# How sample heterogeneity can obscure the signal of microbial interactions

Running title: Heterogeneity obscures microbial interactions

**David W. Armitage***
Department of Biological Sciences
University of Notre Dame
Notre Dame, IN 46556

**Stuart E. Jones**
Department of Biological Sciences
University of Notre Dame
Notre Dame, IN 46556

January 14, 2019

**ABSTRACT**

Compositional data are commonly subjected to computational tools such as correlation networks, null models, and dynamic models, with the goal of identifying the ecological processes structuring microbial communities. Researchers applying these methods assume that the signs and magnitudes of species interactions and vital rates can be reliably parsed from compositional data. However, we contend that this assumption is violated when sample units contain any underlying spatial structure. Here, we show how three phenomena — Simpson's paradox, context-dependence, and nonlinear averaging — can lead to erroneous conclusions about population parameters and species interactions when samples contain heterogeneous mixtures of populations or communities. At the root of this issue is the fundamental mismatch between the spatial scales of species interactions (micrometres) and that of a typical marker gene sample (millimetres to centimetres). These issues can be overcome by measuring and accounting for spatial heterogeneity at very small scales, which will lead to more reliable inference of the ecological mechanisms structuring natural microbial communities.

## 1 Common "pattern-to-process" inferential methods yield troubling results

Advances in sequencing technology offer microbiologists unprecedented access to the composition and dynamics of microbial communities [1]. Marker gene and metagenomic surveys regularly chronicle hundreds to thousands of taxa, many previously unknown, all seemingly coexisting within their respective habitats. In possession of these large observational datasets, microbial ecologists have adapted theory and methods developed from plant and animal ecology to investigate how species interactions — such as competition, predation, and facilitation — structure microbial communities [2, 3].

In environmental marker gene surveys, negative covariation between microbial taxa (OTUs) is commonly attributed to the effects of interspecific competition. However, without experimental systems in which competition (or any other

* *Corresponding author*: `dave.armitage@gmail.com`

interaction) may be directly manipulated and detected, researchers often employ randomization-based null models, correlation networks, and population dynamic models to identify and quantify putative interspecific interactions from observational sequence data [4, 5, 6, 7]. However, the utility of these tools for reliably parsing and quantifying signals of competition from those of trophic interactions, dispersal limitation, and habitat filtering has been disputed for decades [8, 9]. Recently, a number of studies have challenged null model and correlation-based methods to recapitulate known interactions in well-studied marine intertidal habitats [10, 11, 12]. In all cases, these tests revealed troubling inaccuracies and discrepancies among the various methods, calling into question their ability to reliably identify true ecological interactions. For microbial communities, the only successful validations of these methods have occurred in simple, well-mixed liquid cultures [7]. Most natural microbial communities, however, are spatially structured and exhibit marked heterogeneity at multiple spatial scales. Failure to account for this underlying spatial heterogeneity in environmental samples can undermine our conclusions about the ecological processes structuring microbial assemblages [13].

## 2 Causes and consequences of heterogeneity in microbial samples

Typical sample volumes used for environmental marker gene and metagenomics studies are rarely smaller than 0.1 mL, but can be as large as 100 L of seawater and 100 g of soil in low-DNA habitats. Unless these samples come from a well-mixed, completely homogeneous medium, they will contain at least some small amount of spatial structure. For example, a typical 0.25 g sample of soil containing particles 1 mm in diameter (i.e., a very coarse sand) will inevitably contain hundreds to thousands of discrete granules on which microbial communities can assemble. These discrete habitats can represent a heterogeneous array of environments or resources, each selecting for their own unique local microbial communities [14]. However, even a physicochemically homogeneous collection of particles can contain a mosaic of distinct microbial communities owing to the effects of limited or asymmetric dispersal, priority effects, and successional turnover.

Fine-scale heterogeneity in microbial communities appears to be a general property of environmental samples, having been repeatedly documented in aquatic, soil, fecal, leaf surface, and wastewater habitats [14, 15, 16, 17, 18, 19]. Owing to this, marker gene samples commonly represent a sum of sequence reads made over underlying environmental heterogeneity. However, because microbial interactions such as resource competition, phage predation, DNA transfer, and syntrophy are hypothesized to take place at spatial scales much smaller than that of the typical bulk sample, it can be argued that many marker gene samples actually measure the *metacommunity* — a collection of semi-autonomous communities linked through dispersal [20]. In the following sections, we illustrate how sampling at the metacommunity scale can introduce biases or error into computational estimates of interspecific interactions by virtue of three phenomena: *Simpson's paradox*, *context-dependence*, and *nonlinear averaging*.

