanwhile, the total absolute CPSD gives a different network with one giant component and many singletons. This further suggests that the coherence metrics provide a useful predictor of species functional relationships.

III. DISCUSSION

We have shown that by judicious analysis of abundance dynamics data, one can reliably learn patterns of resource competition in microbial communities. Our success emerged from three key perspectives that distinguish this study from prior work.

First, rather than attempting to infer effective interactions between taxa such as those described by a Lotka-Volterra framework [38], we focus on resource-mediated interactions. Consumer resource models have been shown to quantitatively predict abundance and resource dynamics in communities [6,18]. Further, resource-explicit formalisms do not suffer from the ambiguities of effective interactions, especially in the context of microbial communities [39]. Moreover, the resource-centric picture of communities naturally motivates us to focus on coarse patterns of resource-utilization overlap, which are empirically supported by the existence of guilds in communities. Guilds allow us to simplify the problem from trying to infer every interaction in a community to inferring a lower-dimensional structure that aggregates strains together by resource preferences.

Within this dynamic resource-centric picture, our second key innovation was to move beyond simple correlations $\mathbf{C}(0)$ by employing spectral methods that utilize full temporal information. Our method stands in contrast to covariance-based approaches [19,20] that estimate covariance or inverse covariance matrices to infer associations but do not utilize dynamics. One exception is a recent study that employs temporal delays
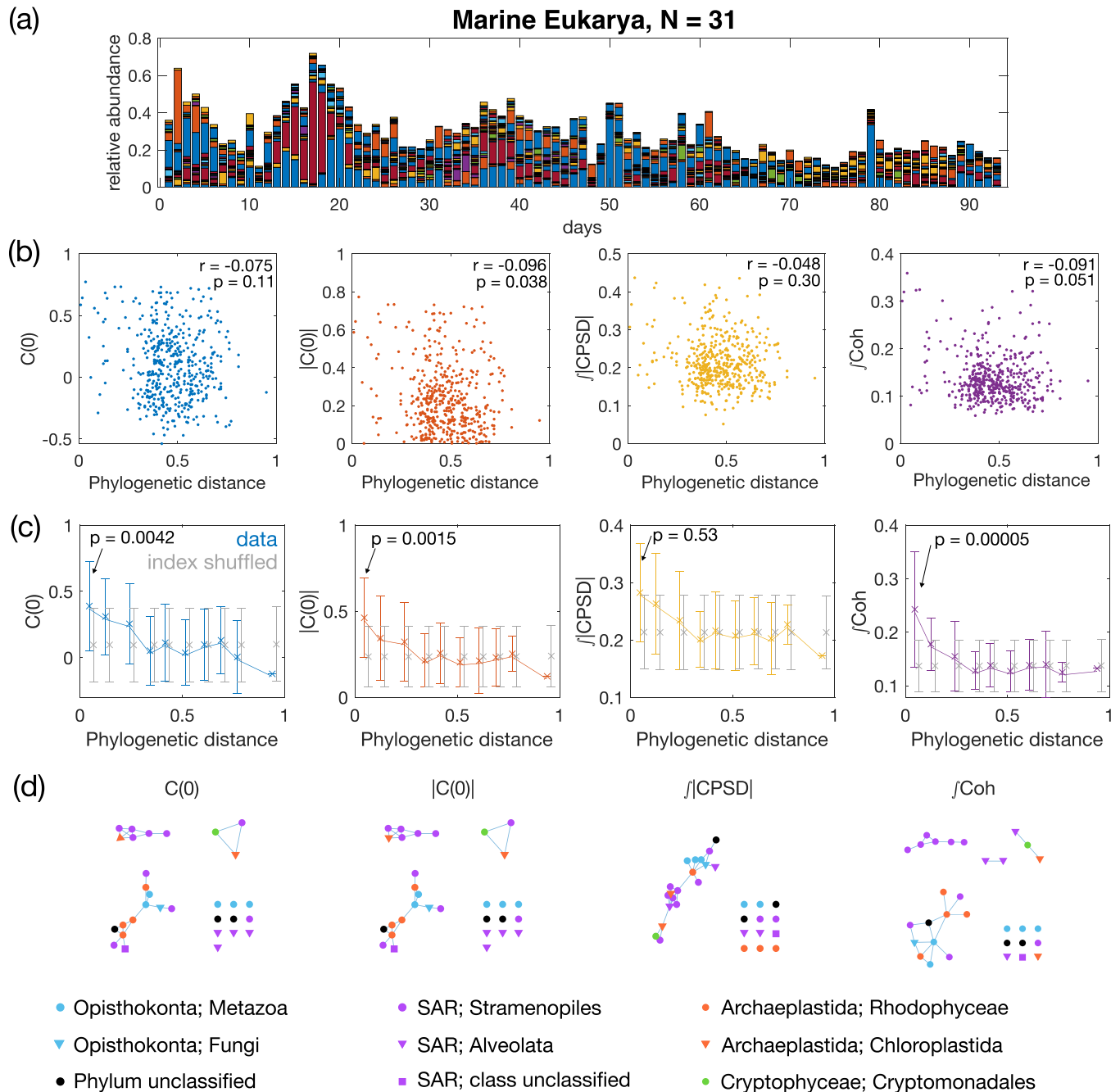
FIG. 7. Pairwise metrics of the relative abundances of marine eukaryotic OTUs correlate with phylogeny. (a) Relative abundance of the marine eukaryotic OTUs (downsampled to $N = 31$; see the text for details) over 93 days. Each color represents one of the 31 OTUs. Data are from [22]. (b) Scatterplot of pairwise metrics of relative abundance exhibiting a weak anticorrelation with the pairwise sequence distance (Jukes-Cantor distance). Pairwise measures include the equal-time correlation $\mathbf{C}(0)$ and its absolute value $|\mathbf{C}(0)|$, the total magnitude of the CPSD, and the total coherence. Spearman's rank correlation coefficient r and the p value are given in each panel. (c) Binned average of the pairwise metrics vs pairwise sequence distances. Binning is taken at regular intervals of the sequence distance. The mean across each bin is plotted with a cross symbol (connected with lines to guide the eye) and error bars give the standard deviation within each bin. Shown in gray are the results from a randomized control, where the order of indices of eukaryotic OTUs are shuffled randomly (number of shuffles equal to 10^5). The p value for the mean of the pairwise metrics in the first bin to be different from the index-shuffled null model is given in each panel (see Sec. IV F for details). (d) Resource-overlap networks inferred using the four pairwise metrics, with threshold set to produce $k = 13$ clusters including singletons. The phylum and class of each OTU is given by the color and the shape of the node, respectively, and in the legend with the format phylum;class.

