

Low-Dimensional Coexistence in Complex Microbial Ecosystems

Karthik Srinivasan¹, Germán Plata², and Purushottam D. Dixit^{1,3,4,5}

¹*Department of Biomedical Engineering, Yale University, New Haven, Connecticut 06511, USA*

²*Computational Sciences, BiomEdit Inc. Indianapolis, Indiana 46202, USA*

³*Systems Biology Institute, Yale University, West Haven, Connecticut 06516, USA*

⁴*Quantitative Biology Institute, Yale University, New Haven, Connecticut 06511, USA*

⁵*Integrated Graduate Program in Physics, Engineering, and Biology, Yale University, New Haven, Connecticut 06511, USA*



(Received 13 November 2025; accepted 3 August 2026; published 15 September 2026)

The statistical physics view of microbial ecosystems often invokes random unstructured interactions, consistent with bacterial coexistence in very high-dimensional niche spaces. While this approach successfully reproduces broad community-level statistics, its central assumption of high-dimensional coexistence remains untested due to a lack of direct empirical estimates. Here we evaluate this assumption by inferring the effective niche dimensionality directly from abundance profile data using a joint species distribution model, a nonlinear low-rank matrix factorization approach. The inferred niche dimensionality, η_D , tracks ecological expectations: It correlates significantly with the dimensionality of the metabolic environment across paired 16S-metabolomics datasets, supporting the abundance-only inference. Applying this framework across nearly 200 human gut microbiome studies, we find that the inferred dimensionality is consistently low (median $\eta_D \approx 4$), markedly lower than median $\eta_D \approx 13$ predicted by diversity-matched disordered consumer-resource models, directly contradicting the high-dimensional assumption underlying current theoretical models. Low dimensionality was associated with stronger interspecies competition and reduced biodiversity, consistent with the long-standing macroecological predictions. Consumer-resource simulations reproduced these empirical relationships only when extrinsic resource availability trade-offs were imposed, with species-intrinsic metabolic trade-offs contributing an additional reduction in the effective dimensionality. Together, our results challenge the high-dimensional assumption underlying current physics-based descriptions of microbiomes and identify low-dimensional coexistence as an essential feature of host associated microbiomes.

DOI: [10.1103/rtht-q5dr](https://doi.org/10.1103/rtht-q5dr)

I. INTRODUCTION

Host-associated microbiomes are large, complex ecosystems comprising thousands of interacting taxa drawn from multiple kingdoms of life [1,2]. Despite extensive empirical and theoretical attention, the basic rules governing their assembly, stability, and function remain unsettled. A foundational and unresolved question is the effective dimensionality of these communities, that is, the number of independent ecological axes required to describe their coexistence.

Different theoretical traditions answer the questions of organization of complex systems such as microbiomes from varied perspectives. Statistical physics has long served as a productive lens for modeling complex biological systems including neural networks [3], bird flocks [4], and protein sequences [5]. It typically treats macroscopic order as emerging from a large collection of randomly interacting degrees of freedom. The high-dimensional perspective deeply influences modern theoretical ecology as well. Much like spin-glass models in other areas of biophysics, most statistical physics models of microbial communities treat them

as disordered systems using generalized Lotka-Volterra or consumer-resource models with random, unstructured interactions [6–15].

The macroscopic predictions of disordered, high-dimensional models (species abundance distributions, diversity-stability relations, and occupancy patterns) can often reproduce patterns observed in real microbiomes [8–12]. However, these are *indirect* tests: Macroscopic statistics can be reproduced by multiple models with very different interaction mechanisms. Notably, the central assumption of high dimensionality of coexistence itself has never been directly tested from empirical data.

We seek to infer a dimensionality of microbiomes that is biologically meaningful rather than a statistical compression of the data. Hutchinson's multidimensional niche theory [16] provides the natural framework. In Hutchinson's theory, a species' niche is the multidimensional set of environmental conditions that supports its persistence, and coexistence is promoted when species occupy distinct regions (niches) of this space, thereby minimizing *niche overlap* (interspecies competition) [17,18]. Although Hutchinson's original formulation of niches was qualitative [19], over the past several decades ecologists have converted the niche framework into a quantitative science, developing formal mathematical structures and empirical methods for estimating niche breadth and niche volume directly from observational data [18–22].

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

A central quantitative parameter in this framework is niche dimensionality (η_D from here onward), defined as the number of effective environmental axes governing community composition [23–25]. In macroecology, higher η_D reduces competition and facilitates biodiversity, while lower η_D constrains it [17,25–28]. Empirical studies additionally link η_D to community stability, invasibility, and evolutionary diversification [23,29–38]. Niche dimensionality is therefore a biologically meaningful observable tied to specific predictions about competition, diversity, and stability. Quantifying it directly from data turns those predictions into empirical tests.

Unfortunately, estimating niche dimensionality is not always straightforward because directly cataloging all relevant environmental axes is often experimentally infeasible. Microbial species’ responses to environmental variables are highly heterogeneous, and the relevant biochemical variables are typically correlated with one another, only partially measured, or entirely unknown. In other areas of theoretical ecology, this issue is resolved through inference of *effective* dimensionality from observed community structure, for example, food-web interaction matrices [39–42], pairwise competition networks [43], or trait-environment data [44].

Phenotyping bacterial taxa comprising microbial ecosystems is challenging. Consequently, community structures such as interaction matrices are often unavailable for microbial ecosystems. In contrast, community composition data (relative abundances of taxa) is relatively easy to obtain. We therefore developed a computational approach to infer η_D from community composition itself.

We used the joint species distribution modeling (JSDMs) from ecology. JSDMs jointly model species occurrences against measured environmental variables or, when such measurements are unavailable, against latent variables acting as effective stand-ins for the environment [45–48]. Specifically, we use a nonlinear latent-variable JSDM that we and others have derived directly from the mechanistic consumer-resource model [49–51]. In this model, the abundance profile is explained as a nonlinear low-rank matrix factorization, in which sample-specific latent factors play the role of effective resources and species-specific loadings the role of consumer preferences [Eq. (1)]. Latent-variable factorizations of this form have direct counterparts in other areas of the physics of complex biological systems, where high-dimensional observations often lie on low-dimensional manifolds parameterized by a small number of unobserved variables [52–54], for example, in models of neural population activity [55,56] and in latent-variable models of protein sequence ensembles [57].

We apply this approach to study several microbial ecosystems across different degrees of taxonomic resolution. In paired 16S rRNA–metabolomics data [58], η_D tracked the dimensionality of the metabolic environment, suggesting that abundance data encode niche structure. Applied to $\sim 10^5$ samples spanning nearly 200 gut microbiome studies [59], inferred η_D was consistently very low (median $\eta_D = 4$), well below the high-dimensional regime predicted by current microbiome theory under random, full-rank interactions. This central finding motivates rethinking microbial coexistence in terms of structured rather than unstructured interactions, in contrast to the assumptions underlying current theoretical models. Across studies, reduced η_D was associated with higher metabolic niche overlap (competition) and lower

biodiversity, consistent with the long-standing macroecological prediction that lower-dimensional niche spaces enforce stronger competition and constrain coexistence [32–34,60]. Consumer-resource simulations recapitulated these empirical patterns only when correlations among nutrient inflows (such as those induced by host diet) were imposed, with species-intrinsic metabolic trade-offs providing a smaller additional reduction in the effective degrees of freedom. Additionally, dimensionality was reduced in stressed or diet-simplified microbiomes and broadly covaried with ecological stress markers [61,62]. Together, these results indicate that microbial ecosystems are low dimensional, challenging the high-dimensional assumptions of current theoretical models.

II. RESULTS

We first introduce a latent variable JSDM [45–48] to infer η_D from relative abundance data. JSDMs operationalize niche theory empirically by relating species coexistence to environmental variables. When direct environmental measurements are unavailable, as in microbial ecosystems, latent-variable JSDMs replace these with latent variables that act as effective stand-ins for the environment.

We find that the inferred dimensionality mirrors environmental complexity and is consistent with macroecological expectations linking dimensionality, competition, and diversity. Next, we identify the mechanistic origins of low dimensionality using consumer-resource simulations with intrinsic (metabolic) and extrinsic (environmental) trade-offs. Finally, we find that host diet and health systematically modulate microbial niche dimensionality.

Before turning to the analysis, a note on terminology. Defining “species” in bacteria is intrinsically challenging [63]: The combination of horizontal gene transfer and asexual reproduction make the boundaries of bacterial species difficult to identify [64]. Consequently, the diverse microbial ecology datasets analyzed here operationalize taxa at different resolutions, ranging from amplicon sequence variants (ASVs) [65] that identify all unique sequences, to operational taxonomic units (OTUs) [66] clustered at fixed sequence-identity thresholds, to coarser genus-level groupings. Throughout this text, we use “taxon” as a unifying shorthand across these different definitions [67].

A. Latent variable JSDM quantifies niche dimensionality

To estimate microbial niche dimensionality from relative abundance data, we employ a latent-variable JSDM formulated as a nonlinear low-rank factorization [49–51] (see Appendix A):

$$x_{so} = \frac{1}{\Omega_s} \exp \left(\sum_{k=1}^K z_{sk} \theta_{ko} \right), \quad (1)$$

where x_{so} is the relative abundance of taxon o in sample s (Ω_s is the normalization factor), z_{sk} are sample-specific latent embeddings representing ecosystem-specific environmental variables, and θ_{ko} are species-specific loadings representing traits that couple to those variables. Notably, we and others have previously shown that this particular form of the JSDM can be mathematically derived from the mechanistic consumer-resource model, in which z_{sk} specifically

correspond to time averages of effective nutrient concentrations and θ_{ko} to consumer preferences [49–51].

Niche dimensionality η_D in this formulation is defined as the characteristic decay scale of reconstruction error with increasing latent dimension K (Appendix A). Conceptually, η_D does not enumerate specific environmental variables governing community composition but rather the number of *effective* dimensions that capture the dominant ecological constraints. Our approach extends to compositional data the inferential paradigm previously applied to food webs and competition networks [42,43].

We first assessed whether the abundance-inferred latent space corresponds to niche dimensions. Specifically, the JSMD identifies z_{sk} as ecosystem-specific environmental variables and θ_{ko} as species-specific traits that couple to those variables. If the inferred latent space recovers true niche axes, then (i) species loadings θ_{ko} should predict species-specific traits, and (ii) ecosystem embeddings z_{sk} should predict environmental measurements, even though both are inferred from relative abundance data alone.

Given that the metabolism of small molecules plays a dominant role in determining microbial community structure [8,11,68,69], we used metabolic phenotypes to support the ecological realism of our low-dimensional representation: metabolite-utilization profiles and enzyme content as species traits and measured metabolite concentrations as environmental variables. We used several paired 16S rRNA–metabolomics datasets [58] that simultaneously characterized community abundance and the metabolic environment for several samples.

To test prediction (i), we mapped bacterial taxa to approximately 7000 genome-scale gut microbial metabolic models [70] and inferred metabolite usage propensities r_{mo} via flux balance analysis (FBA) [71,72], where $m \in [1, 846]$ indexes specific metabolites. We set $r_{mo} = 1$ if organism o can use metabolite m for growth and zero otherwise (Appendixes B and C). We and others have previously shown that while FBA may not accurately capture the quantitative growth rates of taxa on specific metabolites, it can reasonably assess the binary question whether a taxon can use a particular metabolite for growth [72]. We fit logistic regression models predicting r_{mo} from θ_{ko} and analogously predicted genome-annotated enzyme content from θ_{ko} as a robustness check. To test prediction (ii), we fit linear regression models predicting measured metabolite concentrations directly from z_{sk} . The latent features θ_{ko} and z_{sk} are inferred from relative abundance data alone, while the validation targets r_{mo} and measured metabolite concentrations come from independent sources (FBA on genome-scale metabolic models, and mass spectrometry, respectively). Given these independent measurements, a predictive ability for both tests provides evidence that the latent representation carries a biological signal about the ecosystem as well as the species. In all analyses, models using our inferred embeddings consistently outperformed shuffled controls [Figs. 1(c) and 1(d) and Fig. 7], further supporting that abundance-based latent factors may capture biologically meaningful microbe–environment couplings.