### 2.1 Simpson's paradox

Simpson's paradox refers to the reversal or negation of a statistical association between two variables, $X$ and $Y$, when conditioned on a third variable, $Z$ [21]. In ecology, this $Z$ variable might include information on spatial variation among local patches [22] or species identity [23], which, if accounted for, changes the direction of trends at the regional

or community scales. Computational approaches to inferring microbial interactions can be sensitive to the effects of Simpson's paradox. For instance, the inferred signs of interspecific correlation coefficients might change when comparing analytic results obtained from bulk community samples with results that have statistically accounted for underlying variation in microhabitats or resource availability within bulk samples.

To illustrate this point, consider a study employing SSU-rRNA sequences from bulk soil samples to infer the sign of interspecific interaction between two fungal taxa. If the true nature of this interaction is competitive, then our results are anticipated to reveal a negative correlation between the abundances of the two fungi. To add some realism to this scenario, let us assume that each of our samples represent collections of discrete microhabitats on which our focal taxa grow. Finally, we might also make the realistic assumption that both of our fungal taxa respond similarly to these discrete microhabitats such that sub-optimal habitats support fewer individuals of both species. If we populate bulk soil samples with random draws of simulated communities on each of three discrete microhabitat types (Fig. 1a), we find that even slight variation in the frequency distribution of these microhabitats within bulk samples leads to positive correlations between our two taxa, contradictory to their true, competitive local interactions. Furthermore, by repeating this experiment many times, each time re-assembling our bulk samples by populating them with equal numbers of randomly-selected discrete microhabitat particles, we encounter an overwhelming majority of cases where the inferred sign of interaction between our two taxa (positive) is the opposite of its true sign (negative) (Fig. 1b), leading us to erroneously conclude that these species are not strong competitors when, in truth, they are. Because of Simpson's paradox, we contend that unless the assumption of homogeneity within and among microbial community samples is justified, interspecific interaction coefficients derived from correlation or model-based approaches should be interpreted with extreme caution, and should always include a statement concerning the spatial context of the sample including potential sources of underlying spatial heterogeneity.

### 2.2 Context dependence

An common assumption of computational approaches for identifying species interactions is that the sign and strength of interactions are immutable across time and space. This assumption reduces the sample sizes required for estimating correlation coefficients or population parameters, and permits the use of graph theoretic descriptors of network structure (connectace, nestedness, etc.). However, interactions within and among microbial populations can markedly change with population density, community composition, or environmental context, such that interactions at one place and time cannot reliably be extrapolated to new habitats [24, 25, 26, 27, 28]. Context-dependent interactions are commonly associated with spatial heterogeneity [29], meaning that network or population dynamic inferences can be conditional on particular environmental or community states. For instance, it is not unreasonable to expect the sign of species interactions to change across a gradient of habitat quality (Fig. 2). While temporal correlation network approaches might be used to circumvent the static interactions assumption at larger spatial scales or in well-mixed samples, they cannot account for variable interactions arising from underlying spatial heterogeneity within individual samples.

From a theoretical perspective, context-dependence is hypothesized to be be a critical factor for maintaining diversity in spatially-structured communities [30]. For instance, the abilities of two competing microbial strains to coexist will be enhanced if the negative impacts of competition experienced by each strain are stronger in more