and Granger causality to infer associations [40]. Similarly, spectral metrics (CPSD and coherence) provide a high-fidelity picture of resource overlaps in communities by extracting information from multiple temporal delays. We expect these methods will prove powerful for inferring the structure in a complex community where the acquisition of high-resolution time series has become commonplace.

Finally, our spectral methods not only embrace but rely on the inherently dynamic nature of the environment $R_\alpha(t)$. In contrast, time-series methods, including those based on Lotka-Volterra models [41], often assume steady-state dynamics [42,43] (with notable exceptions [24,25]). However, it is empirically clear that microbes live in environments that are inherently dynamic [14]. One of the most striking pieces of evidence of the importance of resource dynamics in the microbial world is recent physiological work showing that bacteria dynamically allocate cellular resources in a manner that optimizes growth under changing environments but not steady-state conditions [44,45]. Therefore, we regard the fact that our method relies on resource-driven abundance dynamics as an empirically motivated strength.

Using this approach, our analysis of eukaryotic microbial dynamics in a high-resolution marine dataset showed promising results in two domains. First, we observed that spectral methods outperform simple correlations in detecting statistically significant resource overlap between phylogenetically related taxa [Fig. 7(b)]. Second, it is enticing that these methods yield distinct network architectures from the same data [Fig. 7(c)]. An important avenue for future work is to more carefully vet the predicted associations from these metrics with synthetic communities.

