Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The latent simplicity of microbial ecological interactions

Álvaro Sánchez

sanchezlabqb@gmail.com

Institute of Functional Biology and Genomics, CSIC - University of Salamanca https://orcid.org/0000-0002-2292-5608

José Camacho-Mateu

Institute of Functional Biology and Genomics, CSIC - University of Salamanca

Isabel Quirós-Rodriguez

Institute of Functional Biology and Genomics, CSIC - University of Salamanca

Giulio Burgio

Institute of Functional Biology and Genomics, CSIC - University of Salamanca

Andrea Arrabal

Institute of Functional Biology and Genomics, CSIC - University of Salamanca

Alfonso Mendaña

Institute of Functional Biology and Genomics, CSIC - University of Salamanca

Miguel D. Fernandez-de-Bobadilla

Institute of Functional Biology and Genomics, CSIC - University of Salamanca

Belén Benítez-Domínguez

Institute of Functional Biology and Genomics, CSIC - University of Salamanca

Biological Sciences - Article

Keywords:

Posted Date: October 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11139241/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

The latent simplicity of microbial ecological interactions

José Camacho-Mateu[1,†,*], Isabel Quirós-Rodriguez[1,†], Giulio Burgio[1,†], Andrea Arrabal[1,†], Alfonso Mendaña[1], Miguel D. Fernandez-de-Bobadilla[1], Belén Benítez-Domínguez[1], Álvaro Sanchez[1,*]

[1]Institute of Functional Biology and Genomics (IBFG), CSIC - University of Salamanca, 37007 Salamanca, Spain

[†] These authors contributed equally to this work

[*] Corresponding authors: JCM: joscamateu@usal.es, AS: sanchezlabqb@gmail.com

Abstract

Microbial communities carry out functions of critical ecological, clinical and biotechnological significance. These functions arise from individual species contributions and from pairwise and higher-order interactions, whose number grows exponentially with community size. Despite this apparent complexity, community function is often surprisingly predictable. Here we show that this predictability does not require weak or sparse interactions. Instead, high-order interactions are strongly coordinated with pairwise interactions and additive effects, following simple linear laws. This coordination causes species effects in community function to vary together across community backgrounds, allowing function to be described by only one or two collective variables. We demonstrate this organization across dozens of combinatorial bacterial experiments and multiple environments. For total biomass, consumer–resource models and experiments identify the dominant collective variable with species yield, a trait that remains stable as abundances change through community dynamics. Microbial community functions are therefore predictable not because interactions are simple, but because their complexity is coherently organized. This provides a theoretical framework for understanding and ultimately designing community function.

Introduction

Microbial communities perform functions that are essential in natural and engineered ecosystems, from nutrient cycling in soils[1] to the production of metabolites[2,3], pharmaceuticals and biofuels[4,5]. These functions are often collective: the contribution of any one species depends on which other species are present. Species compete for shared resources, exchange metabolites and alter the environment they inhabit[6–9,10–13], and these effects are themselves modified when further species are added, producing the higher-order interactions widely documented in microbial consortia[14–17]. Because the number of such

interactions grows exponentially with community size, predicting function from composition would seem to require measuring an exponentially large space of community states. Thus, rationally designing a community to optimize a desired function would appear to be intractable[18,19].

Yet this combinatorial barrier may be less prohibitive than it appears. Accumulating evidence indicates that microbial communities often require only a small number of parameters to describe and predict their function[10,20–23]. One such regularity is that the effect of adding a species on the function of the community (its “functional effect”) often depends linearly on the function of the background community it is added to (Fig. 1A,B), a relationship known as a Functional Effect Equation (FEE)[21,24,25]. If, as we argue above, the functional effect of a species is itself the result of many pairwise and higher-order interactions with other members of the community, how can such simplicity in the effect of a species coexist with such complexity in the interactions that produce it?

In this paper we address this question and show that the two are reconcilable only if species interactions are not independent of one another: a strong FEE for a focal species implies that its higher-order interactions are proportional to the corresponding lower-order ones, with the constant of proportionality set by the slope of the FEE. We test this prediction experimentally and show that it holds across functions and environments. Furthermore, we show that this latent simplicity among interactions gives rise to an emergent simplicity in community function. If the interactions involving each species are proportional to one another, different species cannot affect the same community in idiosyncratic ways: their effects must rise and fall together as the background community changes. Community function is then governed by just one or two ecological collective modes: weighted combinations of species whose functional effects change together across backgrounds, even when the underlying interactions are strong and context-dependent.

The collective modes are inferred from the landscape itself, which leaves open what they represent biologically. Using consumer-resource models and biomass as a model function, we show that the dominant mode typically corresponds to a vector of species traits: the growth yield, or biomass produced per unit of resource consumed. We confirm this experimentally across 31 resource environments. Because yield is a property of a species' metabolism rather than of its abundance, the dominant mode should not be affected by population dynamics: we find it to barely change over ten daily passages, even as species abundances diverge from the inoculum as communities move towards a stable state. The community-function landscape is therefore strikingly robust to population dynamics. Our results reveal a latent simplicity in microbial ecological interactions: community functions are

not simple because interactions are weak, but rather because interactions are proportional to each other and contribute coherently through a small number of collective coordinates, which can be mechanistically interpretable.

Results

Pairwise and Higher-Order Ecological Interactions Follow Simple Statistical Patterns

To study how interactions shape community function, we represent the mapping between community composition and function as a *community-function landscape*[19,23,26]. Each community corresponds to a binary vector $\mathbf{x} \in \{-1, +1\}^N$, where entries $x_i = +1$ and $x_i = -1$ indicate species presence and absence, respectively. The community-function landscape is then a function $F(\mathbf{x})$ assigning a quantitative output, such as biomass or a metabolite production, to each community composition (Fig. 1B, and Supplementary Text S2).

Previous work has shown that the functional effect of adding species i to a background community can often be predicted from the function of that background through a simple linear relationship (see Fig. 1B)[21]. Specifically, if $\Delta_i F(\mathbf{x})$ denotes the change in function caused by adding species i to background $\mathbf{x}$ (see Fig. 1A), then

$$\Delta_i F(\mathbf{x}) = a_i + m_i F(\mathbf{x}^{(i,-)}) + \theta_i(\mathbf{x}), \quad (1)$$

where a_i and m_i are species-specific coefficients, and $\theta_i(\mathbf{x})$ captures deviations from the linear trend. We refer to these relationships as Functional Effect Equations (FEEs) (Fig. 1B, and Supplementary Text S3)[21,23].

The functional effect of a species on a community emerges from pairwise and higher-order interactions with the resident community members[21]. Thus, we reasoned that the success of FEEs should reflect a latent, simple underlying organization of interactions at all orders. We tested this hypothesis by rewriting each community-function landscape in interaction space using the Walsh-Hadamard (WH) transform[18,27,28] (see Supplementary Texts S2-S3 for a full description). The WH transform decomposes $F(\mathbf{x})$ into additive, pairwise and higher-order coefficients,

$$F(\mathbf{x}) = f_0 + \sum_i f_i x_i + \sum_{i<j} f_{ij} x_i x_j + \sum_{i<j<k} f_{ijk} x_i x_j x_k \cdots, \quad (2)$$

providing a one-to-one map between measured community functions for all 2^N communities in the landscape and the 2^N interaction coefficients at all orders (Fig. 1 B,C). As we show in the Supplementary Text S3, this representation yields a simple prediction. If the FEE for species i is exact, then WH coefficients involving i are recursively related to coefficients one order below (Fig. 1C, and Supplementary Texts S2 and S3)

$$f_{\mathbf{s}+\mathbf{i}} = c_i f_{\boldsymbol{s}}, \qquad c_i = \frac{m_i}{2+m_i}, \tag{3}$$

where $\boldsymbol{s}$ is a binary vector denoting a subset of species not containing i, and $\mathbf{i}$ is the unit vector that adds species i to that subset. Thus, the slope m_i of the FEE predicts the slope c_i of a linear relationship in interaction space: pairwise coefficients involving i should scale with the corresponding additive coefficients, third-order coefficients involving i with pairwise coefficients not containing i, and so on, in all cases obeying the same linear law. Eq. (3) is exact only in the limit of an exact FEE. For empirical FEEs, this proportionality is approximate: deviations from the predicted line are determined by the residuals of the FEE (Supplementary Text S3).

To test this prediction we need a fully sampled community-function landscape. Because complete factorial landscapes are experimentally demanding, existing datasets are scarce and often have limited replication. This is a critical limitation for testing the predicted linear law amongst interaction coefficients, as high-order coefficients are especially sensitive to measurement noise[18]. We therefore constructed complete landscapes comprising all possible combinations (all monocultures, pairs, trios, four-member and the five member community) of five bacterial species (*Pseudomonas aeruginosa PA14 (PA), Pseudomonas putida KT2440-GFP (KT), Pseudomonas atacamensis (P1), Pseudomonas putida (P2)* and *Pseudomonas alloputida (P3)*). Each community was constructed in four independent biological replicates. Communities were inoculated by introducing all present species at an $OD = 0.001$, and then grown for $24\,h$ at $T = 32$ºC in $400\,\mu L$ of minimal carbon-limited M9 media supplemented with citrate and acetate as carbon sources, each at $0.014\,Cmol/L$ ($0.028\,Cmol/L$ total). This medium is similar to those that have been previously used to construct synthetic community-function landscapes[21,22,24,29]. For each community, we measured two different community functions: optical density at $600nm$, as a proxy for biomass, and absorbance of the filtered supernatant at $405nm$, as a proxy for pyoverdine concentration[30,31] (see Methods and Supplementary Text S1 for details). This experiment allowed us to determine the function $F(\mathbf{x})$ for each community composition $\mathbf{x}$.

We first analysed the biomass landscape. Consistent with similar findings in other community-function landscapes[21,24,25], all species exhibited strong FEEs, with R^2 values above ~ 0.8 (Fig. 1B). To test the theoretical prediction that a strongly predictive FEE for the focal species i would imply the existence of a strong linear association between interactions $f_{\mathbf{s+i}}$ and f_s, we directly determined all interaction coefficients through the Walsh-Hadamard transform of $F(\mathbf{x})$ (Fig. 1B). Consistent with our theoretical expectation (Fig. 1C), strong FEEs with $R^2 \sim 0.8\text{-}0.9$ do imply that all interaction coefficients are strongly related to one another, following approximate linear relationships, with $R^2 \sim 0.6$ to ~ 0.7, and slopes consistent with those predicted by Eq. (3). Thus, interactions involving a focal species are not independent across orders, but are statistically constrained by lower-order coefficients. Repeating the same analysis for pyoverdine production yielded qualitatively similar results (Supplementary Fig. S4).