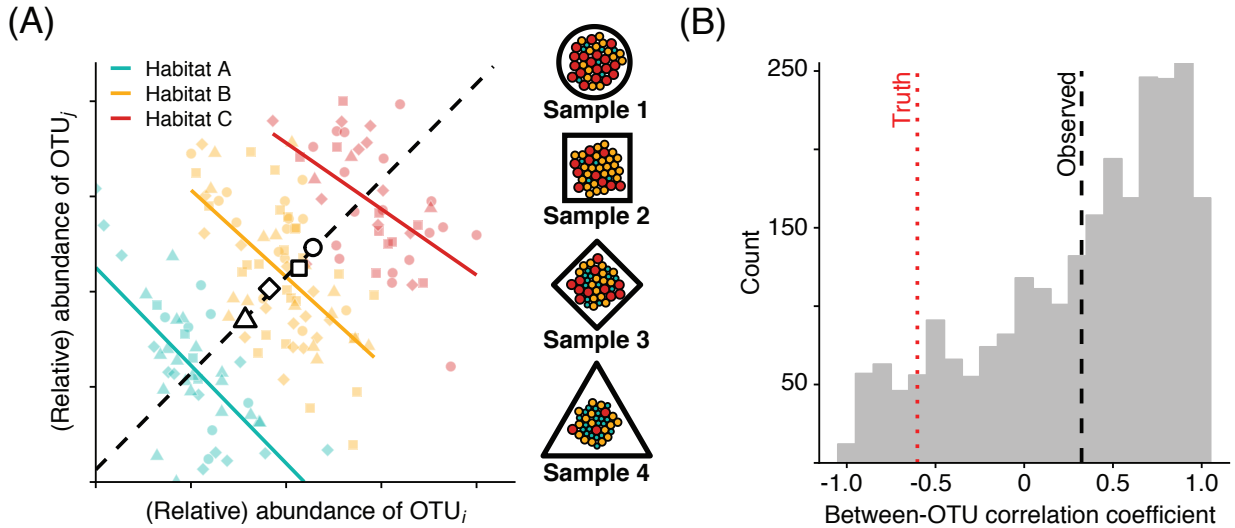


**Figure 1:** (A) Example of how Simpson's paradox can influence the identification of interspecific interactions. Colored points show the abundances of OTUs $i$ and $j$ in samples across three discrete microhabitats. Though the OTUs compete with one another in all three habitat types, their population responses to each habitat are correlated. When bulk samples containing any variation in microhabitat composition are sequenced (denoted by white points), the inferred sign of species interactions can be erroneous. (B) A simulation analysis of 2500 individual OTU correlations taken from samples consisting of 250 randomly-assembled individual particles reveals that the average inferred sign of interspecific interactions is positive, whereas the true sign of these interactions (simulated at the scale of individual particles) is negative.

favourable habitat patches [30]. Given that microbial species richness appears to peak in particulate, heterogeneous habitats (soil, sediments) [1], context-dependent interactions within these habitats may be quite common and important in promoting high levels of diversity. Currently, the extent of context-dependent interactions in spatially-structured microbial communities remains largely unknown. We note, though, that correlation network approaches have been successfully used to identify context-dependent interactions robust to experimental ground-truthing [31]. However, until the prevalence and magnitude of context-dependent microbial interactions are better understood, we encourage researchers to exercise caution when making general statements concerning any local estimates of interspecific interactions, ideally contextualizing results to the specific environment and scale at which measurements were taken.

### 2.3 Nonlinear averaging

In our final example, imagine that we are now able to obtain measurements of microbial populations at the scale of the individual particles comprising a heterogeneous sample. Such data could be obtained, for instance, using a fluorescence in situ hybridization (FISH) approach to directly count cell densities. More commonly, however, microbial time series are collected from colony counts and marker gene amplicon surveys. Regardless of their source, these temporal data can be used to fit models describing the temporal dynamics of one or more populations of OTUs. The generalized Lotka-Volterra (gLV) model is being increasingly employed for this purpose, and returns parameter estimates describing

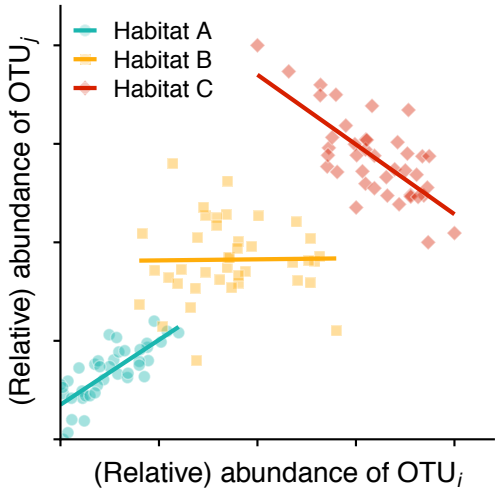


**Figure 2:** Illustration of context-dependent interactions between two microbial OTUs. Here, the sign of the interaction (positive, zero, and negative) is negatively correlated with habitat quality.