Despite these successes, there are important avenues for improvement. First, as discussed above, at present, our method does not naturally deal with missing time points, uneven temporal sampling, or zeros in the data. It is an important avenue for future work to merge these methods with recent principled advances for handling zeros in amplicon sequencing data [46] or utilizing Lomb-Scargle periodogram methods for unevenly sampled data [47].

Finally, our work complements recent efforts to infer consumer-resource models or metabolic parameters directly from sequencing and metabolomic data [6,16,48]. The method of Goyal *et al.* [48] attempts to infer a much more detailed resource-exchange network by using genomic and metabolomic data to pinpoint actual cross-feeding interactions. An exciting avenue for future work would be to combine the sophisticated spectral methods developed here with more explicit genomic or metabolomic datasets.

IV. METHODS

A. Initialization of parameters of ECRMs

For all system sizes, the parameters of the ECRM systems are set as follows. The uptake rate for nonzero resource intake is set to $r = 0.1$. The yield $\gamma_{i\alpha}$ is given, drawn from a Gaussian distribution with mean 1 and standard deviation $1/6$. Redraws are performed to ensure all yields are non-negative. The death rate of the species is set to $d_x = 0.3$, while the depletion rate of the resource is set to $d_R = 0.5$. For sinusoidal drives, the

timescales of the environmental fluctuation ω_α are chosen randomly from a log-uniform distribution with a predetermined range to represent the mixture of multiple timescales in a real environment and to avoid resonance. The phase ϕ_α is chosen from a uniform random distribution between 0 and 2π. For OU environmental drives, due to their stochastic nature, we can choose the intrinsic frequencies ω_α to be the same for all resources α.

B. Finite-size scaling in ECRMs

In order to bridge the toy model with a small number of species and guilds with realistic ecological systems with large numbers of species and guilds, we scale the parameters of the ECRM given different system sizes. As we increase the number of species N, we want to also increase both the number of guilds k and the number of species within each guild N/k. Hence, we set the number of guilds $k \propto \sqrt{N}$. The number of resources should also increase as $M \propto N$. Setting the probability of species i to have a nonzero intake of its preferred resource as p_{in} and the probability of species i to have a nonzero intake of its nonpreferred resource as p_{out}, we can compute the expectation value of the average degree of each species in the effective species-species interaction due to resource-utilization overlap, the expectation value of $(\mathbf{G}\mathbf{G}^\mathsf{T})_{ij}$, etc.

Case 1: Species i and j belong to the same guild. The probability that there is no link between two species is equal to the probability that there is no common resource between the two, which is

$$P((\mathbf{G}\mathbf{G}^\mathsf{T})_{ij} = 0) = \left(1 - p_{\mathrm{in}}^2\right)^{M/k}\left(1 - p_{\mathrm{out}}^2\right)^{M(1-1/k)}. \quad (16)$$

Case 2: Species i and j do not belong to the same guild. The probability that there is no link is

$$P((\mathbf{G}\mathbf{G}^\mathsf{T})_{ij} = 0) = \left(1 - p_{\mathrm{in}}p_{\mathrm{out}}\right)^{2M/k}\left(1 - p_{\mathrm{out}}^2\right)^{M(1-2/k)}. \quad (17)$$

Given those two cases, the average degree for each node is

$$d = \left[1 - (1 - p_{\mathrm{in}}^2)^{M/k}(1 - p_{\mathrm{out}}^2)^{M(1-1/k)}\right]\left(\frac{N}{k} - 1\right)$$
$$+ \left[1 - (1 - p_{\mathrm{in}}p_{\mathrm{out}})^{2M/k}(1 - p_{\mathrm{out}}^2)^{M(1-2/k)}\right]N\left(1 - \frac{1}{k}\right). \quad (18)$$

For proper scaling, we want that, in the limit of large N, the average degree d converges to a constant value. The choice of $p_{\mathrm{in}} \propto N^{-1/2}$ and $p_{\mathrm{out}} \propto N^{-1}$ satisfies this condition, by ensuring that the average degrees towards in-guild and out-of-guild species are both finite. Notice that this scaling results in a $1/N$ scaling for both the expectation value and the variance of elements of the $\mathbf{G}\mathbf{G}^\mathsf{T}$ matrix.