To test the generality of this comparison between theory and experiment, we repeated the entire experiment in a wide range of different growth conditions. Previous work has established that species interactions are sensible to the physico-chemical environment in which species grow[6–9]. Because the dominant form of species interactions in experiments like ours is resource competition and cross-feeding[32], we systematically manipulated the resource composition of the environment, by creating all 31 combinations of 5 different carbon sources: *citrate, acetate, fructose, glucose and glycerol* (Supplementary Text S1). In each of these 31 different growth media we constructed our full factorial community-function landscape in four biological replicates exactly as described above, and measured the same two functions. The good match between theory and experiment was consistent across all environments (Fig. S5-S9), indicating that the existence of simple linear equations constraining interactions at all orders is a robust feature of our community-function landscapes.

To assess to what extent this finding generalizes beyond the interactions that take place in our specific experiments and functions, we re-analyzed six previously published experiments with publicly available data[15,18,24,29,33,34] (Supplementary Table T2). These datasets included both functions similar to those measured in our experiments, such as biomass, and functions extending beyond them. The latter included the population sizes of individual community members in a recent full-factorial study of plant-associated communities[33], which quantify how the presence of one species affects the abundance of another and therefore connect more directly to the population-dynamical interpretation of ecological interactions. We also included community-level functions as diverse as life-history traits of fruit flies grown axenically with every possible combination of five different members of their native

microbiome[34]. Although these prior experiments lacked sufficient replication to estimate higher-order interactions reliably[18], first-, second-, and third-order interaction coefficients frequently followed the same linear co-dependencies observed in our experiments (Supplementary Figs. S24-S26). This suggests that our empirical findings extend beyond our particular experimental system.

Correlated interactions produce ecological collective modes and low-dimensional community-function landscapes

Complexity is often associated with systems that exhibit dense pairwise and higher-order interaction networks. Yet, the fact that interactions at all orders are strongly dependent on each other in community-function landscapes (and that the functional effects of species follow simple linear trends) suggests that such landscapes may be simpler than one might expect based on how many interactions they involve. To formalize this intuition, we construct the vector of functional effects,

$$\mathbf{g}(\mathbf{x}) = \frac{1}{2}(\Delta_1 F(\mathbf{x}), \Delta_2 F(\mathbf{x}), \cdots, \Delta_N F(\mathbf{x})), \tag{4}$$

which captures the effect that every species has on function F on background $\mathbf{x}$, e.g. $\Delta_1 F(\mathbf{x}) = F(1, x_2, x_3, \ldots) - F(-1, x_2, x_3, \ldots)$. From Eq. (1), we see that $\Delta_i F(\mathbf{x})$ is affected by the interactions at all orders involving the focal species i as well as every other member of the community (Supplementary Text S3). If these interactions were all independent, species would produce unrelated functional effects on the same background. By contrast, the effects of different species on the same background are coordinated rather than idiosyncratic when interactions are proportional to one another (Supplementary Text S4-S5). Whether the functional effects of different species vary independently or coherently across backgrounds can be captured by the following matrix (Fig. 2 A, and Supplementary Text S4)

$$M_{ij} = \langle g_i(\mathbf{x}) g_j(\mathbf{x}) \rangle_{\mathbf{x}}. \tag{5}$$

The eigenvectors of $\mathbf{M}$, $\boldsymbol{u}_\alpha$, $\alpha = 1, \cdots, N$, define *ecological collective modes*: linear combinations of species along which functional effects are coherent. If the landscape is effectively low-dimensional, the spectrum of $\mathbf{M}$ should be dominated by a few leading eigenmodes (Fig. 2A). These modes also provide natural coordinates for compressing the original high-dimensional landscape. Projecting a community $\mathbf{x}$ onto an eigenmode $\boldsymbol{u}_\alpha$ defines a latent coordinate,

$$\nu_\alpha(\mathbf{x}) = \boldsymbol{u}_\alpha^{\top}\mathbf{x}. \quad (6)$$

When one mode dominates, community function can be approximated by a one-dimensional non-linear map, $F(\mathbf{x}) \approx q(\nu_1)$. Importantly, as we show on the Supplementary Text S5, this map is a latent projection that can be theoretically calculated from the leading eigenmode $\boldsymbol{u}_1$, the leading eigenvalue λ_1, and the first-order terms (f_i) in the WH expansion of the landscape. In practice, we can evaluate this projection using the second-order truncation of the expansion, motivated by previous work showing that first and pairwise terms capture most of the variance in microbial community-function landscapes [18,22] (Supplementary Text S2 and Supplementary Figs. S2-S3).

When a second mode (ν_2) also contributes significantly, the approximation extends naturally to a response surface $F(\mathbf{x}) \approx Q(\nu_1, \nu_2)$ (Supplementary Text S6). This surface can likewise be evaluated from the two leading eigenmodes of $\mathbf{M}$, their eigenvalues, and the first and second-order interaction coefficients. Importantly, $Q(\nu_1, \nu_2)$ can include nonlinear interaction terms between latent coordinates (Supplementary Text S6). Thus, the collective modes define a natural coarse-graining of the landscape, in which the effects of many species-level interactions are compressed into effective interactions among just two ecological collective modes.

Our experimental data support these predictions. For the community-function landscape represented in Fig. 1, assembled in citrate and acetate as the sole carbon sources, the spectrum of $\mathbf{M}$ was strongly concentrated in two leading collective modes (Fig. 2B,C). For instance, for the biomass function, the leading eigenvalue accounted for $\sim 77\%$, of the total spectral weight and the first two eigenvalues together accounted for $\sim 96\%$ of it. In pyoverdine, the corresponding values were $\sim 80\%$ and $\sim 98\%$. The leading mode was also clearly separated from the second, with $\lambda_1/\lambda_2 = 4.15$ for biomass and $\lambda_1/\lambda_2 = 4.55$ for pyoverdine, indicating a dominant collective axis in both landscapes (Fig. 2B, C). Consistently, the stable rank of $\mathbf{M}$[35]–a measure of the effective number of orthogonal dimensions of a matrix–was close to 1 in both cases (1.06 for biomass and 1.05 for pyoverdine), confirming that the functional-effect matrix is effectively low-rank (see Methods).

Unlike classical biodiversity–ecosystem function relationships, which summarize community structure primarily through species biodiversity (here computed as the number of species in a community; Fig. 2B, C)[29,36–46], the first two modes captured species-specific patterns of functional covariation across community backgrounds. For biomass, we found close

agreement between the theoretical curve $q(\nu_1)$ and the experiments, with a root-mean-square deviation of $RMSD = 0.013$, or $\sim 11.5\%$ of the observed functional range (Fig. 2B). The theoretically expected two-dimensional surface $Q(\nu_1, \nu_2)$ also exhibited a strong agreement with experiments, describing the structure of the landscape in the latent plane with a $RMSD = 0.015$. For pyoverdine, biodiversity was much less informative (Fig. 2C), yet the theoretical $q(\nu_1)$ strongly predicted the empirical pattern with a $RMSD = 0.005$, or $\sim 7.6\%$ of the observed functional range (Fig. 2C). Similarly, the theoretical surface $Q(\nu_1, \nu_2)$ again provided a close representation of the empirical landscape, with $RMSD = 0.0099$.

The resulting latent coordinates $\{\nu_\alpha(\mathbf{x})\}_\alpha$ are analogous to those used in global-epistasis models, where an observed phenotype is represented as a nonlinear function of an underlying additive trait[47,48]. Here, however, the latent trait is not fitted phenomenologically; rather, it emerges from the leading eigenmodes of the functional-effect matrix $\mathbf{M}$. Thus, the existence of predictive FEEs, the fact that interactions at different orders are linked together through simple linear models, and the collapse of the community-function landscape into one or two latent-coordinates, are all different manifestations of the latent simplicity of the community-function landscape (Supplementary Text S7).

This low-dimensional architecture was consistently found in all the other 30 community-function landscapes described in the previous section, each assembled in a different combination of resources. Across all these environments, the first two collective modes consistently captured most of the spectral weight of functional-effects and were separated from the other modes by a pronounced spectral gap (Supplementary Figs. S10–S14). We found similar results for the other previously published full-factorial datasets (Supplementary Figs. S27-S31). Although coefficient-level relationships were often noisier, the spectra of $\mathbf{M}$ were typically concentrated in one or two collective coordinates, and the corresponding latent-coordinate projections provided coherent representations of community function (Supplementary Text S9). This suggests that the latent simplicity observed in our experiments is not an artifact of our particular empirical system, but may be a broader feature of microbial community-function landscapes.

Our results are indeed consistent with recent studies finding that community-function landscapes often have low effective dimensionality, with much of their variation captured by a few collective variables[10,20]. Being inferred through data-driven coarse-graining, dimensionality reduction, or machine learning[10–13], these collective variables typically lack a

clear biological interpretation. To help us answer this question, we must resort to mechanistic ecological models.

Microbial yields give biological meaning to the leading ecological collective mode

To go beyond mere statistical findings, we turned to studying mechanistic models of our microbial communities. These allow us to control species-level traits and potentially connect them to community-level outcomes. Considering the case where the community function is total biomass, we studied Consumer-Resource models (CRMs) in which species compete for either one or multiple public resources[49–51] (Fig. 3A; Supplementary Text S8) as a proxy for our experimental setup. Despite their simplicity, CRMs have proven to capture the behavior of microbial communities reared under similar conditions to the ones we have studied here[52–54], and they exhibit many of the emergent properties of natural microbiomes[49–51, 52–54]. Furthermore, competition for a common resource has been recently proposed as a mechanism for the emergence of FEEs [55]. In such systems, any functional effect or interaction is exclusively mediated by resource depletion, whereby adding a species changes the fraction of the common resources available to the background community. This provides a mechanistic route to coordinated functional effects and low effective rank in $\mathbf{M}$.