microbial species' growth rates and interspecific interactions. The gLV model commonly takes the form

$$\frac{dN_i}{dt} = N_i \left( \mu_i + \sum_{j=1}^{M} \alpha_{ij} N_j \right), \qquad i = 1, \ldots, M, \tag{1}$$

where $N_i$ is the abundance of OTU $i$, $\mu_i$ is its maximum *per capita* growth rate, and $\alpha_{ij}$ is a parameter describing the proportional change in its growth rate with conspecific or heterospecific densities. Values of $\alpha_{ij}$ greater than zero imply that OTU $j$ has a positive effect on OTU $i$, which might stem from interactions such as syntrophy, whereas values less than zero can signify interactions such as competition or chemical inhibition. For illustrative purposes, we will use a simplified gLV model containing a single OTU ($i = j$), and define $\alpha_{ij} = -\mu_i K_i^{-1}$, where $K_i$ represents the local carrying capacity of OTU $i$. This results in the familiar logistic model describing decelerating microbial population growth with increasing population density:

$$\frac{dN}{dt} = \mu N \left( 1 - \frac{N}{K} \right). \tag{2}$$

Importantly, this model describes population dynamics at the spatial scale over which intraspecific interactions occur, which, in a heterogeneous sample experiencing dispersal among particles, is at the scale of individual microhabitat patches. Called the *characteristic scale*, it is the scale which maximizes the ratio of deterministic signal to the influences of stochasticity and spatial heterogeneity [32], making it the optimal scale for characterizing deterministic species interactions using gLV models.

Researchers are often interested in "scaling up" microbial growth parameters or process rates estimated from data collected at local scales in order to compare them across different habitats or derive predictions for large spatial extents. There are two primary ways to estimate scaled-up population dynamics. First, one can fit dynamic models (such as the gLV) to arithmetic averages collected over multiple local habitats (such as particles) at particular time points. These population averages, measured as individual cells or sequence reads per unit area or volume, are a

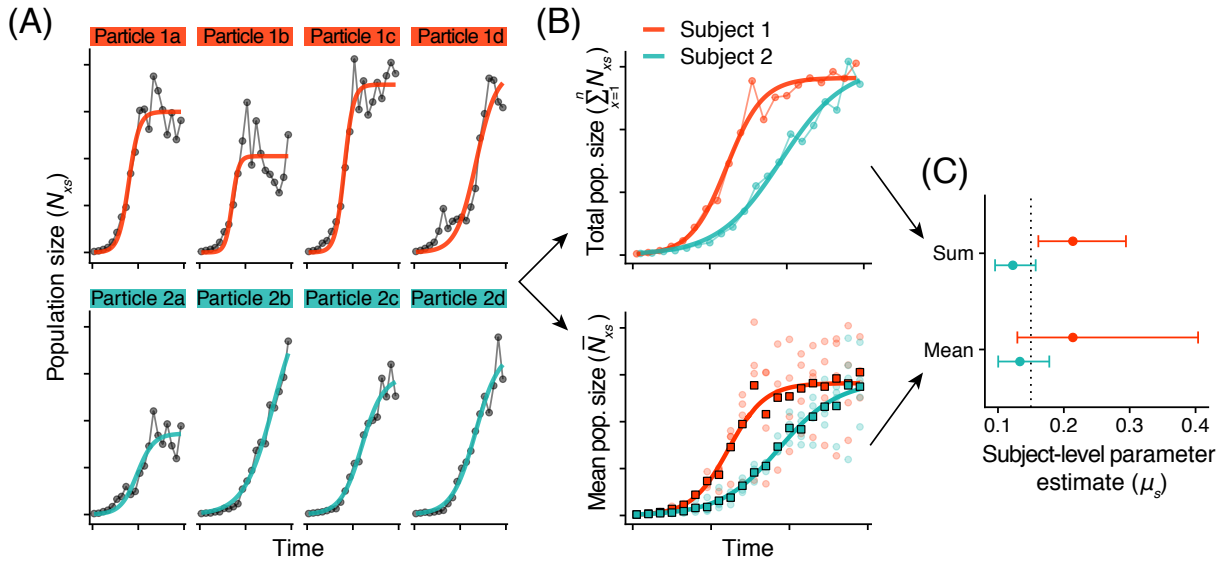


**Figure 3:** (A) Logistic growth curves fit to simulated population time series on independent particles assigned to two subjects. The growth parameters used to generate these populations are identical ($\mu = 0.15, K = 20$) and small amounts of stochastic noise were added to each time point. (B) Choosing to summarize these subjects using either total or mean population sizes results in disagreement over parameter estimates. (C) In the summation aggregation, the 95% confidence intervals around the estimated growth rates do not overlap between subjects, erroneously implying that these rates differ between subjects.

