

Supplementary Information to

The latent simplicity of microbial ecological interactions

José Camacho-Mateu^{1,†,*}, Isabel Quirós-Rodriguez^{1,†}, Giulio Burgio^{1,†}, Andrea Arrabal^{1,†}, Alfonso Mendaña¹, Miguel D. Fernandez-de-Bobadilla¹, Belén Benítez-Domínguez¹, and Álvaro Sanchez^{1,*}

¹Institute 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.

Contents

S1	Experimental methods	3
S1.1	Strains and media	3
S1.2	Five species – Five resources combinatorial assembly experiment	3
S1.3	Long-term stability combinatorial assembly experiment	4
S2	From community-function landscapes to interaction space	4
S2.1	Context-dependent interactions as deviations from additivity	5
S2.2	Walsh–Hadamard coefficients as background-averaged interactions	5
S2.3	Variance decomposition across interaction orders	5
S3	Functional Effect Equations constrain interaction coefficients	6
S4	Ecological collective modes and low-dimensional community-function landscapes	7
S4.1	Discrete gradients	7
S4.2	Second-moment matrix of marginal effects	8
S4.3	Discrete curvature and pairwise interactions	9
S4.4	Second-order WH truncation	9
S4.5	Closed-form expressions in the quadratic WH model	10
S5	Rank–1 limit: single ecological collective mode	10
S5.1	Approximate rank–1 regime	11
S5.2	Functional collapse and latent projection	11
S5.3	Gradient covariance structure	12
S6	Multi-dimensional latent manifolds	12
S7	Unifying global epistasis	13
S7.1	Low-order interpretation of FEE predictability	14
S7.2	From per-component predictability to spectral concentration	15
S7.3	Connection to latent-coordinate models	16
S7.4	Collective epistasis as interaction-induced variability	16
S8	Mechanistic origin of global epistasis in consumer–resource systems	17
S8.1	Biomass landscape	17
S8.1.1	Small yield heterogeneity: yield-driven regime	19
S8.1.2	Strongly correlated yields and per-capita growth rates: hierarchical regime	20
S8.1.3	Multiple resources	21
S8.2	Null distribution for pairwise angles	21
S9	Analysis of previously published datasets	22
S10	Supplementary Tables	25
S11	Supplementary Figures	26

S1 Experimental methods

S1.1 Strains and media

We used two five-species bacterial pools in this study. The first pool was used for the five species–five resources combinatorial assembly experiment, and the second for the long-term serial-dilution experiment.

- **Five species–five resources experiment:** *Pseudomonas atacamensis* (P1), *Pseudomonas putida* (P2), *Pseudomonas allopitida* (P3), *Pseudomonas aeruginosa* PA14 (PA) [13], and *Pseudomonas putida* KT2440-GFP (KT) [16]. Strains P1, P2 and P3 were isolated from distinct geographical locations in Madrid, Spain.
- **Serial-dilution experiment:** *Acinetobacter johnsonii*, *Serratia proteomaculans*, *Pseudomonas defluvi*, *Aeromonas hydrophila* and *Lelliottia amnigena*. These strains were individually isolated from distinct geographical locations in Salamanca, Spain. To test whether these strains coexist, they were mixed and were serially passaged for seven days until stable coexistence of the five strains was observed.

Genomic DNA from isolates was subjected to whole-genome sequencing (WGS) on an Illumina NovaSeq X platform paired-end, 2×150 bp reads). Raw reads were processed with a custom Nextflow (DSL2) pipeline [6], including quality trimming with fastp [4], de novo assembly with SPAdes [1], and assembly quality control with QUAST [12] and CheckM2 [5]. Draft genomes were annotated using Bakta v1.12.0 with the full database (v6.0) [21], and taxonomically classified using the classifier module of PATO, a pangenome analysis toolkit [10]. Strains P3, PA and KT were routinely grown at 37°C, whereas the remaining strains were incubated at 28°C, unless otherwise stated. Cell stocks were stored at –80°C in 40% glycerol.

Minimal growth medium consisted of M9 salts supplemented with MgSO₄ and CaCl₂. Resource mixtures were built from five carbon sources: glucose, fructose, glycerol, citrate and acetate (Table T1). Carbon-source stock solutions were prepared at 5× concentration, sterilized using 0.22 µm filters, and diluted during assembly so that each supplied carbon source reached a final concentration of 0.014 Cmol L^{–1}.

S1.2 Five species – Five resources combinatorial assembly experiment

The five *Pseudomonas* species were assembled in all possible presence–absence combinations and crossed with all possible presence–absence combinations of five carbon sources, generating a complete 2⁵ × 2⁵ species–resource design (Fig. S1).

- **Starting inocula:** each strain was streaked from a glycerol stock onto LB agar and grown overnight at their respective strain-specific temperatures. A single colony of each species was inoculated into a 50 mL conical tube containing 25 mL of LB broth (Condalab) and grown overnight under static conditions at 32°C. Cultures were harvested by centrifugation at 2,500 rpm (1,215 × g; Thermo Scientific SL Plus Centrifuge series) for 25 min. Cultures were washed three times in carbon-free M9 medium and adjusted to optical density at 600 nm (OD₆₀₀) = 0.01 using a Multiskan SkyHigh microplate reader (Thermo Fisher Scientific).
- **Species combinations:** all 2⁵ = 32 species combinations were prepared following the previously described protocol by Diaz-Colunga et al. [8]. Whenever a species was present, its contribution was adjusted to OD₆₀₀ = 0.002.
- **Resource combinations:** all 2⁵ = 32 carbon-source combinations were prepared in 96-deep-well plates. The total carbon concentration ranged from 0.014 to 0.07 Cmol L^{–1}, depending on the number of supplied resources.
- **Assembly:** each species combination was crossed with a full carbon-source landscape. A volume of 200 µL of bacterial inoculum was added to 200 µL of resource medium, yielding a final volume of 400 µL. This 1 : 1 dilution brought the contribution of each species present down to an initial OD₆₀₀ = 0.001 per well.

- **Replication and measurements:** the full landscape was assayed across four independent biological replicates. After 24 h of incubation at 32°C, cultures were homogenized, biomass was quantified as OD₆₀₀. Plates were then centrifuged and pyoverdine production was quantified from the supernatant as OD₄₀₅.

S1.3 Long-term stability combinatorial assembly experiment

The serial-dilution experiment was designed to test whether the leading productivity mode remained stable as community composition changed through population dynamics. We used five species that had previously been enriched under daily serial passaging in citrate-limited M9 medium.

- **Species pool:** *Acinetobacter johnsonii*, *Serratia proteomaculans*, *Pseudomonas defluvi*, *Aeromonas hydrophila* and *Lelliottia amnigena*.
- **Inoculum preparation:** strains were streaked from glycerol stocks onto ESBL Chromogenic Medium agar plates (Condalab) and incubated overnight at 28°C. Single colonies were grown overnight in LB broth under static conditions, washed three times in carbon-free M9, and resuspended in citrate-supplemented M9 medium.
- **Community assembly:** all $2^5 = 32$ species combinations were assembled following the protocol described by [8]. Whenever a species was present, its initial contribution was OD₆₀₀ = 0.001, independent of community composition.
- **Growth and dilution:** communities were grown in M9 supplemented with citrate (0.07 mol C L⁻¹). Every 24 h, cultures were homogenized, biomass was measured as OD₆₀₀, and 3.75 μL of culture were transferred into 371.25 μL of fresh medium, corresponding to a 1:100 dilution.
- **Duration and replicability:** the experiment was continued for 10 daily passages, across two independent biological replicates, each comprising three technical replicates. A schematic representation of the experiment is shown in Fig. 4 of the main text.

The final community composition after 10 daily passages was assessed by plating the 31 assembled communities. Species loss was rare: only two final communities were missing one species, as indicated in Fig. S18.

Yield estimation

The yield of species i is defined as the amount of biomass produced per unit of initially available resource, and is estimated here from the net biomass accumulated during one daily growth cycle. The yield can be estimated by comparing the biomass growth of two consecutive days:

$$\tilde{Y}_i = [B_i^{24h}(t) - B_i^{24h}(t-1)/100] \cdot R_0^{-1} \quad (1)$$

where B_i is the biomass of species i in monoculture, R_0 is the initial resource concentration, and the factor 1/100 accounts for the daily dilution. This estimation is used for Fig. 4D-F of the main text.

For single-batch experiments, the yield has been estimated as $(B_i^{24h} - B_i(0)) \cdot R_0^{-1}$, being $B_i(0)$ the initial biomass at inoculum. This estimation is used in Fig. 3C of the main text.

S2 From community-function landscapes to interaction space

We consider ecological systems composed of N binary microbial species. A community configuration is encoded as a binary vector $\mathbf{x} = (x_1, \dots, x_N)$, where $x_i = +1$ denotes presence of species i and $x_i = -1$ its absence. The space of possible communities is therefore $\mathcal{X} = \{-1, +1\}^N$, the vertex set of an N -dimensional hypercube. Each configuration is assigned a quantitative community-level outcome, such as total biomass, community growth rate, or ecosystem productivity. This defines a *community-function landscape*, a mapping $F : \mathcal{X} \rightarrow \mathbb{R}$ assigning a functional value to each combination of species.

S2.1 Context-dependent interactions as deviations from additivity

We define interactions operationally as deviations from additivity in finite differences of the community-function landscape. For a focal species i , its local first-order, or marginal, effect in a given background configuration is

$$\Delta_i F(\mathbf{x}) = F(\mathbf{x}^{(i,+)}) - F(\mathbf{x}^{(i,-)}), \quad (2)$$

where $\mathbf{x}^{(i,+)}$ and $\mathbf{x}^{(i,-)}$ denote the configurations obtained by setting $x_i = +1$ and $x_i = -1$, respectively, while keeping all other coordinates fixed. Thus, $\Delta_i F(\mathbf{x})$ depends only on the non-focal coordinates, which define the background context.

If species acted additively, the marginal effect of i would not depend on the state of another species j . The local second-order, or pairwise, interaction between i and j is therefore the corresponding difference of marginal effects

$$\varepsilon_{ij}(\mathbf{x}) = F(\mathbf{x}^{(i,+,j,+)}) - F(\mathbf{x}^{(i,+,j,-)}) - F(\mathbf{x}^{(i,-,j,+)}) + F(\mathbf{x}^{(i,-,j,-)}). \quad (3)$$

Higher-order local interactions are defined analogously as alternating finite differences over a set of k focal species [3]. We denote them by $\varepsilon_{\mathbf{s}}(\mathbf{x})$, where $\mathbf{s} \in \{0, 1\}^N$ indicates the subset of focal species. This quantity measures, in the specified background context, the deviation of the joint effect of those species from all lower-order contributions.

S2.2 Walsh–Hadamard coefficients as background-averaged interactions

On the discrete hypercube $\mathcal{X} = \{-1, +1\}^N$, we use the uniform inner product

$$\langle F|G \rangle = 2^{-N} \sum_{\mathbf{x} \in \mathcal{X}} F(\mathbf{x})G(\mathbf{x}). \quad (4)$$

Under this inner product, the Walsh–Hadamard (WH) functions form an orthogonal basis for functions $F : \mathcal{X} \rightarrow \mathbb{R}$. Each basis function is indexed by a binary vector $\mathbf{s} \in \{0, 1\}^N$:

$$\phi_{\mathbf{s}}(\mathbf{x}) = \prod_{i:s_i=1} x_i. \quad (5)$$

Here $\mathbf{x}$ denotes a community configuration, whereas $\mathbf{s}$ selects a subset of species. The WH expansion of the landscape is

$$F(\mathbf{x}) = \sum_{\mathbf{s} \in \{0,1\}^N} f_{\mathbf{s}} \phi_{\mathbf{s}}(\mathbf{x}), \quad f_{\mathbf{s}} = \langle F|\phi_{\mathbf{s}} \rangle, \quad (6)$$

where the WH coefficients $f_{\mathbf{s}}$ correspond to background-averaged local interactions. For an index $\mathbf{s}$ of order $n = |\mathbf{s}| = \sum_{i=1}^N s_i$, we have

$$f_{\mathbf{s}} = 2^{-n} \mathbb{E}_{\text{bg}} [\varepsilon_{\mathbf{s}}(\mathbf{x})], \quad (7)$$

where the average is taken over the non-focal coordinates, with $x_i = -1$ for all indices with $s_i = 1$ before applying the finite-difference operator. Thus, after the normalization 2^{-n} , first-order coefficients encode mean marginal effects, whereas higher-order coefficients encode background-averaged deviations from lower-order additivity.

S2.3 Variance decomposition across interaction orders

We define $\text{Var}(F) = \langle F^2 \rangle - \langle F \rangle^2$, with averages taken uniformly over $\mathcal{X}$. Since $f_0 = \langle F \rangle$, orthogonality of the WH basis $\{\phi_{\mathbf{s}}(\mathbf{x})\}$ implies Parseval’s identity,

$$\text{Var}(F) = \sum_{\mathbf{s} \neq \mathbf{0}} f_{\mathbf{s}}^2. \quad (8)$$

Grouping WH coefficients by order $n = |\mathbf{s}|$ yields the exact order-wise decomposition [3, 22]

$$V(n) = \sum_{|\mathbf{s}|=n} f_{\mathbf{s}}^2, \quad \text{Var}(F) = \sum_{n=1}^N V(n). \quad (9)$$

This orthogonal partition holds only when F is defined on the full hypercube and the inner product is uniform. For incomplete landscapes, WH basis functions are no longer orthogonal under the empirical inner product on the sampled configurations, so $V(n)$ no longer defines an exact variance contribution (see [3] for details).

We used this decomposition to quantify how variance is distributed across interaction orders in the empirical biomass and pyoverdine landscapes. Across resource environments, the first two Walsh–Hadamard orders captured the dominant share of variance, typically accounting for approximately 70%–90% of biomass variance and 85%–95% of pyoverdine variance. This motivated the second-order approximation used in the spectral and latent-coordinate analyses (Figs. S2 and S3).

S3 Functional Effect Equations constrain interaction coefficients

Functional Effect Equations (FEEs) relate the effect of adding species i to the function of the background community through simple linear relationships [9],

$$\Delta_i F(\mathbf{x}) = a_i + b_i F(\mathbf{x}^{(i,-)}) + r_i(\mathbf{x}), \quad (10)$$

where $\Delta_i F$ is the unnormalised finite difference defined in Eq. (2), and $r_i(\mathbf{x})$ captures deviations from linearity. We first consider the exact limit $r_i(\mathbf{x}) = 0$. In this limit, FEEs reduce the effective degrees of freedom of the community-function landscape, suggesting an underlying low-dimensional structure.

To make this constraint explicit, we express the landscape in the WH basis, Eq. (6). The unnormalised finite-difference operator acts on WH basis functions as

$$\Delta_i \phi_{\mathbf{s}}(\mathbf{x}) = \begin{cases} 2\phi_{\mathbf{s}-\mathbf{i}}(\mathbf{x}), & s_i = 1, \\ 0, & s_i = 0, \end{cases} \quad (11)$$

where $\mathbf{i}$ is the unit vector in direction i . Therefore,

$$\Delta_i F(\mathbf{x}) = 2 \sum_{\mathbf{s}: s_i=1} f_{\mathbf{s}} \phi_{\mathbf{s}-\mathbf{i}}(\mathbf{x}). \quad (12)$$

Similarly, the background function entering the FEE can be expanded on the reduced subcube $x_i = -1$ as

$$F(\mathbf{x}^{(i,-)}) = \sum_{\mathbf{s}: s_i=0} (f_{\mathbf{s}} - f_{\mathbf{s}+\mathbf{i}}) \phi_{\mathbf{s}}(\mathbf{x}). \quad (13)$$

For an exact FEE, Eq. (10) defines a linear relationship between two functions expanded in the same WH basis on the reduced subcube. Matching coefficients gives, for all $\mathbf{s}$ with $s_i = 0$,

$$(2 + b_i) f_{\mathbf{s}+\mathbf{i}} = a_i \delta_{\mathbf{s}, \mathbf{0}} + b_i f_{\mathbf{s}}. \quad (14)$$

Thus, for non-constant terms $|\mathbf{s}| \geq 1$,

$$f_{\mathbf{s}+\mathbf{i}} = c_i f_{\mathbf{s}}, \quad c_i = \frac{b_i}{2 + b_i}, \quad s_i = 0, \quad (15)$$

whereas the constant term fixes the first-order coefficient,

$$f_i = \frac{a_i + b_i f_{\mathbf{0}}}{2 + b_i}. \quad (16)$$

For compactness, we write f_i , f_{ij} , and more generally f_S for the WH coefficient whose index vector $\mathbf{s}$ has entries equal to one for components in $\{i\}$, $\{i, j\}$, or S , respectively, and zero otherwise. Equation (15) then reads

$$f_{S \cup \{i\}} = c_i f_S, \quad i \notin S, \quad |S| \geq 1. \quad (17)$$

If exact FEEs hold for all focal species on the same landscape, then, for example,

$$f_{ij} = c_i f_j = c_j f_i. \quad (18)$$

This symmetry imposes a consistency condition on the slopes and first-order coefficients. When these consistency conditions hold for every species i and the relevant first-order coefficients are nonzero, the ratios f_i/c_i are equal across species. Denoting this common ratio by κ repeated application of Eq. (17) gives

$$f_S = \kappa^{|S|-1} \prod_{i \in S} f_i, \quad |S| \geq 2. \quad (19)$$

Thus, exact and mutually consistent FEEs imply that WH coefficients across orders are linked by species-specific multiplicative factors. More generally, the residual can be expanded in the WH basis of the reduced subcube as

$$r_i(\mathbf{x}) = \sum_{\mathbf{s}: s_i=0} \rho_{i,\mathbf{s}} \phi_{\mathbf{s}}(\mathbf{x}). \quad (20)$$

Matching coefficients in Eq. (10) then gives

$$(2 + b_i) f_{\mathbf{s}+\mathbf{i}} = a_i \delta_{\mathbf{s},\mathbf{0}} + b_i f_{\mathbf{s}} + \rho_{i,\mathbf{s}}. \quad (21)$$

Therefore, deviations from the exact cross-order recursion are directly determined by the WH coefficients of the FEE residual. Empirical analyses across resource environments are shown for biomass in Figs. S5–S6, and for pyoverdine in Figs. S7–S8. Consistent with Eq. (21), more predictive FEEs are associated with stronger cross-order relationships among WH coefficients (Fig. S9). These relationships are generally noisier than the corresponding FEEs, as expected from the increasing uncertainty of higher-order WH coefficients [3].

S4 Ecological collective modes and low-dimensional community-function landscapes

FEEs constrain WH coefficients across orders, Eqs. (18)–(19). We now express these constraints at the landscape level by collecting all marginal effects into a discrete gradient field and studying the second-moment matrix of this field. Its dominant eigendirections define *ecological collective modes*.