Having established that the ecosystem embeddings z_{sk} predict measured metabolite concentrations [Fig. 1(d)], we

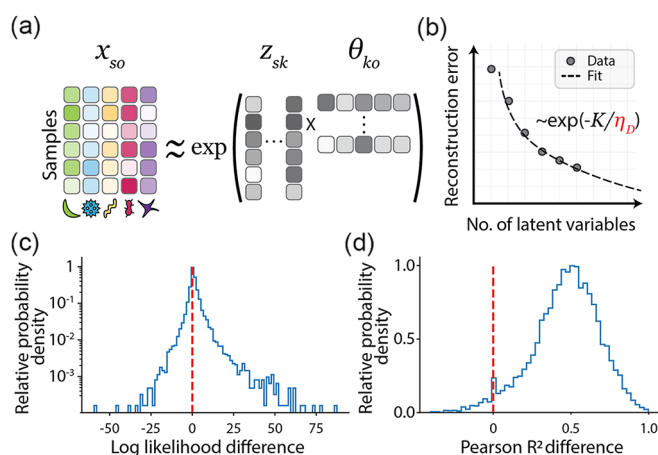


FIG. 1. Latent-variable approach to estimate niche dimensionality η_D . (a) Nonlinear low-rank factorization of relative abundances; z_{sk} are ecosystem-specific latent embeddings representing environmental variables (effective “resources” in the consumer–resource interpretation) and θ_{ko} are species-specific loadings representing traits (“preferences”). (b) η_D is extracted as the characteristic decay scale of the reconstruction error vs K curve. (c) Species loadings θ_{ko} predict imputed metabolite-utilization profiles (histogram of difference in log-likelihood vs shuffled). (d) Ecosystem embeddings z_{sk} predict measured metabolite concentrations (histogram of difference in Pearson R^2 vs shuffled).

asked whether individual niche axes could be assigned a concrete biological identity. Using the paired 16S–metabolomics data of He *et al.* [73], we ordered the inferred niche axes by the abundance variance they capture (via singular value decomposition) and, for each axis, used sparse regression to predict host projections on niche axes from measured metabolite concentrations (Appendix E, Fig. 8). Metabolite concentrations alone explained $\sim 50\%$ of the host-to-host variance along the leading axes, with explanatory power falling off sharply beyond the first few dimensions [Fig. 8(a)]. The metabolites that shaped the dominant niche axes were dominated by short-chain fatty acids (acetate, butyrate) and amino acids [Fig. 8(b)], i.e., the shared products of microbial cross-feeding. The leading niche axes therefore appear to correspond, at least in part, to a small number of biologically coherent metabolites rather than being purely abstract statistical constructs.

The JSMD-based factorization we use is essential for recovering this niche-dimensionality signal. Linear dimensionality reduction methods such as singular value decomposition (SVD) applied to log-transformed abundances, a common practice in microbiome analysis, failed to recover environmental dimensionality (Fig. 5). This is consistent with our prior work showing that multinomial-likelihood factorization is more robust to sequencing noise than SVD-type alternatives [50].

B. Estimated dimensionality reflects environmental complexity and follows the niche dimensionality hypothesis

Given the prominent role of metabolism in organizing microbial communities [8], we hypothesized that η_D inferred from taxon abundances alone should mirror the

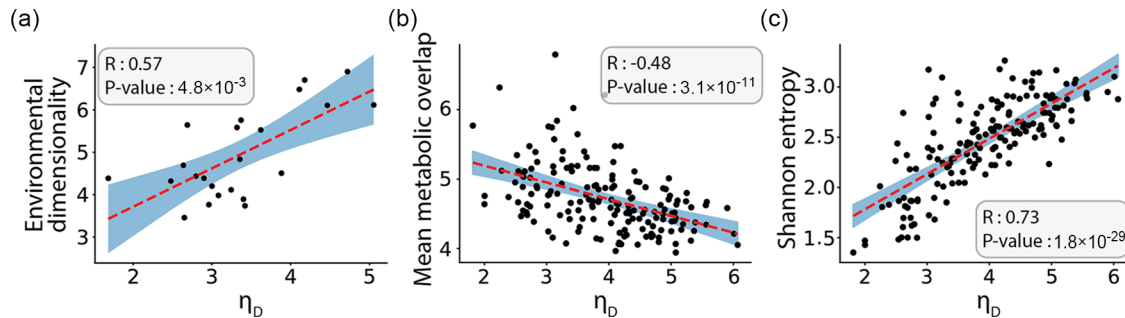


FIG. 2. Niche dimensionality captures environmental complexity and macroecological expectations. (a) η_D estimated from relative abundances correlates with the dimensionality of the metabolic environment. (b) η_D correlates negatively with metabolic niche overlap (competition). (c) η_D correlates positively with α diversity (Shannon index).

complexity of the metabolic environment. Indeed, in a large collection of paired human gut 16S rRNA–metabolomics datasets [58], η_D correlated significantly with the dimensionality of the metabolite concentration space [Spearman $r = 0.57$, $p = 4.8 \times 10^{-3}$; Fig. 2(a)], where environmental dimensionality was quantified by singular value decomposition (Appendix B). Notably, this correlation was significant regardless of the methodology used to compute dimensionality (Fig. 6). These two dimensionalities are derived from completely independent measurements: 16S rRNA gene sequencing (taxonomic abundances) and mass spectrometry (chemical concentrations). A statistically significant relationship across such distinct and noisy modalities is consistent with the inferred abundance-based axes tracking low-dimensional variation in the metabolic environment of the microbiomes.

Having established that η_D tracks environmental complexity, we applied the framework to $\sim 10^5$ samples spanning nearly 200 gut microbiome studies [59]. A critical concern when comparing dimensionalities across heterogeneous datasets is whether observed variations are driven by biological differences or by technical artifacts, such as variations in sequencing depth and number of samples. To mitigate these effects, we implemented a procedure where we standardized both the number of samples per study and the sequencing depth per sample across all datasets (Appendix B).

With these controls in place, η_D estimates (median 4.0, coefficient of variation 0.23) were very low and varied significantly across the datasets. We next tested whether the inferred η_D behaved as expected from macroecological theory. Since niche dimensionality limits niche overlap, lower η_D should coincide with stronger competition and reduced biodiversity [17,23,25–28]. Indeed, across these studies, metabolic niche overlap (imputed via FBA; Appendixes C and B) correlated negatively with η_D [Spearman $r = -0.48$, $p = 3.1 \times 10^{-11}$; Fig. 2(b)]. Conversely, η_D correlated positively with α diversity [Spearman $r = 0.73$, $p = 1.8 \times 10^{-29}$; Fig. 2(c)]. The negative association between η_D and niche overlap remained statistically significant even after controlling for α diversity [partial Spearman $r = -0.26$, $p = 6.8 \times 10^{-4}$; Fig. 3], further suggesting that the inferred niche dimensionality is not merely a statistical artifact but an ecologically meaningful characterization of the community.

C. trade-offs explain low dimensionality of microbiomes

What is the null expectation for the niche dimensionality of complex microbial communities? Conventional statistical-physics treatments simulate disordered consumer-resource communities tuned to reproduce diversity metrics, often treating reproduction of these metrics as sufficient. We adopted this approach as our baseline and asked whether such models recover the empirical η_D . We selected 50 random microbiome studies from the ~ 200 datasets in the Microbiomap collection [59]. The diversity metrics of this subset were statistically indistinguishable from the larger collection (Appendix G 5). We performed Markov chain Monte Carlo sampling in model hyperparameters (e.g., magnitude of the consumer preference matrix, outflow rates) to match empirical mean α diversity (Shannon and richness) and β diversity (average pairwise Jensen-Shannon divergence). While the simulated microbiomes reproduced diversity metrics with an average error of less than 10%, their median niche dimensionality was $\eta_D = 13.2$ [11.7, 14.9] (brackets show 25th and 75th percentiles), significantly higher than the range observed in real microbiomes (median $\eta_D = 4.0$ [3.2, 4.6]) [Fig. 3(b)]. This indicates that real microbiomes are substantially lower dimensional than the disordered picture, in direct tension with the central assumption in many theoretical studies of microbiomes.

How do we explain the low dimensionality of real microbiomes? The *niche dimensionality hypothesis* posits that trade-offs enable high diversity in low-dimensional niche spaces [32–34,60]. In microbiomes, trade-offs may arise from intrinsic constraints, including genomic streamlining (bias toward gene loss) [74] or physiological limitations (restricted capacity to express multiple metabolic pathways) [75]. Following previous work [76–79], we modeled *intrinsic* metabolic trade-offs by limiting the ability of species to excel at utilizing multiple resources towards growth (Appendix G). In simulations tuned to reproduce empirical diversity metrics, intrinsic trade-offs enabled species to coexist in lower-dimensional niches while preserving richness [Fig. 3(b)] (median $\eta_D = 9.5$ [6.4, 13.4]). However, simulated niche dimensionalities remained significantly higher than those inferred from real microbiomes.

A second, often overlooked class of trade-offs relevant to gut microbiomes are *extrinsic* ones imposed by the

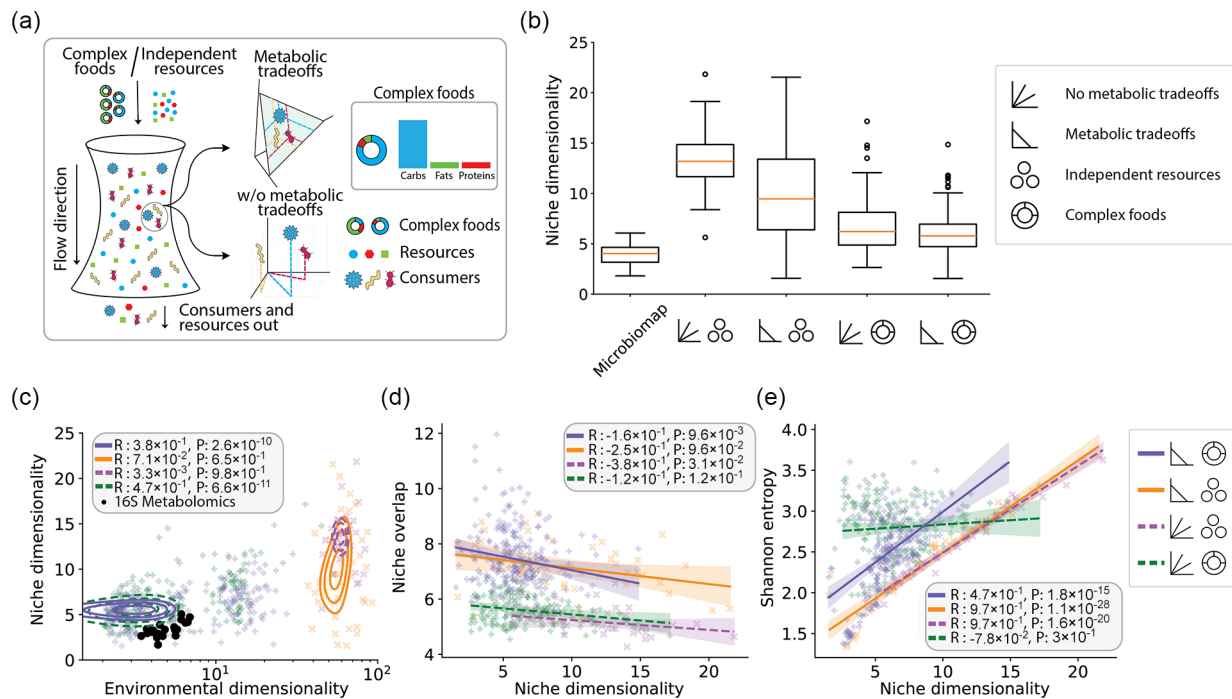


FIG. 3. Trade-offs generate low dimensionality and empirical covariation. (a) Consumer-resource model (CRM) schematic and simulation strategy. (b) Box plot of η_D across data and models: The Microbiomap collection [59], shuffled data, baseline models reproducing diversity metrics, models with intrinsic metabolic trade-offs, models with extrinsic environmental trade-offs, and models with both trade-offs, respectively (brackets represent 25th–75th percentiles). [(c)–(e)] Only the combined-trade-off model recapitulates observed correlations between η_D and (c) environmental complexity, (d) metabolic niche overlap, and (e) α diversity.

environment, such as host diet. Humans consume nutrients in the form of food groups (e.g., starches, meats, and fruits) under roughly constant total calorie intake, so greater consumption of one group reduces intake of others [80,81]. This covariation generates correlations among metabolite concentrations that limit microbial access to independent nutrient axes. In the context of classical niche theory, correlated environments reduce the dimensionality of species niche overlap (see Appendix H). To that end, we investigated the effects of environmental trade-offs by simulating consumer-resource models (CRMs) in which nutrients were supplied as correlated “food groups” [Appendix G, Fig. 3(a)]. Analogous to the intrinsic metabolic trade-offs, the relative abundances of individual food groups were constrained to a simplex, representing a constant total nutrient intake. Incorporating extrinsic environmental trade-offs markedly reduced the inferred dimensionality (median $\eta_D = 6.2$ [4.9, 8.1]). Combining both intrinsic and extrinsic trade-offs further lowered η_D (median $\eta_D = 5.8$ [4.7, 7.0]). Crucially, only models with *both* trade-offs reproduced the full empirical pattern set: η_D increased with environmental dimensionality, decreased with metabolic niche overlap, and increased with α diversity [Figs. 3(c)–3(e)]. The positive coupling between environmental complexity and η_D emerged *only* when environmental trade-offs were present.