To study whether CRMs would display the same general principles of low effective dimensionality of community-function landscapes observed in our experiments, we simulated more than 2.5×10^5 landscapes, systematically accounting for across-species heterogeneity of traits (yield, $\boldsymbol{Y}$, and maximum per-capita growth rate, $\boldsymbol{Z}$) as well as correlations between traits. The full model details are sketched in Fig. 3A and thoroughly described in Supplementary Text S8. The simulations generally reproduce the global epistasis patterns found in experimental data, namely the existence of strongly predictive FEEs, the co-dependence of interactions at all orders following similarly simple trends as those observed empirically, and the collapse of the community-function landscape into one or two latent collective coordinates (Fig. 3B).

Given that CRMs exhibit a similar behavior to our experiments, but they are fully defined mechanistically, we set out to investigate whether we could mathematically explain how the leading mode is connected to species traits. As we elaborate in the Supplementary Text S8, it is possible to show that under empirically realistic limits the leading eigenvector $\boldsymbol{u}_1$of $\mathbf{M}$ is approximately parallel to the vector of yields, $\boldsymbol{Y} = (Y_1, \cdots, Y_N)$ (Fig. 3A). Our theoretical analysis proves that the alignment between the two vectors is primarily controlled by the heterogeneity among yields, as quantified by their coefficient of variation, $CV(\boldsymbol{Y})$. When

yields are narrowly distributed ($CV(\boldsymbol{Y})$ is low), the leading collective mode is strongly aligned with the yield vector ($|cos(\boldsymbol{u}_1, \boldsymbol{Y})|$ approaches 1), independently of how per-capita growth rates ($\boldsymbol{Z}$) are distributed across species. In this regime, the dominant eigenvector is thus mechanistically interpretable in terms of species yields.

This theoretical prediction was first successfully validated through simulations (Supplementary Fig. S16). To test whether, as suggested by the model, the first collective mode $\boldsymbol{u}_1$ does indeed approximately represent the vector of species yields in our experimental communities, we estimated yields from species growth in monoculture in each resource environment (Supplementary Text S1) and then compared the empirical vector of yields with the leading eigenvector inferred from the empirical biomass landscapes. In Fig. 3C we show the comparison between $\boldsymbol{u}_1$ and $\boldsymbol{Y}$ for three representative community-function landscapes with a small $CV(\boldsymbol{Y})$. For all three landscapes we find that the leading eigenvector is very close to the yield vector ($|cos(\boldsymbol{u}_1, \boldsymbol{Y})| = 0.89, 0.96, 0.98$), in agreement with the prediction from the CRM. To assess whether this alignment was significant, we extended this analysis to the remaining 28 community-function landscapes in our previous experiments and compared the empirical cosine similarities against a random-vector null model (Supplementary Text S8). Figure 3C summarizes this comparison across all 31 resource environments, each replicated four times. We find that the empirical distribution of cosine similarities between $\boldsymbol{u}_1$ and $\boldsymbol{Y}$ is strongly shifted toward high values compared to the null expectation, deviating significantly from it (Kolmogorov-Smirnov test, $D(124) = 0.74$, $p < .001$). Consistent with our theoretical predictions, the distribution of $CV(\boldsymbol{Y})$ in our empirical landscapes is highly skewed toward low values (median: 0.31; 95% CI: $[0.06, 1.01]$; skewness: 2.34). This successful test of the theoretical prediction suggests that the leading collective mode in biomass functional landscapes can be directly interpreted from species traits, namely the metabolic efficiencies expressed as biomass production per unit carbon consumed.

This result requires the distribution of yields to be relatively narrow. If yields varied widely across species, the functional effects (and therefore the leading mode $\boldsymbol{u}_1$) would become more sensitive to differences in resource-exploitation ability, which are governed by per-capita growth rates. However, as we show in Supplementary Text S8, when growth rates strongly correlate with yields, species form a competitive hierarchy that aligns $\boldsymbol{u}_1$ to the vector of focal effects $\mathbf{f} = (f_1, \cdots, f_N)$, and $\mathbf{f}$ is in this case entirely determined by $\boldsymbol{Y}$.

Therefore, within CRMs' domain of validity, the leading collective mode can be accurately reconstructed from species traits also in this hierarchical regime.

The leading biomass mode remains stable through population dynamics

An important caveat of the datasets analyzed in the previous sections is that they are all single-batch experiments. All species are always inoculated at an arbitrarily fixed initial density, which allows us to define the composition as a binary vector $\mathbf{x}$. This implies, however, that the results will generally depend on the chosen inoculum density. One may argue that, if communities had been serially passaged instead, a vector containing presence/absence in the initial inoculum would not be informative of community function, as the composition will change at the onset of every passage until communities reach a stable state.

However, an intriguing hypothesis emerges from the results of the previous section. There, we found that the leading collective mode of biomass-production landscapes was strongly aligned to the vector of yields. The growth yield is a species trait related to its metabolic efficiency, which should be relatively robust to changes in the size of its population. Therefore, we hypothesized that the leading biomass mode ($\boldsymbol{u}_1$) should be roughly insensitive to the shifts in community composition that result from ecological interactions and population dynamics as communities approach a stable equilibrium.

To test this hypothesis, we followed a top-down community assembly approach to find a pool of five stably coexisting species in citrate ($0.07\ Cmol/L$) minimal M9 media under daily serial passaging (with a dilution factor $1{:}\,100$) every $24\ h$ at 28ºC. We found a community of five species that coexisted together under this regime: *Acinetobacter johnsonii, Serratia proteomaculans, Pseudomonas defluvii, Aeromonas hydrophila and Lelliottia amnigena* (Fig. 4A). We then assembled all 31 species assemblages we could form with these 5 species, by inoculating together every monoculture, pair, trio, 4-member and 5-member communities at equal initial abundances (Fig. 4A). We grew these assemblages in the same M9 medium supplemented with citrate as the sole carbon source and subjected them to the same daily serial-passaging regime used during the assembly stage (Methods and Supplementary Text S1). We measured total biomass every $24\ h$ for 10 days in two independent biological replicates, with three technical replicates in one experiment and two in the other (Fig. 4A). This experiment generated a fully sampled community-function landscape where the function $F(\mathbf{x}, t)$ maps the function measured on day t of serial passaging to the presence/absence composition vector $\mathbf{x}$ on the day we inoculated our communities. Species

extinctions were rare, and most $(29/31)$ of the communities contained all inoculated species on day 10, even though species abundances differed among the communities and their composition diverged over time from that in the starting inocula (Supplementary Fig. S18).

Remarkably, we found that the binary community-composition vector $\mathbf{x}$, which merely accounts for species presence/absence in the inoculum, remains highly informative of community function despite the underlying population dynamics. FEEs based solely on inoculum composition retained substantial predictive power throughout serial passaging, although their performance changed differently across species. For instance, for *P. defluvii*, the R^2 of the FEE increased from 0.62 after the first single batch (an experiment equivalent to those analyzed in Figs. 1-3), to 0.88 after four days of serial passaging, remaining around 0.9 on days 8-10 (Fig. 4B). A similar, although weaker and less monotonic, improvement was observed for *A. johnsonii* and *L. amnigena*, whereas predictive performance remained approximately stable or more variable for *S. proteomaculans* and *A. hydrophila* (Supplementary Figs. S19-S20).

Consistent with our hypothesis, the leading biomass mode of $F(\mathbf{x},t)$, $\boldsymbol{u}_1(t)$, barely changed over serial passaging and remained aligned with its day-1 orientation, with $|cos(\boldsymbol{u}_1(Day\,1)\,,\boldsymbol{u}_1(t))|\sim[0.96,0.99]$ (Fig. 4C). By contrast, the second mode was less conserved, with $|cos(\boldsymbol{u}_2(Day\,1)\,,\boldsymbol{u}_2(t))|\sim[0.76,0.89]$ at later passages (Fig. 4C). Also consistent with the foundation of our hypothesis, $\boldsymbol{u}_1(t)$ remained strongly aligned with the yield vector throughout the experiment, with $|cos(\boldsymbol{u}_1(t)\,,\boldsymbol{Y})|\sim[0.95,0.98]$ (Fig. 4D). This alignment was visible at the species level when comparing the components of $\boldsymbol{u}_1(t)$ with species yields at representative early and late passages (Fig. 4E and Supplementary Fig. S23).

The strong conservation of $\boldsymbol{u}_1(t)$ over time has an important consequence. If the leading collective coordinate, $\nu_1(\mathbf{x},t)=\boldsymbol{u}_1(t)\cdot\mathbf{x}$, varies little over time, then the projection of the landscape, $F(\mathbf{x},t)\approx q(\nu_1(\mathbf{x},t))$, should follow suit, revealing a temporally stable organization of biomass production (Fig. 4F). Accordingly, the projection of $F(\mathbf{x},t)$ on $\boldsymbol{u}_1$ after a single batch would provide a reasonable approximation of this projection even after several days of serial passaging. As we observed in our experiments, this holds true despite the population dynamics clearly causing the landscape to collapse into an apparently bimodal distribution of biomass over time (Fig. 4F, top panel). Thus, while community dynamics will push different communities towards different stable equilibria, and shift species abundances over time, the

leading collective mode will remain stable and continue to organize variation in biomass production across community backgrounds.

Discussion

The functions of microbial communities emerge from pairwise and higher-order interactions among their members. Simplicity in microbial community function has generally been understood as an absence of interactions: function is predictable when higher-order effects are weak enough to neglect[22], or sparse enough that few of them matter[17]. In the landscapes analyzed here, interactions are reasonably strong at every order and function is nonetheless governed by one or two variables, because interactions at different orders are not independent from each other. For each species, its higher-order interactions are proportional to the corresponding lower-order ones, with a constant of proportionality fixed by the slope of its Functional Effect Equation[21]. Therefore, community functions are simple not because interactions are absent, but because interactions at all orders are proportional to one another.