S4.1 Discrete gradients

To study all marginal effects simultaneously, we introduce a discrete gradient field. We define the normalised discrete derivative along coordinate i as

$$D_i F(\mathbf{x}) = \frac{F(\mathbf{x}^{(i,+)}) - F(\mathbf{x}^{(i,-)})}{2} = \frac{\Delta_i F(\mathbf{x})}{2}, \quad (22)$$

where the factor $1/2$ accounts for the spin displacement $\Delta x_i = 2$. This differs from the unnormalised finite difference $\Delta_i F$ used so far. This convention keeps the local finite differences consistent with previous literature, while using the normalised derivative only for the gradient representation. The discrete gradient field is

$$\mathbf{g}(\mathbf{x}) = (D_1 F(\mathbf{x}), \dots, D_N F(\mathbf{x})) \in \mathbb{R}^N. \quad (23)$$

It collects the normalised local marginal effects of all species and plays the role of a gradient field on the hypercube. The normalisation removes the factor of 2 in the finite-difference action on WH basis functions, Eq. (11):

$$D_i \phi_{\mathbf{s}}(\mathbf{x}) = \begin{cases} \phi_{\mathbf{s}-\mathbf{i}}(\mathbf{x}), & s_i = 1, \\ 0, & s_i = 0. \end{cases} \quad (24)$$

Therefore,

$$g_i(\mathbf{x}) = D_i F(\mathbf{x}) = \sum_{\mathbf{s}: s_i=1} f_{\mathbf{s}} \phi_{\mathbf{s}-\mathbf{i}}(\mathbf{x}) = \sum_{\mathbf{s}: s_i=0} f_{\mathbf{s}+\mathbf{i}} \phi_{\mathbf{s}}(\mathbf{x}). \quad (25)$$

Averaging over configurations and using WH orthogonality gives

$$\mathbb{E}_{\mathbf{x}}[D_i F(\mathbf{x})] = f_i. \quad (26)$$

Thus, the background-averaged gradient equals the vector of first-order WH coefficients,

$$\mathbb{E}_{\mathbf{x}}[\mathbf{g}(\mathbf{x})] = (f_1, \dots, f_N)^\top \equiv \mathbf{f}. \quad (27)$$

S4.2 Second-moment matrix of marginal effects

We now introduce the central geometric object of this work,

$$\mathbf{M} = \mathbb{E}_{\mathbf{x}} [\mathbf{g}(\mathbf{x})\mathbf{g}(\mathbf{x})^\top]. \quad (28)$$

The matrix $\mathbf{M}$ measures the second moments and alignments of marginal effects across backgrounds. Let $\mathcal{G}$ be the matrix whose rows are configurations and whose columns are marginal effects,

$$\mathcal{G}_{\mathbf{x}i} = g_i(\mathbf{x}). \quad (29)$$

Then

$$\mathbf{M} = \mathbb{E}_{\mathbf{x}}[\mathbf{g}(\mathbf{x})\mathbf{g}(\mathbf{x})^\top] = 2^{-N} \mathcal{G}^\top \mathcal{G}. \quad (30)$$

Thus, diagonalizing $\mathbf{M}$ is equivalent to performing an uncentred PCA, or equivalently an SVD, of the marginal-effect matrix $\mathcal{G}$. Because $\mathbf{M}$ is symmetric and positive semidefinite, it admits the spectral decomposition

$$\mathbf{M} = \sum_{\alpha=1}^N \lambda_\alpha \mathbf{u}_\alpha \mathbf{u}_\alpha^\top, \quad (31)$$

where $\{\mathbf{u}_\alpha\}$ form an orthonormal basis and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N \geq 0$. Each eigenvector $\mathbf{u}_\alpha$ defines a collective direction in species space along which marginal effects are coordinated, corresponding to an *ecological collective mode*. The associated eigenvalue λ_α quantifies the mean-square magnitude of the marginal-effect field along that direction. Using the WH expansion of g_i , Eq. (25), we obtain, for $i \neq j$,

$$M_{ij} = \sum_{\mathbf{s}: s_i = s_j = 0} f_{\mathbf{s}+\mathbf{i}} f_{\mathbf{s}+\mathbf{j}}, \quad (32)$$

and

$$M_{ii} = \sum_{\mathbf{s}: s_i = 0} f_{\mathbf{s}+\mathbf{i}}^2. \quad (33)$$

Equivalently, $\mathbf{M}$ can be decomposed by WH order,

$$\mathbf{M} = \sum_{n=1}^N \mathbf{M}^{(n)}, \quad (34)$$

where $\mathbf{M}^{(n)}$ contains the contribution of order- n WH coefficients. Its entries are

$$M_{ij}^{(n)} = \sum_{\substack{\mathbf{s}: s_i = s_j = 0 \\ |\mathbf{s}| = n-1}} f_{\mathbf{s}+\mathbf{i}} f_{\mathbf{s}+\mathbf{j}}, \quad i \neq j, \quad (35)$$

and

$$M_{ii}^{(n)} = \sum_{\substack{\mathbf{s}: s_i = 0 \\ |\mathbf{s}| = n-1}} f_{\mathbf{s}+\mathbf{i}}^2. \quad (36)$$

Thus, although $\mathbf{M}$ is built from first-order derivatives, it contains contributions from interactions of all orders. In components,

$$M_{ij} = f_i f_j + \sum_{k \neq i, j} f_{ik} f_{jk} + \sum_{\substack{k < \ell \\ k, \ell \neq i, j}} f_{ik\ell} f_{j\ell k} + \dots \quad (37)$$

Taking the trace gives

$$\text{Tr}(\mathbf{M}^{(n)}) = nV(n), \quad \text{Tr}(\mathbf{M}) = \sum_{\mathbf{s} \neq \mathbf{0}} |\mathbf{s}| f_{\mathbf{s}}^2 = \sum_{n=1}^N nV(n). \quad (38)$$

The total spectral weight of $\mathbf{M}$ is therefore the first moment of the variance spectrum. The empirical spectra of $\mathbf{M}$ across resource environments are shown in Fig. S14.

S4.3 Discrete curvature and pairwise interactions

For species $i \neq j$, we define the normalised second discrete derivative

$$D_{ij}F(\mathbf{x}) = \frac{F(\mathbf{x}^{(i,+,j,+)}) - F(\mathbf{x}^{(i,+,j,-)}) - F(\mathbf{x}^{(i,-,j,+)}) + F(\mathbf{x}^{(i,-,j,-)})}{4} = \frac{\varepsilon_{ij}(\mathbf{x})}{4}. \quad (39)$$

This quantity measures how the normalised marginal effect of i depends on the state of j . It vanishes for additive landscapes and provides a discrete notion of curvature on elementary faces of the hypercube. We define the global Hessian matrix as the background-averaged discrete curvature,

$$H_{ij} \equiv \mathbb{E}_{\mathbf{x}} [D_{ij}F(\mathbf{x})], \quad i \neq j, \quad (40)$$

and set $H_{ii} = 0$. Using the WH expansion, second discrete derivatives act on WH basis functions as

$$D_{ij}\phi_{\mathbf{s}}(\mathbf{x}) = \begin{cases} \phi_{\mathbf{s}-\mathbf{i}-\mathbf{j}}(\mathbf{x}), & s_i = s_j = 1, \\ 0, & \text{otherwise.} \end{cases} \quad (41)$$

It follows that

$$D_{ij}F(\mathbf{x}) = \sum_{\mathbf{s}: s_i=s_j=1} f_{\mathbf{s}} \phi_{\mathbf{s}-\mathbf{i}-\mathbf{j}}(\mathbf{x}). \quad (42)$$

Averaging over configurations and using WH orthogonality gives

$$H_{ij} = f_{ij}. \quad (43)$$

Thus, the global Hessian matrix coincides exactly with the matrix of second-order WH coefficients.

S4.4 Second-order WH truncation

The WH expansion Eq. (6) can be grouped by interaction order as

$$F(\mathbf{x}) = f_0 + \sum_{i=1}^N f_i \phi_i(\mathbf{x}) + \sum_{i < j} f_{ij} \phi_{ij}(\mathbf{x}) + \dots \quad (44)$$

In community-function landscapes studied to date, the variance spectrum is typically dominated by first- and second-order terms, $V(1) + V(2) \gg \sum_{n \geq 3} V(n)$, so the landscape can be approximated by retaining only additive and pairwise WH coefficients [3, 22]. Since $\phi_i(\mathbf{x}) = x_i$ and $\phi_{ij}(\mathbf{x}) = x_i x_j$, the second-order truncation is

$$F(\mathbf{x}) \approx f_0 + \mathbf{f}^\top \mathbf{x} + \frac{1}{2} \mathbf{x}^\top \mathbf{H} \mathbf{x}, \quad (45)$$

where $\mathbf{f} = (f_1, \dots, f_N)^\top$ and $H_{ij} = f_{ij}$ for $i \neq j$, with $H_{ii} = 0$. Thus, the landscape is approximated by a quadratic function on the hypercube. The vector $\mathbf{f}$ gives the background-averaged slopes, whereas $\mathbf{H}$ gives the background-averaged pairwise curvature.

S4.5 Closed-form expressions in the quadratic WH model

When the WH spectrum is dominated by first- and second-order coefficients, the landscape can be approximated by its second-order WH truncation. The gradient field then has the form

$$\mathbf{g}(\mathbf{x}) \approx \mathbf{f} + \mathbf{H}\mathbf{x}. \quad (46)$$

Equivalently, for each species i ,

$$D_i F(\mathbf{x}) \approx f_i + \sum_{j \neq i} H_{ij} x_j. \quad (47)$$

Taking the mean and variance across configurations, and using $\mathbb{E}[\mathbf{x}] = \mathbf{0}$ and $\mathbb{E}[\mathbf{x}\mathbf{x}^\top] = \mathbf{I}$, gives

$$\mathbb{E}_{\mathbf{x}}[D_i F(\mathbf{x})] = f_i, \quad \text{Var}_{\mathbf{x}}(D_i F) = \sum_{j \neq i} H_{ij}^2. \quad (48)$$

Thus, first-order coefficients set the mean marginal effects, whereas pairwise curvature terms determine their variability across backgrounds. The corresponding second-moment matrix is

$$\mathbf{M}^{(\leq 2)} = \mathbb{E}_{\mathbf{x}} [(\mathbf{f} + \mathbf{H}\mathbf{x})(\mathbf{f} + \mathbf{H}\mathbf{x})^\top], \quad (49)$$

which yields

$$\mathbf{M} \approx \mathbf{M}^{(\leq 2)} = \mathbf{f}\mathbf{f}^\top + \mathbf{H}\mathbf{H}^\top. \quad (50)$$

The term $\mathbf{f}\mathbf{f}^\top$ is the contribution of the mean gradient, whereas $\mathbf{H}\mathbf{H}^\top$ is the covariance of the gradient field induced by pairwise curvature.

S5 Rank-1 limit: single ecological collective mode

We now examine the idealized case in which $\mathbf{M}$ is exactly rank-1,

$$\mathbf{M} = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^\top, \quad \|\mathbf{u}_1\| = 1. \quad (51)$$

Geometrically, this means that all marginal-effect vectors are aligned with a single collective direction. Indeed, for any vector $\mathbf{v}$ orthogonal to $\mathbf{u}_1$,

$$\mathbf{v}^\top \mathbf{M} \mathbf{v} = \mathbb{E}_{\mathbf{x}} [(\mathbf{v}^\top \mathbf{g}(\mathbf{x}))^2] = \lambda_1 (\mathbf{v}^\top \mathbf{u}_1)^2 = 0. \quad (52)$$

Since the integrand is nonnegative and the uniform measure assigns positive weight to every configuration, it follows that

$$\mathbf{v}^\top \mathbf{g}(\mathbf{x}) = 0 \quad \forall \mathbf{x}. \quad (53)$$

Thus, every gradient vector lies in the one-dimensional subspace spanned by $\mathbf{u}_1$. Equivalently, there exists a scalar field $\alpha(\mathbf{x})$ such that

$$\mathbf{g}(\mathbf{x}) = \alpha(\mathbf{x}) \mathbf{u}_1 \quad \forall \mathbf{x}. \quad (54)$$

Since $\|\mathbf{u}_1\| = 1$, this scalar field is the projection of the gradient onto the collective mode,

$$\alpha(\mathbf{x}) = \mathbf{u}_1^\top \mathbf{g}(\mathbf{x}). \quad (55)$$

Its second moment is fixed by the nonzero eigenvalue:

$$\mathbb{E}_{\mathbf{x}} [\alpha^2(\mathbf{x})] = \mathbf{u}_1^\top \mathbf{M} \mathbf{u}_1 = \lambda_1. \quad (56)$$

In the quadratic WH model,

$$\mathbf{g}(\mathbf{x}) \approx \mathbf{f} + \mathbf{H}\mathbf{x}, \quad \mathbf{M} \approx \mathbf{f}\mathbf{f}^\top + \mathbf{H}\mathbf{H}^\top. \quad (57)$$

Exact rank-1 structure of $\mathbf{M}$ implies

$$\mathbf{f} \in \text{span}\{\mathbf{u}_1\}, \quad \text{Im}(\mathbf{H}) \subseteq \text{span}\{\mathbf{u}_1\}. \quad (58)$$

Thus, there exists a scalar B such that

$$\mathbf{f} = B\mathbf{u}_1, \quad \mathbf{H}\mathbf{H}^\top = (\lambda_1 - B^2)\mathbf{u}_1\mathbf{u}_1^\top. \quad (59)$$

Here B is the additive projection along the leading mode, whereas $\lambda_1 - B^2$ is the spectral weight of the background-dependent contribution to the gradient field. If $\mathbf{H}$ were an unconstrained symmetric matrix, $\text{Im}(\mathbf{H}) \subseteq \text{span}\{\mathbf{u}_1\}$ would imply $\mathbf{H} = C\mathbf{u}_1\mathbf{u}_1^\top$, with $C^2 = \lambda_1 - B^2$. However, in the WH representation the pairwise matrix has zero diagonal, $H_{ii} = 0$. Therefore, a nonzero symmetric rank-1 $\mathbf{H}$ is incompatible with this convention, since $Cu_{1,i}^2 = 0$ for all i forces $C = 0$ unless the mode is trivial. Hence, in a strict quadratic WH landscape, exact rank-1 $\mathbf{M}$ is only possible for purely additive landscapes, or as an idealized limiting case.

S5.1 Approximate rank-1 regime

When $\lambda_1 \gg \lambda_2, \dots, \lambda_N$, the gradient field is predominantly aligned with the leading collective mode $\mathbf{u}_1$. We decompose it as

$$\mathbf{g}(\mathbf{x}) = \alpha(\mathbf{x})\mathbf{u}_1 + \mathbf{g}_\perp(\mathbf{x}), \quad \alpha(\mathbf{x}) = \mathbf{u}_1^\top \mathbf{g}(\mathbf{x}), \quad (60)$$

where $\mathbf{g}_\perp(\mathbf{x})$ is orthogonal to $\mathbf{u}_1$. The spectral identity gives

$$\mathbb{E}_{\mathbf{x}} [\alpha^2(\mathbf{x})] = \lambda_1, \quad \mathbb{E}_{\mathbf{x}} [\|\mathbf{g}_\perp(\mathbf{x})\|^2] = \sum_{\beta=2}^N \lambda_\beta. \quad (61)$$

Thus, when $\sum_{\beta=2}^N \lambda_\beta \ll \lambda_1$, most of the mean-square gradient magnitude is captured by a single collective direction. This motivates projecting configurations onto

$$\nu = \mathbf{u}_1^\top \mathbf{x} \quad (62)$$

and seeking an approximate functional collapse of the form $F(\mathbf{x}) \approx q(\nu)$. Because eigenvectors are defined up to an overall sign, we oriented $\mathbf{u}_1$ so that $F(\mathbf{x})$ varied with $\mathbf{u}_1^\top \mathbf{x}$ in the same direction as with species biodiversity.

S5.2 Functional collapse and latent projection

The approximate rank-1 regime motivates a one-dimensional representation of the landscape in terms of the collective coordinate ν defined in Eq. (62). In the quadratic WH approximation, we consider a functional collapse of the form

$$F(\mathbf{x}) \approx q(\nu) = F_0 + B\nu + C_q\nu^2. \quad (63)$$

This corresponds to additive and pairwise coefficients organized around the same collective direction. Indeed,

$$\nu^2 = \sum_i u_{1,i}^2 x_i^2 + 2 \sum_{i < j} u_{1,i} u_{1,j} x_i x_j = 1 + 2 \sum_{i < j} u_{1,i} u_{1,j} x_i x_j, \quad (64)$$

where $\|\mathbf{u}_1\| = 1$ and $x_i^2 = 1$. The diagonal contribution is constant on the spin hypercube and can be absorbed into the intercept, so

$$F_0 = f_0 - C_q. \quad (65)$$

Thus, the zero diagonal of the WH pairwise matrix is compatible with a quadratic functional collapse; it only removes a constant contribution to ν^2 . For empirical landscapes, the linear coefficient along the collective coordinate is obtained by projection,

$$B = \mathbf{f}^\top \mathbf{u}_1. \quad (66)$$

The total second moment of the gradient field along the same direction is

$$\lambda_1 = \mathbf{u}_1^\top \mathbf{M} \mathbf{u}_1. \quad (67)$$

We define the quadratic coefficient used in the one-dimensional spectral prediction as

$$C_q = \rho \sqrt{\lambda_1 - B^2}, \quad (68)$$

where the sign $\rho = \pm 1$ is chosen according to the observed orientation of the projected response. Thus, C_q is fixed in magnitude by the spectral weight of the leading mode and parameterizes the predicted response $q(\nu)$. It should be distinguished from C_H , the coefficient of a specific aligned pairwise WH matrix.

At the coefficient level, a literal aligned pairwise WH structure can be written as

$$H_{ij} = 2C_H u_{1,i} u_{1,j} \quad (i \neq j), \quad H_{ii} = 0. \quad (69)$$

For such a structure,

$$C_q^2 = \lambda_1 - B^2 = \mathbf{u}_1^\top \mathbf{H} \mathbf{H}^\top \mathbf{u}_1 = 4C_H^2 \sum_i u_{1,i}^2 (1 - u_{1,i}^2)^2. \quad (70)$$

Therefore, the spectral coefficient C_q used in the projected response and the WH coefficient C_H are related, but not identical in general, because the pairwise WH matrix has zero diagonal. One-dimensional latent-coordinate projections for biomass and pyoverdine landscapes across resource environments are shown in Figs. S10 and S11.

S5.3 Gradient covariance structure

The exact rank-1 analysis above shows that low rank of $\mathbf{M}$ reflects alignment of the full gradient field, but does not by itself distinguish additive from background-dependent structure. To separate these contributions, we define the centered second moment of the gradient field,

$$\mathbf{\Sigma} = \mathbf{M} - \mathbf{f} \mathbf{f}^\top = \mathbb{E}_{\mathbf{x}} \left[(\mathbf{g}(\mathbf{x}) - \mathbf{f}) (\mathbf{g}(\mathbf{x}) - \mathbf{f})^\top \right]. \quad (71)$$

This matrix captures the background-dependent variability of marginal effects. For a purely additive landscape, $F(\mathbf{x}) = \mathbf{a}^\top \mathbf{x}$, the gradient is constant, $\mathbf{g}(\mathbf{x}) = \mathbf{a}$. Therefore,

$$\mathbf{M} = \mathbf{a} \mathbf{a}^\top, \quad \mathbf{\Sigma} = \mathbf{0}. \quad (72)$$

Thus, $\mathbf{M}$ can be rank-1 even in the absence of background-dependent interactions. In contrast, $\mathbf{\Sigma}$ vanishes for additive landscapes and isolates the component of gradient variation induced by interactions.

The two matrices therefore encode complementary notions of low dimensionality. Low rank of $\mathbf{M}$ indicates that the full marginal-effect field is aligned with a small number of collective directions. Low rank of $\mathbf{\Sigma}$ indicates that the deviations from the mean gradient are low-dimensional. A nontrivial one-dimensional functional collapse requires both the mean gradient and the background-dependent variability to be organized along the same collective coordinate. The corresponding covariance spectra across resource environments are shown in Fig. S15.

S6 Multi-dimensional latent manifolds

When the spectrum of $\mathbf{M}$ is concentrated but not strictly rank-1, the alignment of marginal effects extends to a K -dimensional dominant subspace. Let $\{\lambda_\alpha, \mathbf{u}_\alpha\}_{\alpha=1}^N$ denote the eigenpairs of $\mathbf{M}$, ordered by decreasing eigenvalue, and let $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_K)$ collect the leading eigenvectors. We define the latent coordinates as the orthogonal projections

$$\nu = \mathbf{U}^\top \mathbf{x}, \quad \nu_\alpha = \mathbf{u}_\alpha^\top \mathbf{x}, \quad \alpha = 1, \dots, K. \quad (73)$$

We then approximate the landscape by a quadratic function on this latent subspace,

$$F(\mathbf{x}) \approx Q(\nu) = F_0 + \mathbf{a}^\top \nu + \nu^\top \mathbf{B} \nu. \quad (74)$$

The projected coefficients are defined as

$$\mathbf{a} = \mathbf{U}^\top \mathbf{f}, \quad \mathbf{B} = \mathbf{U}^\top \mathbf{H} \mathbf{U}. \quad (75)$$

Thus, additive effects and pairwise curvature are reorganized in terms of collective coordinates rather than individual species. Because the configurations are uniformly distributed on the spin hypercube, $\mathbb{E}_{\mathbf{x}}[\boldsymbol{\nu}] = \mathbf{0}$ and

$$\mathbb{E}_{\mathbf{x}}[\boldsymbol{\nu} \boldsymbol{\nu}^\top] = \mathbf{U}^\top \mathbb{E}_{\mathbf{x}}[\mathbf{x} \mathbf{x}^\top] \mathbf{U} = \mathbf{I}_K. \quad (76)$$

Therefore, the intercept is chosen so that the projected landscape has the same mean as the WH landscape:

$$\mathbb{E}_{\mathbf{x}}[Q(\boldsymbol{\nu})] = F_0 + \text{Tr}(\mathbf{B}) = f_0. \quad (77)$$

Hence

$$F_0 = f_0 - \text{Tr}(\mathbf{B}). \quad (78)$$

The matrix $\mathbf{B}$ encodes curvature within the latent manifold: diagonal entries describe curvature along individual collective coordinates, whereas off-diagonal entries describe coupling between collective coordinates. Thus, the relevant interaction structure of the landscape is reorganized from taxon/resource-level interactions to interactions among collective ecological modes. The spectral weight captured by the first K modes is

$$R_K = \frac{\sum_{\alpha=1}^K \lambda_\alpha}{\text{Tr}(\mathbf{M})}. \quad (79)$$

Choosing K such that R_K exceeds a prescribed threshold yields a low-dimensional geometric representation of the landscape. Projecting the quadratic expression for $\mathbf{M}$ onto an eigenvector gives

$$\lambda_\alpha = (\mathbf{u}_\alpha^\top \mathbf{f})^2 + \mathbf{u}_\alpha^\top \mathbf{H} \mathbf{H}^\top \mathbf{u}_\alpha = a_\alpha^2 + \|\mathbf{H} \mathbf{u}_\alpha\|^2. \quad (80)$$

Expanding the second term in the complete eigenbasis of $\mathbf{M}$ gives

$$\lambda_\alpha = a_\alpha^2 + \sum_{\beta=1}^N (\mathbf{u}_\beta^\top \mathbf{H} \mathbf{u}_\alpha)^2. \quad (81)$$

If the dominant subspace captures most of the curvature, this relation is approximated by

$$\lambda_\alpha \approx a_\alpha^2 + \sum_{\beta=1}^K B_{\beta\alpha}^2. \quad (82)$$

Thus, the variability captured by each collective direction is partitioned between the projected mean gradient and curvature within the latent subspace. For $K = 1$, this construction reduces to a one-dimensional projected quadratic collapse,

$$Q(\nu_1) = F_0 + a_1 \nu_1 + B_{11} \nu_1^2. \quad (83)$$

Note that $Q(\nu_1, \nu_2 = 0)$ is a slice of the two-dimensional projected manifold and is not, in general, identical to the one-dimensional response $q(\nu_1)$. More generally, for $K > 1$, the landscape is represented as a quadratic latent manifold embedded in the original component space. Directions associated with large eigenvalues define stiff collective modes of community function, whereas directions associated with small eigenvalues correspond to weakly constrained, sloppy modes. Two-dimensional latent manifolds for biomass and pyoverdine landscapes are shown in Figs. [S12](#) and [S13](#).