Conceptually, these two types of trade-offs play distinct but complementary roles. Environmental (extrinsic) trade-offs are conceptually transparent: Correlations among resource availabilities literally shrink the number of independent ecological axes, directly reducing the effective dimensionality of the niche space. In contrast, a classical result in ecol-

ogy demonstrates that intrinsic trade-offs among species’ resource preferences are required to maintain species diversity within low-dimensional niches [60]. In our framework, intrinsic trade-offs therefore make coexistence in low-dimensional niche spaces possible at all, while extrinsic trade-offs explain why those niche spaces are low dimensional in the first place. Mathematically, both trade-offs reduce the effective rank of the species-environment coupling matrix: Intrinsic trade-offs constrain species preferences to a low-dimensional simplex, while extrinsic trade-offs introduce correlations among the environmental axes themselves. Importantly, this unified picture is fully consistent with both the classical niche dimensionality hypothesis and our empirical observations: Even with intrinsic trade-offs enabling coexistence, lower dimensionality remains strongly associated with lower observed biodiversity [Fig. 2(c)].

D. Host diet and health modulate microbial niche dimensionality

Because niche dimensionality reflects the complexity of the microbial environment, we hypothesized that host-level factors such as diet and health [82] should systematically influence η_D .

To directly test dietary effects on niche dimensionality (Appendix B), we analyzed longitudinal microbiome data from mice fed either a high-fat, high-sugar (HFHS) diet or a low-fat, plant-polysaccharide (LFPP) diet [83]. The HFHS diet provides a narrow set of rapidly digestible nutrients, whereas the LFPP diet supplies a diverse pool of

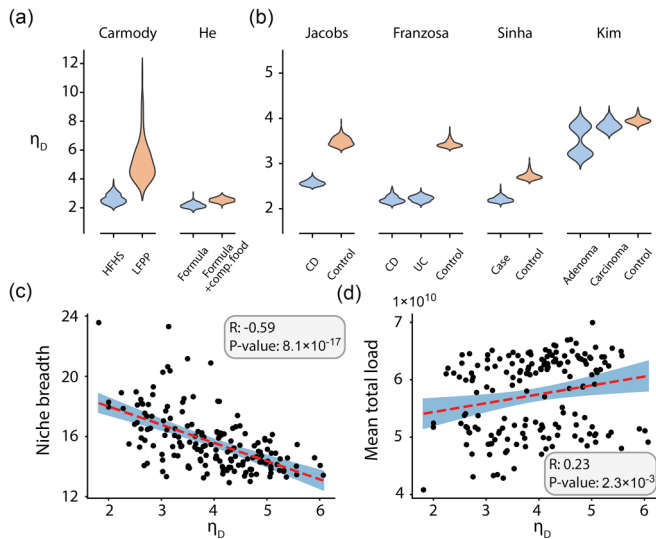


FIG. 4. Niche dimensionality tracks diet complexity and health. (a) Longitudinal mouse microbiomes: Higher η_D under LFPP (diverse) vs HFHS (narrow) diets; human infants: higher η_D with complementary foods vs formula-only. (b) Higher η_D in healthy vs inflammatory bowel disease or colorectal cancer. (c) η_D correlates negatively with niche breadth (generalists' prevalence). (d) η_D correlates positively with predicted total microbial load.

fermentable polysaccharides that yield a chemically rich oligosaccharide environment. Consistent with this difference in nutrient complexity, LFPP-fed mice exhibited significantly higher η_D (Carmody [83], Mann-Whitney U test p value: 5.8×10^{-35}) than HFHS-fed mice [Fig. 4(a)]. Controlled human dietary studies show similar effects: Animal-based diets reduce metabolomic complexity, whereas plant-rich diets expand it [83,84]. In agreement, cross-sectional analysis of human infant microbiomes [73] [Fig. 4(a)] revealed that infants consuming complementary plant-based foods in addition to formula had higher η_D than infants fed exclusively on formula (He [73], Mann-Whitney U test p value: 7.3×10^{-79}). Together, these results indicate that dietary complexity is associated with an increase in the dimensionality of microbial niche space.

Health status showed similar modulation. Across cross-sectional studies of inflammatory bowel disease [85,86] and colorectal cancer [87,88], healthy controls consistently exhibited higher η_D than patients [Fig. 4(b)] (Mann-Whitney U test p values, Jacobs [85]: 1.1×10^{-99} ; Franzosa [86]: 1.1×10^{-99} in both cases; Sinha [87]: 1.1×10^{-99} ; Kim [88]: 1.2×10^{-34} for control versus adenoma and 7.3×10^{-78} for control versus carcinoma). More broadly, dimensionality covaried with indicators of stress: It correlated negatively with niche breadth (mean number of resources consumed by bacteria) [61] [Spearman $r = -0.59$, $p = 2.4 \times 10^{-17}$; Fig. 4(c)] and positively with total microbial load [62], estimated using a recently developed machine learning model [89] [Spearman $r = 0.23$, $p = 2.3 \times 10^{-3}$; Fig. 4(d)] (Appendixes C and B). Together, these analyses identify reduced η_D as a quantitative signature of stressed or diet-simplified microbiomes and suggest that host-level factors systematically shape the effective ecological complexity experienced by gut communities.

III. DISCUSSION

A central challenge in ecology is to explain the origins of biodiversity. Within niche theory, two contrasting coexistence regimes are often invoked [90,91]: ordered coexistence maintained in low-dimensional niche spaces [32,60] and coexistence emerging from very high-dimensional interactions [92,93]. The former predicts a direct link among dimensionality, competition, and biodiversity, the so-called *niche dimensionality hypothesis* [32–34], while the latter has dominated physicists' view of microbiomes [6–15].

Here we set out to empirically test the assumption of high-dimensional coexistence in microbial ecosystems. We used the niche dimensionality η_D as our metric and developed a nonlinear low-rank latent-variable approach to infer it directly from relative abundance data. The inferred η_D was consistent with long-standing ecological expectations: It tracked the dimensionality of the metabolic environment and correlated negatively with metabolic niche overlap and positively with α diversity. Across ~ 200 gut microbiome studies, η_D was consistently low (median ≈ 4), well below the value $\eta_D \approx 13$ predicted by diversity-matched disordered consumer-resource models. This contradicts the high-dimensional assumption underlying current theoretical models of microbiomes [6–15]. Consumer-resource simulations recapitulated the empirical low dimensionality and its full pattern of couplings only when both intrinsic metabolic trade-offs and extrinsic environmental trade-offs (correlated nutrient supplies, as imposed by host diet) were imposed.

Although intrinsic and extrinsic trade-offs together reduced the simulated η_D from 13.2 to 5.8, a small residual gap remained relative to empirical microbiomes (median $\eta_D \approx 4$), suggesting that additional mechanisms further compress the effective dimensionality. Metabolic cross-feeding and syntrophy couple the production of secondary metabolites to the consumption of primary resources, generating highly correlated nutrient pools [68,94–96]. Phylogenetic conservation of traits [97] may potentially reduce the effective dimensionality of competition. Coupled abiotic variables such as pH and temperature [98] introduce further correlations across environmental axes. Finally, gut microbiomes experience dynamic feast-famine cycles [99] driven by episodic nutrient intake [100]; out-of-steady-state phenomena, correlated cell death during famine, and synchronized growth during nutrient pulses [95,101] can also act as temporal bottlenecks that further collapse the effective dimensionality. Disentangling these contributions empirically is a natural direction for future work.

Our findings challenge a broad class of models that has dominated theoretical microbial ecology. Many current frameworks [8,11–15] reproduce community-level statistics by drawing interactions or consumer preferences independent and identically distributed from a fixed distribution, effectively assuming extremely high-dimensional niches [90,91]. Consistent with this, our randomly initialized consumer-resource models captured species abundance distributions but failed to reproduce the low η_D observed empirically. Future theoretical descriptions must therefore explicitly incorporate dimensionality constraints, low-rank or correlated structure in resource supplies for consumer-resource models,

and low-rank interaction matrices for generalized Lotka-Volterra systems [27–29,90,91]. Notably, empirical analyses in plants [43] and microbes [102] likewise reveal structured competition matrices, suggesting that dimensional constraints on coexistence represent a general ecological principle rather than a microbial peculiarity.

Powerful analytical tools such as dynamic mean-field theory [103,104] have illuminated the typical dynamics of disordered consumer-resource systems by mapping an N -species community onto an effective single-species description. This approach is well suited to species-level quantities but does not directly address collective community-level properties such as niche dimensionality, which depend on how coexistence is organized across correlated environmental axes shared by species. Extending such methods to capture collective structure, as has recently been done in neuroscience [105], is an open and important theoretical direction.

These findings also motivate experimental tests. Targeted measurement of metabolic trade-offs in environmentally relevant taxa, characterization of correlated environmental variables in host-associated environments, and identification of structured competition matrices in defined synthetic communities [43,102] would directly probe the dimensional constraints we identify here.

More broadly, the low dimensionality we observe in microbial communities is part of an emerging pattern across high-dimensional biophysics. Efforts to coarse-grain microbial community composition, whether by linking it to aggregate functions [106–108] or by capturing its fluctuations through a small number of dominant modes [109,110], reflect a growing recognition that ecological complexity often collapses onto a few effective axes. Comparable dimensional compression has been documented across biological scales, from correlated neural population activity to low-dimensional protein mutational landscapes [54,55]. Notably, our own work has shown that the consumer-resource model can be recast as a latent-variable system [50,51], suggesting that low dimensionality may represent a unifying organizational principle of complex biological systems [53].

In summary, the assumption that microbial coexistence is governed by effectively infinite, independent ecological axes is not borne out empirically. Both observations and mechanistic simulations indicate that microbial communities are organized along a small number of effective dimensions, shaped by intrinsic metabolic trade-offs and extrinsic environmental correlations. By inferring niche dimensionality directly from abundance data, our framework provides a quantitative bridge between statistical-physics models of microbiomes and the long-standing traditions of niche theory in macroecology.

ACKNOWLEDGMENTS

This work was supported by the Bezos Earth Fund through the AI for Climate and Nature Grand Challenge and by the Gates Foundation Grant INV-062437 to BiomEdit. P.D.D. Acknowledges NIH Grant R35GM142547.

DATA AVAILABILITY

The data that support the findings of this article are openly available in previously published articles.

APPENDIX A: JOINT SPECIES DISTRIBUTION MODEL

Traditional JSDMs model species abundances as functions of measured environmental covariates [45–48], capturing the portion of ecological variation explained by known or measured factors. In contrast, their latent-variable extensions [48–51] invoke unobserved environmental axes that capture shared, unmeasured, or unknown environmental influences shaping community composition.

Here we use a latent variable JSDM that can be derived from the consumer-resource model [49–51]. Specifically, the relative abundance of taxa are modeled as a nonlinear low-rank matrix factorization:

$$x_{so} = \frac{1}{\Omega_s} \exp \left(\sum_k z_{sk} \theta_{ko} \right), \quad (\text{A1})$$

where x_{so} is the relative abundance of taxon o in sample s , z_{sk} are sample-specific latent embeddings (latent “resources”), and θ_{ko} are species-specific latent embeddings (latent preferences).

1. Fitting the model to data

We consider that we have measured abundances n_{so} of species $o \in [1, O]$ in samples $s \in [1, S]$. We assume that the total read count $N = \sum_o n_{so}$ is constant across all samples. The multinomial likelihood of the data $\mathbf{n}$ given the model parameters $\mathbf{z}$ and $\boldsymbol{\theta}$ is given by

$$p(\mathbf{n}|\mathbf{z}, \boldsymbol{\theta}) = \prod_s \frac{N!}{\prod_o n_{so}!} \prod_o \frac{1}{\Omega_s} \exp \left(- \sum_k z_{sk} \theta_{ko} \right). \quad (\text{A2})$$

The log-likelihood $\mathcal{L}$ (up to an additive and a multiplicative constant) is given by

$$\mathcal{L} = \sum q_{so} \log x_{so}, \quad (\text{A3})$$

where $q_{so} = n_{so}/N$ are the observed relative abundances. We obtain the latent embeddings using a maximum likelihood approach. The gradients are given by

$$\frac{\partial \mathcal{L}}{\partial z_{sk}} = \sum_o (q_{so} - x_{so}) \theta_{ko}, \quad (\text{A4})$$

$$\frac{\partial \mathcal{L}}{\partial \theta_{ko}} = \sum_k (q_{so} - x_{so}) z_{sk}. \quad (\text{A5})$$

We fit the model to the data using simple gradient descent. The learning procedure is stopped when the sum of the magnitudes of the relative gradients falls below a threshold, $\frac{|\partial \mathcal{L}/\partial \mathbf{z}|}{|\mathbf{z}|} + \frac{|\partial \mathcal{L}/\partial \boldsymbol{\theta}|}{|\boldsymbol{\theta}|} \leq 10^{-2}$.