This finding suggests that the number of interactions in a community provides a poor measure of its complexity, which is set instead by the number of independent ways in which its species affect function. In the dozens of community-function landscapes studied here this number is essentially one or two, each corresponding to a combination of species whose functional effects vary coherently across community backgrounds. These collective modes are inferred from the functional effects they organize, which leaves open what they represent biologically. Variables of this kind are ordinarily obtained by fitting[47,48] or by data-driven dimensionality reduction[10,20], and are correspondingly difficult to interpret. As we have shown for biomass production, the leading mode can instead be generally understood and derived mechanistically as the vector of species yields. Because yields are measured in monoculture, this coordinate can be determined before any community is assembled. Moreover, being species traits, yields are largely unaffected by population dynamics. This robustness is notable because ecological and evolutionary dynamics can restrict the community states that are accessible and thereby constrain responses to artificial selection[56,57], whereas the dominant functional axis identified here remains stable as community composition changes.

Our findings must grapple with several limitations. First, our framework is most directly applicable to fully sampled or densely sampled community-function landscapes, whereas natural microbiomes are usually only sparsely observed, motivating sparse or structure-aware approaches for learning microbiome function[15,17]. Second, we considered

binary community composition in the initial inoculum, although natural communities vary continuously in abundance. While we find that presence/absence at the time of inoculation does a good job at describing community composition even for serially passed communities, it is possible that our results will not extend to every community and for functions beyond biomass. More work needs to be done to explore the limits of this finding. On this note, complementary low-dimensional approaches are being developed for abundance data, including methods based on low-variance ecological constraints in microbial communities [58] and annotation-free inference of functional groups from patterns of taxonomic covariation [59]. Third, the biological interpretation of collective modes is likely function-dependent: while the leading biomass mode can be typically linked to yield, modes associated with other functions, such as secondary-metabolite production, may require different species traits or mechanistic models.[46].

Future work should establish the limits and boundaries of our findings, identifying what ecological mechanisms will create low- versus high-dimensional community-function landscapes. It is plausible that strong niche partitioning, facilitation, or environment-specific metabolic switches may generate multiple competing axes and weaken the dominance of one or two collective modes[17,49]. Consistent with this possibility, replicate communities assembled under similar conditions may converge in core metabolic functions while retaining composition-dependent differences in auxiliary functions[60,61].

More broadly, our theoretical analyses can be applied to any system in which a function is defined over combinations of binary components, as is the case of genetics where these ideas originated[47,48 62,63]. We hope that our work will motivate the ongoing search for unifying principles between ecological community-function landscapes and genotype–phenotype maps[28,62,64–67] and stimulate a similar convergence with recent work on evolutionary collective modes[68], where long-term evolutionary dynamics are organized by collective directions in gene space.

Acknowledgements

We are especially grateful to Alba Rodriguez Carreño for logistic and administrative assistance at all stages of this project.

Methods

Five species–five resources combinatorial assembly experiment

We constructed a fully sampled species-resources community-function landscape using five *Pseudomonas* species and five carbon sources: glycerol, glucose, fructose, citrate and

acetate (Supplementary Table T1 and Fig. S1). All 32 presence-absence combinations of species were crossed with all 32 presence-absence combinations of resources, generating 1,024 species-resources combinations. Communities were assembled in M9 minimal medium and grown for 24 h. Biomass was measured as OD_{600}, and pyoverdine production was measured from culture supernatants as OD_{405}. Combinations with no species, no supplied carbon source, or both, were assigned zero biomass and zero pyoverdine production. The full experiment was performed across four independent biological replicates. Details of strains, media preparation and combinatorial assembly are provided in Supplementary Text S1.

Serial-dilution experiment

To test whether the dominant productivity mode remains stable through ecological dynamics, we assembled all 31 non-empty communities from a second pool of five independently isolated bacterial species enriched for stable coexistence in citrate-limited M9 medium. Communities were grown with citrate as the sole carbon source and propagated by daily 1:100 serial dilution for ten passages. Biomass was measured every 24 h, generating a complete biomass landscape at each passage. Species-level productivity was estimated from monoculture dynamics and compared with the leading collective mode inferred from each daily landscape. Details of the enrichment, community assembly, serial-passaging protocol and yield estimation are provided in Supplementary Text S1.

Landscape analysis

Community-function landscapes were represented on a binary hypercube, with species encoded by presence–absence variables. Interaction coefficients were computed using the Walsh-Hadamard transform. Functional Effect Equations were fitted for each focal species by regressing the effect of adding that species against the function of the corresponding background community. Relationships in interaction space were obtained by comparing coefficients involving the focal species with the corresponding lower-order coefficients not containing it. Mathematical definitions of the Walsh-Hadamard coefficients, Functional Effect Equations and interaction-coefficient relationships are provided in Supplementary Texts S2–S3. All landscape analyses were implemented using custom Python routines developed in the `epistasia2.0` package.

Ecological collective modes

For each landscape, we computed the functional effects of all species across community backgrounds and used them to construct the second-moment matrix of marginal effects. The

eigenvectors of this matrix define ecological collective modes, and the associated eigenvalues quantify the spectral weight captured by each mode. As a complementary measure of effective dimensionality, we computed the stable rank of $\mathbf{M}$[35],

$$R_{stable}(\mathbf{M}) = \frac{\sum_{\alpha=1}^{N} \lambda_\alpha^2}{\lambda_1^2},$$

where λ_1 is the largest eigenvalue of $\mathbf{M}$. Values close to 1 indicate that most of the squared spectral weight is concentrated in a single collective mode. Communities were projected onto latent coordinates defined by the leading modes. One- and two-dimensional reconstructions were obtained by projecting the second-order Walsh-Hadamard approximation of each landscape onto these coordinates. Reconstruction accuracy was quantified using root mean squared deviation. Details of the marginal-effect matrix, spectral decomposition and latent-coordinate reconstructions are provided in Supplementary Texts S4–S6.

Previously published datasets

The same analysis pipeline was applied to previously published fully sampled community-function landscapes spanning different microbial systems, community sizes and measured functions. Dataset descriptions and preprocessing details are provided in Supplementary Text S9 and Supplementary Table T2.

Funding

This work was funded by the European Union (ERC, ECOPROSPECTOR, 101088469) and by projects PID2021-125478NA-I00 and PID2024-160628NB-I00 funded by MICIU/AEI/10.13039/501100011033 and FEDER, UE. B.B.D. acknowledges support from grant PRE2022-103063 funded by MICIU/AEI/10.13039/501100011033 and FSE+. G.B. acknowledges support from grant JDC2024-054148-I funded by MICIU/AEI/10.13039/501100011033 and FSE+. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Code availability

All analyses and figure-generating routines are implemented in the open-source Python package `epistasia2.0`, enabling the computation of FEEs, Walsh-Hadamard interaction

spectra, second-moment matrices, collective modes and latent-coordinate projections from community-function landscapes. Codes and tutorials will be publicly available upon publication.

Data availability

The data will be made publicly available upon official publication of the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Falkowski, P. G., Fenchel, T. & Delong, E. F. The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science* https://doi.org/10.1126/science.1153213 (2008) doi:10.1126/science.1153213.
2. Wang, E.-X., Ding, M.-Z., Ma, Q., Dong, X.-T. & Yuan, Y.-J. Reorganization of a synthetic microbial consortium for one-step vitamin C fermentation. *Microb Cell Fact* **15**, 21 (2016).
3. Wang, Y., Liu, L., Jin, Z. & Zhang, D. Microbial Cell Factories for Green Production of Vitamins. *Front Bioeng Biotechnol* **9**, 661562 (2021).
4. Senne de Oliveira Lino, F., Bajic, D., Vila, J. C. C., Sánchez, A. & Sommer, M. O. A. Complex yeast–bacteria interactions affect the yield of industrial ethanol fermentation. *Nature Communications* **12**, 1498 (2021).
5. Minty, J. J. *et al.* Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. *Proc Natl Acad Sci U S A* **110**, 14592–14597 (2013).
6. Niehaus, L. *et al.* Microbial coexistence through chemical-mediated interactions. *Nat Commun* **10**, 2052 (2019).
7. Lopes, W., Amor, D. R. & Gore, J. Cooperative growth in microbial communities is a driver of multistability. *Nat Commun* **15**, 4709 (2024).
8. Hoek, T. A. *et al.* Correction: Resource Availability Modulates the Cooperative and Competitive Nature of a Microbial Cross-Feeding Mutualism. *PLoS Biol* **15**, e1002606 (2017).
9. Ratzke, C. & Gore, J. Modifying and reacting to the environmental pH can drive bacterial interactions. *PLoS Biol* **16**, e2004248 (2018).
10. Schmitt, M. S. *et al.* Learning functional groups in complex microbiomes. *bioRxiv* (2026) doi:10.64898/2026.03.03.709366.

11. Thompson, J. C., Zavala, V. M. & Venturelli, O. S. Integrating a tailored recurrent neural network with Bayesian experimental design to optimize microbial community functions. *PLOS Computational Biology* **19**, e1011436 (2023).
12. Clark, R. L. *et al.* Design of synthetic human gut microbiome assembly and butyrate production. *Nat Commun* **12**, 3254 (2021).
13. Baranwal, M. *et al.* Recurrent neural networks enable design of multifunctional synthetic human gut microbiome dynamics. *Elife* **11**, (2022).
14. Gallardo-Navarro, O., Aguilar-Salinas, B., Rocha, J. & Olmedo-Álvarez, G. Higher-order interactions and emergent properties of microbial communities: The power of synthetic ecology. *Heliyon* **10**, e33896 (2024).
15. Sanchez-Gorostiaga, A., Bajić, D., Osborne, M. L., Poyatos, J. F. & Sanchez, A. High-order interactions distort the functional landscape of microbial consortia. *PLoS Biol* **17**, e3000550 (2019).
16. Mayfield, M. M. & Stouffer, D. B. Higher-order interactions capture unexplained complexity in diverse communities. *Nat Ecol Evol* **1**, 62 (2017).
17. Arya, S., George, A. B. & O'Dwyer, J. P. Sparsity of higher-order landscape interactions enables learning and prediction for microbiomes. *Proc Natl Acad Sci U S A* **120**, e2307313120 (2023).
18. Camacho-Mateu, J., Burgio, G., Quirós-Rodriguez, I., Fernandez-de-Bobadilla, M. D. & Sanchez, A. On the limits of detection of epistatic higher-order interactions. *bioRxiv* 2026.03.05.709494 (2026) doi:10.64898/2026.03.05.709494.
19. George, A. B. & Korolev, K. S. Ecological landscapes guide the assembly of optimal microbial communities. *PLoS Comput Biol* **19**, e1010570 (2023).
20. Moran, J., Graham, L. C. & Tikhonov, M. Emergent predictability in microbial ecosystems. *Science* **392**, eadr1440 (2026).
21. Diaz-Colunga, J., Skwara, A., Vila, J. C. C., Bajic, D. & Sanchez, A. Global epistasis and the emergence of function in microbial consortia. *Cell* **187**, 3108–3119.e30 (2024).