S7 Unifying global epistasis

The term *global epistasis* is used to describe related but distinct forms of structured background dependence. In evolutionary genetics, global epistasis often denotes systematic relationships between mutational effects and background fitness, such as diminishing returns or increasing costs.

These relationships can be formulated as regression-based models in which the effect of a mutation depends approximately linearly on the fitness of its genetic background [19]. The FEEs used above are the ecological analogue of this formulation for community-function landscapes [7, 9].

A second formulation arises in biophysics and genotype–phenotype mapping, where the landscape is described as a nonlinear function of an additive latent coordinate,

$$F(\mathbf{x}) = q(\nu), \quad \nu = \mathbf{u}^\top \mathbf{x}. \quad (84)$$

In this representation, species combine additively in the latent variable, while nonlinear background dependence is induced by the map q [18, 14]. The framework developed in this work connects these two perspectives through the geometry of the marginal-effect field. The central object is

$$\mathbf{M} = \mathbb{E}_{\mathbf{x}} [\mathbf{g}(\mathbf{x})\mathbf{g}(\mathbf{x})^\top]. \quad (85)$$

Spectral concentration of $\mathbf{M}$ implies that marginal effects vary primarily along one, or a few, collective directions. These directions define latent coordinates for functional collapse and, at the same time, constrain how marginal effects vary across backgrounds.

This connection can be made at different levels of approximation. The exact FEE constraint derived in Section S3 is an all-order statement about WH coefficients, Eq. (17). It is therefore a constraint on background-averaged interaction coefficients, rather than on the full distribution of local microscopic effects [19]. To connect this all-order recursion with the interaction-based interpretation used in previous work, and with the empirical dominance of low-order variance in microbial landscapes, we first specialize to the quadratic WH approximation. In this lowest-order limit, FEE predictability, factorized pairwise interactions, latent-coordinate nonlinearities, and spectral concentration of $\mathbf{M}$ become different projections of the same underlying organization.

S7.1 Low-order interpretation of FEE predictability

Section S3 showed that an exact FEE, imposed as an identity on the reduced subcube, recursively constrains WH coefficients at all orders,

$$f_{\mathbf{s}+\mathbf{i}} = c_i f_{\mathbf{s}}. \quad (86)$$

We now connect this all-order identity with the empirical regression diagnostics used to quantify global epistasis. Previous work derived low-order approximations for the variance of the background function, the variance of the focal effect, and their covariance, leading to approximate expressions for the FEE slope and the coefficient of determination R_i^2 in terms of average effects and average pairwise interactions [9]. Here we rewrite the corresponding R_i^2 approximation in WH notation and show that it has a simple geometric interpretation as an alignment condition.

The coefficient of determination for the FEE of component i is

$$R_i^2 = \frac{\text{Cov}(\Delta_i F, F(\mathbf{x}^{(i,-)}))^2}{\text{Var}(\Delta_i F) \text{Var}(F(\mathbf{x}^{(i,-)}))}. \quad (87)$$

Keeping only the leading terms in the low-order expansion, the background-dependent part of $\Delta_i F$ is determined by the pairwise WH coefficients f_{ij} , whereas the leading variation of the background function on the reduced subcube is determined by $f_j - f_{ij}$. This gives

$$R_i^2 \approx \frac{\left(\sum_{j \neq i} f_{ij} (f_j - f_{ij})\right)^2}{\left(\sum_{j \neq i} f_{ij}^2\right) \left(\sum_{j \neq i} (f_j - f_{ij})^2\right)}. \quad (88)$$

Defining

$$\mathbf{p}^{(i)} = (f_{ij})_{j \neq i}, \quad \mathbf{q}^{(i)} = (f_j - f_{ij})_{j \neq i}, \quad (89)$$

Eq. (88) becomes

$$R_i^2 \approx \frac{(\mathbf{p}^{(i)} \cdot \mathbf{q}^{(i)})^2}{\|\mathbf{p}^{(i)}\|^2 \|\mathbf{q}^{(i)}\|^2} = \cos^2 \theta_i. \quad (90)$$

Thus, $R_i^2 \approx 1$ means that the vector of pairwise WH coefficients involving component i is approximately collinear with the linear component of the background function on the reduced subcube. Equivalently, there exists a scalar α_i such that

$$\mathbf{q}^{(i)} = \alpha_i \mathbf{p}^{(i)}. \quad (91)$$

In components,

$$f_j - f_{ij} = \alpha_i f_{ij}, \quad j \neq i. \quad (92)$$

Solving for f_{ij} gives

$$f_{ij} = c_i f_j, \quad c_i = \frac{1}{1 + \alpha_i}, \quad j \neq i. \quad (93)$$

Therefore, perfect FEE predictability at leading order implies that the row of pairwise WH coefficients associated with component i is proportional to the first-order WH profile of the remaining components. This is the lowest-order, regression-based counterpart of the all-order recursive constraint in Eq. (15).

Conversely, if the pairwise WH coefficients factorize as

$$f_{ij} = \kappa f_i f_j, \quad i \neq j, \quad (94)$$

with $\kappa f_i \neq 0$. For a fixed component i , the vectors entering Eq. (90) become

$$p_j^{(i)} = f_{ij} = \kappa f_i f_j, \quad q_j^{(i)} = f_j - f_{ij} = (1 - \kappa f_i) f_j, \quad j \neq i. \quad (95)$$

Hence

$$\mathbf{q}^{(i)} = \frac{1 - \kappa f_i}{\kappa f_i} \mathbf{p}^{(i)}. \quad (96)$$

The two vectors are therefore collinear, and Eq. (90) gives

$$R_i^2 \approx 1. \quad (97)$$

Thus, factorized pairwise WH coefficients are sufficient to generate perfect FEE predictability at leading order. Together with the previous result, this shows that high FEE predictability and factorized pairwise interaction profiles are equivalent low-order descriptions of the same structure.

S7.2 From per-component predictability to spectral concentration

We now impose Eq. (93) simultaneously for all species and examine the resulting consistency constraints. Assume that, for every component i , there exists a scalar c_i such that

$$f_{ij} = c_i f_j \quad \forall j \neq i. \quad (98)$$

Using symmetry of pairwise coefficients, $f_{ij} = f_{ji}$, we obtain

$$c_i f_j = c_j f_i \quad \forall i \neq j. \quad (99)$$

In the non-degenerate case $f_i \neq 0$, this implies

$$\frac{c_i}{f_i} = \frac{c_j}{f_j} \quad \forall i, j, \quad (100)$$

and therefore there exists a global scalar κ such that

$$c_i = \kappa f_i \quad \forall i. \quad (101)$$

Substituting back gives the factorized pairwise structure

$$f_{ij} = \kappa f_i f_j \quad (i \neq j). \quad (102)$$

Thus, simultaneous FEE predictability at leading order constrains the off-diagonal pairwise coefficients to be organized by the additive profile $\mathbf{f}$.

Within the quadratic WH approximation, this structure gives

$$M_{ii} = f_i^2 [1 + \kappa^2(\|\mathbf{f}\|^2 - f_i^2)], \quad (103)$$

and, for $i \neq j$,

$$M_{ij} = f_i f_j [1 + \kappa^2(\|\mathbf{f}\|^2 - f_i^2 - f_j^2)]. \quad (104)$$

The leading coherent contribution to $\mathbf{M}$ is therefore organized by the outer product $\mathbf{f}\mathbf{f}^\top$. If the additive effects are distributed over many species, so that no single f_i^2 dominates $\|\mathbf{f}\|^2$, the diagonal-exclusion corrections are small and

$$M_{ij} \approx (1 + \kappa^2\|\mathbf{f}\|^2) f_i f_j. \quad (105)$$

In this regime,

$$\mathbf{M} \approx (1 + \kappa^2\|\mathbf{f}\|^2) \mathbf{f}\mathbf{f}^\top. \quad (106)$$

Thus, the coherent part of $\mathbf{M}$ is approximately rank-1, with leading eigenvector aligned with $\mathbf{f}$. In the strict quadratic WH representation, exact rank-1 structure with nonzero pairwise curvature is generally obstructed by the zero diagonal of $\mathbf{H}$. The factorized interaction structure therefore induces spectral concentration, rather than exact rank-1 degeneracy. In this sense, global epistasis appears geometrically as a low-dimensional organization of marginal effects.

In empirical landscapes, R_i^2 is typically high but not exactly one. Then the proportionality condition in Eq. (93) holds only approximately, and the pairwise coefficients can be written as

$$f_{ij} = \kappa f_i f_j + \eta_{ij}, \quad i \neq j. \quad (107)$$

When the residual structure η_{ij} is small or itself low-dimensional, the spectrum of $\mathbf{M}$ remains concentrated along one or a few collective directions. Thus, empirical global epistasis corresponds to spectral concentration of the marginal-effect field rather than exact rank-1 structure.

S7.3 Connection to latent-coordinate models

Classical latent-coordinate models of global epistasis write the observed function as a nonlinear transformation of an additive latent trait,

$$F(\mathbf{x}) = q(\varphi), \quad \varphi = \mathbf{h}^\top \mathbf{x}. \quad (108)$$

In such models, the latent direction $\mathbf{h}$ is either inferred statistically from genotype-phenotype data [18] or associated with a collective physical mode of the system [14]. The framework developed here provides a geometric interpretation of this construction in terms of the marginal-effect field.

In Section S5.2, we showed that when the spectrum of $\mathbf{M}$ is concentrated around a single dominant mode, configurations can be projected onto $\nu_1 = \mathbf{u}_1^\top \mathbf{x}$, where $\mathbf{u}_1$ is the leading eigenvector of the second-moment matrix of marginal effects. In the second-order WH approximation, this gives the quadratic collapse in Eq. (63). Thus, the additive latent trait of classical global-epistasis models is identified with the leading collective coordinate of the marginal-effect field.

This interpretation extends naturally to multiple latent coordinates, as described in Section S6. When higher-order WH coefficients are also organized along the same dominant coordinate, the projected response is no longer restricted to a quadratic polynomial, but becomes a more general nonlinear function $q(\nu_1)$.

S7.4 Collective epistasis as interaction-induced variability

We have focused on $\mathbf{M}$ because its spectral concentration constrains the full marginal-effect field. This is the structure that connects the regression-based view of global epistasis, through FEEs, with the latent-coordinate view, $F(\mathbf{x}) \simeq q(\nu)$. If marginal effects are aligned along one, or a few, collective directions, then they can be predicted from low-dimensional background information and the landscape admits a low-dimensional collapse.

However, as discussed in Section S5.3, $\mathbf{M}$ does not distinguish whether this organization is generated by interactions or by purely additive effects. A purely additive landscape has perfectly

aligned, but background-independent, marginal effects. To isolate the interaction-induced component of marginal-effect variation, we use the centered covariance matrix Σ .

We refer to low-dimensional structure in Σ as *collective epistasis*: interactions generate context dependence, but this context dependence is itself coordinated along a small number of collective directions. Thus, $\mathbf{M}$ captures the global organization underlying FEEs and latent-coordinate collapse, whereas Σ isolates the background-dependent component of that organization.

S8 Mechanistic origin of global epistasis in consumer–resource systems

The geometric framework above applies to any community function F . In this section, we specialize to the case where F is total biomass. Our goal is to identify a mechanistic origin for the leading ecological collective mode $\mathbf{u}_1$ for this function (or any other that is strongly correlated with it), and to determine when marginal effects align across configurations, giving rise to a low-rank structure of $\mathbf{M}$.

As a proxy to our experiments, we model biomass production through Consumer–Resource models (CRMs), these providing a minimal setting in which collective modes can acquire a trait-based interpretation. In their simplest form, the models assume a single public resource accessed by all species. Multiple resources are considered in Sec. S8.1.3.

For any given community $\mathbf{x} \in \{-1, 1\}^S$, species biomasses ($\{N_i\}_{i=1, \dots, S}$) and resource abundance (R) obey

$$\frac{dN_i}{dt} = Z_i N_i q(R), \quad \text{for } i : x_i = 1 \quad (109)$$

$$\frac{dR}{dt} = -q(R) \sum_{i: x_i=1} \frac{Z_i}{Y_i} N_i, \quad (110)$$

where Z_i and Y_i are the (maximum) per-capita growth rate and yield of species i , respectively, and $q : \mathbb{R} \rightarrow [0, 1]$ is a function modulating the per-capita growth rates based on the currently available resource, thereby specifying a particular CRM. We can nonetheless make q disappear from Eqs. (109)-(110) through the reparameterization $t \mapsto qt \equiv \tau(\mathbf{x})$ of the time parameter, which means all CRMs above have the same formal solution, independently of the functional form of q . Equations (109)-(110) then admit the general solution

$$N_i(\tau(\mathbf{x}); \mathbf{x}) = N_i(0) e^{Z_i \tau(\mathbf{x})}, \quad \text{for } i : x_i = 1 \quad (111)$$

$$R(\tau(\mathbf{x}); \mathbf{x}) = R(0) - \sum_{i: x_i=1} C_i(\tau(\mathbf{x}); \mathbf{x}), \quad (112)$$

where

$$C_i(\tau(\mathbf{x}); \mathbf{x}) = N_i(0) \frac{e^{Z_i \tau(\mathbf{x})} - 1}{Y_i} = \frac{N_i(\tau(\mathbf{x})) - N_i(0)}{Y_i} \quad (113)$$

is the total amount of resource i consumed up to the effective time $\tau(\mathbf{x})$. Notice the latter is connected to physical time t through $\tau(\mathbf{x}) = \int_0^t q(R(u; \mathbf{x})) du$.

S8.1 Biomass landscape

We define the community function as the total final biomass,

$$F(\mathbf{x}) = \sum_{i: x_i=1} N_i^*(\mathbf{x}), \quad (114)$$

where $N_i^*(\mathbf{x})$ is the biomass species i reached at the end of the growth cycle. We henceforth assume the cycle is stopped when a fixed amount R^* of resource has been depleted (potentially all of it, $R^* = R(0)$) and that species have equal inocula, $N_i(0) = n_0$. If $C_i(\mathbf{x})$ is the total resource consumed by species i , then $R^* = \sum_{i: x_i=1} C_i(\mathbf{x})$.

Let us now build the second-moment matrix of functional effects, $\mathbf{M} = \langle \Delta F \Delta F^\top \rangle_{\mathbf{x}}$ (notice this is just 4 times the matrix $\langle \mathbf{g} \mathbf{g}^\top \rangle_{\mathbf{x}}$ considered elsewhere in this work). The local effect $\Delta_i F(\mathbf{x})$ reads

$$\Delta_i F(\mathbf{x}) = n_0 + Y_i R^* c_i(\mathbf{x}^{(i,+)}) + \sum_{j: x_j^{(i,-)}=1} Y_j R^* \Delta_i c_j(\mathbf{x}), \quad (115)$$

where we introduced the fraction of resource depleted by species i , $c_i = C_i/R^*$, and the shorthand $\Delta_i c_j(\mathbf{x}) = c_j(\mathbf{x}^{(i,+)}) - c_j(\mathbf{x}^{(i,-)})$. Isolating the average yield from Y_i , as $Y_i = \bar{Y} + \delta Y_i$ (in vectorial form, $\mathbf{Y} = \bar{Y} \mathbf{1} + \delta \mathbf{Y}$), and using the constraint $\sum_{j: x_j=1} \Delta_i c_j(\mathbf{x}) = -c_i(\mathbf{x}^{(i,+)})$ (from R^* being fixed), we can recast Eq. (115) into

$$\Delta_i F(\mathbf{x}) = n_0 + R^* \left[\underbrace{\delta Y_i c_i(\mathbf{x}^{(i,+)}) + \sum_{j: x_j^{(i,-)}=1} \delta Y_j \Delta_i c_j(\mathbf{x})}_{\gamma_i(\mathbf{x})} + \underbrace{\bar{Y} \mathbf{1}_{[\mathbf{x}^{(i,-)}=-\mathbf{1}]}}_{\mu_i(\mathbf{x})} \right], \quad (116)$$

where $\gamma_i(\mathbf{x})$ accounts for resource-mediated species interactions, which are determined by yield deviations ($\delta \mathbf{Y}$), while $\mu_i(\mathbf{x})$ accounts for the remaining contribution from monocultures ($\mathbf{1}_{[\mathbf{x}^{(i,-)}=-\mathbf{1}]}$ checks that $\mathbf{x}^{(i,-)}$ is the empty community). In vectorial form,

$$\Delta F(\mathbf{x}) = n_0 \mathbf{1} + R^* [\boldsymbol{\gamma}(\mathbf{x}) + \boldsymbol{\mu}(\mathbf{x})], \quad (117)$$

with

$$\boldsymbol{\mu}(\mathbf{x}) = \begin{cases} \bar{Y} \mathbf{1} & \text{for } \mathbf{x} = -\mathbf{1} \\ \bar{Y}(\mathbf{1}_k + \mathbf{1})/2 & \text{for } \mathbf{x} = \mathbf{1}_k \text{ } (k = 1, \dots, S) \\ \mathbf{0} & \text{otherwise} \end{cases}, \quad (118)$$

where $\mathbf{1}_k$ is a monoculture vector, having 1 at entry k and -1 elsewhere, and $(\mathbf{1}_k + \mathbf{1})/2$ is thus 1 at k and 0 elsewhere.

It is convenient to express the consumed fraction as $c_i(\mathbf{x}) = G_i(\mathbf{x})/(R^* Y_i)$, being $G_i(\mathbf{x}) = n_0(e^{Z_i \tau(\mathbf{x})} - 1)$ the total biomass produced by species i in community $\mathbf{x}$. By introducing the vector $\mathbf{y}$ of entries $y_i = \delta Y_i/Y_i$, and the matrices $\mathbf{K}(\mathbf{x})$, of elements $K_{ij}(\mathbf{x}) = \mathbf{1}_{[x_j^{(i,-)}=1]} \Delta_i G_j(\mathbf{x})$, and $\mathbf{A}(\mathbf{x}) = [\text{diag}(\{G_i(\mathbf{x}^{(i,+)}\})_i) + \mathbf{K}(\mathbf{x})]/R^*$, we can express the interaction term $\boldsymbol{\gamma}$ as linear transformation of $\mathbf{y}$, i.e., $\boldsymbol{\gamma}(\mathbf{x}) = \mathbf{A}(\mathbf{x})\mathbf{y}$. In this form, the dependence of $\boldsymbol{\gamma}$ on per-capita growth rates ($\mathbf{Z}$) is isolated within $\mathbf{A}$.