natural way to describe community dynamics in heterogeneous samples. More commonly, however, time series of entire, heterogeneous samples are collected, and the quantity being modeled becomes the sum of their constituent subpopulations. Aggregating our data using summation, however, masks underlying population variation, and thereby shrinks confidence intervals around parameter estimates. Because of this, researchers are at an increased risk of erroneously rejecting the null hypothesis that population vital rates are equal between samples (type I error). Small-scale variation in population dynamics could be a natural consequence of substrate heterogeneity or dispersal among local patches, which can both generate out-of-phase population dynamics and contribute deterministic noise to our measurements. For example, after simulating population time series from identical gLV process models containing small amounts of noise, we can encounter disagreement in population growth rates when comparing summed versus averaged populations (Fig. 3). In our example, the summation aggregation results in our incorrect rejection of the (true) null hypothesis that the population growth parameters are equal between hosts, whereas aggregation by averaging sub-populations correctly predicts the true population parameters. This issue highlights the need for improved methods for measuring within-sample or within-subject spatial variation in microbial community dynamics at their characteristic scales.

Averaging over nonlinear functions can also bias estimates of population dynamics when extrapolating across spatial scales. Because the average of a nonlinear population function is not equal to the function of population averages (i.e., $\overline{f(N)} \neq f(\overline{N})$), to scale up microbial population dynamics (which are almost unanimously nonlinear) by averaging across local populations will result in biases proportional to the magnitude of nonlinearity of the model — a principle known as *Jensen's inequality*. Moreover, spatial heterogeneity arising from asynchronous local populations can interact with and amplify the effects of nonlinear averaging, potentially biasing our estimates of population growth

parameters. This has important consequences for our ability to extract general properties and make predictions from dynamic microbial population models. To illustrate this point, we modify our logistic growth model (eq. 2) to describe a spatially heterogeneous mixture of $n$ sub-populations of an OTU at spatial locations $x$.

$$\frac{dN_x}{dt} = \mu N_x \left(1 - \frac{N_x}{K}\right), \qquad x = 1, \ldots, n. \tag{3}$$

Such a model might describe a population of individual axenic host organisms that have been experimentally inoculated with a microbial strain, and from which the strains' population time series have been obtained. We may be interested in drawing general conclusions concerning the differences in our strain's growth behaviors between host genotypes or experimental treatments, requiring representative population parameters to be averaged across our groups of interest. To proceed with our calculations, we need to account for the effects of averaging over spatial heterogeneity and nonlinear population dynamics.

Scale transition theory [33, 34] provides a recipe for scaling up population dynamics by accounting for the effects of nonlinear averaging over spatial heterogeneity. Below, we introduce the scale transition using two simple models, but refer interested readers to the original papers for general scale transition approaches [33, 34]. For notational simplicity, let us replace the growth function in equation 3, $\mu N_x(1 - N_x/K)$, with $G(N_x)$. We can obtain the scaled-up population dynamics by accounting for the nonlinearity in $G(N_x)$, proportional to its second derivative, $G''(N_x)$, as well as the spatial variation in $N_x$, measured by the spatial variance, $\mathrm{Var}(N)$. The full, spatially-averaged population dynamics can be approximated as

$$\begin{aligned} \frac{d\overline{N}}{dt} = \overline{G(N)} &\approx G(\overline{N}) + \frac{1}{2}\, G''(\overline{N})\mathrm{Var}(N) \\ &\approx \overline{g(N)}\,\overline{N} + \frac{1}{2}\, g''(\overline{N})\mathrm{Var}(N)\,\overline{N} + g'(\overline{N})\mathrm{Var}(N), \end{aligned} \tag{4}$$

where $\frac{1}{2}G''(\overline{N}) = g'(\overline{N}) = -\mu/K$. This approximation is exact when the growth function is quadratic (as is the case for logistic growth). The consequences of averaging population dynamics over spatial variation and nonlinearities is illustrated in Fig. 4a. We see that calculating our population dynamic model using the spatial population average, $G(\overline{N})$ overestimates the scaled-up population growth rate, $\overline{G(N)}$. Inclusion of the second part of the right hand side of equation 4 corrects for this. Furthermore, we see that a consequence of increasing the spatial variation of population densities in our samples (Fig. 4b), has the effect of reshaping our spatially-averaged growth function (Fig. 4c), such that our estimates of carrying capacity are inversely related to the spatial variation of our local populations.