22. Skwara, A. *et al.* Statistically learning the functional landscape of microbial communities. *Nat Ecol Evol* **7**, 1823–1833 (2023).
23. Sanchez, A. *et al.* The community-function landscape of microbial consortia. *Cell Syst* **14**, 122–134 (2023).
24. Diaz-Colunga, J., Catalan, P., San Roman, M., Arrabal, A. & Sanchez, A. Full factorial construction of synthetic microbial communities. *Elife* **13**, (2026).
25. Ruiz, J. *et al.* Predictability of the community‐function landscape in wine yeast ecosystems. *Molecular Systems Biology* **19**, MSB202311613 (2023).
26. San Román, M., Arrabal, A., Benitez-Dominguez, B., Quirós-Rodríguez, I. & Diaz-Colunga, J. Towards synthetic ecology: strategies for the optimization of microbial community functions. *Front. Synth. Biol.* **3**, 1532846 (2025).
27. Poelwijk, F. J., Krishna, V. & Ranganathan, R. The Context-Dependence of Mutations: A Linkage of Formalisms. *PLoS Comput Biol* **12**, e1004771 (2016).
28. Reddy, G. & Desai, M. M. Global epistasis emerges from a generic model of a complex trait. *Elife* **10**, (2021).
29. Langenheder, S., Bulling, M. T., Solan, M. & Prosser, J. I. Bacterial biodiversity-ecosystem functioning relations are modified by environmental complexity. *PLoS One* **5**, e10834 (2010).
30. Taguchi, F. *et al.* The Siderophore Pyoverdine of Pseudomonas syringae pv. tabaci 6605 Is an Intrinsic Virulence Factor in Host Tobacco Infection. *Journal of Bacteriology* https://doi.org/10.1128/jb.00689-09 (2010) doi:10.1128/jb.00689-09.
31. Kang, D. *et al.* Pyoverdine-Dependent Virulence of Isolates From Cystic Fibrosis Patients. *Front Microbiol* **10**, 2048 (2019).
32. Estrela, S. *et al.* Functional attractors in microbial community assembly. *Cell Syst* **13**, 29–42.e7 (2022).
33. Ishizawa, H. *et al.* Duckweed as an Emerging Model System for Plant–Microbiome Interactions. *Environmental Microbiology* **27**, e70181 (2025).

34. Gould, A. L. *et al.* Microbiome interactions shape host fitness. *Proc Natl Acad Sci U S A* **115**, E11951–E11960 (2018).
35. Rudelson, M. & Vershynin, R. Sampling from large matrices. *Journal of the ACM (JACM)* https://doi.org/10.1145/1255443.1255449 (2007) doi:10.1145/1255443.1255449.
36. Cardinale, B. J. *et al.* Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature* **443**, 989–992 (2006).
37. Bell, T., Newman, J. A., Silverman, B. W., Turner, S. L. & Lilley, A. K. The contribution of species richness and composition to bacterial services. *Nature* **436**, 1157–1160 (2005).
38. Naeem, S., Hahn, D. R. & Schuurman, G. Producer-decomposer co-dependency influences biodiversity effects. *Nature* **403**, 762–764 (2000).
39. Pennekamp, F. *et al.* Biodiversity increases and decreases ecosystem stability. *Nature* **563**, 109–112 (2018).
40. Tilman, D. *et al.* The Influence of Functional Diversity and Composition on Ecosystem Processes. *Science* https://doi.org/10.1126/science.277.5330.1300 (1997) doi:10.1126/science.277.5330.1300.
41. Loreau, M. & Hector, A. Partitioning selection and complementarity in biodiversity experiments. *Nature* **412**, 72–76 (2001).
42. Tilman, D., Wedin, D. & Knops, J. Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature* **379**, 718–720 (1996).
43. Tilman, D. *et al.* Diversity and Productivity in a Long-Term Grassland Experiment. *Science* https://doi.org/10.1126/science.1060391 (2001) doi:10.1126/science.1060391.
44. Hector, A. *et al.* Plant diversity and productivity experiments in european grasslands. *Science* **286**, 1123–1127 (1999).
45. Cardinale, B. J. *et al.* Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67 (2012).
46. Ardichvili, A. N. *et al.* Beyond Biomass: How Interactions Shape Species' Contribution to Ecosystem Functioning. *Ecol Lett* **29**, e70370 (2026).

47. Otwinowski, J., McCandlish, D. M. & Plotkin, J. B. Inferring the shape of global epistasis. *Proc Natl Acad Sci U S A* **115**, E7550–E7558 (2018).
48. Husain, K. & Murugan, A. Physical Constraints on Epistasis. *Mol Biol Evol* **37**, 2865–2874 (2020).
49. Marsland, R., 3rd, Cui, W. & Mehta, P. A minimal model for microbial biodiversity can reproduce experimentally observed ecological patterns. *Sci Rep* **10**, 3308 (2020).
50. MacArthur, R. Species packing and competitive equilibrium for many species. *Theor Popul Biol* **1**, 1–11 (1970).
51. Tilman, D. *Resource Competition and Community Structure*. (Princeton University Press, Princeton, 1983).
52. Corral López, R. *et al.* Imbalance in gut microbial interactions as a marker of health and disease. *Science* **391**, 890–895 (2026).
53. Goldford, J. E. *et al.* Emergent simplicity in microbial community assembly. *Science* **361**, 469–474 (2018).
54. Robert Marsland, Wenping Cui, Joshua Goldford, Pankaj Mehta. The Community Simulator: A Python package for microbial ecology. *PLOS ONE* https://doi.org/10.1371/journal. pone.0230430 (2020) doi:10.1371/journal. pone.0230430.
55. Kuehn, S. Global epistasis in ecosystems arises from resource constraints. *biorXiv* doi:10.64898/2026.05.12.724736.
56. Xie, L. & Shou, W. Steering ecological-evolutionary dynamics to improve artificial selection of microbial communities. *Nat Commun* **12**, 6799 (2021).
57. Xie, L., Yuan, A. E. & Shou, W. Simulations reveal challenges to artificial community selection and possible strategies for success. *PLoS Biol* **17**, e3000295 (2019).
58. Víctor Peris-Yagüe, Stella Perez, Daniel R. Amor, Simona Cocco, Rémi Monasson. Least Component Analysis reveals ecological constraints in microbial communities. *biorXiv* doi:10.64898/2026.05.22.727221.

59. Shan, X., Goyal, A., Gregor, R. & Cordero, O. X. Annotation-free discovery of functional groups in microbial communities. *Nat Ecol Evol* **7**, 716–724 (2023).
60. Bittleston, L. S., Gralka, M., Leventhal, G. E., Mizrahi, I. & Cordero, O. X. Context-dependent dynamics lead to the assembly of functionally distinct microbial communities. *Nature Communications* **11**, 1440 (2020).
61. George, A. B., Wang, T. & Maslov, S. Functional convergence in slow-growing microbial communities arises from thermodynamic constraints. *ISME J* **17**, 1482–1494 (2023).
62. Starr, T. N. & Thornton, J. W. Epistasis in protein evolution. *Protein science : a publication of the Protein Society* **25**, (2016).
63. Kryazhimskiy, S., Rice, D. P., Jerison, E. R. & Desai, M. M. Microbial evolution. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344**, 1519–1522 (2014).
64. Poelwijk, F. J., Socolich, M. & Ranganathan, R. Learning the pattern of epistasis linking genotype and phenotype in a protein. *Nat Commun* **10**, 4213 (2019).
65. Domingo, J., Baeza-Centurion, P. & Lehner, B. The Causes and Consequences of Genetic Interactions (Epistasis). *Annu Rev Genomics Hum Genet* **20**, 433–460 (2019).
66. Bajić, D., Vila, J. C. C., Blount, Z. D. & Sánchez, A. On the deformability of an empirical fitness landscape by microbial evolution. *Proceedings of the National Academy of Sciences* **115**, 11286–11291 (2018).
67. Sailer, Z. R. & Harms, M. J. Detecting High-Order Epistasis in Nonlinear Genotype-Phenotype Maps. *Genetics* **205**, 1079–1088 (2017).
68. Brown, A., Datta, S., Mehta, P. & Cleary, B. From genes to collective modes: biological constraints shape metabolic evolution. *bioRxiv* (2026) doi:10.64898/2026.03.11.711104.