Let us now compute $\mathbf{M} = \langle \Delta F \Delta F^\top \rangle_{\mathbf{x}}$ by averaging across communities. The vector $\boldsymbol{\mu}$ has first and second moments $\langle \boldsymbol{\mu} \rangle_{\mathbf{x}} = 2^{1-S} \bar{Y} \mathbf{1}$ and $\langle \boldsymbol{\mu} \boldsymbol{\mu}^\top \rangle_{\mathbf{x}} = 2^{-S} \bar{Y}^2 (\mathbf{1} \mathbf{1}^\top + \mathbf{I})$, being $\mathbf{I}$ the identity matrix. We thus get

$$\begin{aligned} \mathbf{M} &= [n_0^2 + 2^{1-S} n_0 \bar{Y} R^* + 2^{-S} \bar{Y}^2 R^{*2}] \mathbf{1} \mathbf{1}^\top + n_0 R^* [\mathbf{1} (\langle \mathbf{A} \rangle_{\mathbf{x}} \mathbf{y})^\top + \langle \mathbf{A} \rangle_{\mathbf{x}} \mathbf{y} \mathbf{1}^\top] \\ &\quad + R^{*2} [\langle \boldsymbol{\mu} (\mathbf{A} \mathbf{y})^\top \rangle_{\mathbf{x}} + \langle (\mathbf{A} \mathbf{y}) \boldsymbol{\mu}^\top \rangle_{\mathbf{x}}] + R^{*2} \langle \mathbf{A} \rangle_{\mathbf{x}} \mathbf{y} \mathbf{y}^\top \langle \mathbf{A}^\top \rangle_{\mathbf{x}} + 2^{-S} \bar{Y}^2 R^{*2} \mathbf{I}. \end{aligned} \quad (119)$$

Equation (119) provides an exact expression that sets the stage to connect the leading eigenmode $\mathbf{u}_1$ of $\mathbf{M}$ to the yield vector $\mathbf{Y}$. In fact, it already reveals that the structure of $\mathbf{M}$ is primarily governed by yield heterogeneity ($\mathbf{y}$), whereas variation in per-capita growth rate alone (entering only through $\mathbf{A}$) does not affect $\mathbf{M}$. In Sec. S8.1.1, we use perturbation theory to make this statement precise in the limit of small yield variation. Then, in Sec. S8.1.2, we show how $\mathbf{u}_1$ can be accurately predicted from $\mathbf{Y}$ also in a regime where yields and per-capita growth rates are strongly correlated.

Numerical integration of the model. Equations (109)-(110) were integrated numerically using the 4th-order Runge-Kutta method. Integration was stopped when R reduced to 1% of its initial value $R(0)$, i.e., $R^* = 0.01 \cdot R(0)$. The integration was performed for each of the 2^S species combinations and the biomass landscape computed. Species present in a community were all assigned a starting biomass $n_0 = 0.001$. The function q was chosen according to a standard Michaelis-Menten kinetics, $q(R) = R/(K + R)$, being $K = 0.01$ the half-saturation constant. Species traits $\mathbf{Y}$ and

$\mathbf{Z}$ were sampled from uniform distributions of fixed means and variable standard deviations to allow us to systematically adjust their coefficients of variation. The means were fixed to a scale consistent with that observed experimentally, respectively, $\bar{Y} = 5.0$ and $\bar{Z} = 0.2$, upon setting $R(0) = 0.05$. Standard deviations of both were instead varied through steps of 0.15 from 0 to, respectively, $1.5 \times \bar{Y}$ and $1.5 \times \bar{Z}$, so that both $\text{CV}(\mathbf{Y})$ and $\text{CV}(\mathbf{Z})$ took 51 equally-spaced values between 0 and 1.5. For each of these values, we sampled 10 times the vectors $\mathbf{Y}$ and $\mathbf{Z}$, resulting in a total of $510 \times 510 = 260100$ simulated landscapes. Results for $S = 5$ species are shown in Fig. 3B of the main text and in Fig. S16. It is worth noticing that the values chosen for the parameters entering the model are not critical from a qualitative standpoint.

S8.1.1 Small yield heterogeneity: yield-driven regime

Let $y_i \ll 1$ for any species i and expand Eq. (119) with respect to the small parameter $\|\mathbf{y}\| \ll 1$ ($\|\cdot\|$ denotes vector norm). At first order in $\|\mathbf{y}\|$, $\mathbf{y} = \delta\mathbf{Y}/\bar{Y}$, which is orthogonal to $\mathbf{1}$. Correspondingly, $\mathbf{M} = \mathbf{M}^{(0)} + \mathbf{M}^{(1)}$, with $[\mathbf{M}^{(0)}]_{ij} \in \mathcal{O}(1)$ and $[\mathbf{M}^{(1)}]_{ij} \in \mathcal{O}(\|\mathbf{y}\|)$, $\forall (i, j)$; precisely,

$$\mathbf{M}^{(0)} = [n_0^2 + 2^{1-S}n_0\bar{Y}R^* + 2^{-S}\bar{Y}^2R^{*2}] \mathbf{1}\mathbf{1}^\top + 2^{-S}\bar{Y}^2R^{*2}\mathbf{I}, \quad (120)$$

$$\mathbf{M}^{(1)} = n_0R^*[\mathbf{1}(\langle \mathbf{A}^{(0)} \rangle_{\mathbf{x}} \mathbf{y})^\top + \langle \mathbf{A}^{(0)} \rangle_{\mathbf{x}} \mathbf{y} \mathbf{1}^\top] + R^{*2}[\langle \boldsymbol{\mu}(\mathbf{A}^{(0)} \mathbf{y})^\top \rangle_{\mathbf{x}} + \langle (\mathbf{A}^{(0)} \mathbf{y}) \boldsymbol{\mu}^\top \rangle_{\mathbf{x}}]. \quad (121)$$

Since any $\mathcal{O}(\|\mathbf{y}\|^2)$ is discarded, matrix $\mathbf{A}$ only enters in its unperturbed form $\mathbf{A}^{(0)}$, which exclusively depends on growth rates.

The unperturbed matrix, $\mathbf{M}^{(0)}$, has leading eigenvector $\mathbf{u}_1^{(0)} = \mathbf{1}/\sqrt{S}$ and eigenvalue $\lambda_1^{(0)} = \sum_j M_{ij}^{(0)} (\forall i)$, while the remaining $S - 1$ eigenvectors span the degenerate orthogonal subspace associated to the eigenvalue $\lambda_\perp^{(0)} = M_{ii}^{(0)} - M_{ij}^{(0)} (\forall i \neq j)$, leading to a large spectral gap $\lambda_1^{(0)} - \lambda_\perp^{(0)} = S[n_0^2 + 2^{1-S}n_0\bar{Y}R^* + 2^{-S}\bar{Y}^2R^{*2}]$ that increases roughly linearly with pool size S . We can thus resort to non-degenerate perturbation theory to estimate $\mathbf{u}_1$ through the first-order expansion $\mathbf{u}_1 = \mathbf{u}_1^{(0)} + \mathbf{u}_1^{(1)}$. The first-order correction reads

$$\mathbf{u}_1^{(1)} = \sum_{k=2}^S \frac{[\mathbf{u}_k^{(0)\top} \mathbf{M}^{(1)} \mathbf{1}]\mathbf{u}_k^{(0)}}{(\lambda_1^{(0)} - \lambda_\perp^{(0)})\sqrt{S}} = \frac{P_\perp \mathbf{M}^{(1)} \mathbf{1}}{(\lambda_1^{(0)} - \lambda_\perp^{(0)})\sqrt{S}}, \quad (122)$$

where in the second step we inserted the orthogonal projector $\mathbf{P}_\perp = \sum_{k \geq 2} \mathbf{u}_k^{(0)} \mathbf{u}_k^{(0)\top} = \mathbf{I} - \mathbf{1}\mathbf{1}^\top/S$, from the identity $\sum_k \mathbf{u}_k^{(0)} \mathbf{u}_k^{(0)\top} = \mathbf{I}$. $\mathbf{P}_\perp$ filters out the component of $\mathbf{M}^{(1)} \mathbf{1}$ parallel to $\mathbf{1}$, while acting as the identity on the orthogonal subspace. Naming $c_\parallel \mathbf{v}_\parallel$ such component and $c_\perp \mathbf{v}_\perp$ the orthogonal one ($\|\mathbf{v}_\parallel\| = \|\mathbf{v}_\perp\| = 1$), $\mathbf{P}_\perp \mathbf{M}^{(1)} \mathbf{1} = c_\perp \mathbf{v}_\perp$. Importantly, since $\|\mathbf{M}^{(1)} \mathbf{1}\| \in \mathcal{O}(\|\mathbf{y}\|)$, $c_\perp \in \mathcal{O}(\|\mathbf{y}\|)$ as well.

Upon the perturbation, $\mathbf{u}_1$ thus reads

$$\mathbf{u}_1 = \frac{1}{\sqrt{S}} (\mathbf{1} + \rho \mathbf{v}_\perp) + \mathcal{O}\left(\frac{\|\delta\mathbf{Y}\|^2}{\bar{Y}^2}\right), \quad (123)$$

with $\rho = c_\perp/(\lambda_1^{(0)} - \lambda_\perp^{(0)}) \in \mathcal{O}(\|\delta\mathbf{Y}\|/\bar{Y})$. The yield vector, on the other hand, tilts as $\mathbf{Y} = \bar{Y}(\mathbf{1} + \delta\mathbf{Y}/\bar{Y})$. Denoting with $\hat{\mathbf{Y}} = \mathbf{Y}/\|\mathbf{Y}\| = (\mathbf{1} + \delta\mathbf{Y}/\bar{Y})/\sqrt{S} + \mathcal{O}(\|\delta\mathbf{Y}\|^2/\bar{Y}^2)$ the normalized yield vector, we can thus state the main result of this section:

$$\mathbf{u}_1 = \hat{\mathbf{Y}} + \underbrace{\frac{1}{\sqrt{S}} \left(\frac{\rho \bar{Y}}{\|\delta\mathbf{Y}\|} \mathbf{v}_\perp - \delta\hat{\mathbf{Y}} \right)}_{\mathcal{O}(1)} \frac{\|\delta\mathbf{Y}\|}{\bar{Y}} = \hat{\mathbf{Y}} + \mathcal{O}\left(\frac{\|\delta\mathbf{Y}\|}{\bar{Y}}\right); \quad (124)$$

and accordingly,

$$\cos(\mathbf{u}_1, \mathbf{Y}) = 1 - \mathcal{O}(\text{CV}(\mathbf{Y})^2), \quad (125)$$

Like $\mathbf{Y}$, $\mathbf{u}_1$ rotates towards the subspace orthogonal to $\mathbf{1}$ (specifically, in the direction picked up by $\mathbf{M}^{(1)} \mathbf{1}$) by an angle whose magnitude is of the order of $\|\delta\mathbf{Y}\|/\bar{Y}$. The cosine of the angle between

$\mathbf{Y}$ and $\mathbf{u}_1$ thus decreases by an amount of the order of $\text{CV}(\mathbf{Y})^2$, that is, only quadratically in yield heterogeneity. As shown in Figs. S16(A)-(B), numerical simulations corroborate our analytical prediction.

As a special case, we show below that the perturbation $\mathbf{v}_\perp$ to $\mathbf{u}_1$ becomes exactly $\delta\mathbf{Y}/\bar{Y}$ when the per-capita growth rate Z either takes the same value for all species or, more generally, it is drawn from a permutation-symmetric distribution, so that species becomes indistinguishable (except for their yield) upon ensemble averaging.

In a nutshell, the across-species variation in per-capita growth rates determines the direction of the perturbation to $\mathbf{u}_1$, yet it is the across-species variation in yields to establish the order of magnitude of such perturbation. As a consequence, the misalignment between $\mathbf{u}_1$ and $\mathbf{Y}$ is small as long as $\text{CV}(\mathbf{Y})$ is, no matter how growth rates are assigned or distributed.

Permutation-symmetric distribution of per-capita growth rates. Special cases of such distributions are i.i.d. Z s or identical Z s. Under ensemble expectation, species are indistinguishable and matrix $\mathbf{A}^{(0)}$ collapses into the permutation-symmetric form $\mathbf{A}^{(0)}(\mathbf{x}) = (G(\mathbf{x})/R^*)\mathbf{I} + (K(\mathbf{x})/R^*)(\mathbf{1}\mathbf{1}^\top - \mathbf{I})$; respectively, $\langle \mathbf{A}^{(0)} \rangle_{\mathbf{x}} = (\bar{G}/R^*)\mathbf{I} + (\bar{K}/R^*)(\mathbf{1}\mathbf{1}^\top - \mathbf{I})$. Observing that $\mathbf{y}$ is an eigenvector of both $\mathbf{P}_\perp$ and $\mathbf{A}^{(0)}$, the four terms in Eq. (121) acts on $\mathbf{1}$ to give:

1. $[\mathbf{1}(\langle \mathbf{A}^{(0)} \rangle_{\mathbf{x}} \mathbf{y})^\top] \mathbf{1}$
2. $[\langle \mathbf{A}^{(0)} \rangle_{\mathbf{x}} \mathbf{y} \mathbf{1}^\top] \mathbf{1} = S \langle \mathbf{A} \rangle_{\mathbf{x}} \mathbf{y} = [S(\bar{G} + |\bar{K}|)/R^*] \mathbf{y}$ (where we used $\bar{K} < 0$)
3. $[\langle \mu(\mathbf{A}^{(0)} \mathbf{y})^\top \rangle_{\mathbf{x}}] \mathbf{1} = \langle [S(G + |K|)/R^*] \mu \rangle_{\mathbf{x}} (\mathbf{y}^\top \mathbf{1}) = 0$
4. $[\langle (\mathbf{A}^{(0)} \mathbf{y}) \mu^\top \rangle_{\mathbf{x}}] \mathbf{1} = [2^{1-S} S \bar{Y}^2] \mathbf{y}$

Applying $\mathbf{P}_\perp$ on the left, the first term is annihilated, while the second and fourth are unaffected. Putting all together, we get

$$\mathbf{u}_1^{(1)} = \frac{1}{\sqrt{S}} \underbrace{\frac{n_0(\bar{G} + |\bar{K}|) + 2^{1-S} \bar{Y}^2 R^{*2}}{n_0^2 + 2^{1-S} n_0 \bar{Y} R^* + 2^{-S} \bar{Y}^2 R^{*2}}}_{\rho \in \mathcal{O}(1)} \mathbf{y} = \frac{\rho}{\sqrt{S}} \frac{\delta\mathbf{Y}}{\bar{Y}}; \quad (126)$$

and finally

$$\mathbf{u}_1 = \frac{1}{\sqrt{S}} \left(\mathbf{1} + \rho \frac{\delta\mathbf{Y}}{\bar{Y}} \right) + \mathcal{O} \left(\frac{\|\delta\mathbf{Y}\|^2}{\bar{Y}^2} \right). \quad (127)$$

Comparing with Eq. (123), we see that $\mathbf{v}_\perp = \delta\mathbf{Y}/\bar{Y}$ in this case. The leading eigenmode $\mathbf{u}_1$ thus rotates towards the same direction $+\delta\mathbf{Y}$ (since $\rho > 0$) as $\mathbf{Y}$. Specifically, $\mathbf{u}_1 = \hat{\mathbf{Y}} + (\rho - 1) \delta\mathbf{Y}/\bar{Y}$ and $\cos(\mathbf{u}_1, \mathbf{Y}) = 1 - [(\rho - 1)^2/2] \text{CV}(\mathbf{Y})^2$.

S8.1.2 Strongly correlated yields and per-capita growth rates: hierarchical regime

Let us assume that yields and per-capita growth rates are strongly correlated, namely $\mathbf{Z} = a\mathbf{Y} + \mathbf{V}$, with $\mathbf{V} \perp \mathbf{Y}$ such that $\|\mathbf{V}\| \ll a\|\mathbf{Y}\|$. Conveniently, let us label species from most to least dominant, i.e., in descending order of per-capita growth rate (therefore of yield), so that $Z_1 > \dots > Z_S$ ($Y_1 > \dots > Y_S$). Let $i_1(\mathbf{x}) = \min_{j: x_j=1} j$ denote the dominant species in community $\mathbf{x}$.

Assuming that $\text{CV}(\mathbf{Z})$ (hence $\text{CV}(\mathbf{Y})$) is large and there is enough resource to guarantee $Z_{i_1(\mathbf{x})} \tau(\mathbf{x}) \gg 1$ (equivalently, $N_{i_1(\mathbf{x})} \gg n_0$), from $\sum_{i: x_i=1} c_i(\mathbf{x}) = 1$ we obtain the approximation $\tau(\mathbf{x}) \approx \ln(Y_{i_1(\mathbf{x})} R^*/n_0)/Z_{i_1(\mathbf{x})}$, with a relative error smaller than $|\ln(|\mathbf{x}|)/\ln(Y_{i_1(\mathbf{x})} R^*/n_0)| \ll 1$ (notice $|\mathbf{x}|$ indicates the number of species in $\mathbf{x}$). Consequently, $c_{i_1(\mathbf{x})}(\mathbf{x}) \approx 1$, and $c_j(\mathbf{x}) \approx 0$ for any (non-dominant) species $j > i_1(\mathbf{x})$ in $\mathbf{x}$. That is, virtually all the available resource is depleted by the dominant species in the community, while the others barely grow from the inoculum. The matrix $\mathbf{A}(\mathbf{x})$ becomes

$$A_{ij}(\mathbf{x}) \approx \begin{cases} Y_i 1_{[i < i_1(\mathbf{x})]} & \text{for } i = j \\ -Y_j 1_{[j = i_1(\mathbf{x})]} & \text{for } i < j \\ 0 & \text{for } i > j \end{cases} \implies \langle A_{ij}(\mathbf{x}) \rangle_{\mathbf{x}} \approx \begin{cases} Y_i 2^{1-i} & \text{for } i = j \\ -Y_j 2^{1-j} & \text{for } i < j \\ 0 & \text{for } i > j \end{cases} \quad (128)$$

being 2^{1-j} the fraction of communities in which species j is dominant (species j is dominant if all species from 1 to $j-1$ are absent). The factor 2^{1-j} reflects the competitive hierarchy among species, weighting them geometrically from the most to the less dominant one.

Recalling the assumption of abundant resource, which implies $Y_i R^* \gg n_0$ for any i , and assuming $2^{-S} \ll 1$, Eq. (119) reduces to

$$\mathbf{M} = R^{*2} (\langle \mathbf{A} \rangle_{\mathbf{x}} \mathbf{y}) (\langle \mathbf{A} \rangle_{\mathbf{x}} \mathbf{y})^\top + \mathcal{O}(n_0/(\bar{Y} R^*)) + \mathcal{O}(2^{-S}). \quad (129)$$

At leading order, $\mathbf{M}$ is thus a rank-1 matrix whose principal eigenvector $\mathbf{u}_1$ is parallel to the approximate focal effects vector $\mathbf{f}^{(\text{rank})} = R^* \langle \mathbf{A} \rangle_{\mathbf{x}} \mathbf{y}$. Using Eq. (128),

$$u_{1,i} \propto f_i^{(\text{rank})} = 2^{1-i} \left(Y_i - \sum_{k=1}^{S-i} 2^{-k} Y_{i+k} \right) + \mathcal{O}(2^{-S}), \quad (130)$$

which is entirely determined by yields. Therefore, once measured the yield of each species from monocultures, we can accurately reconstruct the leading collective mode from Eq. (130). As shown in Fig. S16(C), numerical simulations confirm this prediction.

S8.1.3 Multiple resources

The results in Secs. S8.1.1-S8.1.2 extend directly to the case of multiple shared resources. Given L resources, $\{R_r\}_{r=1,\dots,L}$, the dynamical equations read

$$\frac{dN_i}{dt} = N_i \sum_{r=1}^m Z_{i,r} q(R_r), \quad \text{for } i : x_i = 1 \quad (131)$$

$$\frac{dR_r}{dt} = -q(R_r) \sum_{i:x_i=1} \frac{Z_{i,r}}{Y_i} N_i, \quad (132)$$

where $Z_{i,r}$ is the (maximum) per-capita growth rate of species i when consuming resource r , and Y_i is the yield of species i in the environment specified by the m present resources. By the latter we mean that L resources together define an environment and in such environment we measure a community-function landscape. Then, the S yields are computed from the S monocultures in such environment. Therefore, we can always assign to species a measurable, ecologically meaningful value of yield ($\{Y_i\}_i$) within each landscape (environment), while keeping the heterogeneity across resources in the values of the growth rates ($\{Z_{i,r}\}_{i,r}$).

The solution to Eqs. (131)-(132) takes the same form as Eqs. (109)-(110), namely

$$N_i(\boldsymbol{\tau}(\mathbf{x}); \mathbf{x}) = N_i(0) e^{\mathbf{Z}_i^\top \boldsymbol{\tau}}, \quad \text{for } i : x_i = 1 \quad (133)$$

$$R_r(\boldsymbol{\tau}(\mathbf{x}); \mathbf{x}) = R_r(0) - \sum_{i:x_i=1} C_{i,r}(\boldsymbol{\tau}(\mathbf{x}); \mathbf{x}), \quad (134)$$

being $\boldsymbol{\tau} = (\tau_1, \dots, \tau_L)$ the vector of effective times, with $\tau_l = \int_0^t q(R_l(u)) du$, $\mathbf{Z}_i = (Z_{i,1}, \dots, Z_{i,L})$ the vector of per-capita growth rates of species i , and $C_{i,r}$ the amount of resource r consumed by this species. Assuming that all species are inoculated with the same biomass n_0 and growth is stopped when a fixed total abundance R^* of resources has been depleted, a formally identical analysis to the one made in Sec. S8.1 follows in terms of the total amount of resource consumed by each species ($\{C_i^*\}_i$, $C_i^* = \sum_r C_{i,r}^*$).