C. Initiating guild structure with guild-size variability

To introduce more realistic heterogeneity in the ECRMs, we introduce a tuning parameter q_N, which sets geometrically spaced guild sizes $N_k = q_N^{k-1}N_1$ and $\sum N_k = N$. Furthermore, to allow variability in species abundance, especially across different guilds, we introduce another tuning parameter q_M, which accordingly adjusts the number of resources associated within each guild, again by setting a geometric series,

where $M_k = q_M^{k-1} M_1$ and $\sum M_k + N = M$. The N_k and M_k are rounded up to the nearest integer. If the sum of the rounded up number is smaller than the total number of species N, the missing species is added to the largest guild. If the sum of the rounded up number is greater than the total number of species N, the additional species is subtracted from the smallest guild, unless the smallest guild has only two species, in which case the additional species is subtracted from the second smallest guild such that the minimum size of each guild is 2.

To give intuition, if $q_N = 1$, all guild sizes are equal. The further q_N is from 1, the greater the differences in guild sizes. If $q_M = q_N$, then on average each species is exposed to the same number of resources and the species are even.

D. Local measure of detectability

We introduce a local measure of detectability. More specifically, for guild k, the true positive rate is the ratio between the number of detected adjacency links between all species present in guild k and the number of possible pairs, $N_k(N_k - 1)/2$. The guild-specific false positive rate is defined as the probability to falsely classify a link between a species i in guild k with a species j outside of guild k.

E. Measuring sequence distance in marine eukaryotic OTUs

To measure sequence distance, we first align the 18S eukaryotic rRNA sequences of the $N_{sub} = 31$ eukaryotic OTUs using the multiple alignment program for amino acid or nucleotide sequences [49]. Then the sequence distance is computed for each pair of OTUs, using the built-in MATLAB function seqpdist, namely, the Jukes-Cantor distance.

F. Statistical tests for binned pairwise metrics against the index-shuffled null model

For analyzing the structure in the marine eukaryotic OTU time-series data, we conduct statistical tests for the binned pairwise metrics against the index-shuffled null model. Specifically, we shuffle the index of species randomly and then compute the pairwise measurements. This null model preserves the set of all elements of the pairwise measures, but destroys any association with the species relationships.

The statistical test conducted in Fig. 7(b) checks whether the mean of the eight pairwise measurements in the first bin are significantly larger than in the index-shuffled randomization. We perform index shuffling 10^5 times. For each randomization, we compute the mean in the first bin. We compute the p value as the proportion of shuffling experiments that yield a larger mean in the first bin than the observed one.

DATA AVAILABILITY

Codes and data generating all figures in this paper are available from [50]. The coastal plankton dataset is from [22].

[1] S. D. Wilson and D. Tilman, Plant competition and resource availability in response to disturbance and fertilization, Ecology **74**, 599 (1993).

[2] K. R. Foster and T. Bell, Competition, not cooperation, dominates interactions among culturable microbial species, Curr. Biol. **22**, 1845 (2012).

[3] M. S. Datta, E. Sliwerska, J. Gore, M. F. Polz, and O. X. Cordero, Microbial interactions lead to rapid micro-scale successions on model marine particles, Nat. Commun. **7**, 11965 (2016).

[4] S. K. B. Shabat, G. Sasson, A. Doron-Faigenboim, T. Durman, S. Yaacoby, M. E. Berg Miller, B. A. White, N. Shterzer, and I. Mizrahi, Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants, ISME J. **10**, 2958 (2016).

[5] J. M. Tiedje, A. J. Sexstone, D. D. Myrold, and J. A. Robinson, Denitrification: Ecological niches, competition and survival, Antonie van Leeuwenhoek **48**, 569 (1982).

[6] K. Gowda, D. Ping, M. Mani, and S. Kuehn, Genomic structure predicts metabolite dynamics in microbial communities, Cell **185**, 530 (2022).

[7] S. Louca, L. W. Parfrey, and M. Doebeli, Decoupling function and taxonomy in the global ocean microbiome, Science **353**, 1272 (2016).