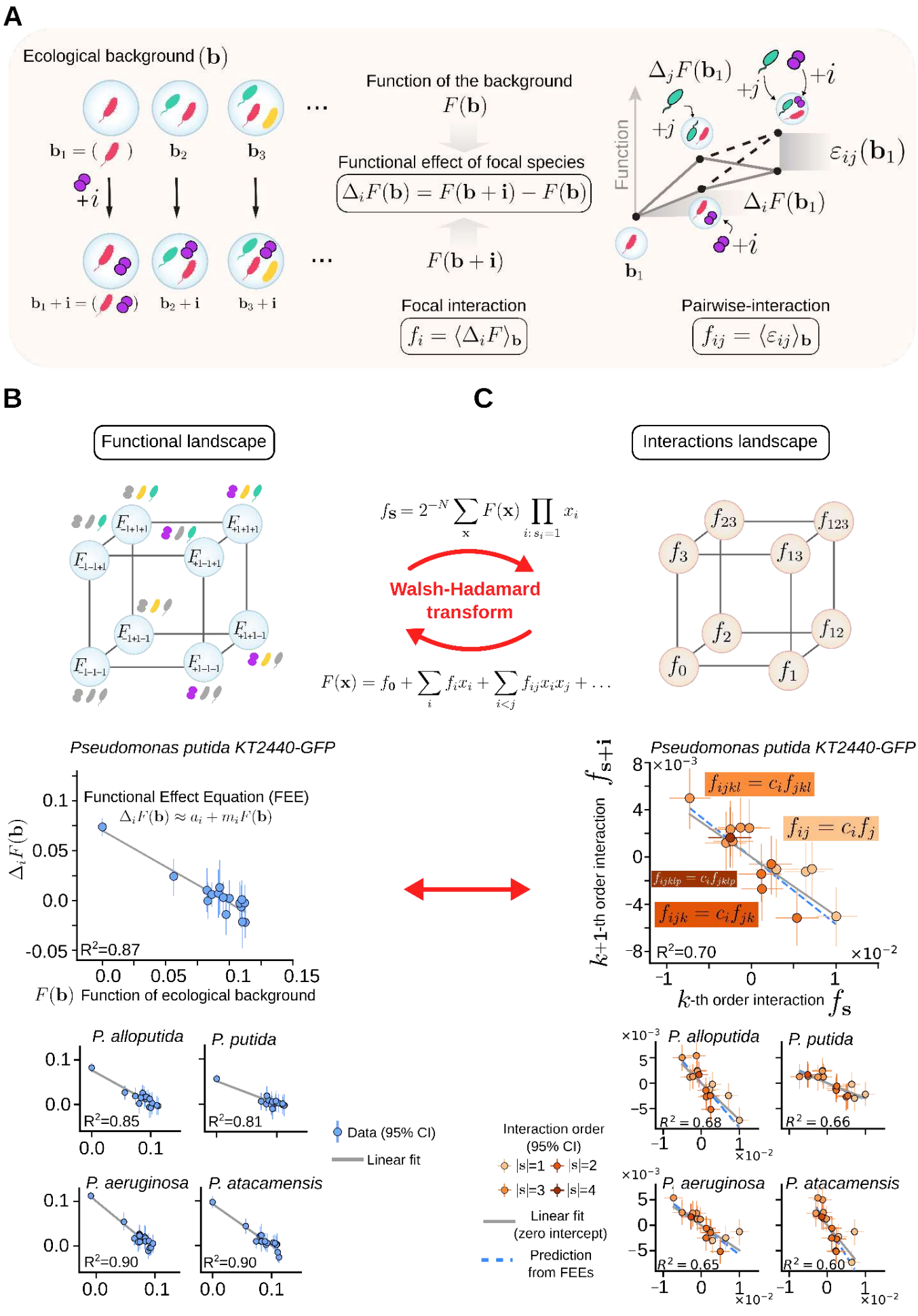
A
Ecological background $(\mathbf{b})$
$\mathbf{b}_1 = (\ )$
$\mathbf{b}_2$
$\mathbf{b}_3$
$+i$
$\mathbf{b}_1 + \mathbf{i} = (\ )$
$\mathbf{b}_2 + \mathbf{i}$
$\mathbf{b}_3 + \mathbf{i}$
Function of the background
$F(\mathbf{b})$
Functional effect of focal species
$\Delta_i F(\mathbf{b}) = F(\mathbf{b}+\mathbf{i}) - F(\mathbf{b})$
$F(\mathbf{b}+\mathbf{i})$
Focal interaction
$f_i = \langle \Delta_i F \rangle_{\mathbf{b}}$
Function
$\Delta_j F(\mathbf{b}_1)$
$+j$
$+i$
$\varepsilon_{ij}(\mathbf{b}_1)$
$\Delta_i F(\mathbf{b}_1)$
$\mathbf{b}_1$
Pairwise-interaction
$f_{ij} = \langle \varepsilon_{ij} \rangle_{\mathbf{b}}$
B
Functional landscape
C
Interactions landscape
$f_{\mathbf{s}} = 2^{-N} \sum_{\mathbf{x}} F(\mathbf{x}) \prod_{i:\, s_i=1} x_i$
Walsh-Hadamard transform
$F(\mathbf{x}) = f_0 + \sum_i f_i x_i + \sum_{i<j} f_{ij} x_i x_j + \dots$
f_{23}
f_{123}
f_3
f_{13}
f_2
f_{12}
f_0
f_1
Pseudomonas putida KT2440-GFP
Functional Effect Equation (FEE)
$\Delta_i F(\mathbf{b}) \approx a_i + m_i F(\mathbf{b})$
R^2=0.87
$\Delta_i F(\mathbf{b})$
$F(\mathbf{b})$ Function of ecological background
Pseudomonas putida KT2440-GFP
$f_{ijkl} = c_i f_{jkl}$
$f_{ij} = c_i f_j$
$f_{ijklp} = c_i f_{jklp}$
$f_{ijk} = c_i f_{jk}$
R^2=0.70
k-th order interaction $f_{\mathbf{s}}$
$k+1$-th order interaction $f_{\mathbf{s}+\mathbf{i}}$
P. alloputida
R^2=0.85
P. putida
R^2=0.81
P. aeruginosa
R^2=0.90
P. atacamensis
R^2=0.90
Data (95% CI)
Linear fit
Interaction order (95% CI)
|s|=1 |s|=2
|s|=3 |s|=4
Linear fit (zero intercept)
Prediction from FEEs
P. alloputida
R^2 = 0.68
P. putida
R^2 = 0.66
P. aeruginosa
R^2 = 0.65
P. atacamensis
R^2 = 0.60

Figure 1. Functional Effect Equations imply linear relationships among ecological interaction coefficients. (A) For a focal species i, the functional effect in an ecological background $\boldsymbol{b}$ is defined as the change in community function upon addition of that species, $\Delta_i F(\boldsymbol{b}) = F(\boldsymbol{b} + \mathbf{i}) - F(\boldsymbol{b})$. Averaging this effect across backgrounds gives the first-order interaction coefficient of the focal species, $f_i = \langle \Delta_i F \rangle_{\mathbf{b}}$. A pairwise interaction ε_{ij} quantifies how the functional effect of species i changes in the presence of another species j, and averaging this context-dependent change across backgrounds gives the pairwise interaction coefficient $f_{ij} = \langle \varepsilon_{ij} \rangle_{\mathbf{b}}$. Higher-order interaction coefficients are defined analogously (Supplementary Text S2). **(B, C; Top)** The same community-function landscape can be represented either in ecological-function space or, after the Walsh-Hadamard transform, Eq. (2), in interaction space. Here, $\mathbf{s}$ is a binary vector indicating the species involved in an interaction, and $|\mathbf{s}| = k$ gives its interaction order. In this interactions representation, an exact FEE predicts that interaction coefficients involving a focal species i, $f_{\mathbf{s}+\mathbf{i}}$, are proportional to the corresponding lower-order coefficients not containing that species, $f_{\mathbf{s}}$, with slope $c_i = m_i/(2 + m_i)$ (Supplementary Text S3). **(B; Bottom)** FEEs measured in the biomass landscape of five bacterial species grown in citrate + acetate medium. Points show functional effects across community backgrounds with *95%* confidence intervals, gray lines show linear fits. All focal species show strong FEE predictability, with $R^2 = 0.81$ to 0.90. **(C; Bottom)** Interaction-coefficient relationships in the same biomass landscape. WH coefficients involving each focal species collapse against the corresponding lower-order coefficients not containing that species. Dashed blue lines show the slopes predicted from the FEEs, and gray lines show empirical linear fits. The interaction-coefficient relationships are strongly linear across focal species, with $R^2 = 0.60$ to 0.70. The corresponding analyses for biomass in the remaining resource environments and for pyoverdine production are shown in Supplementary Figs. S4–S9.

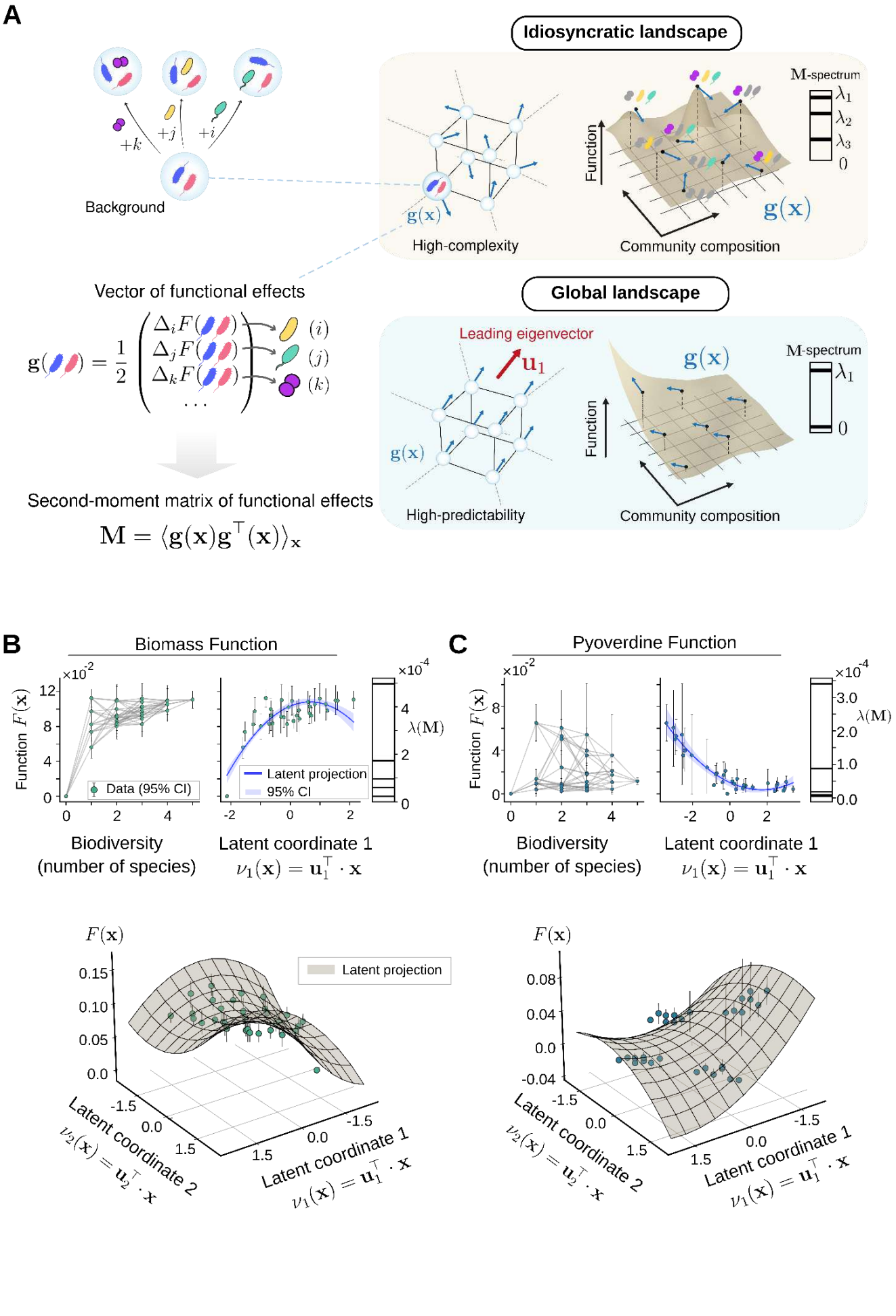
A
Idiosyncratic landscape
+k
+j
+i
Background
g(x)
High-complexity
Function
M-spectrum
λ1
λ2
λ3
0
g(x)
Community composition
Vector of functional effects
g( ) = 1/2 (ΔiF( ) (i)
ΔjF( ) (j)
ΔkF( ) (k)
...
Global landscape
Leading eigenvector
u1
g(x)
High-predictability
Function
g(x)
M-spectrum
λ1
0
Community composition
Second-moment matrix of functional effects
M = ⟨g(x)g⊤(x)⟩x
B
Biomass Function
Function F(x)
×10-2
Data (95% CI)
Latent projection
95% CI
×10-4
λ(M)
Biodiversity
(number of species)
Latent coordinate 1
ν1(x) = u1⊤ · x
C
Pyoverdine Function
Function F(x)
×10-2
×10-4
λ(M)
Biodiversity
(number of species)
Latent coordinate 1
ν1(x) = u1⊤ · x
F(x)
Latent projection
Latent coordinate 2
ν2(x) = u2⊤ · x
Latent coordinate 1
ν1(x) = u1⊤ · x
F(x)
Latent coordinate 2
ν2(x) = u2⊤ · x
Latent coordinate 1
ν1(x) = u1⊤ · x