S8.2 Null distribution for pairwise angles

To assess whether the observed alignment between the leading eigenvector $\mathbf{u}_1$ and the yield $\mathbf{Y}$ exceeds what is expected by chance, we compare empirical angle distributions to a geometric null model corresponding to random directions. In an S -dimensional space, the cosine of the angle ω between two normalized vectors chosen uniformly at random follows a scaled Beta distribution [2]

$$\cos(\omega) \sim 2 \text{B} \left(\frac{S-1}{2}, \frac{S-1}{2} \right) - 1. \quad (135)$$

S9 Analysis of previously published datasets

In addition to the empirical landscapes generated in this work, we applied our framework to six previously published full-factorial community-function datasets [3, 8, 11, 15, 17, 20]. These datasets span distinct microbial systems, community sizes, and measured functions, these including biomass, population size, metabolic activity, and host life-history traits. Community size ranged from $N = 5$ to $N = 10$ components (Table T2). We used these data to test whether the organization of marginal effects observed in our experiments also appears across different ecological systems and functions.

For each focal component i , we first evaluated the corresponding Functional Effect Equation defined in Eq. (10). We then transformed each complete landscape into WH interaction space and compared coefficients involving the focal component, f_{s+i} , with the corresponding lower-order coefficients, f_s . The slopes predicted from the FEEs were calculated using Eq. (15) and compared with the empirical relationships among WH coefficients.

Several focal components exhibited approximately linear FEEs together with the corresponding cross-order organization in interaction space (Figs. S24, S25 and S26). These relationships were observed for functions distinct from those measured in our experiments. For example, the population-size landscapes of Ishizawa et al. [15] quantify how the presence of one community member modifies the abundance of another, bringing the measured function closer to the classical population-dynamical definition of ecological interactions. Similar patterns were also observed for species richness, biomass, and host-associated traits.

The coefficient-level relationships were generally less pronounced than in our replicated experiments. Most published datasets contained a single measurement for each community composition. WH coefficients estimated from unreplicated landscapes are highly sensitive to measurement noise, particularly at high interaction orders [3]. We therefore restricted the interaction-space analysis to first- and second-order background coefficients. These correspond to pairwise and third-order coefficients after addition of the focal component. The resulting comparisons should be interpreted as qualitative evidence of cross-order organization rather than as precise tests of the slopes predicted by the FEEs.

Departures from a single global FEE also revealed structured local rules. In the biomass landscape of Sanchez-Gorostiaga et al. [20], the functional effect of *Paenibacillus polymyxa* separated into two approximately linear branches depending on whether *Bacillus thuringiensis* was present or absent from the ecological background (Fig. S24). Thus, the presence of one species can modify the functional rule followed by another and partition the landscape into distinct local regimes. Such conditional FEEs are not inconsistent with an organized landscape. Instead, they suggest that some systems may be described by a small number of locally valid rules rather than by a single relationship applying across all ecological backgrounds.

For each published landscape, we next constructed the second-moment matrix of marginal effects defined in Eq. (28) and calculated its ecological collective modes. The corresponding spectra were generally concentrated in the first one or two modes (Fig. S30 and S31). We then projected community composition onto the leading latent coordinate defined in Eq. (62). The resulting one-dimensional projections, obtained using Eq. (63), generally organized community function more effectively than species biodiversity alone (Fig. S27 and Fig. S29). When the second mode carried appreciable spectral weight, the two-dimensional latent manifolds defined in Eq. (74) captured additional landscape structure (Fig. S28).

The *Drosophila* microbiome dataset of Gould et al. [11] included several host traits measured over the same set of microbial communities. Total fecundity, daily fecundity, and development time displayed distinct community-function landscapes, but all three showed concentrated marginal-effect spectra and could be represented using one or two latent coordinates (Fig. S29). This illustrates that collective modes are function-dependent, while different functions defined over the same compositional space can nevertheless exhibit comparable low-dimensional organization.

Together, these reanalyses reveal a clear separation between coefficient-level and landscape-level inference. Approximate FEEs and their interaction-space counterparts were detectable across systems, but were often obscured by the lack of replication and the resulting uncertainty in individual WH coefficients. By contrast, the spectral organization of the marginal-effect matrix $\mathbf{M}$ was consistently recovered: marginal effects concentrated along a small number of collective directions, and the resulting one- or two-dimensional projections provided coherent representations of every

published landscape examined.

References

- [1] Anton Bankevich et al. “SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing”. In: *Journal of computational biology* 19.5 (2012), pp. 455–477.
- [2] Tony T. Cai, Jianqing Fan, and Tiefeng Jiang. “Distributions of Angles in Random Packing on Spheres”. In: *Journal of Machine Learning Research* 14 (2013), pp. 1837–1864.
- [3] Jose Camacho-Mateu et al. “On the limits of detection of epistatic higher-order interactions”. In: *bioRxiv* (2026), pp. 2026–03.
- [4] Shifu Chen et al. “fastp: an ultra-fast all-in-one FASTQ preprocessor”. In: *Bioinformatics* 34.17 (2018), pp. i884–i890.
- [5] Alex Chklovski et al. “CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning”. In: *BioRxiv* (2022), pp. 2022–07.
- [6] Paolo Di Tommaso et al. “Nextflow enables reproducible computational workflows”. In: *Nature biotechnology* 35.4 (2017), pp. 316–319.
- [7] Juan Diaz-Colunga, Alvaro Sanchez, and C Brandon Ogbunugafor. “Environmental modulation of global epistasis in a drug resistance fitness landscape”. In: *Nature Communications* 14.1 (2023), p. 8055.
- [8] Juan Diaz-Colunga et al. “Full factorial construction of synthetic microbial communities”. In: *ELife* 13 (2026), RP101906.
- [9] Juan Diaz-Colunga et al. “Global epistasis and the emergence of function in microbial consortia”. In: *Cell* 187.12 (2024), pp. 3108–3119.
- [10] Miguel D Fernández-de-Bobadilla et al. “PATO: pangenome analysis toolkit”. In: *Bioinformatics* 37.23 (2021), pp. 4564–4566.
- [11] Alison L Gould et al. “Microbiome interactions shape host fitness”. In: *Proceedings of the National Academy of Sciences* 115.51 (2018), E11951–E11960.
- [12] Alexey Gurevich et al. “QUAST: quality assessment tool for genome assemblies”. In: *Bioinformatics* 29.8 (2013), pp. 1072–1075.
- [13] Sara Hernando-Amado et al. “Mutational background influences *P. aeruginosa* ciprofloxacin resistance evolution but preserves collateral sensitivity robustness”. In: *Proceedings of the National Academy of Sciences* 119.15 (2022), e2109370119.
- [14] Kabir Husain and Arvind Murugan. “Physical constraints on epistasis”. In: *Molecular Biology and Evolution* 37.10 (2020), pp. 2865–2874.
- [15] Hidehiro Ishizawa et al. “Duckweed as an Emerging Model System for Plant–Microbiome Interactions”. In: *Environmental Microbiology* 27.9 (2025), e70181.
- [16] Birgit Koch, Linda Elise Jensen, and Ole Nybroe. “A panel of Tn7-based vectors for insertion of the gfp marker gene or for delivery of cloned DNA into Gram-negative bacteria at a neutral chromosomal site”. In: *Journal of microbiological methods* 45.3 (2001), pp. 187–195.
- [17] Silke Langenheder et al. “Bacterial biodiversity-ecosystem functioning relations are modified by environmental complexity”. In: *PloS one* 5.5 (2010), e10834.
- [18] Jakub Otwinowski, David M McCandlish, and Joshua B Plotkin. “Inferring the shape of global epistasis”. In: *Proceedings of the National Academy of Sciences* 115.32 (2018), E7550–E7558.
- [19] Gautam Reddy and Michael M Desai. “Global epistasis emerges from a generic model of a complex trait”. In: *Elife* 10 (2021), e64740.
- [20] Alicia Sanchez-Gorostiaga et al. “High-order interactions distort the functional landscape of microbial consortia”. In: *PLoS biology* 17.12 (2019), e3000550.
- [21] Oliver Schwengers et al. “Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification”. In: *Microbial genomics* 7.11 (2021), p. 000685.
- [22] Abigail Skwara et al. “Statistically learning the functional landscape of microbial communities”. In: *Nature ecology & evolution* 7.11 (2023), pp. 1823–1833.

S10 Supplementary Tables

Table T1: **List of carbon sources used in this study.**

Reagent type	Carbon source	1× C-mol/L	5× C-mol/L	Supplier	Reference
Chem. compound	CH ₃ COONa	0.014	0.07	Sigma-Aldrich	W302406
Chem. compound	NaC ₆ H ₇ O ₇	0.014	0.07	Sigma-Aldrich	71498
Chem. compound	D-(–)-Fructose	0.014	0.07	Sigma-Aldrich	F0127
Chem. compound	D-(+)-Glucose	0.014	0.07	Sigma-Aldrich	G7021
Chem. compound	Glycerol	0.014	0.07	Sigma-Aldrich	G9012

Table T2: **Previously published community-function landscapes analyzed in this work.**

Source	<i>N</i>	Community function
Langenheder et al. (2010)	6	Substrate oxidation (medium: glucose and xylose)
Sánchez-Gorostiaga et al. (2019)	6	Starch hydrolysis
Díaz-Colunga et al. (2026)	8	Biomass
Ishizawa et al. (2025)	7	Population size
Camacho-Mateu et al. (2026)	10	Biomass
Gould et al. (2018)	5	Development time, total fecundity and daily fecundity

S11 Supplementary Figures

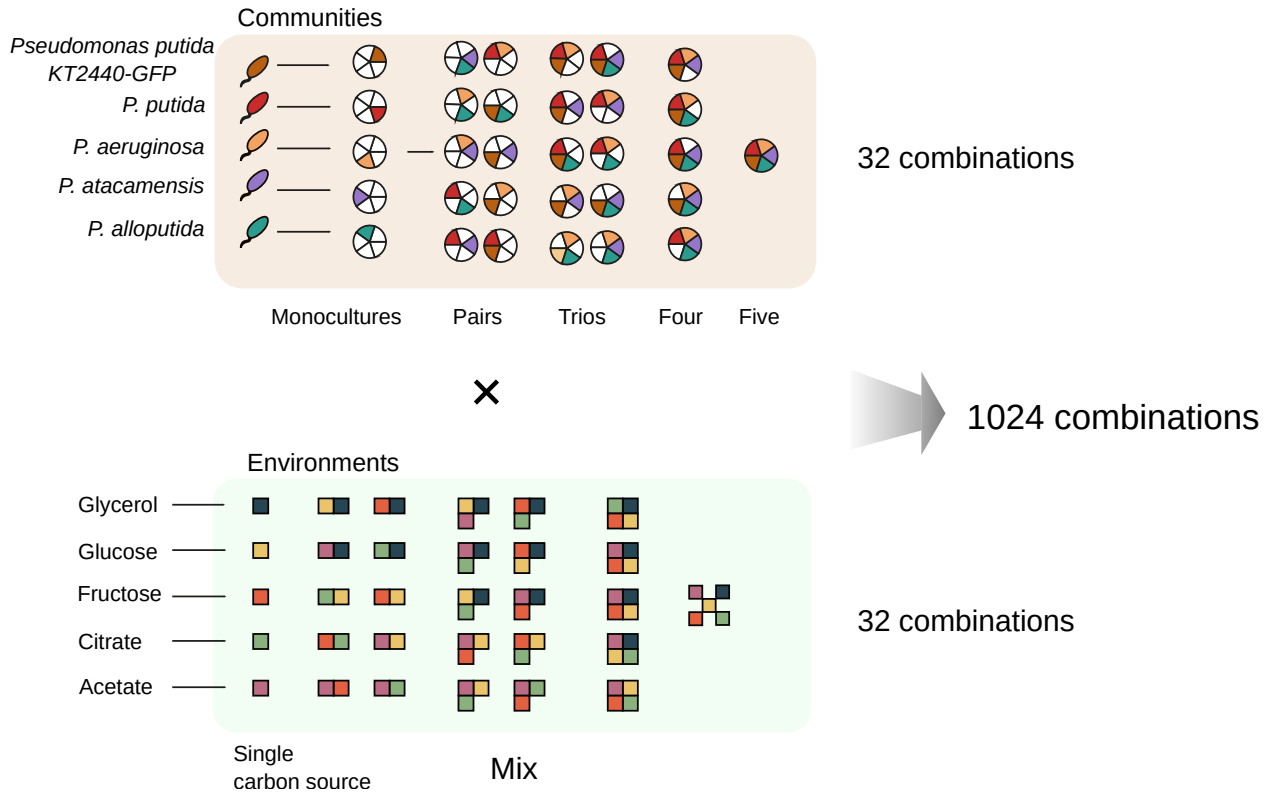


Figure S1: Full-factorial design of the five-species-five-resource combinatorial assembly experiment. See Sec. S1.2. Five *Pseudomonas* species were assembled in all $2^5 = 32$ possible presence-absence combinations, ranging from the empty community to the complete five-species community. Independently, five carbon sources—glycerol, glucose, fructose, citrate, and acetate—were supplied in all $2^5 = 32$ possible presence-absence combinations, ranging from no-resource controls to mixtures containing all five resources. Crossing the species and resource landscapes generated $2^5 \times 2^5 = 1,024$ community-environment combinations. Community function was set to zero for conditions containing no species or no supplied carbon source.

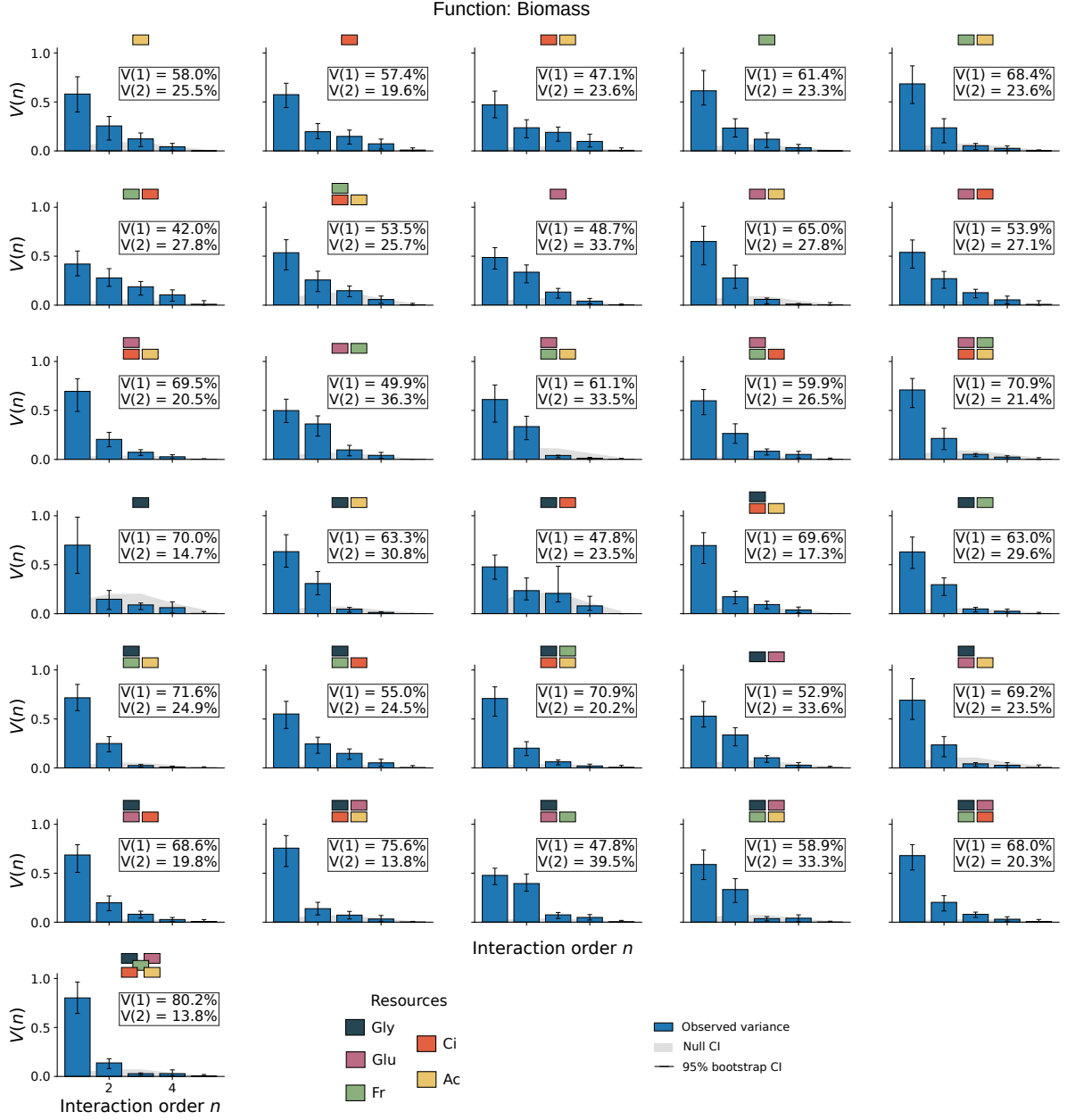


Figure S2: **Variance decomposition of biomass community-function landscapes across WH interaction orders.** For each resource environment, the biomass community-function landscape was decomposed into WH coefficients grouped by interaction order n . Bars show the fraction of total landscape variance contributed by coefficients of order n , $V(n)$, as defined in Eq. (9). Error bars indicate 95% bootstrap confidence intervals, and gray bands show the corresponding null expectation (both the null model and the bootstrap confidence intervals are computed following [3]). Resource composition is indicated above each panel by colored squares. Labels report the contributions of first- and second-order coefficients, $V(1)$ and $V(2)$. Across environments, the first two interaction orders together accounted for most of the total variance, supporting the second-order WH approximation used in the spectral and latent-coordinate analyses.

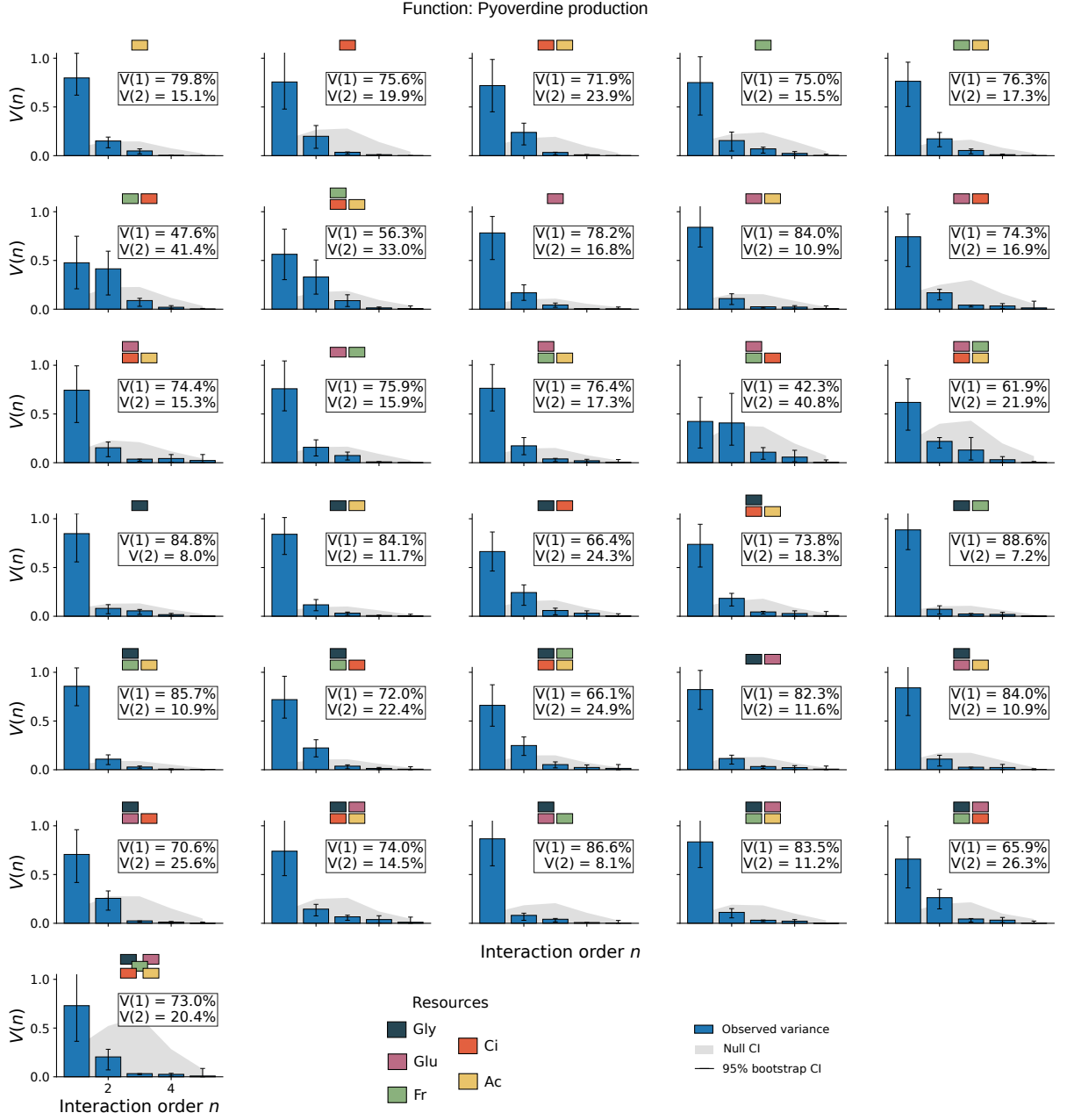


Figure S3: Variance decomposition of pyoverdine community-function landscapes across WH interaction orders. For each resource environment, the pyoverdine community-function landscape was decomposed into WH coefficients grouped by interaction order n . Bars show the fraction of total landscape variance contributed by coefficients of order n , $V(n)$, as defined in Eq. (9). Error bars indicate 95% bootstrap confidence intervals, and gray bands show the corresponding null expectation. Resource composition is indicated above each panel by colored squares. Labels report the contributions of first- and second-order coefficients, $V(1)$ and $V(2)$. Across most environments, these two orders together accounted for approximately 90% or more of the total variance, with higher-order coefficients contributing only a small residual fraction. This concentration supports the second-order Walsh–Hadamard approximation used in the spectral and latent-coordinate analyses of pyoverdine production.