2. Estimating niche dimensionality

Niche dimensionality was calculated based on the rate of decrease of the Kullback-Leibler (KL) divergence between data and fit (error function for the gradient descent) with increasing dimension of the latent space (K). To capture the rate of decrease, KL divergence for $K = 1$ to 5 was fit with an exponential function:

$$\text{KL}(k) = A \exp(-k/\eta_D). \quad (\text{A6})$$

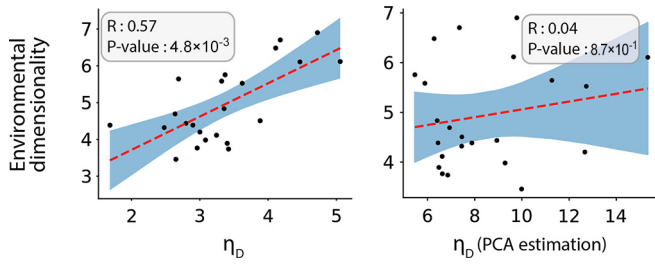


FIG. 5. PCA-based dimensionality does not recover environmental dimensionality. (a) JSDM-based η_D correlates strongly with environmental dimensionality (computed from metabolomics data); Spearman $R = 0.57$, $p = 4.8 \times 10^{-3}$. (b) PCA-based dimensionality, computed as the participation ratio of PCA eigenvalues on log-transformed abundances, shows no significant correlation; $R = 0.04$, $p = 8.7 \times 10^{-1}$.

The coefficient η_D in the exponent was defined to be the niche dimensionality. The fit to the exponential function was performed by log-transforming the KL divergences and using linear regression based on L2-norm.

3. Comparison with linear dimensionality reduction

We tested whether linear dimensionality reduction can also recover environmental dimensionality from microbiome relative abundance data. We estimated a PCA-based dimensionality η_D^{PCA} using the PCA eigenvalues on log-transformed abundances following the same procedure as used to estimate environmental dimensionality and compared both estimates against environmental dimensionality computed from matching mass-spectrometry metabolomics. The JSDM-based η_D correlates strongly with environmental dimensionality [Fig. 5(a), Spearman $R = 0.57$, $p = 4.8 \times 10^{-3}$]. The PCA-based estimate does not [Fig. 5(b), $R = 0.04$, $p = 8.7 \times 10^{-1}$]. The choice of dimensionality-reduction method therefore materially affects whether the inferred dimensionality recovers the underlying environmental structure.

APPENDIX B: DATA CURATION

1. The Microbiomap collection

Microbiomap is a large collection of human microbiome datasets, processed uniformly across ~ 400 different datasets comprising $\sim 1.7 \times 10^5$ independent samples [59]. Individual datasets had varying samples (ranging from ~ 10 to ~ 1000), very different number of bacterial ASVs, and total read count. To reduce potential biases related to sample size, number of bacteria, and read counts, we randomly subsampled all datasets to 5000 reads 10 times to create 10 variants. To eliminate the bias arising from varying number of samples, we subsampled all datasets to 100 samples (we only included datasets with more than 100 samples). This process was also repeated 10 times. Therefore, there were 100 total replicates for each project. To minimize technical noise in abundance measurements [111], we only included bacteria whose average relative abundance across samples was $> 0.1\%$. After this filtering, only the datasets with > 70 ASVs were used for further analysis.

The total load for each sample [Fig. 4(d)] was estimated using a recent machine learning method [89]. The total load for a dataset was reported as the average across samples.

2. Paired 16S-metabolomics data

We used a large curated collection of paired 16S microbiome composition and metabolomics datasets [58] to investigate the relationship between metabolomics and niche dimensionality. Metabolomics data were log-transformed and z scored to standardize different metabolites. Metabolomics datasets often have many zeros. To avoid taking logarithms of zeros, in every dataset, we offset all metabolite measurements by the 10th percentile of the nonzero abundance prior to log-transformation.

Niche dimensionality for the microbiome was computed as described in the main text. The dimensionality of the metabolic environment was computed by performing singular value decomposition on the log-transformed metabolomics data and fitting an exponential decay to the singular value spectrum. To avoid artifacts related to sample size, we subsampled each dataset to 50 samples (or 70% of the sample size if the sample size was lower than 50). Each dataset was subsampled randomly to the value determined by the above procedure 100 times. The quantities of interest (niche dimensionality and environmental complexity) for each dataset was then calculated as the mean of that across the 100 replicates.

The taxonomic resolution varied across studies: ASV-level for DADA2-processed 16S studies, species-level for kraken+bracken-processed shotgun studies, and genus-level for one legacy OTU study (Sinha 2016), all unified under the GTDB taxonomic reference.

3. Other studies

Other studies analyzed in this work included mouse longitudinal microbiomes [83] and infant microbiomes [73], and both were reported at the taxonomic resolution of OTUs.

a. Robustness of the dimensionality correlation

To ensure that the observed correlation between the niche dimensionality and the dimensionality of the metabolic environment (Spearman $r = 0.57$, $p = 4.7 \times 10^{-3}$) was not an artifact of the chosen fitting cutoffs or the specific metric used, we performed two robustness checks.

First, we varied the maximum number of latent components (K) used to fit the exponential decay of the KL divergence. The correlation between the two dimensionalities remained consistently strong and statistically significant across a wide range of cutoffs: $K = 4$ ($r = 0.50$, $p = 0.016$), $K = 6$ ($r = 0.67$, $p = 4.1 \times 10^{-4}$), $K = 7$ ($r = 0.69$, $p = 2.7 \times 10^{-4}$), $K = 8$ ($r = 0.66$, $p = 6.4 \times 10^{-4}$), $K = 9$ ($r = 0.65$, $p = 7.3 \times 10^{-4}$), and $K = 10$ ($r = 0.66$, $p = 5.5 \times 10^{-4}$) [Fig. 6(a)].

Second, we tested whether alternative dimensionality metrics could recover the same qualitative relationship without relying on an exponential decay fit. To do so, we calculated the participation ratio (PR) and the Shannon entropy directly from the list of KL divergences obtained when fitting the JSDM with an increasing number of latent dimensions. We then correlated these metrics with the PR and Shannon

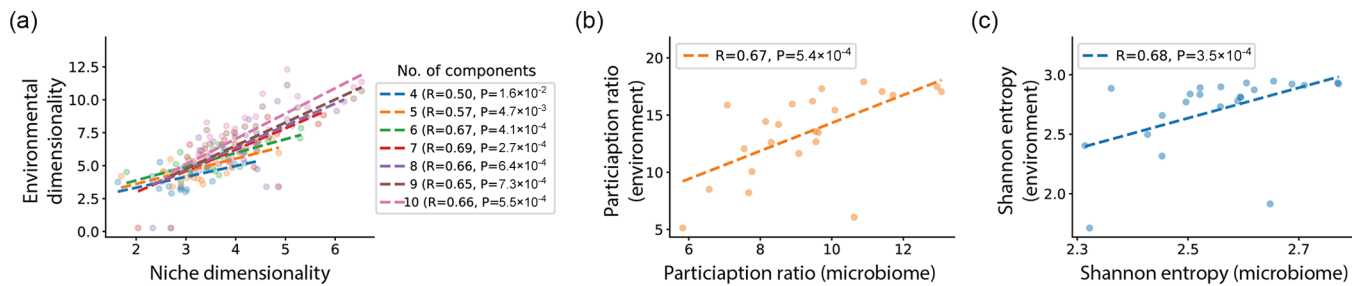


FIG. 6. Robustness of the microbiome-metabolome dimensionality correlation. (a) Scatter plot of environmental dimensionality versus niche dimensionality computed across different fitting cutoffs (K). (b) Correlation between the participation ratio of the metabolomics singular values and the participation ratio of the JSDM KL divergences. (c) Correlation between the Shannon entropy of the metabolomics singular values and the Shannon entropy of the JSDM KL divergences.

entropy computed from the singular values of the corresponding metabolomics data. Notably, we found strong, statistically significant correlations between the two domains using both PR [Spearman $r = 0.67$, $p = 5.4 \times 10^{-4}$; Fig. 6(b)] and Shannon entropy [Spearman $r = 0.68$, $p = 3.5 \times 10^{-4}$; Fig. 6(c)]. This confirms that the observed coupling between microbial and environmental complexity is a robust biological feature of the data, independent of the specific mathematical formulation or cutoffs used to quantify dimensionality.

b. Predicting metabolomics using learned ecosystem embeddings

Each dataset was subsampled 20 times wherein each time 50 random samples (or 70% of the total number of samples if the total number of samples was less than 50) were selected. This dataset was fit with the latent variable JSDM with $K = 10$ latent variables. The following procedure was performed on this learned set of latent variables and a shuffled set where the shuffling was performed for each latent variable separately. This preserved the distribution of each latent variable across samples. The latent variables were then split into training and testing sets with a ratio of 80% and 20% respectively. A linear model was trained on the training set to fit to the corresponding metabolite concentrations. The performance of latent variable predicting the metabolite concentrations was quantified as the difference in out of sample Pearson R^2 of the top 10% of best predicted metabolites. The Pearson R^2 for the same metabolites predicted by the shuffled latent variables with their own linear model served as the null model.

4. Host health and diet

The datasets studying host health and diet were processed as follows. To eliminate sample size effect, each group in a study was randomly subsampled to 20 samples. Niche dimensionality was calculated on this smaller dataset of 20 samples and the process was repeated 300 times. The infant milk study [73] and the murine microbiome diet study [83] were a mix of longitudinal and cross-sectional data. This was handled in infant milk study by grouping samples from a subject and replacing it with its mean, effectively making the dataset cross sectional. In the murine diet study, niche dimensionality was determined for the longitudinal microbiome data for each mouse separately.

APPENDIX C: EVALUATING NICHE OVERLAP USING FLUX BALANCE ANALYSIS

1. Obtaining the metabolite preference matrix

Microbial species survival in the gut largely depends on the ability of species to metabolize available nutrients for growth [8,69]. To approximate the metabolic niches of gut bacteria, we analyzed thousands of manually and computationally curated genome-scale metabolic models [70] using FBA [71]. Although FBA does not yield quantitative preference parameters, it reliably predicts whether a species can use a particular resource [72].

To that end, we identified *resources* as annotated exchange reactions. In any particular model (representing a bacterial species and often a specific bacterial strain), we first allowed all exchange reactions to have an influx of a fixed amount and maximized the growth rate (biomass production). We then individually increased the upper limit of influx for each exchange reaction and calculated maximum growth rate. The nutrients that were able to increase the growth rate were deemed to be those that the model could use for growth. All calculations were performed anaerobically (setting influx of molecular oxygen to zero). We did not differentiate between resources that contributed carbon, nitrogen, sulfur, etc. At the end of the calculation, we obtained a binary matrix that represented predicted abilities (presence or absence) of individual models to use individual metabolite resources.

2. Mapping flux balance analysis models to taxa

For every subsampled dataset (see above), we aimed to match the ASVs to the ~ 7000 genome scale metabolic models. To do so, we used taxonomy in a hierarchical manner. For every ASV, we matched the closest fine-grained taxonomic label to that of the genome scale models. We only allow mapping to the class level. Any mapping at a broader taxonomic resolution was discarded. This way, each ASV was potentially mapped to multiple metabolic models (corresponding to multiple consumer and preference vectors). For calculating the niche overlap and mean number of resources consumed, we randomly chose one of these vectors.

3. Niche overlap calculation

Niche overlap in the consumer-resource model was calculated as follows. Each organism (consumer) was represented

as a binary vector of resource preferences. For an average k -tuple of organisms $k \in [2, 6]$, we first estimated the number of resources consumed by *all* k organisms by first randomly drawing k organisms from the ecosystem and then averaging over multiple draws. The upper limit of $k = 6$ was chosen due to computational constraints arising from combinatorial nature of this calculation.