Figure 2. Correlated ecological interactions generate low-dimensional collective modes in community-function landscapes. **(A)** For each community background, the functional effects of adding each focal species define a functional-effect vector $\mathbf{g}(\mathbf{x})$, **Eq. (4)**. The second-moment matrix of these vectors, $\mathbf{M} = \langle \boldsymbol{g}(\mathbf{x})\boldsymbol{g}^{\top}(\mathbf{x})\rangle_{\mathbf{x}}$, quantifies how the functional effects are coordinated across backgrounds. Idiosyncratic landscapes produce functional-effect vectors pointing in many directions and a broadly distributed $\lambda(\mathbf{M})$-spectrum, whereas structured or global landscapes have aligned functional effects and a spectrum dominated by one or a few collective modes. **(B)** Biomass landscape measured in *citrate* + *acetate* medium. Projection onto community biodiversity (computed as the total number of species) captures the overall increase in biomass with the number of species, but leaves substantial variation among communities with the same biodiversity. Projection onto the leading eigenmode $\boldsymbol{u}_1$ reveals a latent coordinate, $\nu_1(\mathbf{x}) = \boldsymbol{u}_1^{\top} \cdot \mathbf{x}$, that captures the dominant trend. The spectrum of $\mathbf{M}$ is strongly concentrated: the first eigenvalue, accounts for $\sim 77\%$ of the total spectral weight and the first two for $\sim 96\%$. The second-order WH, **Eq. (2)**, projection onto $\nu_1(\mathbf{x})$ yields a one-dimensional latent map, $q(\nu_1)$, that captures the dominant trend with $RMSD = 0.013$, or $\sim 11.5\%$ of the observed functional range (see Supplementary Text S5). Including the second collective coordinate, $\nu_2(\mathbf{x})$, gives a two-dimensional latent surface, $Q(\nu_1, \nu_2)$, that describes the structure of the landscape in the latent plane (see Supplementary Text S6). **(C)** Pyoverdine landscape measured in the same medium. Pyoverdine production is poorly explained by species biodiversity, whereas projection onto $\nu_1(\mathbf{x})$ reveals a clear functional collapse. Here, the first eigenvalue accounts for $\sim 80\%$ of the total spectral weight and including the second eigenvalue for $\sim 98\%$. The latent map $q(\nu_1)$ captures the empirical structure with $RMSD = 0.005$, or $\sim 7.6\%$ of the observed functional range, and $Q(\nu_1, \nu_2)$ provides a compact two-dimensional representation.

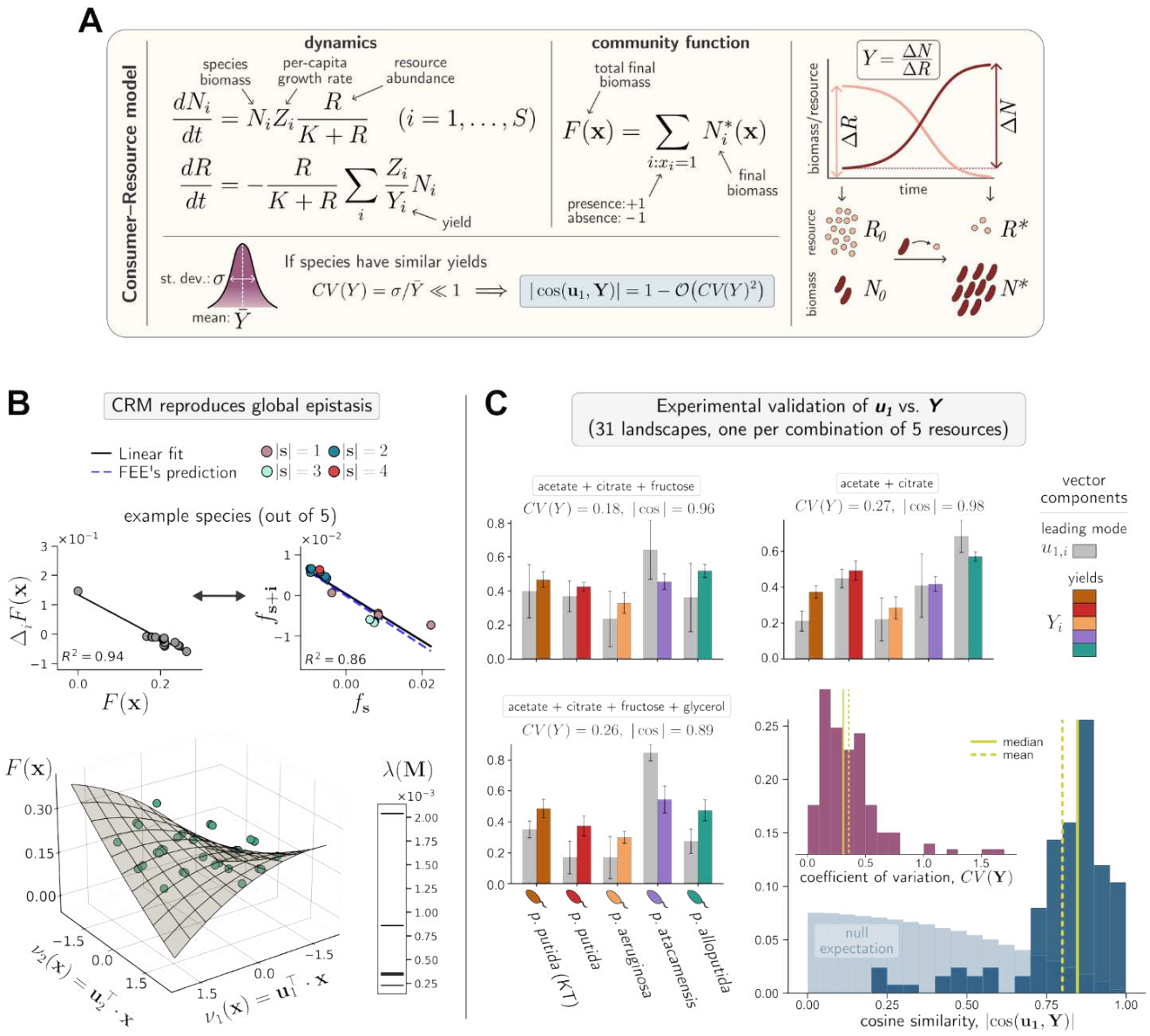
A
Consumer–Resource model
dynamics
species biomass
per-capita growth rate
resource abundance
$\frac{dN_i}{dt} = N_i Z_i \frac{R}{K+R} \quad (i = 1, \ldots, S)$
$\frac{dR}{dt} = -\frac{R}{K+R} \sum_i \frac{Z_i}{Y_i} N_i$
yield
community function
total final biomass
$F(\mathbf{x}) = \sum_{i:x_i=1} N_i^*(\mathbf{x})$
final biomass
presence: +1
absence: −1
$Y = \frac{\Delta N}{\Delta R}$
biomass/resource
time
resource
biomass
R_0
R^*
N_0
N^*
st. dev.: σ
mean: $\bar{Y}$
If species have similar yields
$CV(Y) = \sigma/\bar{Y} \ll 1 \implies |\cos(\mathbf{u}_1, \mathbf{Y})| = 1 - \mathcal{O}(CV(Y)^2)$
B
CRM reproduces global epistasis
Linear fit
FEE's prediction
|s| = 1
|s| = 2
|s| = 3
|s| = 4
example species (out of 5)
$R^2 = 0.94$
$R^2 = 0.86$
$\Delta_i F(\mathbf{x})$
$F(\mathbf{x})$
f_{s+i}
f_s
$\lambda(\mathbf{M})$
$\nu_1(\mathbf{x}) = \mathbf{u}_1^\top \cdot \mathbf{x}$
$\nu_2(\mathbf{x}) = \mathbf{u}_2^\top \cdot \mathbf{x}$
C
Experimental validation of $\boldsymbol{u_1}$ vs. $\boldsymbol{Y}$
(31 landscapes, one per combination of 5 resources)
acetate + citrate + fructose
$CV(Y) = 0.18$, $|\cos| = 0.96$
acetate + citrate
$CV(Y) = 0.27$, $|\cos| = 0.98$
acetate + citrate + fructose + glycerol
$CV(Y) = 0.26$, $|\cos| = 0.89$
vector components
leading mode $u_{1,i}$
yields Y_i
p. putida (KT)
p. putida
p. aeruginosa
p. atacamensis
p. alloputida
median
mean
coefficient of variation, $CV(\mathbf{Y})$
null expectation
cosine similarity, $|\cos(\mathbf{u}_1, \mathbf{Y})|$