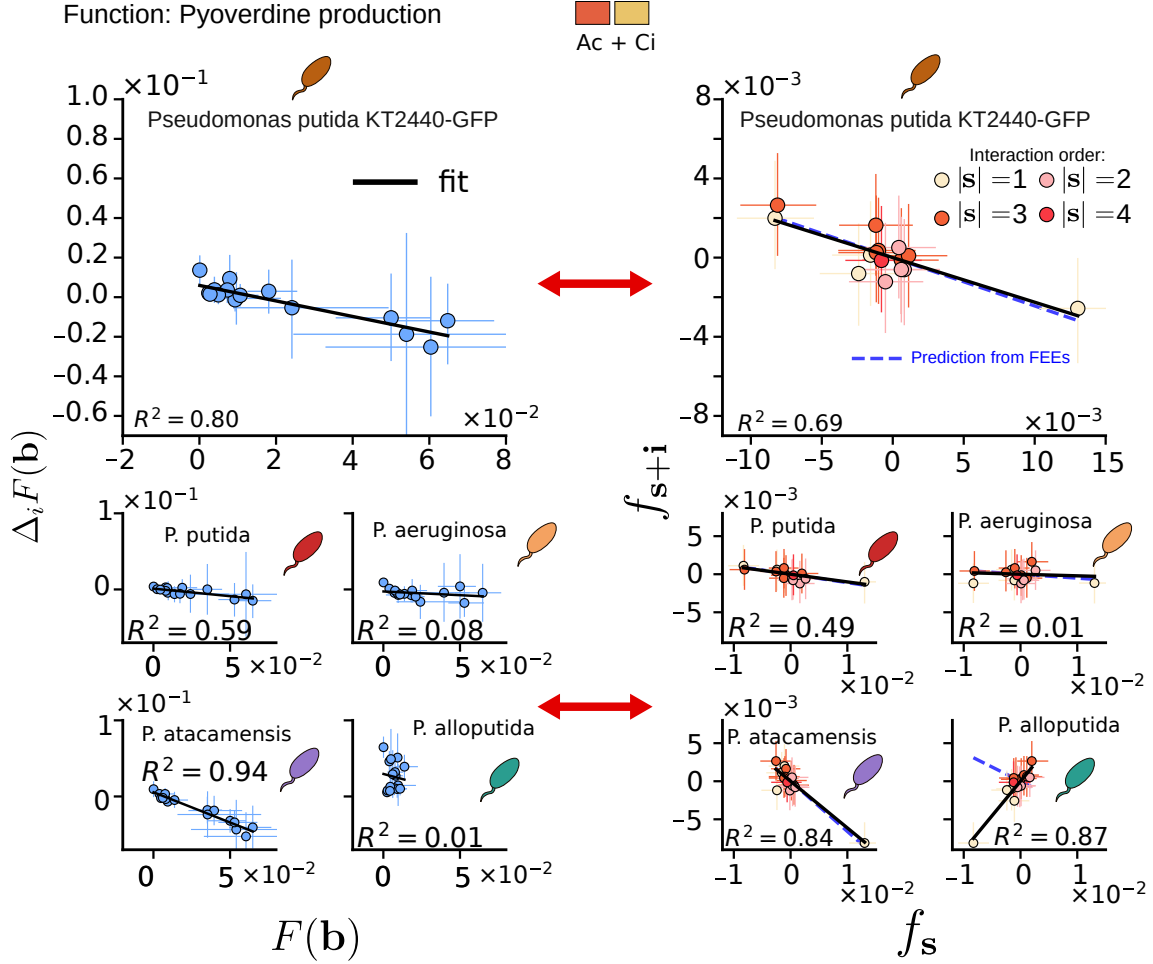


Figure S4: **FEEs predict a hierarchical organization of interaction coefficients for pyoverdine production.** This figure complements Fig. 1 of the main text, which presents the corresponding analysis for biomass. Here, the same analysis is shown for pyoverdine production in citrate + acetate medium. Left panels show, for each focal species i , its functional effect, $\Delta_i F(\mathbf{b})$, as a function of the background community function, $F(\mathbf{b})$, as defined in Eq. (10). Right panels show the corresponding relationships in interaction space by comparing WH coefficients involving the focal species, f_{s+i} , with the associated coefficients one order below, f_s , as predicted by Eq. (15). Solid black lines show empirical linear fits, whereas dashed blue lines indicate the slopes predicted from the fitted FEE slope b_i . Error bars denote 95% bootstrap confidence intervals, computed as described in [3].

Biomass: Functional Effect Equations

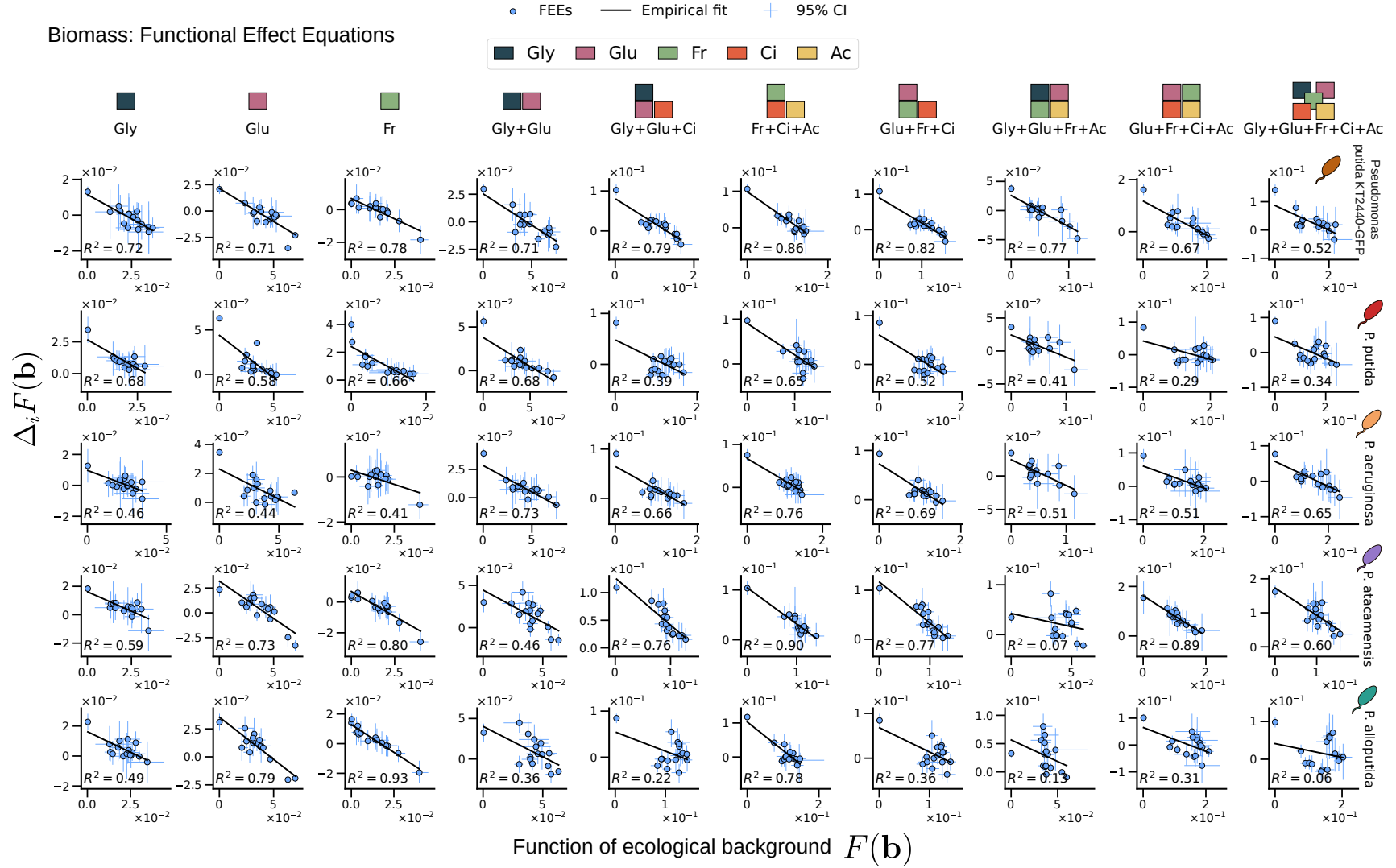


Figure S5: **Functional Effect Equations for biomass production across randomly resource environments.** This figure extends the FEE analysis shown in Fig. 1B to additional resource environments. Each column corresponds to a different resource environment, and each row to a focal species. Resource composition is indicated above the columns by colored rectangles denoting the carbon sources present in each medium—glycerol, glucose, fructose, citrate, and acetate—. In each panel, the functional effect of adding focal species i , $\Delta_i F(\mathbf{b})$, is plotted against the function of the corresponding background community, $F(\mathbf{b})$. Points show measured functional effects, vertical error bars denote 95% bootstrap confidence intervals, and solid black lines show empirical linear fits. The coefficient of determination, R_i^2 , is reported in each panel.

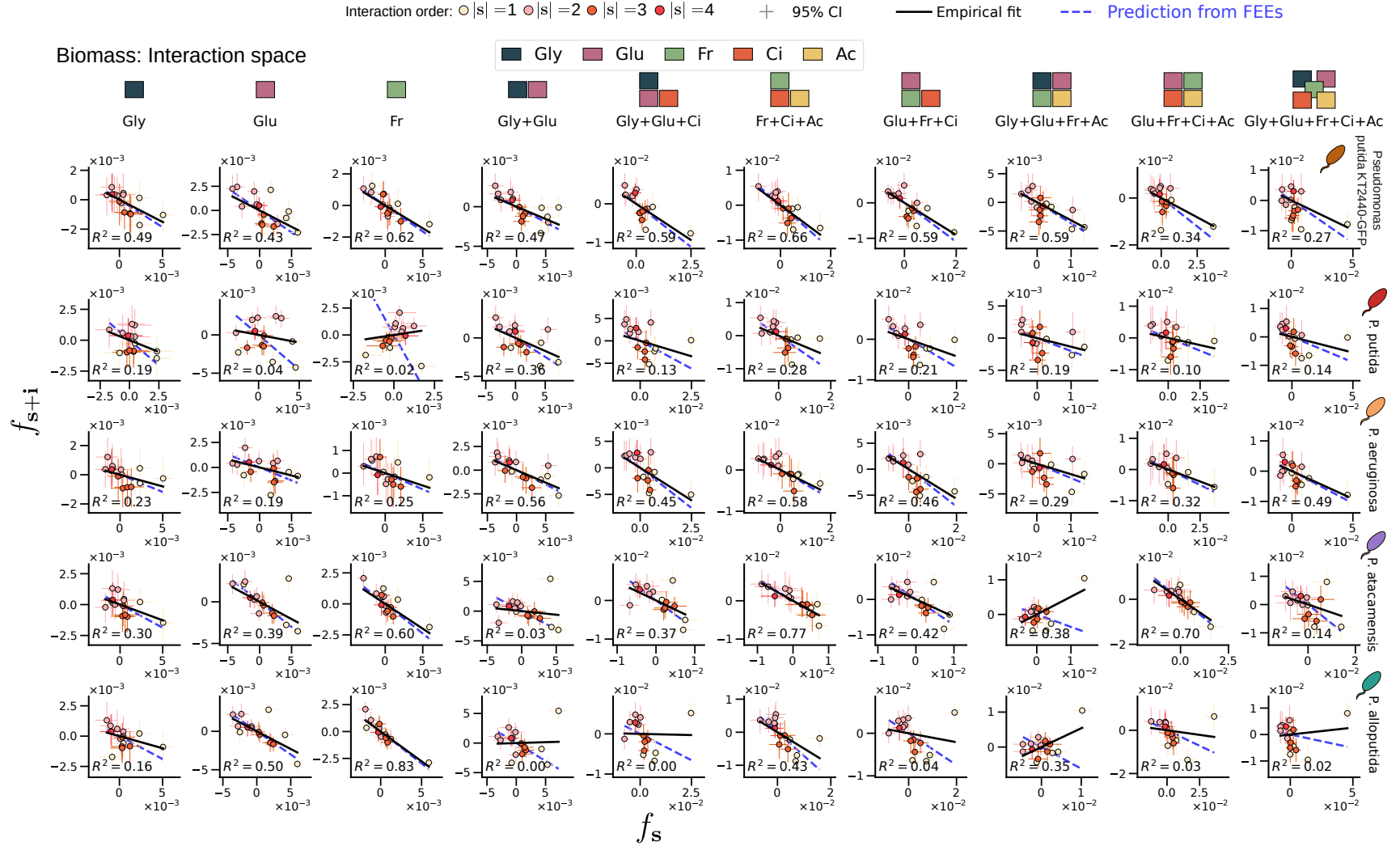


Figure S6: **WH interaction coefficients follow simple cross-order relationships for biomass production across randomly selected resource environments.** This figure extends the WH coefficient analysis shown in Fig. 1C to a representative subset of resource environments. Each column corresponds to a different resource environment, and each row to a focal species. Resource composition is indicated above the columns by colored rectangles denoting the carbon sources present in each medium—glycerol, glucose, fructose, citrate, and acetate. In each panel, WH coefficients involving focal species i , f_{s+i} , are plotted against the corresponding coefficients one order below, f_s , for subsets s not containing species i . Points show estimated WH coefficients and are colored by interaction order. Solid black lines show empirical linear fits, whereas dashed blue lines indicate the slopes predicted from the corresponding fitted FEE slope b_i , according to Eq. (15). Vertical error bars denote 95% bootstrap confidence intervals. The coefficient of determination of the interaction-space fit, R_i^2 , is reported in each panel.

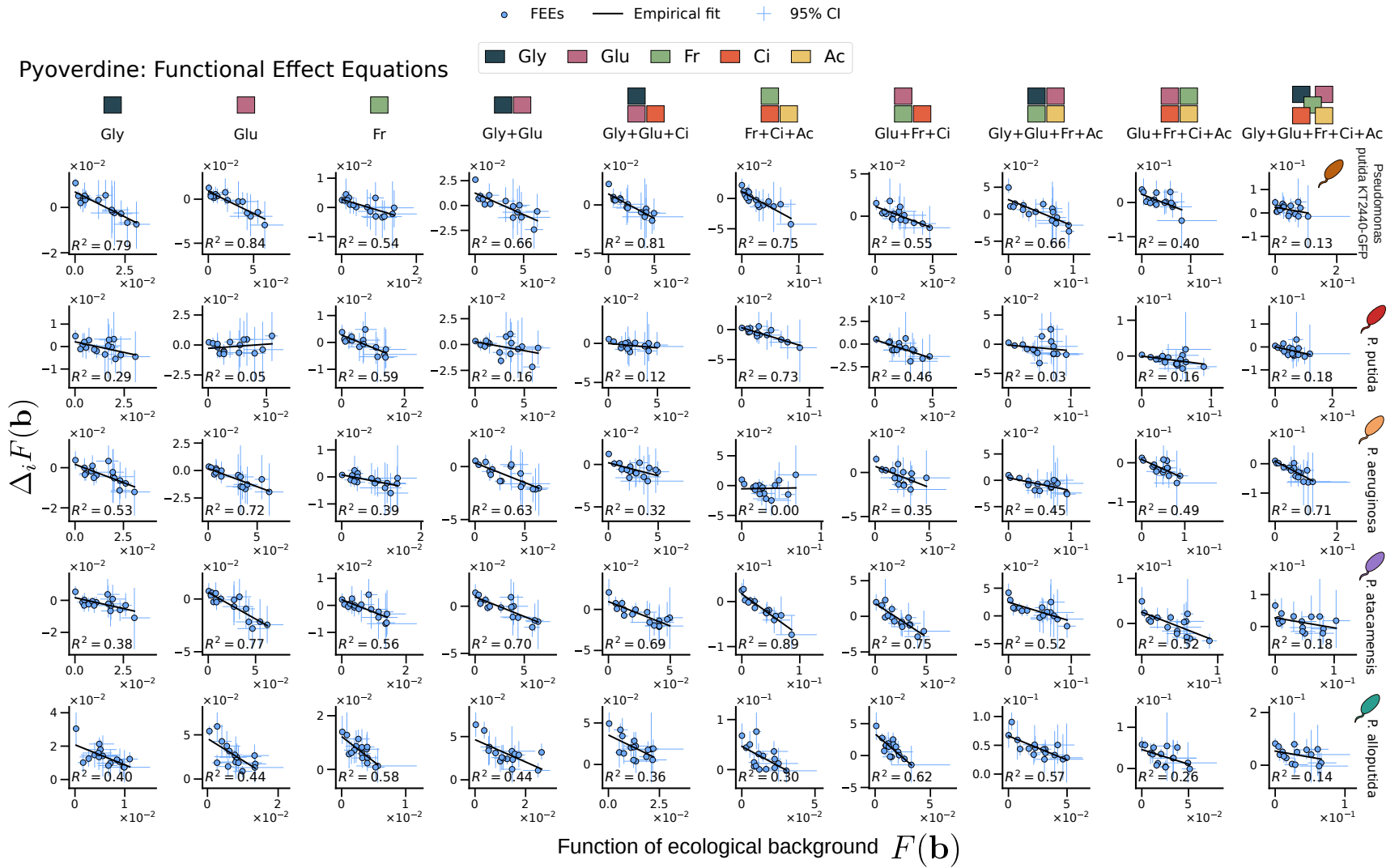


Figure S7: **Functional Effect Equations for pyoverdine production across randomly selected resource environments.** This figure presents the same analysis as Fig. S5, using pyoverdine production as the community function. Each column corresponds to a different resource environment, and each row to a focal species. In each panel, the functional effect of adding focal species i , $\Delta_i F(\mathbf{b})$, is plotted against the function of the corresponding background community, $F(\mathbf{b})$. Points show measured functional effects, vertical error bars denote 95% bootstrap confidence intervals, and solid black lines show empirical linear fits. The coefficient of determination, R_i^2 , is reported in each panel.