A similar, albeit more complicated scale transition can be calculated for a multispecies gLV model (eq. 1) [33]. By defining $W_i = \sum_{j=1}^{M} \alpha_{ij} N_j$ and $g(W_i) = \mu_i + W_i$, the scaled up version of equation 1 can be written as a function of mean field terms, a nonlinearity term, and spatial variances and covariances:

$$\begin{aligned} \frac{d\overline{N}}{dt} &\approx \overline{g(W_i)}\,\overline{N_i} + \left[\frac{1}{2}\, g''(\overline{W_i})\mathrm{Var}(W_i) + g'(\overline{W_i})\mathrm{Cov}(W_i, v_i)\right]\overline{N_i} \\ &\approx \left(\mu_i + \sum_{j=1}^{M} \alpha_{ij}\overline{N_j}\right)\overline{N_i} - \sum_{j=1}^{M} \alpha_{ij}\mathrm{Cov}(N_i, N_j), \end{aligned} \tag{5}$$

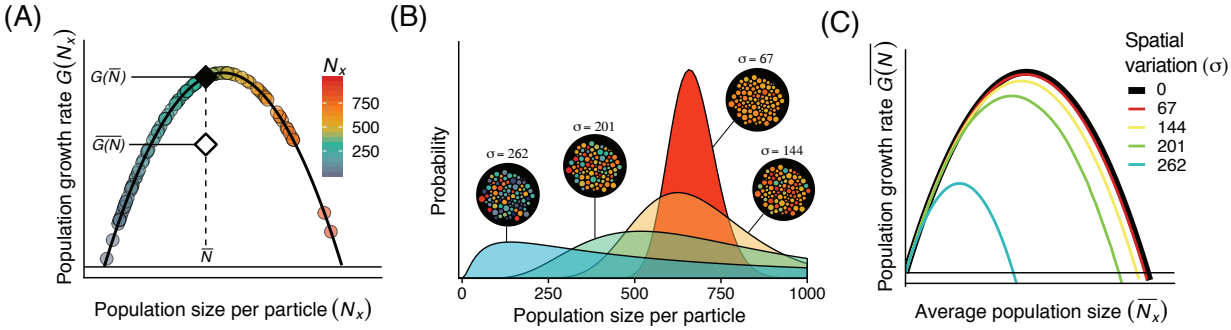


**Figure 4:** (A) Illustration of the effects of averaging over nonlinear, heterogeneous populations. Points lying along the concave-down growth function denote populations on individual particles within a sample, whose growth rates are governed by the logistic function $G(N_x)$ (the black curve). Note the difference between the scaled up growth function (white diamond) and growth function fit to average population sizes (black diamond). This difference can be corrected using scale transition theory. (B) We populated four samples (black circles) with particles whose population distributions follow truncated lognormal distributions ($\mu = 665, a = 0, b = 1000$) with various standard deviations, $\sigma$. (C) Increasing the standard deviations — and thereby the spatial variation — of populations within samples results in vastly different spatially-averaged population dynamics.

where $v_i = N_{i,x}/\overline{N_i}$. Once again, we see that the spatially-averaged population dynamics are not simply a function of average populations across space. However, the only extra information needed to calculate the scale transition are the spatial variances and covariances of the populations, which we can approximate by measuring local population densities across a sufficient number of particles comprising our sample. Thus, the calculation of scale transition terms is straightforward once they are defined for a particular dynamic model.

## 3 Recommendations moving forward

Despite the various ways in which spatial heterogeneity can subvert the interpretation of microbial community interactions and dynamics, we are optimistic that these issues can be surmounted with prudent data collection, analysis, and interpretation. The lurking effects of habitat heterogeneity are most effectively mitigated by quantifying microbial populations or communities at the spatial scales over which cell-cell interactions occur, which is on the scale of micrometers to millimeters. Sampling at this scale has successfully been accomplished using individual grains of sand [14], aquatic organic particles [35], and sludge granules [36] — all of which encountered marked heterogeneity among particles. Sampling at this scale is facilitated by technologies such as fluorescence-activated cell sorting and laser-assisted microdissection, which offer the opportunity to precisely and efficiently capture individual microscopic particles for sequencing.