Due to the sparsity of the resource preference matrix, the average overlap between k species decreases with increasing k . The null expectation for this decrease for a given sparsity ρ of the consumer preference matrix and the number of resources M is simply $M\rho^k$. For each k , we computed the competition as the ratio of the estimated overlap to the null expectation. The total niche overlap was defined as the sum over all values of k .

The same procedure was followed for niche overlap calculation in microbiomap with only the null model being calculated differently. The null model for microbiomap was simply the niche overlap calculation performed from a random set of bacteria selected to have a similar consumer preference matrix sparsity.

a. Niche breadth

In any given dataset, the niche breadth [Fig. 4(c)] was estimated as the mean number of resources consumed by bacteria (after mapping bacteria to resources, see above). As above, only bacteria with mean abundance $>0.1\%$ were included in the analysis.

b. Predicting species-preferences using learned taxon loadings

For each replicate of a dataset, the latent variable JSMD was trained with 20 latent variables. The species loadings θ_{ko} learned from this procedure was used as a regressor in a logistic regression model to predict the average consumer preference matrix constructed using the mapping to FBA models described above. Specifically, we fit the utilization profiles for 846 distinct metabolites. The plot is a histogram of the difference in the log-likelihood of the model fit of all resources across all taxa used the learnt θ and corresponding shuffled θ .

APPENDIX D: PREDICTING GENOME-ANNOTATED ENZYME CONTENT

To ensure our results were not artifacts of the assumptions inherent to FBA, we performed a complementary analysis using genome-annotated enzyme content. Instead of simulating metabolic fluxes, we constructed a binary matrix indicating the presence or absence of specific metabolic enzymes for each bacterial taxon based on genomic annotations.

Using the same logistic regression approach applied to the FBA preferences (Appendix C), we tested whether the inferred species loadings θ_{ko} could predict this enzyme gene content. In total, this complementary analysis evaluated the presence or absence of 7780 individual metabolic genes. As with the FBA-derived preferences, the JSMD embeddings significantly outperformed shuffled controls in predicting the presence of metabolic enzymes (Fig. 7), demonstrating that the latent space captures real biological structure independent of FBA-specific simulation parameters.

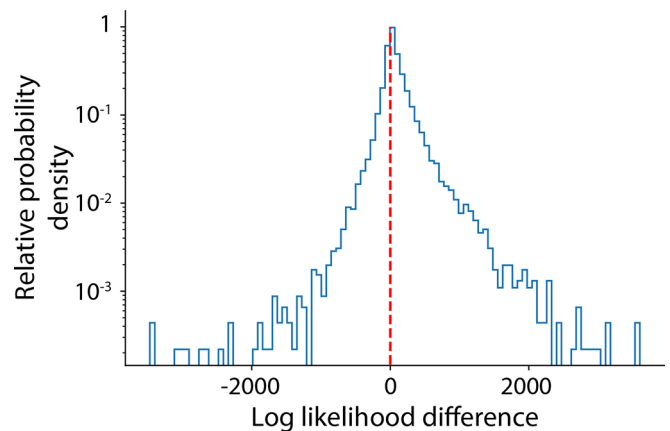


FIG. 7. Species loadings predict genome-annotated enzyme content. Histogram of the difference in log-likelihood of the logistic regression model predicting enzyme presence or absence using the learned species loadings (θ_{ko}) versus shuffled controls.

APPENDIX E: IDENTIFYING METABOLITES THAT SHAPE NICHE AXES

To provide an explicit biological interpretation of the learned latent dimensions [Fig. 1(d)], we utilized the paired 16S-metabolomics dataset from He *et al.* [73]. In Fig. 8, we show the results for the study group of infants consuming complementary plant-based foods in addition to formula. The dataset was preprocessed by taking the natural logarithm of metabolite concentrations after offsetting zeros to 10% of the minimum nonzero concentration. To ensure robustness, we performed the analysis across 100 random subsamples of the data, drawing 70% samples per iteration.

In each iteration, we inferred the ecosystem embeddings (z_{sk}) and species loadings (θ_{ko}) with $K = 10$ latent dimensions. We reoriented and ordered the learned latent dimensions by performing SVD on the reconstructed exponent matrix ($z\theta$), yielding axes ordered by the variance they capture.

We then applied L_1 -regularized (Lasso) regression to predict each resulting latent dimension from the measured metabolite concentrations. A binary search mechanism was employed to find the optimal regularization penalty (α) that selected approximately a target of 10 nonzero metabolic features, enforcing sparsity for biological interpretability. The predictive performance along each axis was quantified by calculating the mean Pearson R^2 across the 100 iterations. As demonstrated in Fig. 8(a), the leading niche axes that capture the most variance are also those that are best predicted by specific metabolite concentrations. Furthermore, by evaluating the specific Lasso coefficients on the top dimensions [Fig. 8(b)], we can isolate the nonzero predictors, identifying metabolites such as short-chain fatty acids (e.g., acetate and butyrate) and amino acids as the principal biochemical drivers defining these microbial niche spaces.

APPENDIX F: ALTERNATIVE METRICS TO COMPUTE NICHE DIMENSIONALITY

To further validate our approach to compute niche dimensionality (Fig. 1), we compared estimated η_D to other standard metrics to infer matrix dimensionality. Specifically, we chose

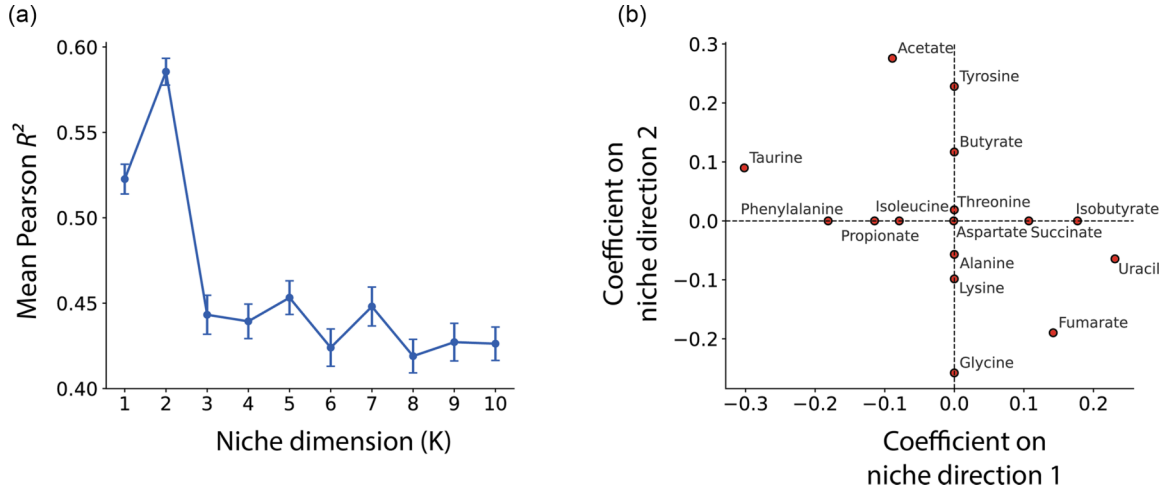


FIG. 8. Identifying specific metabolic drivers of the niche space. (a) Predictive performance (mean Pearson R^2) of sparse Lasso regression models predicting individual niche axes from measured metabolites. Niche dimensions (x axis) are ordered by the amount of variance they capture in the niche space (via SVD). Error bars represent the standard error of the mean across 100 random subsamples of the data. (b) Scatterplot of the Lasso regression coefficients of the metabolites for the first two dominant niche dimensions. Metabolites selected by the sparse regression (having nonzero coefficients) are highlighted in red and labeled, explicitly identifying the specific biological variables that define these primary niche axes.

participation ratio [112] and spectrum entropy [113]. The first step in determining these two metrics for any matrix is computing the singular value decomposition to obtain singular values $\{\lambda_i\}$. The PR and spectrum entropy (SE) are then defined as

$$\text{PR} = \frac{(\sum_i \lambda_i)^2}{\sum_j \lambda_j^2}, \quad (\text{F1})$$

$$\text{SE} = - \sum_i \frac{\lambda_i}{\sum_j \lambda_j} \log \frac{\lambda_i}{\sum_j \lambda_j}. \quad (\text{F2})$$

We compute the two metrics by first performing latent variable inference with a large latent dimension ($K = 20$) and then performing singular value decomposition on the matrix $M = Z \times \Theta$.

Both these metrics were corrected by a null model. Corrected metrics correlated very strongly with our estimate of the niche dimensionality (with a Spearman R of 0.76 and 0.88, respectively, Fig. 9). Null model correction was done as follows. The metric of choice is calculated for the regular relative abundance data and for the relative abundance data

shuffled such that the relative abundance of each taxon is shuffled across samples after which each sample is renormalized to sum up to 1. The null-model-corrected metric is then the difference between the two.

APPENDIX G: CONSUMER-RESOURCE MODEL SIMULATIONS

1. Solving the CRM as a convex optimization problem

The consumer resource model simulations were performed as convex optimization problems similar to the community simulator package by Marsland *et al.* [114].

We first start with the equations describing the CRM:

$$\frac{1}{n_o} \frac{dn_o}{dt} = \mu_o(t) = \sum_{k=1}^K r_k(t) \theta_{ko} - \delta, \quad (\text{G1})$$

$$\frac{dr_k}{dt} = J_k - r_k(t) \sum_o n_o(t) \theta_{ko} - \frac{r_k(t)}{\tau}. \quad (\text{G2})$$

The convex optimization problem to determine community composition is stated as follows. We want to minimize the

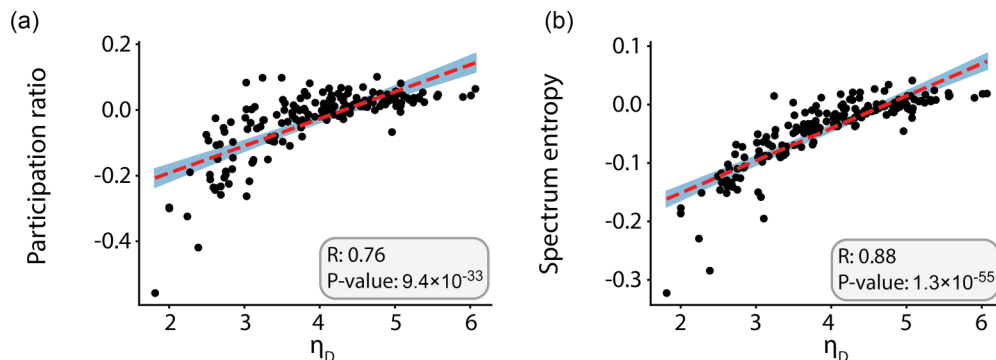


FIG. 9. Participation ratio and spectrum entropy correlate strongly with estimated η_D .

Kullback-Leibler divergence between J_k/δ and r_k given by

$$\text{KL}(\tilde{J}/\delta, \tilde{r}) = \sum_k \frac{J_k}{\delta} \ln \left(\frac{J_k}{\delta r_k} \right) - (J_k - r_k), \quad (\text{G3})$$

subject to the constraint that $\mu_o(t) \leq 0$. The consumer populations are the Lagrange multipliers that impose the constraints.

2. Parameter selection

To explore the parameter space of the consumer resource models, we first look at the set of equations satisfied at steady state is given by

$$\begin{aligned} n_o \left[\sum_k \theta_{ko} r_k - \delta \right] &= 0, \\ \frac{1}{\tau} (r'_k - r_k) - \sum_o \theta_{ko} n_o r_k &= 0, \end{aligned} \quad (\text{G4})$$

where we have rewritten the inflow rate in terms of the dilution rate, $J_k = r'_k/\tau$. r'_k is the steady-state value the resources would take in the absence of consumers. We define $\hat{\theta}_{ko} = \theta_{ko}/\Theta$, $\hat{r}'_k = r'_k/R$, and $\hat{r}_k = r_k/R$ which are dimensionless quantities,

$$\begin{aligned} n_o \left[\sum_k \hat{\theta}_{ko} \hat{r}_k - \frac{\delta}{\Theta R} \right] &= 0, \\ \hat{r}'_k - \hat{r}_k - \Theta \tau \sum_o \hat{\theta}_{ko} n_o \hat{r}_k &= 0. \end{aligned} \quad (\text{G5})$$

We see that just two combinations of parameters, $\delta/(\Theta R)$ and $\Theta \tau$, capture the entirety of the variation at steady state once $\hat{r}'_k$ and $\hat{\theta}_{ko}$ are fixed. $\hat{\theta}_{ko}$ is chosen to be a uniform sparse matrix with a density of 10% and $\hat{r}'_k$ is sampled from a uniform distribution. To simulate the variation between samples within a dataset, we generate each sample's inflow, r'_{ks} where s denotes the sample index, from a Gaussian distribution centered at r'_k with a variance of σr_k^2 . σ , Θ , and δ/Θ along with ρ , which will be introduced in Appendix G 4, are parameters we optimize with simulated annealing.