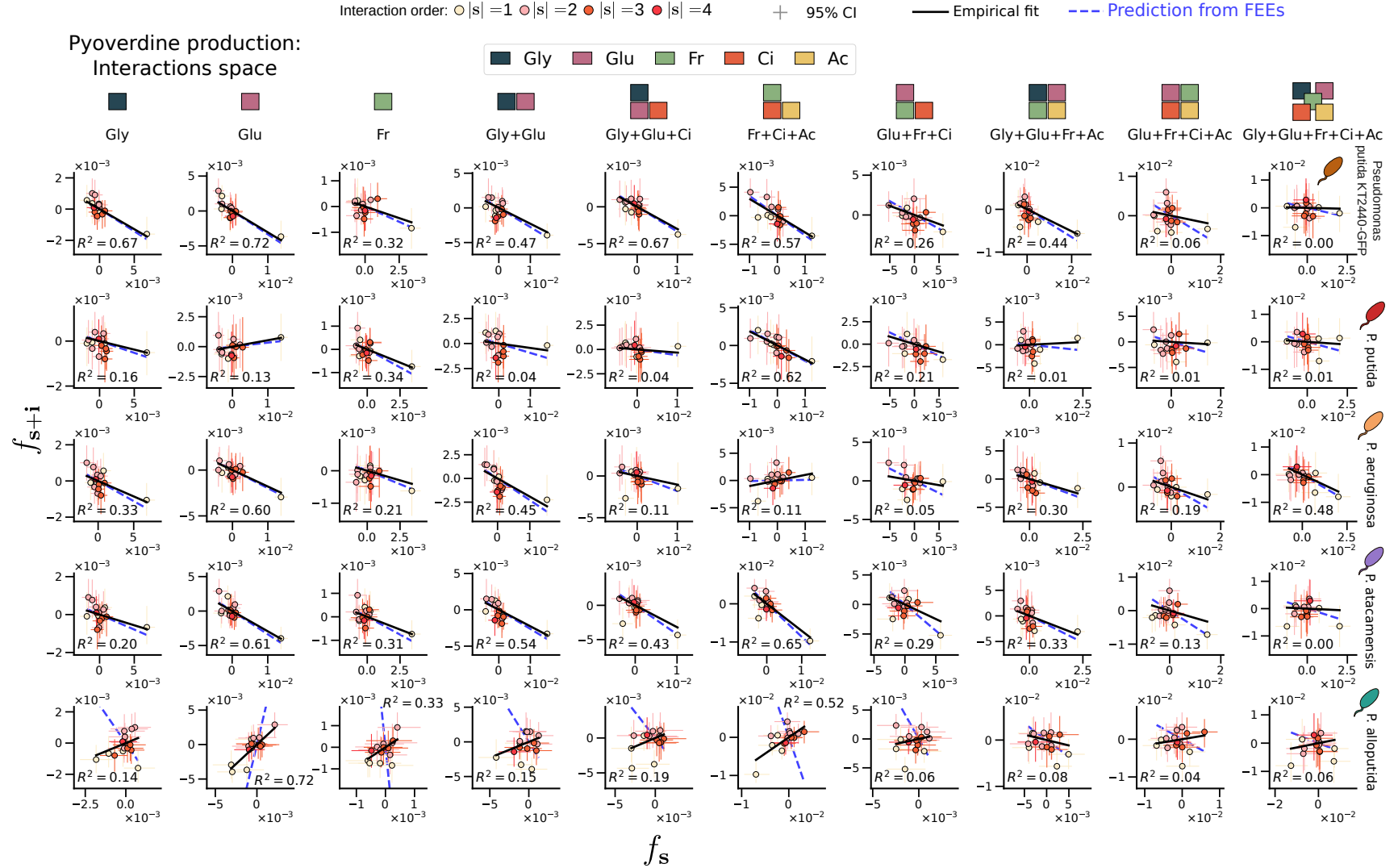


Figure S8: **WH interaction coefficients follow simple cross-order relationships for pyoverdine production across randomly selected resource environments.** This figure presents the same analysis as Fig. S6, using pyoverdine production as the community function. Each column corresponds to a different resource environment, and each row to a focal species. In each panel, WH coefficients involving focal species i , f_{s+i} , are plotted against the corresponding coefficients one order below, f_s , for subsets s not containing species i . Points show estimated WH coefficients and are colored by interaction order. Solid black lines show empirical linear fits, whereas dashed blue lines indicate the slopes predicted from the corresponding fitted FEE slope b_i , using $c_i = b_i/(2 + b_i)$ as given by Eq. (15). Vertical error bars denote 95% bootstrap confidence intervals. The coefficient of determination of the interaction-space fit, R_i^2 , is reported in each panel.

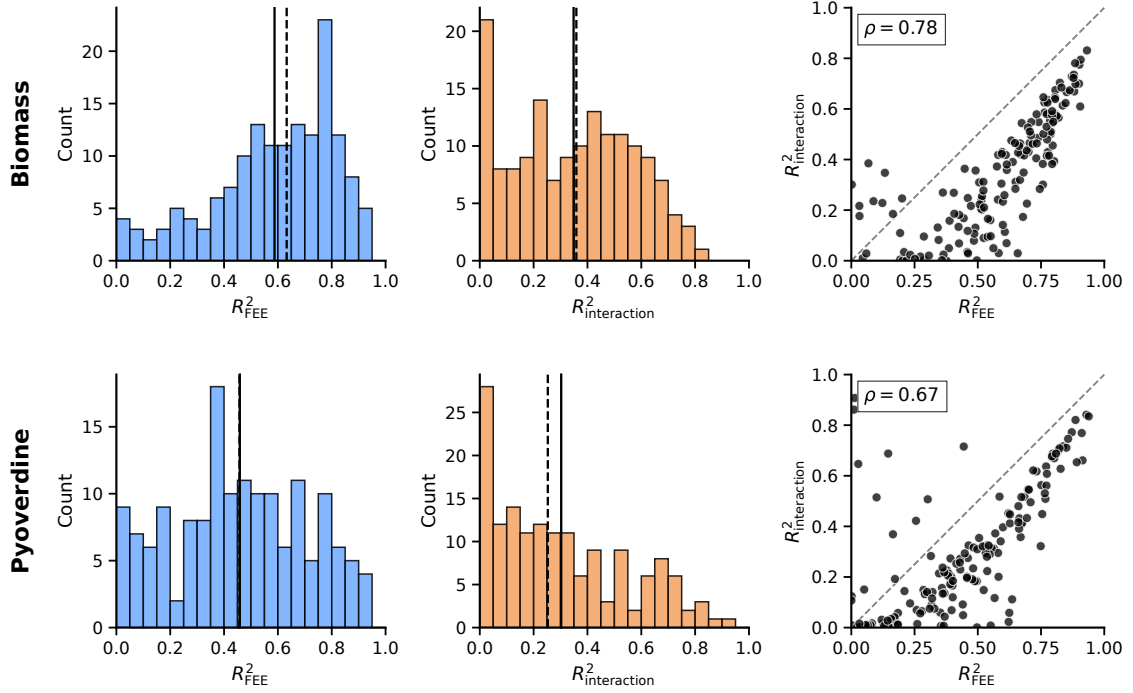


Figure S9: **FEE predictability is associated with stronger cross-order relationships among WH interaction coefficients.** Distributions of the coefficients of determination for Functional Effect Equations, R_{FEE}^2 , and for the corresponding cross-order relationships among WH interaction coefficients, $R_{\text{interaction}}^2$, across focal species and resource environments. Top panels show biomass production and bottom panels pyoverdine production. Left and middle panels show the distributions of R_{FEE}^2 and $R_{\text{interaction}}^2$, respectively; solid and dashed vertical lines indicate the mean and median. Right panels compare both quantities for each focal species–environment combination. The two measures are strongly positively correlated for biomass ($\rho = 0.78$) and pyoverdine ($\rho = 0.67$), showing that more predictive FEEs are associated with more pronounced cross-order organization in interaction space. Most points fall below the identity line, consistent with the greater statistical uncertainty involved in estimating WH coefficients, particularly at higher interaction orders.

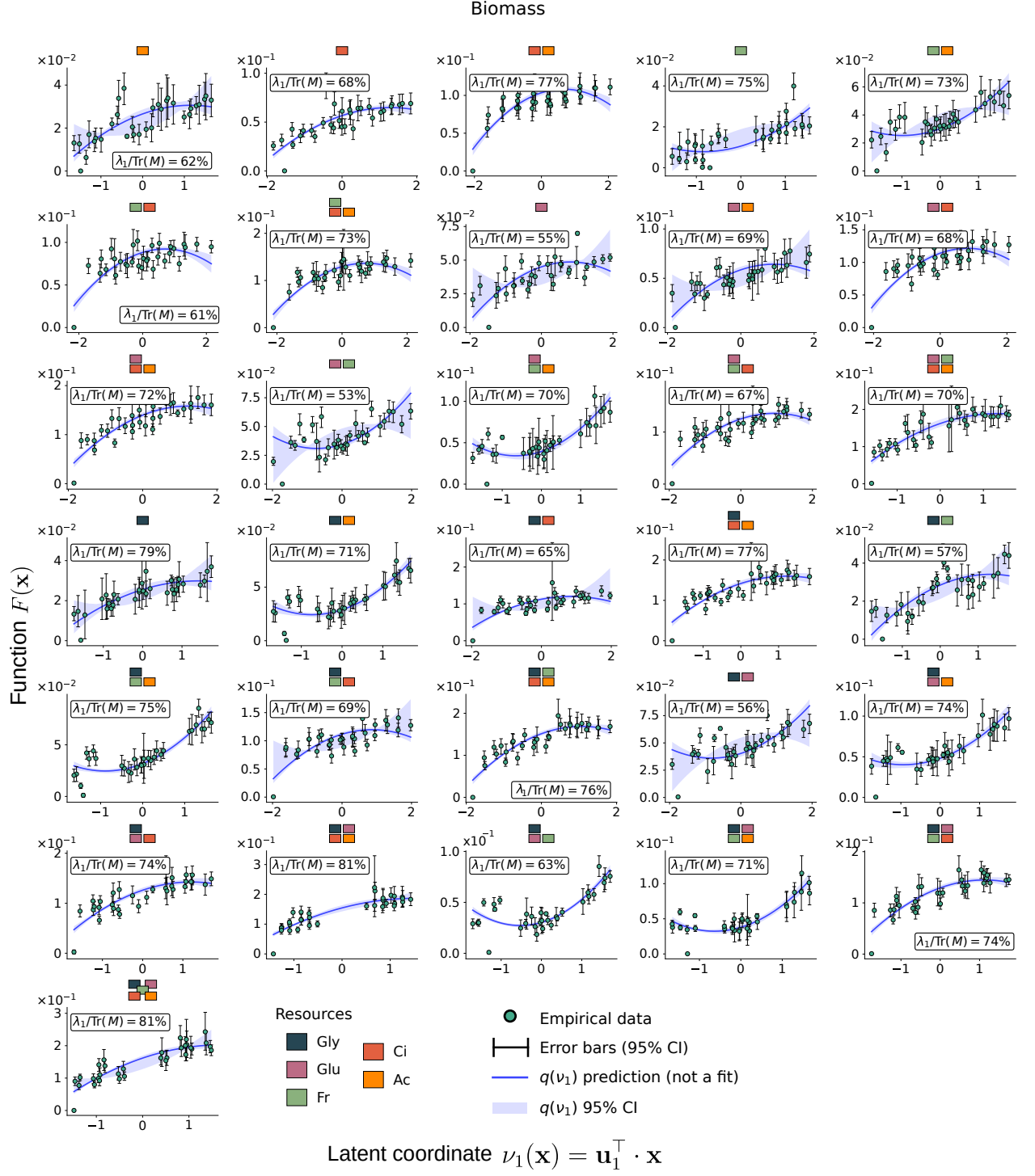


Figure S10: **One-dimensional latent-coordinate projections of biomass community-function landscapes across resource environments.** Each panel shows a biomass community-function landscape projected onto its leading latent coordinate, $\nu_1(\mathbf{x}) = \mathbf{u}_1^\top \mathbf{x}$. Points show measured community functions, $F(\mathbf{x})$, and error bars denote 95% bootstrap confidence intervals. Solid blue lines show the latent-coordinate prediction, $q(\nu_1)$, defined in Eq. (63), and shaded blue regions denote the corresponding 95% bootstrap confidence intervals. The label in each panel reports the fraction of the total spectral weight captured by the leading mode, $\lambda_1/\text{Tr}(\mathbf{M})$.

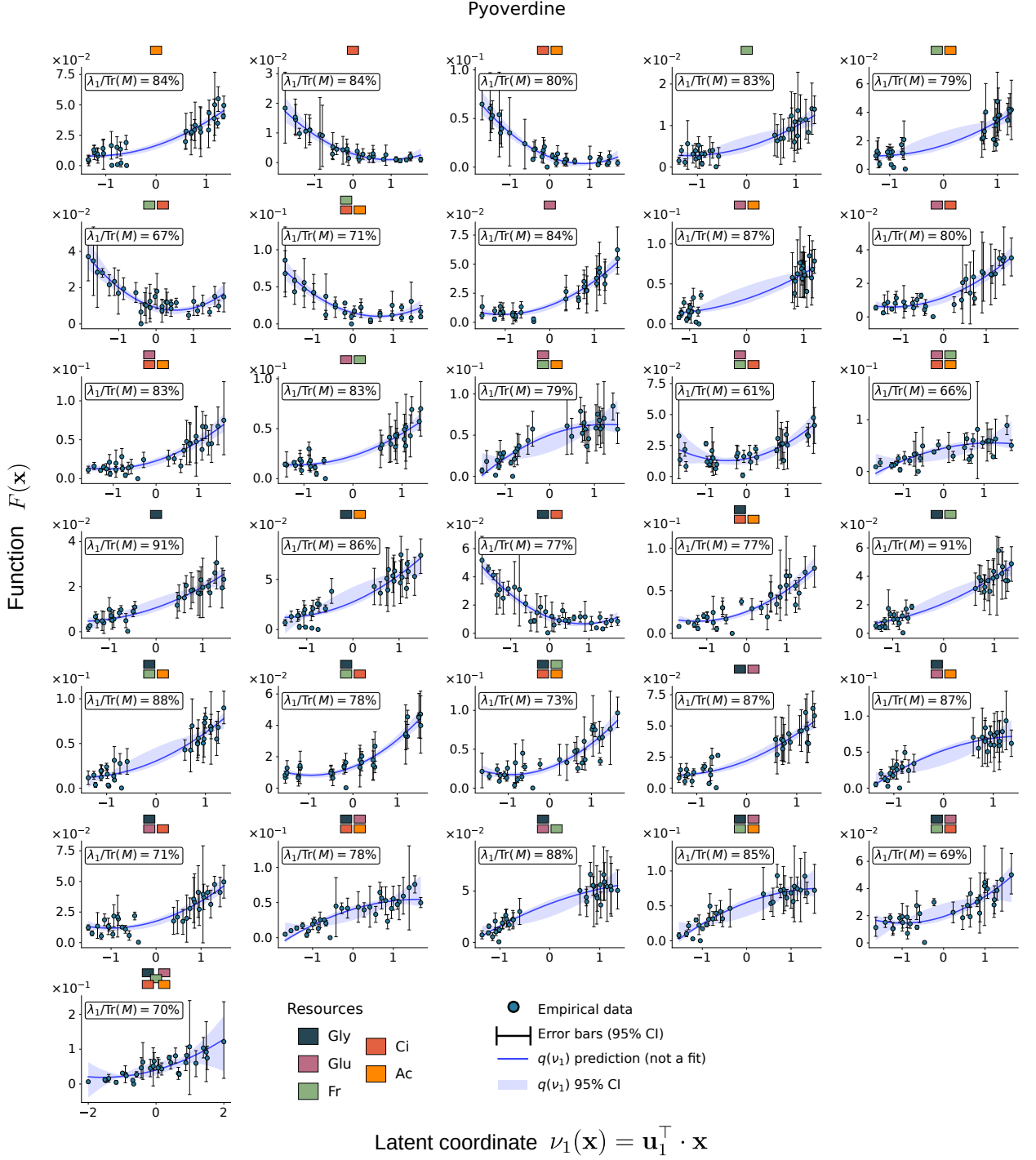


Figure S11: **One-dimensional latent-coordinate projections of pyoverdine community-function landscapes across resource environments.** Each panel shows a pyoverdine community-function landscape projected onto its leading latent coordinate, $\nu_1(\mathbf{x}) = \mathbf{u}_1^\top \mathbf{x}$. Points show measured community functions, $F(\mathbf{x})$, and error bars denote 95% bootstrap confidence intervals. Solid blue lines show the latent-coordinate prediction, $q(\nu_1)$, defined in Eq. (63), and shaded blue regions denote the corresponding 95% bootstrap confidence intervals. The label in each panel reports the fraction of the total spectral weight captured by the leading mode, $\lambda_1/\text{Tr}(\mathbf{M})$.

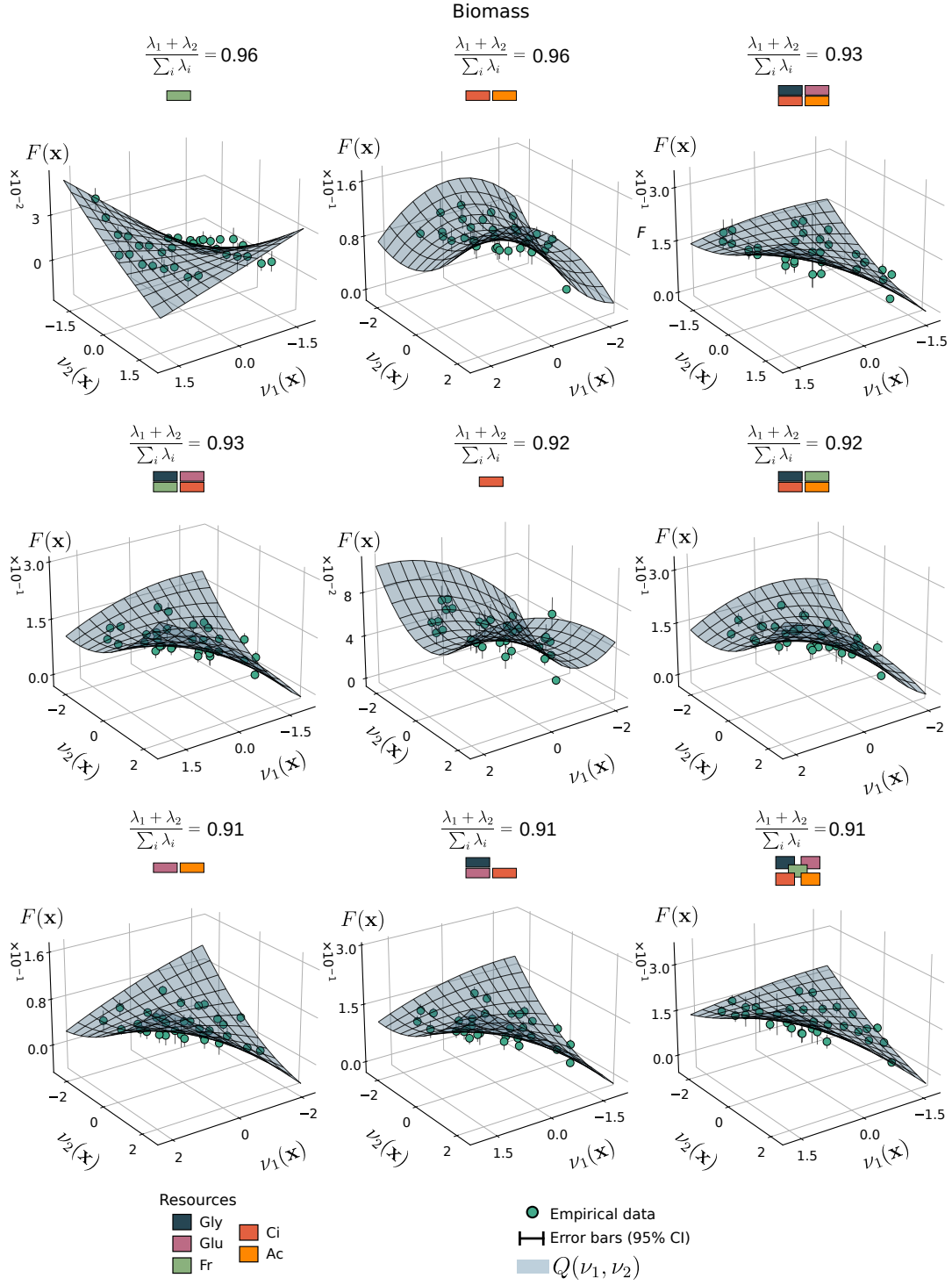


Figure S12: **Two-dimensional latent-coordinate manifolds of biomass community-function landscapes across selected resource environments.** Each panel shows a biomass community-function landscape projected onto its first two latent coordinates, $\nu_1(\mathbf{x}) = \mathbf{u}_1^\top \mathbf{x}$ and $\nu_2(\mathbf{x}) = \mathbf{u}_2^\top \mathbf{x}$. Points show measured community functions, $F(\mathbf{x})$, after 24 h of growth, and error bars denote 95% bootstrap confidence intervals. Blue surfaces show the predicted quadratic latent manifold, $Q(\nu_1, \nu_2)$, defined in Eq. (74). The nine panels correspond to the resource environments for which the $K = 2$ manifold captured the largest fraction of biomass variation.