[8] M. B. Nelson, A. C. Martiny, and J. B. H. Martiny, Global biogeography of microbial nitrogen-cycling traits in soil, Proc. Natl. Acad. Sci. USA **113**, 8033 (2016).

[9] J. J. Kharbush, H. G. Close, B. A. S. Van Mooy *et al.*, Particulate organic carbon deconstructed: Molecular and chemical composition of particulate organic carbon in the ocean, Front. Mar. Sci. **7**, 518 (2020).

[10] M. Tramontano, S. Andrejev, M. Pruteanu *et al.*, Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies, Nat. Microbiol. **3**, 514 (2018).

[11] S. Louca, M. F. Polz, F. Mazel *et al.*, Function and functional redundancy in microbial systems, Nat. Ecol. Evol. **2**, 936 (2018).

[12] C. Staley, A. P. Ferrieri, M. M. Tfaily *et al.*, Diurnal cycling of rhizosphere bacterial communities is associated with shifts in carbon metabolism, Microbiome **5**, 65 (2017).

[13] A. Salari and J. Cremer, Diurnal variations in digestion and flow drive microbial dynamics in the gut, PRX Life **3**, 023012 (2025).

[14] H. F. Birch, The effect of soil drying on humus decomposition and nitrogen availability, Plant Soil **10**, 9 (1958).

[15] N. Klitgord and D. Segrè, Environments that induce synthetic microbial ecosystems, PLoS Comput. Biol. **6**, e1001002 (2010).

[16] Z. Li, A. Selim, and S. Kuehn, Statistical prediction of microbial metabolic traits from genomes, PLoS Comput. Biol. **19**, e1011705 (2023).

[17] R. MacArthur, Species packing and competitive equilibrium for many species, Theor. Popul. Biol. **1**, 1 (1970).

[18] P. Y. Ho, T. H. Nguyen, J. M. Sanchez, B. C. DeFelice, and K. C. Huang, Resource competition predicts assembly of gut bacterial communities *in vitro*, Nat. Microbiol. **9**, 1036 (2024).

[19] Z. D. Kurtz, C. L. Müller, E. R. Miraldi, D. R. Littman, M. J. Blase, and R. A. Bonneau, Sparse and compositionally robust inference of microbial ecological networks, PLoS Comput. Biol. **11**, e1004226 (2015).

[20] J. Friedman and E. J. Alm, Inferring correlation networks from genomic survey data, PLoS Comput. Biol. **8**, e1002687 (2012).

[21] B. M. Robicheau, J. Tolman, E. M. Bertrand, and J. LaRoche, Highly-resolved interannual phytoplankton community dynamics of the coastal Northwest Atlantic, ISME Commun. **2**, 38 (2022).

[22] A. M. Martin-Platero, B. Cleary, K. Kauffman, S. P. Preheim, D. J. McGillicuddy, E. J. Alm, and M. F. Polz, High resolution time series reveals cohesive but short-lived communities in coastal plankton, Nat. Commun. **9**, 266 (2018).

[23] A. Goyal, L. S. Bittleston, G. E. Leventhal, L. Lu, and O. X. Cordero, Interactions between strains govern the eco-evolutionary dynamics of microbial communities, eLife **11**, e74987 (2022).

[24] W. Cui, R. Marsland, and P. Mehta, Effect of resource dynamics on species packing in diverse ecosystems, Phys. Rev. Lett. **125**, 048101 (2020).

[25] R. Marsland, W. Cui, J. Goldford, and P. Mehta, The Community Simulator: A Python package for microbial ecology, PLoS One **15**, e0230430 (2020).

[26] M. Advani, G. Bunin, and P. Mehta, Statistical physics of community ecology: A cavity solution to MacArthur's consumer resource model, J. Stat. Mech. (2018) 033406.

[27] D. Liptzin, W. L. Silver, and M. Detto, Temporal dynamics in soil oxygen and greenhouse gases in two humid tropical forests, Ecosystems **14**, 171 (2011).