3. Implementing intrinsic metabolic trade-offs

For the simulations with intrinsic metabolic trade-offs [76,77], we normalized the randomly sampled consumer preference matrix, $\hat{\theta}$, by the sum of each species' preference vector, $\sum_k \hat{\theta}_{ko}$. As a result, the preference vectors for all species summed up to 1. This matrix was then scaled by Θ chosen by the simulated annealing procedure.

4. Implementing environmental trade-offs

We capture the systematic variation of environmental influence on samples by grouping the presence or absence of resources. As an example, a typical person's diet comprises multiple food groups (starches, proteins, fiber, etc.). These food groups have their own set of metabolites where some of these metabolites are present in only one group and some are shared by both groups.

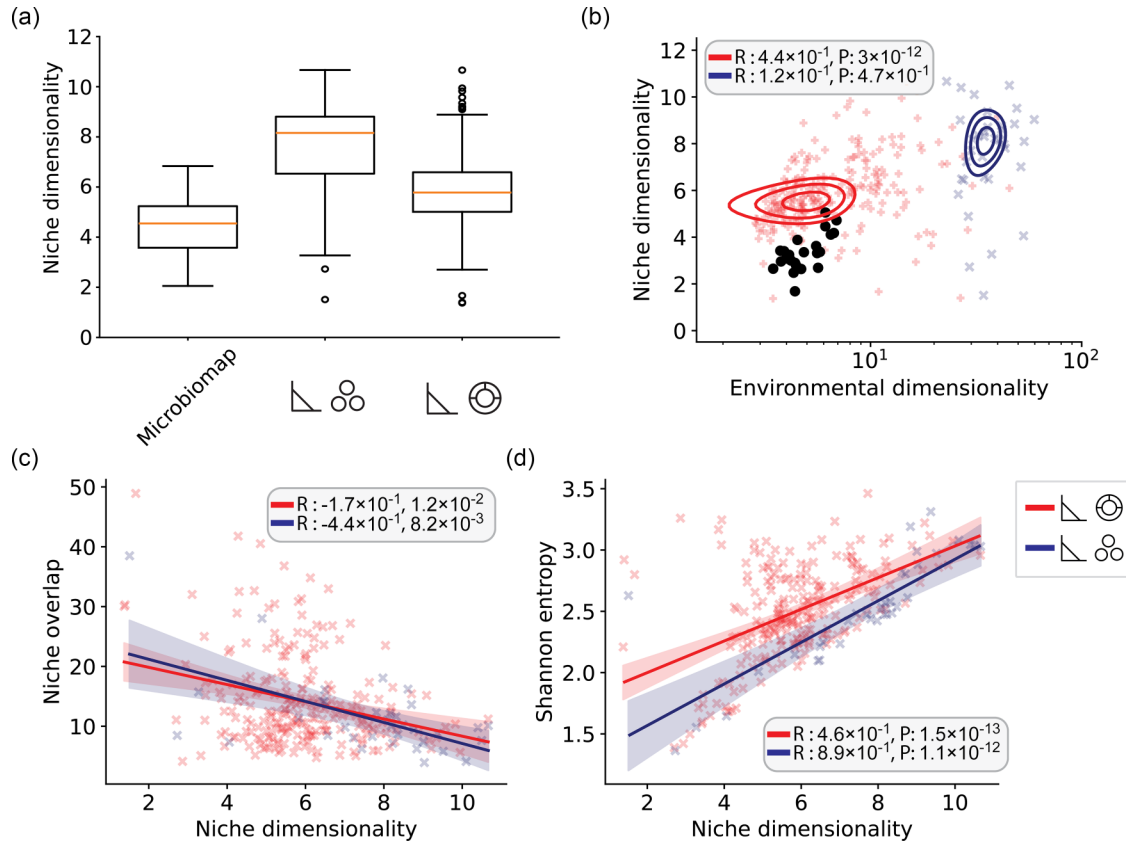


FIG. 10. Niche dimensionality and correlations with consumer or resource simulations with 50 resources.

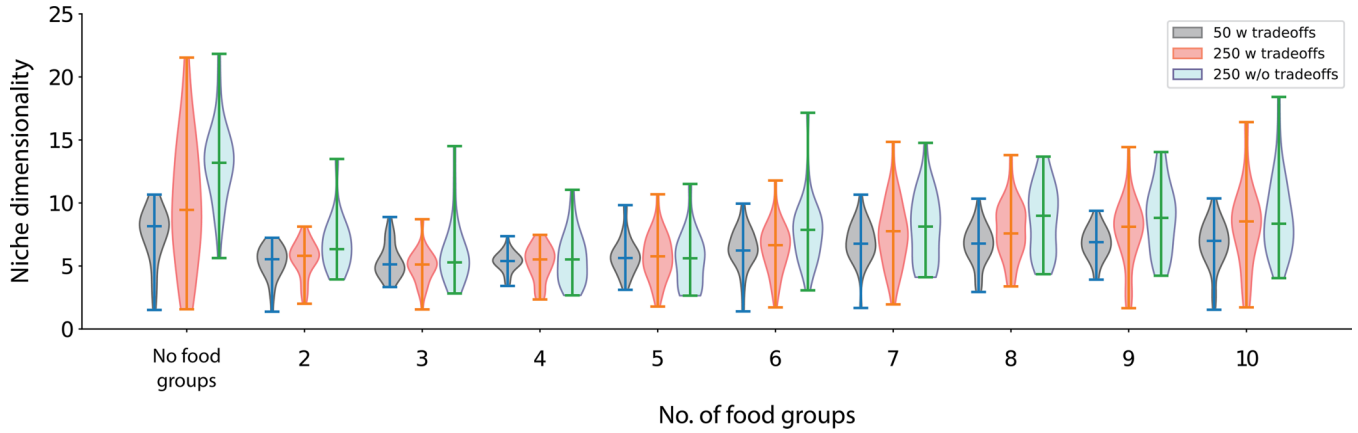


FIG. 11. Violin plots of niche dimensionality versus number of food groups in the consumer or resource simulations.

In our simulations we enforce this structure by generating scaling vectors that alters the magnitude of the resource inflow rates across samples. These scaling vectors are generated as random linear combinations of different masking vectors $\tilde{m}_i$, which are binary vectors that denote the presence or absence of a resource in a particular food group i . The scaling vector of sample s , denoted by $\tilde{\zeta}_s$, is given by

$$\tilde{\zeta}_s = \frac{1}{\rho} \sum_{i=1}^{N_m} \frac{\alpha_{si}}{\sum_j \alpha_{sj}} \tilde{m}_i, \quad (\text{G6})$$

where N_m is the total number of masking vectors, ρ is the density of the masking vectors, $\tilde{m}_i$ is the i th masking vector, and α_{si} are the relative abundances of individual food groups ($\sum_i \alpha_{si} = 1$), which are sampled from a uniform distribution for each sample and then normalized.

5. Simulation details

To investigate the role of metabolic and extrinsic trade-offs, we randomly selected 50 datasets from the Microbiomap collection [59]. The niche dimensionality, α diversity, species richness, and β diversity of this subset was statistically indistinguishable compared to the larger set, with p values of 0.26, 0.59, 0.51, and 0.26 respectively.

All the simulations in the main text started with 250 resources, 700 consumers, and a consumer preference matrix with 10% sparsity. The preferences values were uniformly distributed between 0 and Θ (a hyperparameter of the model). The mean resource inflow values were also sampled from a uniform distribution from 0 to R (another hyperparameter) which was set to $R = 5$. For the simulations with environmental trade-offs (see above), the masking vectors together were sampled from a random sparse matrix with density ρ and the α matrix was populated with random samples from a uniform random distribution between 0 and 1.

Notably, the total number of resources did not affect our overall conclusions. Model simulations with 50 resources also reproduced low niche dimensionality distributions with trade-offs and the observed empirical relationships between niche dimensionality and community structure. Simulations with 50 resources without metabolic trade-offs can only support

at best 50 consumers. Therefore, we only performed these calculations with metabolic trade-offs (Fig. 10).

We pooled the results from two to seven food groups. The distribution of niche dimensionality remained largely consistent above the seven food group threshold (Fig. 11).

APPENDIX H: RESOURCE COVARIATION AND NICHE DIMENSIONALITY

In this section we will show how environmental covariation reduces the dimensionality of niche overlaps. The calculation follows classic treatment by MacArthur [17,25,26] that estimates niche overlap as Gaussian overlap integrals. Assume that there is a D -dimensional resource space in which the *niche* of an organism is expressed a normal distribution,

$$\begin{aligned} C_i(\mathbf{r}) &= \mathcal{N}(\mathbf{r}; \mathbf{r}_i, \Sigma_i) \\ &= \frac{1}{\sqrt{(2\pi)^D |\Sigma_i|}} \exp\left(-\frac{1}{2}(\mathbf{r} - \mathbf{r}_i)^T \Sigma_i^{-1} (\mathbf{r} - \mathbf{r}_i)\right). \end{aligned} \quad (\text{H1})$$

Here $C_i(\mathbf{r})$ is the population size (or growth rate) of organism i if the niche variables take the value $\mathbf{r}$.

Most standard niche overlap calculations assume that $\mathbf{r}$ are uniformly distributed [17,25,26]. Here we explicitly account for correlations in the environment. To that end, let $S(\mathbf{r}) = \mathcal{N}(\mathbf{r}; \mathbf{r}_R, \Sigma_R)$ denote the distribution over environmental variables, assumed to be a multivariate Gaussian distribution with mean $\mathbf{r}_R$ and a covariance matrix Σ_R .

We define the competition between two species i and j to be the overlap integral

$$\chi_{ij} = \int C_i(\mathbf{r}) C_j(\mathbf{r}) S(\mathbf{r}) d\mathbf{r}. \quad (\text{H2})$$

To evaluate the integral, we shift the coordinate frame so that the mean of the environmental variable distribution $\mathbf{r}_R$ is at the origin. We define means of the organism's niche preference distribution from this frame to be $\mathbf{x}_i = \mathbf{r}_i - \mathbf{r}_R$ and

$\mathbf{x}_j = \mathbf{r}_j - \mathbf{r}_R$, the pooled inverse covariance as $\Sigma_P^{-1} = \Sigma_i^{-1} + \Sigma_j^{-1} + \Sigma_R^{-1}$, and the pooled mean $\mu_P = \Sigma_P(\Sigma_i^{-1}\mathbf{x}_i + \Sigma_j^{-1}\mathbf{x}_j)$. The overlap integral can be expressed as

$$\chi_{ij} = \frac{|\Sigma_P|^{1/2}}{\sqrt{(2\pi)^{2D}|\Sigma_i||\Sigma_j||\Sigma_R|}} \times \exp\left(-\frac{1}{2}[\mathbf{x}_i^T \Sigma_i^{-1} \mathbf{x}_i + \mathbf{x}_j^T \Sigma_j^{-1} \mathbf{x}_j - \mu_P^T \Sigma_P^{-1} \mu_P]\right). \quad (\text{H3})$$

We now impose structure in the environment by assuming that the resources are sampled from a lower-dimensional space (of dimension $k < D$). To do so, we assume $\mathbf{r} = \mathbf{B}\mathbf{u}$, where $\mathbf{B}$ is a $D \times k$ matrix whose columns form an orthonormal basis. To work in the subspace we transform distributions in the k -dimensional space of $\mathbf{u}$. We have $\mathbf{u}_R = \mathbf{0}$, $\Omega_R = \mathbf{B}^T \Sigma_R \mathbf{B}$, $\mathbf{u}_i = \mathbf{B}^T(\mathbf{r}_i - \mathbf{r}_R)$, and $\Omega_i = \mathbf{B}^T \Sigma_i \mathbf{B}$. The transformed k -dimensional Gaussian PDFs are given

by

$$C_{i,k}(\mathbf{u}) = \frac{1}{\sqrt{(2\pi)^k |\Omega_i|}} \exp\left(-\frac{1}{2}(\mathbf{u} - \mathbf{u}_i)^T \Omega_i^{-1} (\mathbf{u} - \mathbf{u}_i)\right). \quad (\text{H4})$$

With these definitions, the niche overlap in the lower-dimensional space is given by

$$\chi_{ij,k} = \frac{|\Omega_P|^{1/2}}{\sqrt{(2\pi)^{2k} |\Omega_i||\Omega_j||\Omega_R|}} \times \exp\left(-\frac{1}{2}[\mathbf{u}_i^T \Omega_i^{-1} \mathbf{u}_i + \mathbf{u}_j^T \Omega_j^{-1} \mathbf{u}_j - \mu_P^T \Omega_P^{-1} \mu_P]\right), \quad (\text{H5})$$

where $\Omega_P^{-1} = \Omega_i^{-1} + \Omega_j^{-1} + \Omega_R^{-1}$ is the pooled inverse covariance and $\mu_P = \Omega_P(\Omega_i^{-1}\mathbf{u}_i + \Omega_j^{-1}\mathbf{u}_j)$ is the pooled mean.