The restrictive assumptions of most correlation network and null models hinder our reliable assessment of microbial interactions in all but the most homogeneous samples. While creative new approaches for identifying nonlinear and context-dependent species interactions are becoming available [37], we suggest these methods be ground-truthed with more complex and realistic data than are currently in use. For example, rather than using time series simulated from equilibrial Lotka-Volterra equations to ground truth a new method, a more powerful validation routine could use data simulated from spatially-explicit agent-based models, which can test methods' robustness to spatial heterogeneity,

scale-dependence, and demographic stochasticity. We also encourage the inclusion of dynamic parameters in generalised Lotka-Volterra models. While it is challenging to estimate these parameters from observational data, experiments consistently show that microbial growth rate, carrying capacity, and interaction parameters are functions of their underlying environments. A benefit of including environmentally-dependent growth parameters in gLV models is that such models can then be used to partition the effects of various coexistence-promoting mechanisms [30]. Environmentally-dependent parameters also allow us to investigate the effects of environmental change on microbial populations and communities.

Given the potential for biases and errors stemming from the joint effects of underlying spatiotemporal heterogeneity and methodological choices [9], it may seem like the inference of species interactions from observational microbial data represents an *underdetermination* problem. That is, there may be multiple, or even infinite potential mechanisms capable of generating an observed community pattern. However, this problem, like many in ecology and evolution, can more precisely be described as an example of *contrast failure* [38]. Instead of a solution-free, underdetermined system, we instead have one where our failure to parse competing hypotheses is a transient consequence of data insufficiency. Access to better, more *contrastive* data, derived either experimentally or observationally at the appropriate spatiotemporal scales, will refine our ability to discriminate among alternative hypotheses. In the meantime, we do not advocate for the abandonment of 'pattern-to-process' approaches for deciphering microbial interactions. On the contrary, we are optimistic about continued methodological development in this area. In the meantime, we implore researchers to consider and confront the lurking effects of spatial structure on their inferred microbial interaction networks and growth parameters. At minimum, this could simply comprise a comment on the spatiotemporal scale over which the results are anticipated to hold and a description of the spatial structure of the sample unit.

## References


[1] Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017;551:457–463.

[2] Prosser JI, Bohannan BJM, Curtis TP, Ellis RJ, Firestone MK, Freckleton RP, et al. The role of ecological theory in microbial ecology. Nature Reviews Microbiology. 2007;5:384–392.

[3] Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF, et al. Patterns and processes of microbial community assembly. Microbiol Mol Biol Rev. 2013;77:342–356.

[4] Faust K, Raes J. Microbial interactions: from networks to models. Nature Reviews Microbiology. 2012;10:538–550.

[5] Bálint M, Bahram M, Eren AM, Faust K, Fuhrman JA, Lindahl B, et al. Millions of reads, thousands of taxa: microbial community structure and associations analyzed via marker genes. FEMS Microbiology Reviews. 2016;40:686–700.

[6] Layeghifard M, Hwang DM, Guttman DS. Disentangling interactions in the microbiome: A network perspective. Trends in Microbiology. 2017;25:217–228.

[7] Xiao Y, Angulo MT, Friedman J, Waldor MK, Weiss ST, Liu YY. Mapping the ecological networks of microbial communities. Nature Communications. 2017;8:2042.

[8] Roughgarden J. Competition and theory in community ecology. The American Naturalist. 1983;122:583–601.

[9] Weiss S, Treuren WV, Lozupone C, Faust K, Friedman J, Deng Y, et al. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. The ISME Journal. 2016;10:1669–1681.

[10] Sander EL, Wootton JT, Allesina S. Ecological network inference from long-term presence-absence data. Scientific Reports. 2017;7:7154.

[11] Barner AK, Coblentz KE, Hacker SD, Menge BA. Fundamental contradictions among observational and experimental estimates of non-trophic species interactions. Ecology. 2018;99:557–566.

[12] Freilich MA, Wieters E, Broitman BR, Marquet PA, Navarrete SA. Species co-occurrence networks: Can they reveal trophic and non-trophic interactions in ecological communities? Ecology. 2018;99:690–699.

[13] Berry D, Widder S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Frontiers in Microbiology. 2014;5.

[14] Probandt D, Eickhorst T, Ellrott A, Amann R, Knittel K. Microbial life on a sand grain: from bulk sediment to single grains. The ISME Journal. 2018;12:623–633.

[15] Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF. Resource partitioning and sympatric differentiation among closely related bacterioplankton. Science. 2008;320:1081–1085.

[16] Sessitsch A, Weilharter A, Gerzabek MH, Kirchmann H, Kandeler E. Microbial population structures in soil particle size fractions of a long-term fertilizer field experiment. Appl Environ Microbiol. 2001;67:4215–4224.