[28] K. Crocker, A. Skwara, R. Kannan, A. Murugan, and S. Kuehn, Microbial functional guilds respond cohesively to rapidly fluctuating environments, bioRxiv (2025), doi:10.1101/2025.01.30.635766.

[29] T. Platt and K. L. Denman, Spectral analysis in ecology, Annu. Rev. Ecol. Syst. **6**, 189 (1975).

[30] D. Ramírez, I. Santamaría, and L. Scharf, *Coherence: In Signal Processing and Machine Learning* (Springer, Cham, 2023).

[31] L. Barnett and A. K. Seth, The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference, J. Neurosci. Methods **223**, 50 (2014).

[32] X. Shan, A. Goyal, R. Gregor, and O. X. Cordero, Annotation-free discovery of functional groups in microbial communities, Nat. Ecol. Evol. **7**, 716 (2023).

[33] K. K. Lee *et al.*, Functional regimes define the response of the soil microbiome to environmental change, bioRxiv (2024), doi:10.1101/2024.03.15.584851.

[34] See Supplemental Material at http://link.aps.org/supplemental/10.1103/8gkj-rdzy for figures.

[35] M. Sireci, M. A. Muñoz, and J. Grilli, Environmental fluctuations explain the universal decay of species-abundance correlations with phylogenetic distance, Proc. Natl. Acad. Sci. USA **120**, e2217144120 (2023).

[36] A. C. Martiny, K. Treseder, and G. Pusch, Phylogenetic conservatism of functional traits in microorganisms, ISME J. **7**, 830 (2013).

[37] P. Lemos-Costa, Z. R. Miller, and S. Allesina, Phylogeny structures species' interactions in experimental ecological communities, Ecol. Lett. **27**, e14490 (2024).

[38] R. R. Stein, V. Bucci, N. C. Toussaint, C. G. Buffie, G. Rätsch, E. G. Pamer, C. Sander, and J. B. Xavier, Ecological modeling from time-series inference: Insight into dynamics and stability of intestinal microbiota, PLoS Comput. Biol. **9**, e1003388 (2013).

[39] B. Momeni, L. Xie, and W. Shou, Lotka-Volterra pairwise modeling fails to capture diverse pairwise microbial interactions, eLife **6**, e25051 (2017).

[40] K. Mainali, S. Bewick, B. Vecchio-Pagan, D. Karig, and W. F. Fagan, Detecting interaction networks in the human microbiome with conditional Granger causality, PLoS Comput. Biol. **15**, e1007037 (2019).

[41] C. K. Fisher and P. Mehta, Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression, PLoS One **9**, e102451 (2014).

[42] A. R. Ives, B. Dennis, K. L. Cottingham, and S. R. Carpenter, Estimating community stability and ecological interactions from time-series data, Ecol. Monogr. **73**, 301 (2003).

[43] D. S. Maynard, Z. R. Miller, and S. Allesina, Predicting coexistence in experimental ecological communities, Nat. Ecol. Evol. **4**, 91 (2020).

[44] G. Chure and J. Cremer, An optimal regulation of fluxes dictates microbial growth in and out of steady state, eLife **12**, e84878 (2023).

[45] R. Balakrishnan and J. Cremer, Conditionally unutilized proteins and their profound effects on growth and adaptation across microbial species, Curr. Opin. Microbiol. **75**, 102366 (2023).

[46] B. J. Callahan, P. J. McMurdie, and S. P. Holmes, Exact sequence variants should replace operational taxonomic units in marker-gene data analysis, ISME J. **11**, 2639 (2017).

[47] J. D. Scargle, Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data, Astrophys. J. **263**, 835 (1982).

[48] A. Goyal, T. Wang, V. Dubinkina, and S. Maslov, Ecology-guided prediction of cross-feeding interactions in the human gut microbiome, Nat. Commun. **12**, 1335 (2021).

[49] K. Katoh and D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability, Mol. Biol. Evol. **30**, 772 (2013).

[50] X. Chen, K. Crocker, S. Kuehn, A. M. Walczak, and T. Mora (Center for Open Science, Washington, DC, 2025), https://osf.io/hvnjt/.