A comparison of Eqs. (H3) and (H5) shows that restricting the environments to a lower dimension leads to overlap matrices that are computed in a lower-dimensional space as well [27,28,43,102].

-
- [1] P. J. Turnbaugh, R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, and J. I. Gordon, The human microbiome project, *Nature (London)* **449**, 804 (2007).
 - [2] A. K. Nash, T. A. Auchtung, M. C. Wong, D. P. Smith, J. R. Gesell, M. C. Ross, C. J. Stewart, G. A. Metcalf, D. M. Muzny, R. A. Gibbs, *et al.*, The gut mycobiome of the human microbiome project healthy cohort, *Microbiome* **5**, 153 (2017).
 - [3] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, Weak pairwise correlations imply strongly correlated network states in a neural population, *Nature (London)* **440**, 1007 (2006).
 - [4] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. M. Walczak, Statistical mechanics for natural flocks of birds, *Proc. Natl. Acad. Sci. USA* **109**, 4786 (2012).
 - [5] S. Cocco, C. Feinauer, M. Figliuzzi, R. Monasson, and M. Weigt, Inverse statistical physics of protein sequences: A key issues review, *Rep. Prog. Phys.* **81**, 032601 (2018).
 - [6] K. Z. Coyte, J. Schluter, and K. R. Foster, The ecology of the microbiome: Networks, competition, and stability, *Science* **350**, 663 (2015).
 - [7] M. Tikhonov and R. Monasson, Collective phase in resource competition in a highly diverse ecosystem, *Phys. Rev. Lett.* **118**, 048103 (2017).
 - [8] R. Marsland, III, W. Cui, and P. Mehta, A minimal model for microbial biodiversity can reproduce experimentally observed ecological patterns, *Sci. Rep.* **10**, 3308 (2020).
 - [9] W. Cui, R. Marsland III, and P. Mehta, Diverse communities behave like typical random ecosystems, *Phys. Rev. E* **104**, 034416 (2021).
 - [10] J. Hu, D. R. Amor, M. Barbier, G. Bunin, and J. Gore, Emergent phases of ecological diversity and dynamics mapped in microcosms, *Science* **378**, 85 (2022).
 - [11] P.-Y. Ho, B. H. Good, and K. C. Huang, Competition for fluctuating resources reproduces statistics of species abundance over time across wide-ranging microbiotas, *eLife* **11**, e75168 (2022).
 - [12] J. Pasqualini, A. Maritan, A. Rinaldo, S. Facchin, E. V. Savarino, A. Altieri, and S. Suweis, Linking complex microbial interactions and dysbiosis through a disordered Lotka–Volterra model, *eLife* **14**, RP105948 (2026).
 - [13] X.-W. Wang and Y.-Y. Liu, Origins of scaling laws in microbial dynamics, *Phys. Rev. Res.* **5**, 013004 (2023).
 - [14] J. Camacho-Mateu, A. Lampo, M. Sireci, M. A. Muñoz, and J. A. Cuesta, Sparse species interactions reproduce abundance correlation patterns in microbial communities, *Proc. Natl. Acad. Sci. USA* **121**, e2309575121 (2024).
 - [15] R. Maskawa, H. Takayasu, L. Takayasu, W. Suda, and M. Takayasu, Stochastic spatiotemporal growth model reproducing the universal statistical laws of the gut microbiome, *Phys. Rev. Res.* **7**, 013269 (2025).
 - [16] G. E. Hutchinson, Concluding remarks, *Cold Spring Harb. Symp. Quant. Biol.* **22**, 415 (1957).
 - [17] R. MacArthur and R. Levins, The limiting similarity, convergence, and divergence of coexisting species, *Am. Nat.* **101**, 377 (1967).
 - [18] J. M. Chase and M. A. Leibold, *Ecological Niches: Linking Classical and Contemporary Approaches* (University of Chicago Press, Chicago, 2009).
 - [19] M. Leibold and V. F. Frans, Re-revisiting the niche concept, *EcoEvoRxiv* (2025).
 - [20] R. D. Holt, Bringing the Hutchinsonian niche into the 21st century: Ecological and evolutionary perspectives, *Proc. Natl. Acad. Sci. USA* **106**, 19659 (2009).
 - [21] H. K. Swanson, M. Lysy, M. Power, A. D. Stasko, J. D. Johnson, and J. D. Reist, A new probabilistic method for quantifying n-dimensional ecological niches and niche overlap, *Ecology* **96**, 318 (2015).
 - [22] B. Blonder, C. B. Morrow, B. Maitner, D. J. Harris, C. Lamanna, C. Violle, B. J. Enquist, and A. J. Kerkhoff, New approaches for delineating n-dimensional hypervolumes, *Methods Ecol. Evol.* **9**, 305 (2018).

- [23] E. R. Pianka, Niche overlap and diffuse competition, *Proc. Natl. Acad. Sci. USA* **71**, 2141 (1974).
- [24] H. J. Morowitz, The dimensionality of niche space, *J. Theor. Biol.* **86**, 259 (1980).
- [25] R. H. MacArthur, *Geographical Ecology: Patterns in the Distribution of Species* (Princeton University Press, Princeton, NJ, 1984).
- [26] R. MacArthur, Species packing and competitive equilibrium for many species, *Theor. Popul. Biol.* **1**, 1 (1970).
- [27] R. M. Yoshiyama and J. Roughgarden, Species packing in two dimensions, *Am. Nat.* **111**, 107 (1977).
- [28] C. Rappoldt and P. Hogeweg, Niche packing and number of species, *Am. Nat.* **116**, 480 (1980).
- [29] T. J. Case, Niche packing and coevolution in competition communities, *Proc. Natl. Acad. Sci. USA* **78**, 5021 (1981).
- [30] E. R. Pianka, The structure of lizard communities, *Annu. Rev. Ecol. Syst.* **4**, 53 (1973).
- [31] R. Costa-Pereira, M. S. Araújo, F. L. Souza, and T. Ingram, Competition and resource breadth shape niche variation and overlap in multiple trophic dimensions, *Proc. R. Soc. B* **286**, 20190369 (2019).
- [32] W. S. Harpole and D. Tilman, Grassland species loss resulting from reduced niche dimension, *Nature (London)* **446**, 791 (2007).
- [33] W. S. Harpole, L. L. Sullivan, E. M. Lind, J. Firn, P. B. Adler, E. T. Borer, J. Chase, P. A. Fay, Y. Hautier, H. Hillebrand, *et al.*, Addition of multiple limiting resources reduces grassland diversity, *Nature (London)* **537**, 93 (2016).
- [34] E. T. Borer, J. B. Grace, W. S. Harpole, A. S. MacDougall, and E. W. Seabloom, A decade of insights into grassland ecosystem responses to global environmental change, *Nat. Ecol. Evol.* **1**, 0118 (2017).
- [35] N. Eisenhauer, W. Schulz, S. Scheu, and A. Jousset, Niche dimensionality links biodiversity and invasibility of microbial communities, *Funct. Ecol.* **27**, 282 (2013).
- [36] P. Nosil and C. P. Sandoval, Ecological niche dimensionality and the evolutionary diversification of stick insects, *PLoS ONE* **3**, e1907 (2008).
- [37] M. J. Larcombe, G. J. Jordan, D. Bryant, and S. I. Higgins, The dimensionality of niche space allows bounded and unbounded processes to jointly influence diversification, *Nat. Commun.* **9**, 4258 (2018).
- [38] L.-M. Chevin, G. Decorzent, and T. Lenormand, Niche dimensionality and the genetics of ecological speciation, *Evolution* **68**, 1244 (2014).
- [39] J. E. Cohen, Food webs and the dimensionality of trophic niche space, *Proc. Natl. Acad. Sci. USA* **74**, 4533 (1977).
- [40] A. G. Rossberg, *Food Webs and Biodiversity: Foundations, Models, Data* (John Wiley & Sons, New York, 2013).
- [41] R. J. Williams and D. W. Purves, The probabilistic niche model reveals substantial variation in the niche structure of empirical food webs, *Ecology* **92**, 1849 (2011).
- [42] A. Eklöf, U. Jacob, J. Kopp, J. Bosch, R. Castro-Urgal, N. P. Chacoff, B. Dalsgaard, C. de Sassi, M. Galetti, P. R. Guimarães, *et al.*, The dimensionality of ecological networks, *Ecol. Lett.* **16**, 577 (2013).
- [43] D. B. Stouffer, O. Godoy, G. V. Dalla Riva, and M. M. Mayfield, The dimensionality of plant–plant competition, *bioRxiv* (2021).
- [44] D. C. Laughlin, The intrinsic dimensionality of plant traits and its relevance to community assembly, *J. Ecol.* **102**, 186 (2014).
- [45] O. Ovaskainen and N. Abrego, *Joint Species Distribution Modelling: With Applications in R* (Cambridge University Press, Cambridge, UK, 2020).
- [46] L. J. Pollock, R. Tingley, W. K. Morris, N. Golding, R. B. O'Hara, K. M. Parris, P. A. Vesik, and M. A. McCarthy, Understanding co-occurrence by modelling species simultaneously with a joint species distribution model (JSDM), *Methods Ecol. Evol.* **5**, 397 (2014).
- [47] F. K. C. Hui, S. Taskinen, S. Pledger, S. D. Foster, and D. I. Warton, Model-based approaches to unconstrained ordination, *Methods Ecol. Evol.* **6**, 399 (2015).
- [48] D. I. Warton, F. G. Blanchet, R. B. O'Hara, O. Ovaskainen, S. Taskinen, S. C. Walker, and F. K. C. Hui, So many variables: Joint modeling in community ecology, *Trends Ecol. Evol.* **30**, 766 (2015).
- [49] C. K. Fisher, T. Mora, and A. M. Walczak, Variable habitat conditions drive species covariation in the human microbiota, *PLoS Comput. Biol.* **13**, e1005435 (2017).
- [50] M. Shahin, B. Ji, and P. D. Dixit, Embed: Essential microbiome dynamics, a dimensionality reduction approach for longitudinal microbiome studies, *npj Syst. Biol. Appl.* **9**, 26 (2023).
- [51] G. Plata, K. Srinivasan, M. Krishnamurthy, L. Herron, and P. Dixit, Designing host-associated microbiomes using the consumer/resource model, *mSystems* **10**, e01068-24 (2025).
- [52] D. J. Schwab, I. Nemenman, and P. Mehta, Zipf's law and criticality in multivariate data without fine-tuning, *Phys. Rev. Lett.* **113**, 068102 (2014).
- [53] V. Thibeault, A. Allard, and P. Desrosiers, The low-rank hypothesis of complex systems, *Nat. Phys.* **20**, 294 (2024).
- [54] C. J. Russo, K. Husain, and A. Murugan, Soft modes as a predictive framework for low-dimensional biological systems across scales, *Annu. Rev. Biophys.* **54**, 401 (2025).
- [55] M. C. Morrell, A. J. Sederberg, and I. Nemenman, Latent dynamical variables produce signatures of spatiotemporal criticality in large biological systems, *Phys. Rev. Lett.* **126**, 118302 (2021).
- [56] X. Zhao, G. Plata, and P. D. Dixit, Sigmoid: A super-statistical generative model for binary data, *PLoS Comput. Biol.* **17**, e1009275 (2021).
- [57] H. Akl, B. Emison, X. Zhao, A. Mondal, A. Perez, and P. D. Dixit, Generalist: A latent space based generative model for protein sequence families, *PLoS Comput. Biol.* **19**, e1011655 (2023).
- [58] E. Muller, Y. M. Algavi, and E. Borenstein, The gut microbiome-metabolome dataset collection: A curated resource for integrative meta-analysis, *npj Biofilms Microb.* **8**, 79 (2022).
- [59] R. J. Abdill, S. P. Graham, V. Rubinetti, M. Ahmadian, P. Hicks, A. Chetty, D. McDonald, P. Ferretti, E. Gibbons, M. Rossi, *et al.*, Integration of 168,000 samples reveals global patterns of the human gut microbiome, *Cell* **188**, 1100 (2025).
- [60] D. Tilman, *Resource Competition and Community Structure* (Princeton University Press, Princeton, NJ, 1982).
- [61] I. Veseli, Y. T. Chen, M. S. Schechter, C. Vanni, E. C. Fogarty, A. R. Watson, B. Jabri, R. Blekhman, A. D. Willis, M. K. Yu, *et al.*, Microbes with higher metabolic independence are