Figure 3. Microbial yields give biological meaning to the leading ecological collective mode. (A) Box illustrating the Consumer-Resource model (CRM) we used as a proxy for a community of S species competing for one (multiple) shared resource(s), together with our main analytical prediction. We predict that when yields are relatively similar, as quantified by a low coefficient of variation $CV(\boldsymbol{Y})$, the leading collective mode $\boldsymbol{u}_1$ of matrix $\mathbf{M}$ approximately aligns with the yield vector $\boldsymbol{Y}$ ($|cos(\boldsymbol{u}_1, \boldsymbol{Y})|$ approaches 1; see Supplementary Text S8 for derivation). **(B)** Results from a representative simulated landscape show that the model reproduces the existence of FEEs (for clarity, here shown for one species only) and of collective latent coordinates ($\boldsymbol{u}_1$,$\boldsymbol{u}_2$). **(C)** Comparison of $\boldsymbol{u}_1$ versus $\boldsymbol{Y}$ (normalized) across the 4 biological replicates of the 31 empirical 5-species landscapes we built from all possible combinations of 5 carbon sources (124 data points). The distribution of cosine similarity between $\boldsymbol{u}_1$ and $\boldsymbol{Y}$ (median: 0.85; 95% CI: $[0.26, 0.98]$; skewness: -1.87) is strongly shifted toward high values compared to the null expectation between random normalized vectors in 5 dimensions (Kolmogorov-Smirnov test, $D(124) = 0.74$, $p < .001$). Correspondingly, $CV(\boldsymbol{Y})$ is concentrated at low values (median: 0.31; 95% CI: $[0.06, 1.01]$; skewness: 2.34). The panel also reports the vectors $\boldsymbol{u}_1$ and $\boldsymbol{Y}$ averaged over the 4 replicates (and their standard errors), broken down by components, for three different mixes of carbon sources. The displayed cosine similarities refer to the averaged vectors.

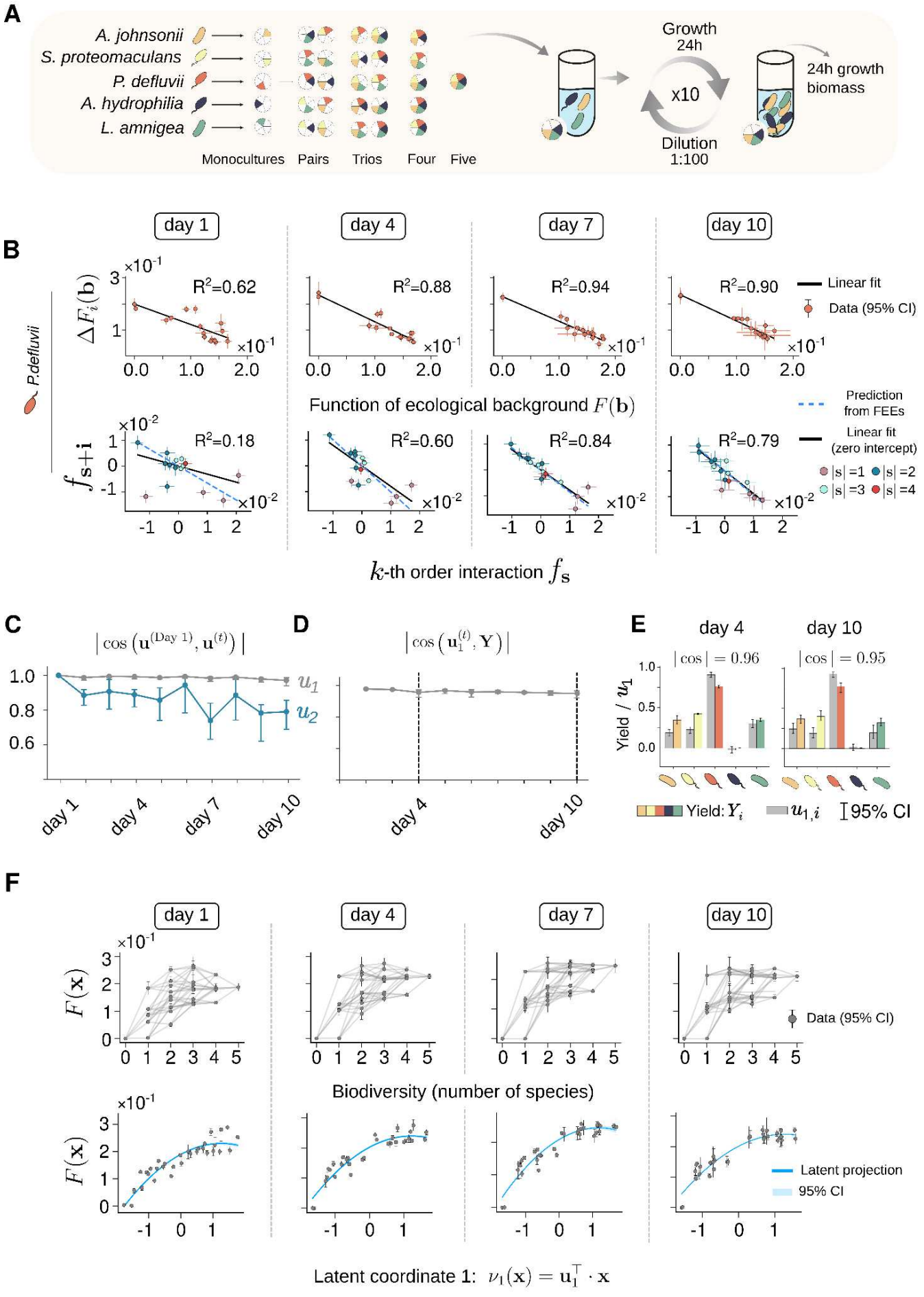
A
A. johnsonii
S. proteomaculans
P. defluvii
A. hydrophilia
L. amnigea
Monocultures
Pairs
Trios
Four
Five
Growth 24h
x10
Dilution 1:100
24h growth biomass
B
P.defluvii
day 1
day 4
day 7
day 10
R²=0.62
R²=0.88
R²=0.94
R²=0.90
R²=0.18
R²=0.60
R²=0.84
R²=0.79
Linear fit
Data (95% CI)
Prediction from FEEs
Linear fit (zero intercept)
|s| =1
|s| =2
|s| =3
|s| =4
Function of ecological background F(b)
k-th order interaction f_s
C
day 1
day 4
day 7
day 10
u1
u2
D
day 4
day 10
E
day 4
day 10
|cos| = 0.96
|cos| = 0.95
Yield / u1
Yield: Y_i
u1,i
95% CI
F
day 1
day 4
day 7
day 10
Data (95% CI)
Biodiversity (number of species)
Latent projection
95% CI
Latent coordinate 1: ν1(x) = u1^T · x

Figure 4. The dominant productivity mode is stable through population dynamics. (A) Experimental design. Complete biomass landscapes were reconstructed during serial dilution by assembling all $(2^5 - 1)$ non-empty communities from five stably coexisting species. Communities were grown for $24\,h$ in citrate-limited M9 medium and serially diluted for 10 daily passages. Biomass was measured at each passage. **(B)** Statistical relationships organizing the biomass landscape for a representative focal species at selected passages (the complete analysis across all focal species and passages is shown in Supplementary Figs. S19-S20). Top: FEEs, Eq. (1), relate the functional effect of adding the focal species to the biomass of the background community. Bottom: interaction-space relationships compare interaction coefficients involving the focal species with the corresponding lower-order coefficients not containing it, as defined in Eq. (3). Both relationships show an initial transient and then become strongly linear through time, increasing from $R^2 = 0.62$ to $R^2 = 0.90$ for FEEs and from $R^2 = 0.18$ to $R^2 = 0.84$ for the interaction-coefficient relationship in the representative example shown. **(C)** Temporal stability of the first two collective modes, quantified as the absolute cosine similarity, $|\boldsymbol{u}_1(Day\,1) \cdot \boldsymbol{u}_1(t)|$, between each mode at passage (t) and its orientation at the first passage $(t = Day\,1)$. The leading mode remains strongly conserved, with $|\boldsymbol{u}_1(Day\,1) \cdot \boldsymbol{u}_1(t)| \sim [0.95, 1.00]$, whereas the second mode is less conserved, $|\boldsymbol{u}_2(Day\,1) \cdot \boldsymbol{u}_2(t)| \sim [0.74, 0.78]$ at later passages. **(D)** Alignment between the leading mode $\boldsymbol{u}_1$ and the yield vector $\boldsymbol{Y}$, estimated from monoculture dynamics across passages (Supplementary Text S1). The high cosine similarity, $|\boldsymbol{u}_1(t) \cdot \boldsymbol{Y}| \sim [0.94, 0.98]$, indicates that $\boldsymbol{u}_1$ remains aligned with a productivity-related trait axis. **(E)** Species-level comparison between the components of $\boldsymbol{u}_1$ and $\boldsymbol{Y}$ at representative passages. The two vectors remain closely aligned at both early and late passages, with $|\boldsymbol{u}_1(t) \cdot \boldsymbol{Y}| = 0.96$ at day 4 and 0.95 at Day10, showing that the conserved leading mode reflects species-specific productivity differences. **(F)** Biomass landscapes projected onto species biodiversity (computed as the total number of species) and onto the leading latent coordinate, $\nu_1(\mathbf{x}) = \boldsymbol{u}_1^{\top} \cdot \mathbf{x}$, across passages. Projection onto $\nu_1(\mathbf{x})$ reveals a stable functional organization of the biomass landscape through time. The corresponding projections for all daily passages are shown in Supplementary Figs. S21-S22.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.pdf