[17] Swidsinski A, Loening–Baucke V, Verstraelen H, Osowska S, Doerffel Y. Biostructure of fecal microbiota in healthy subjects and patients with chronic idiopathic diarrhea. Gastroenterology. 2008;135:568–579.e2.

[18] Remus-Emsermann MNP, Tecon R, Kowalchuk GA, Leveau JHJ. Variation in local carrying capacity and the individual fate of bacterial colonizers in the phyllosphere. The ISME Journal. 2012;6:756–765.

[19] Gonzalez-Gil G, Holliger C. Aerobic granules: microbial landscape and architecture, stages, and practical implications. Applied and Environmental Microbiology. 2014;80:3433–3441.

[20] Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, et al. The metacommunity concept: a framework for multi-scale community ecology. Ecology Letters. 2004;7:601–613.

[21] Blyth CR. On Simpson's paradox and the sure-thing principle. Journal of the American Statistical Association. 1972;67:364–366.

[22] Scheiner SM, Cox SB, Willig MR, Mittelbach GG, Osenberg CW, Kaspari M. Species richness, species–area curves and Simpson's paradox. Evolutionary Ecology Research. 2000;2:791–802.

[23] Allison VJ, Goldberg DE. Species-level versus community-level patterns of mycorrhizal dependence on phosphorus: an example of Simpson's paradox. Functional Ecology. 2002;16:346–352.

[24] Muinck EJd, Stenseth NC, Sachse D, Roost Jv, Rønningen KS, Rudi K, et al. Context-dependent competition in a model gut bacterial community. PLOS ONE. 2013;8:e67210.

[25] Hoek TA, Axelrod K, Biancalani T, Yurtsev EA, Liu J, Gore J. Resource availability modulates the cooperative and competitive nature of a microbial cross-feeding mutualism. PLOS Biology. 2016;14:e1002540.

[26] Liu A, Archer AM, Biggs MB, Papin JA. Growth-altering microbial interactions are responsive to chemical context. PLOS ONE. 2017;12:e0164919.

[27] Tecon R, Ebrahimi A, Kleyer H, Levi SE, Or D. Cell-to-cell bacterial interactions promoted by drier conditions on soil surfaces. Proceedings of the National Academy of Sciences. 2018;115:9791–9796.

[28] Gould AL, Zhang V, Lamberti L, Jones EW, Obadia B, Gavryushkin A, et al. High-dimensional microbiome interactions shape host fitness. bioRxiv. 2018;:232959.

[29] Chamberlain SA, Bronstein JL, Rudgers JA. How context dependent are species interactions? Ecology Letters. 2014;17:881–890.

[30] Chesson P. General theory of competitive coexistence in spatially-varying environments. Theoretical Population Biology. 2000;58:211–237.

[31] Agler MT, Ruhe J, Kroll S, Morhenn C, Kim ST, Weigel D, et al. Microbial hub taxa link host and abiotic factors to plant microbiome variation. PLOS Biology. 2016;14:e1002352.

[32] Pascual M, Levin SA. From individuals to population densities: Searching for the intermediate scale of nontrivial determinism. Ecology. 1999;80:2225–2236.

[33] Chesson P, Donahue MJ, Melbourne BA, Sears ALW. Scale transition theory for understanding mechanisms in metacommunities. In: Holyoak M, Leibold MA, Holt RD, editors. Metacommunities: spatial dynamics and ecological communities. Chicago: University Of Chicago Press; 2005. 279–306.

[34] Chesson P. Scale transition theory: Its aims, motivations and predictions. Ecological Complexity. 2012;10:52–68.

[35] Bižić-Ionescu M, Ionescu D, Grossart HP. Organic particles: Heterogeneous hubs for microbial interactions in aquatic ecosystems. Frontiers in Microbiology. 2018;9.

[36] Kuroda K, Nobu MK, Mei R, Narihiro T, Bocher BTW, Yamaguchi T, et al. A single-granule-level approach reveals ecological heterogeneity in an upflow anaerobic sludge blanket reactor. PLOS ONE. 2016;11:e0167788.

[37] Harris DJ. Inferring species interactions from co-occurrence data with Markov networks. Ecology. 2016;97:3308–3314.

[38] Forber P. Spandrels and a pervasive problem of evidence. Biology & Philosophy. 2008;24:247.