- enriched in human gut microbiomes under stress, *eLife* **12**, RP89862 (2025).
- [62] D. Vandeputte, G. Kathagen, K. D'hoë, S. Vieira-Silva, M. Valles-Colomer, J. Sabino, J. Wang, R. Y. Tito, L. De Commer, Y. Darzi, *et al.*, Quantitative microbiome profiling links gut community variation to microbial load, *Nature (London)* **551**, 507 (2017).
- [63] C. Fraser, E. J. Alm, M. F. Polz, B. G. Spratt, and W. P. Hanage, The bacterial species challenge: Making sense of genetic and ecological diversity, *Science* **323**, 741 (2009).
- [64] P. D. Dixit, T. Y. Pang, and S. Maslov, Recombination-driven genome evolution and stability of bacterial species, *Genetics* **207**, 281 (2017).
- [65] 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).
- [66] M. Blaxter, J. Mann, T. Chapman, F. Thomas, C. Whitton, R. Floyd, and E. Abebe, Defining operational taxonomic units using DNA barcode data, *Philos. Trans. R. Soc. B* **360**, 1935 (2005).
- [67] D. Gevers, F. M. Cohan, J. G. Lawrence, B. G. Spratt, T. Coenye, E. J. Feil, E. Stackebrandt, Y. V. De Peer, P. Vandamme, F. L. Thompson, *et al.*, Re-evaluating prokaryotic species, *Nat. Rev. Microbiol.* **3**, 733 (2005).
- [68] M. Dal Bello, H. Lee, A. Goyal, and J. Gore, Resource–diversity relationships in bacterial communities reflect the network structure of microbial metabolism, *Nat. Ecol. Evol.* **5**, 1424 (2021).
- [69] 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).
- [70] A. Heinken, J. Hertel, G. Acharya, D. A. Ravcheev, M. Nyga, O. E. Okpala, M. Hogan, S. Magnúsdóttir, F. Martinelli, B. Nap, *et al.*, Genome-scale metabolic reconstruction of 7,302 human microorganisms for personalized medicine, *Nat. Biotechnol.* **41**, 1320 (2023).
- [71] J. D. Orth, I. Thiele, and B. Ø. Palsson, What is flux balance analysis?, *Nat. Biotechnol.* **28**, 245 (2010).
- [72] G. Plata, C. S. Henry, and D. Vitkup, Long-term phenotypic evolution of bacteria, *Nature (London)* **517**, 369 (2015).
- [73] X. He, M. Parenti, T. Grip, B. Lönnerdal, N. Timby, M. Domellöf, O. Hernell, and C. M. Slupsky, Fecal microbiome and metabolome of infants fed bovine MFGM supplemented formula or standard formula with breast-fed infants as reference: A randomized controlled trial, *Sci. Rep.* **9**, 11589 (2019).
- [74] A. Mira, H. Ochman, and N. A. Moran, Deletional bias and the evolution of bacterial genomes, *Trends Genet.* **17**, 589 (2001).
- [75] M. Basan, S. Hui, H. Okano, Z. Zhang, Y. Shen, J. R. Williamson, and T. Hwa, Overflow metabolism in *Escherichia coli* results from efficient proteome allocation, *Nature (London)* **528**, 99 (2015).
- [76] A. Posfai, T. TAILLEFUMIER, and N. S. Wingreen, Metabolic trade-offs promote diversity in a model ecosystem, *Phys. Rev. Lett.* **118**, 028103 (2017).
- [77] R. Caetano, Y. Ispolatov, and M. Doebeli, Evolution of diversity in metabolic strategies, *eLife* **10**, e67764 (2021).
- [78] Z. R. Miller and J. P. O'Dwyer, Metabolic trade-offs can reverse the resource-diversity relationship, *Am. Nat.* **204**, E85 (2024).
- [79] E. Blumenthal and P. Mehta, Geometry of ecological coexistence and niche differentiation, *Phys. Rev. E* **108**, 044409 (2023).
- [80] K. L. Keller, J. Kirzner, A. Pietrobelli, M.-P. St-Onge, and M. S. Faith, Increased sweetened beverage intake is associated with reduced milk and calcium intake in 3-to 7-year-old children at multi-item laboratory lunches, *J. Am. Diet. Assoc.* **109**, 497 (2009).
- [81] A. B. Bontrager Yoder and D. A. Schoeller, Fruits and vegetables displace, but do not decrease, total energy in school lunches, *Childhood Obesity* **10**, 357 (2014).
- [82] J. A. Gilbert, M. J. Blaser, J. G. Caporaso, J. K. Jansson, S. V. Lynch, and R. Knight, Current understanding of the human microbiome, *Nat. Med.* **24**, 392 (2018).
- [83] R. N. Carmody, G. K. Gerber, J. M. Luevano, D. M. Gatti, L. Somes, K. L. Svenson, and P. J. Turnbaugh, Diet dominates host genotype in shaping the murine gut microbiota, *Cell Host Microbe* **17**, 72 (2015).
- [84] L. A. David, C. F. Maurice, R. N. Carmody, D. B. Gootenberg, J. E. Button, B. E. Wolfe, A. V. Ling, A. S. Devlin, Y. Varma, M. A. Fischbach, *et al.*, Diet rapidly and reproducibly alters the human gut microbiome, *Nature (London)* **505**, 559 (2014).
- [85] J. P. Jacobs, M. Goudarzi, N. Singh, M. Tong, I. H. McHardy, P. Ruegger, M. Asadourian, B.-H. Moon, A. Ayson, J. Borneman, *et al.*, A disease-associated microbial and metabolomics state in relatives of pediatric inflammatory bowel disease patients, *Cell. Molec. Gastroenterol. Hepatol.* **2**, 750 (2016).
- [86] E. A. Franzosa, A. Sirota-Madi, J. Avila-Pacheco, N. Fornelos, H. J. Haiser, S. Reinker, T. Vatanen, A. B. Hall, H. Mallick, L. J. McIver, *et al.*, Gut microbiome structure and metabolic activity in inflammatory bowel disease, *Nat. Microbiol.* **4**, 293 (2019).
- [87] R. Sinha, J. Ahn, J. N. Sampson, J. Shi, G. Yu, X. Xiong, R. B. Hayes, and J. J. Goedert, Fecal microbiota, fecal metabolome, and colorectal cancer interrelations, *PLoS ONE* **11**, e0152126 (2016).
- [88] M. Kim, E. Vogtmann, D. A. Ahlquist, M. E. Devens, J. B. Kisiel, W. R. Taylor, B. A. White, V. L. Hale, J. Sung, N. Chia, *et al.*, Fecal metabolomic signatures in colorectal adenoma patients are associated with gut microbiota and early events of colorectal cancer pathogenesis, *MBio* **11**, e03186 (2020).
- [89] S. Nishijima, E. Stankevic, O. Aasmets, T. S. B. Schmidt, N. Nagata, M. I. Keller, P. Ferretti, H. B. Juel, A. Fullam, S. M. Robbani, *et al.*, Fecal microbial load is a major determinant of gut microbiome variation and a confounder for disease associations, *Cell* **188**, 222 (2025).
- [90] M. Barbier, C. de Mazancourt, M. Loreau, and G. Bunin, Fingerprints of high-dimensional coexistence in complex ecosystems, *Phys. Rev. X* **11**, 011009 (2021).
- [91] M. Leibold and M. Barbier, Disordered and partially structured models in community ecology: What are they? And how do we use them? *EcoEvoRxiv* (2025).
- [92] N. J. B. Kraft, O. Godoy, and J. M. Levine, Plant functional traits and the multidimensional nature of species coexistence, *Proc. Natl. Acad. Sci. USA* **112**, 797 (2015).
- [93] J. S. Clark, D. Bell, C. Chu, B. Courbaud, M. Dietze, M. Hersch, J. HilleRisLambers, I. Ibáñez, S. LaDeau, S. McMahon, *et al.*, High-dimensional coexistence based on individual variation: A synthesis of evidence, *Ecol. Monogr.* **80**, 569 (2010).

- [94] J. E. Goldford, N. Lu, D. Bajić, S. Estrela, M. Tikhonov, A. Sanchez-Gorostiaga, D. Segrè, P. Mehta, and A. Sanchez, Emergent simplicity in microbial community assembly, *Science* **361**, 469 (2018).
- [95] M. Gralka, R. Szabo, R. Stocker, and O. X. Cordero, Trophic interactions and the drivers of microbial community assembly, *Curr. Biol.* **30**, R1176 (2020).
- [96] S. Estrela, J. Diaz-Colunga, J. C. C. Vila, A. Sanchez-Gorostiaga, and A. Sanchez, Diversity begets diversity under microbial niche construction, *bioRxiv* (2022).
- [97] J. B. Losos, Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species, *Ecol. Lett.* **11**, 995 (2008).
- [98] 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).
- [99] Y. Himeoka and N. Mitarai, Dynamics of bacterial populations under the feast-famine cycles, *Phys. Rev. Res.* **2**, 013372 (2020).
- [100] A. Zarrinpar, A. Chaix, S. Yooseph, and S. Panda, Diet and feeding pattern affect the diurnal dynamics of the gut microbiome, *Cell Metab.* **20**, 1006 (2014).
- [101] A. L. Koch, The adaptive responses of *Escherichia coli* to a feast and famine existence, in *Advances in Microbial Physiology* (Elsevier, Amsterdam, 1971), Vol. 6, pp. 147–217.
- [102] S. P. Mizrahi, H. Lee, A. Goyal, E. Owen, and J. Gore, Structured interactions explain the absence of keystone species in synthetic microcosms, *ISME J.* **19**, wraf211 (2025).
- [103] G. Biroli, G. Bunin, and C. Cammarota, Marginally stable equilibria in critical ecosystems, *New J. Phys.* **20**, 083051 (2018).
- [104] F. Roy, G. Biroli, G. Bunin, and C. Cammarota, Numerical implementation of dynamical mean field theory for disordered systems: Application to the Lotka–Volterra model of ecosystems, *J. Phys. A: Math. Theor.* **52**, 484001 (2019).
- [105] D. G. Clark, L. F. Abbott, and A. Litwin-Kumar, Dimension of activity in random neural networks, *Phys. Rev. Lett.* **131**, 118401 (2023).
- [106] J. Bergelson, M. Kreitman, D. A. Petrov, A. Sanchez, and M. Tikhonov, Functional biology in its natural context: A search for emergent simplicity, *eLife* **10**, e67646 (2021).
- [107] J. Moran and M. Tikhonov, Emergent predictability in microbial ecosystems, *arXiv:2403.19372*.
- [108] K. K. Lee, S. Liu, K. Crocker, J. Wang, D. R. Huggins, M. Tikhonov, M. Mani, and S. Kuehn, Functional regimes define soil microbiome response to environmental change, *Nature (London)*, **644**, 1028 (2025).
- [109] D. R. Hekstra and S. Leibler, Contingency and statistical laws in replicate microbial closed ecosystems, *Cell* **149**, 1164 (2012).
- [110] A. V. Narla, T. Hwa, and A. Murugan, Dynamic coexistence driven by physiological transitions in microbial communities, *Proc. Natl. Acad. Sci. USA* **122**, e2405527122 (2025).
- [111] B. W. Ji, R. U. Sheth, P. D. Dixit, Y. Huang, A. Kaufman, H. H. Wang, and D. Vitkup, Quantifying spatiotemporal variability and noise in absolute microbiota abundances using replicate sampling, *Nat. Methods* **16**, 731 (2019).
- [112] P. Gao, E. Trautmann, B. Yu, G. Santhanam, S. Ryu, K. Shenoy, and S. Ganguli, A theory of multineuronal dimensionality, dynamics and measurement, *bioRxiv* (2017).
- [113] O. Roy and M. Vetterli, The effective rank: A measure of effective dimensionality, in *Proceedings of the 15th European Signal Processing Conference* (IEEE, New York, 2007), pp. 606–610.
- [114] R. Marsland, W. Cui, J. Goldford, and P. Mehta, The community simulator: A Python package for microbial ecology, *PLoS ONE* **15**, e0230430 (2020).