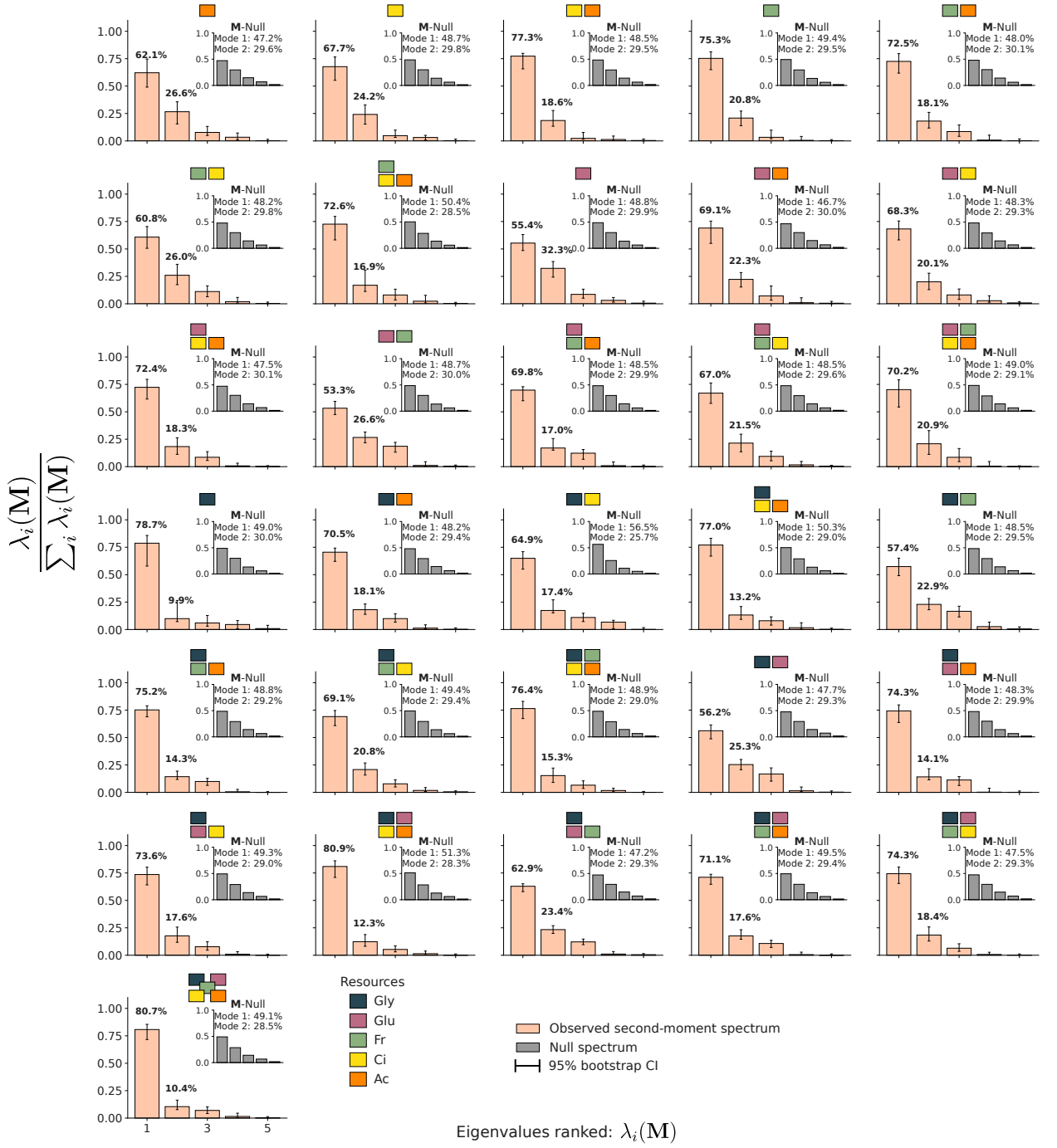


Figure S14: **Low-rank second-moment spectra across resource environments.** For each resource environment, we approximated the second-moment matrix of marginal effects, $\mathbf{M}$, by retaining only first- and second-order WH coefficients. Under this approximation, $\mathbf{M} \approx \mathbf{f}\mathbf{f}^\top + \mathbf{H}\mathbf{H}^\top$, as given by Eq. (50), where $\mathbf{f}$ is the vector of first-order WH coefficients and $\mathbf{H}$ is the matrix of second-order WH coefficients. Bars show the normalized eigenvalue spectrum, $\lambda_i(\mathbf{M}) / \sum_j \lambda_j(\mathbf{M})$, and error bars denote 95% bootstrap confidence intervals. Insets show the corresponding spectra obtained under the null model, $\mathbf{M}_{\text{null}}$. Across resource environments, the observed spectra are dominated by one or two leading modes, whereas the null spectra are more broadly distributed. Resource composition is indicated above each panel by colored rectangles denoting the carbon sources present in each medium.

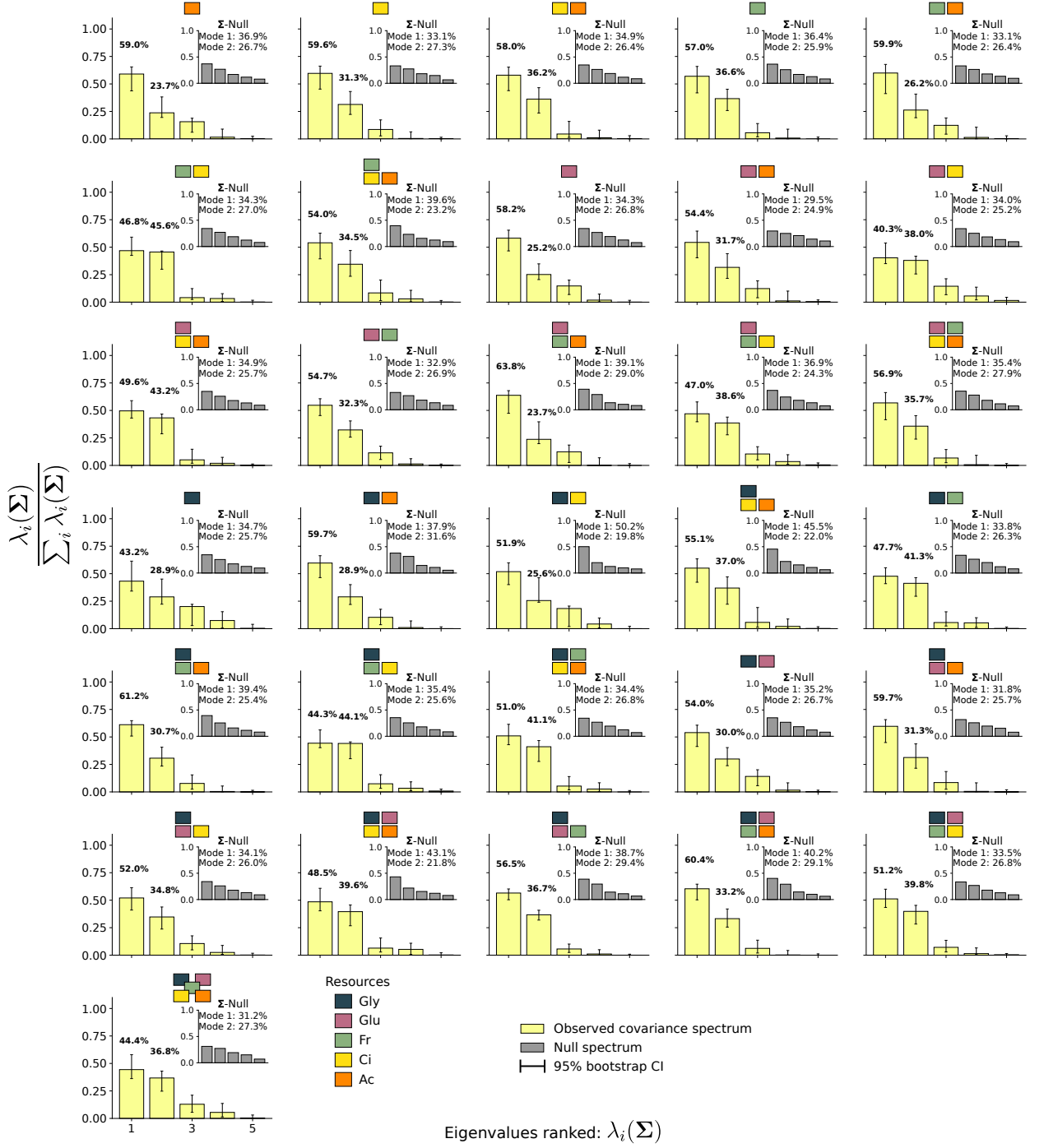


Figure S15: Low-rank covariance spectra across resource environments. For each resource environment, we approximated the covariance matrix of marginal effects, Σ , by retaining only first- and second-order WH coefficients. Since $\Sigma = \mathbf{M} - \mathbf{f}\mathbf{f}^\top$, Eq. (71), the second-order approximation gives $\Sigma \approx \mathbf{H}\mathbf{H}^\top$, where $\mathbf{H}$ is the matrix of second-order WH coefficients. Unlike the uncentered second-moment matrix $\mathbf{M}$, Σ removes the contribution of the mean marginal-effect vector and therefore isolates background-dependent variation. Bars show the normalized eigenvalue spectrum, $\lambda_i(\Sigma) / \sum_j \lambda_j(\Sigma)$, and error bars denote 95% bootstrap confidence intervals. Insets show the corresponding spectra obtained under the null model, Σ_{null} . Across resource environments, the observed covariance spectra are dominated by one or two leading modes, indicating that the background-dependent component of marginal effects is also low-dimensional. Resource composition is indicated above each panel by colored rectangles denoting the carbon sources present in each medium.

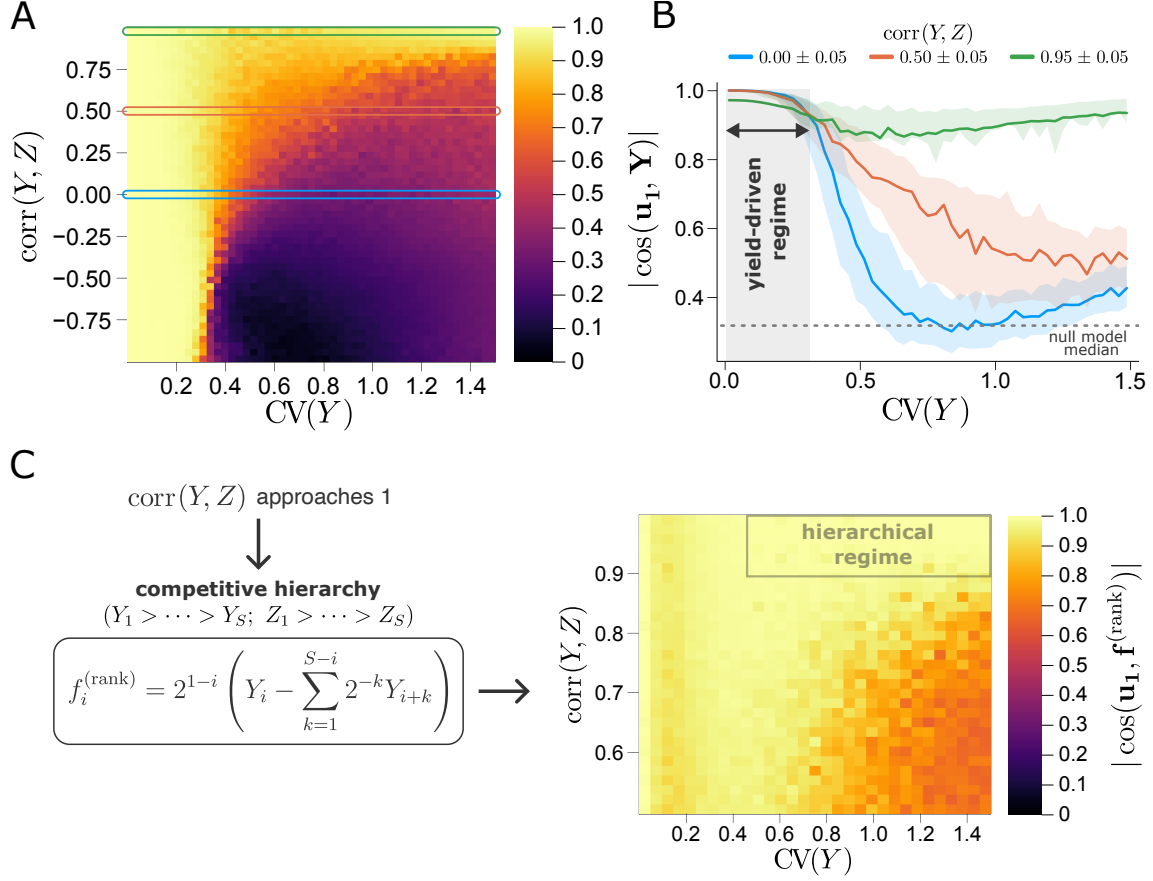


Figure S16: Consumer-Resource models predict the leading collective mode from species traits. Results from the numerical integration of the CRM with one public resource and 5 species (Eqs. (109)-(110); see Sec. S8.1 for details). Yields ($\mathbf{Y}$) and per-capita growth rates ($\mathbf{Z}$) were sampled from uniform distributions of fixed means and variable standard deviations to produce coefficient of variations $\text{CV}(\mathbf{Y})$ and $\text{CV}(\mathbf{Z})$ varying from 0 to 1.5 in 51 equally-spaced steps. The traits vectors $\mathbf{Y}$ and $\mathbf{Z}$ were sampled 10 times for each value of $\text{CV}(\mathbf{Y})$ and $\text{CV}(\mathbf{Z})$, resulting in a total of 260100 simulated landscapes. The Pearson correlation coefficient between $\mathbf{Y}$ and $\mathbf{Z}$, denoted as $\text{corr}(\mathbf{Y}, \mathbf{Z})$, was also computed. **(A)** Heatmap showing the median of $|\cos(\mathbf{u}_1, \mathbf{Y})|$, the cosine similarity between the leading collective mode $\mathbf{u}_1$ and the yields vector $\mathbf{Y}$, against $\text{CV}(\mathbf{Y})$ and $\text{corr}(\mathbf{Y}, \mathbf{Z})$. Simulations confirm the yield-driven regime at low $\text{CV}(\mathbf{Y})$ (independent of growth rates) predicted in Eq. (122). **(B)** Median and IQR of $|\cos(\mathbf{u}_1, \mathbf{Y})|$ are shown against $\text{CV}(\mathbf{Y})$ for uncorrelated ($\text{corr}(\mathbf{Y}, \mathbf{Z}) = 0.00 \pm 0.05$), moderately correlated ($\text{corr}(\mathbf{Y}, \mathbf{Z}) = 0.50 \pm 0.05$), and strongly correlated ($\text{corr}(\mathbf{Y}, \mathbf{Z}) = 0.95 \pm 0.05$) yields and growth rates. The median of the cosine similarity between two random normalized 5-dimensional vectors (see Sec. S8.2) is also reported for comparison. **(C)** As $\text{corr}(\mathbf{Y}, \mathbf{Z})$ approaches 1, the system enters a regime of competitive hierarchy where the leading eigenmode $\mathbf{u}_1$ aligns with the vector of focal effects $\mathbf{f}^{(\text{rank})}$, computed in Eq. (130) and fully determined by yields.

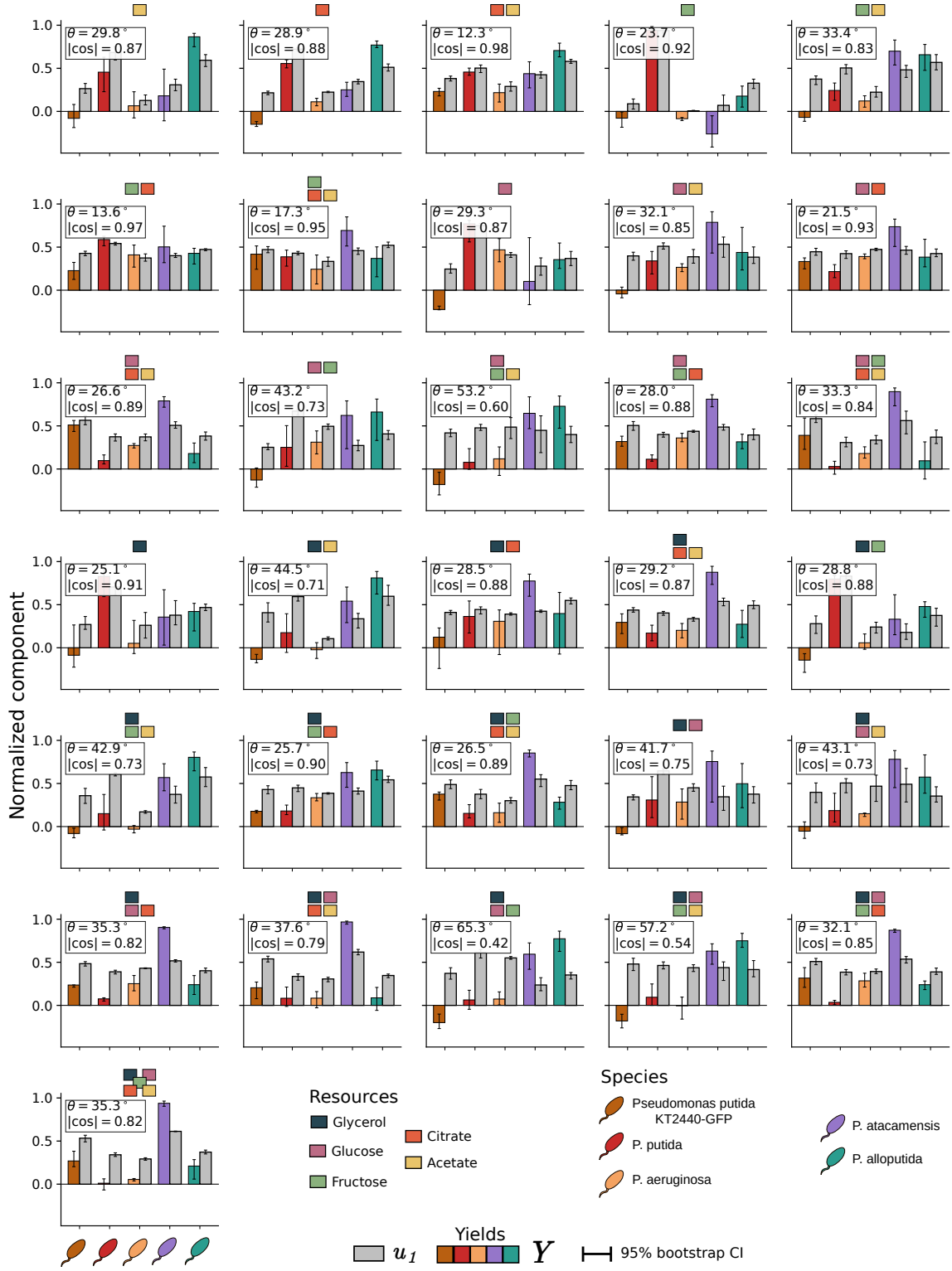
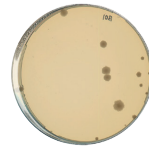
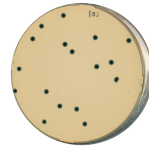


Figure S17: **Yield productivity provides a proxy for the leading collective mode.** Each panel corresponds to a different resource environment and compares the normalized components of the leading collective mode, $\mathbf{u}_1$, with the normalized yield productivity vector, $\mathbf{Y} \approx \mathbf{B}_{\text{mono}}$, computed from the monoculture biomass as proxy. The sign of $\mathbf{u}_1$ was chosen to maximize its overlap with $\mathbf{Y}$. Insets report the sign-invariant angle, $\theta = \arccos(|\mathbf{u}_1^\top \mathbf{Y}|)$, and the corresponding absolute cosine similarity, $|\mathbf{u}_1^\top \mathbf{Y}|$. Error bars denote 95% bootstrap confidence intervals.

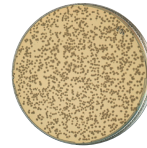
MONOCULTURES



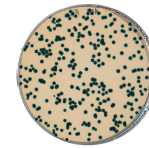
Pseudomonas defluvii



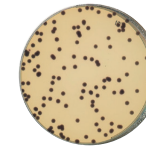
Aeromonas hydrophila



Acinetobacter johnsonii

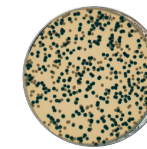
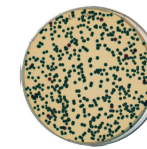
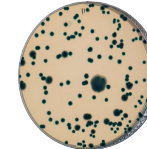
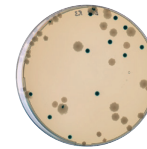
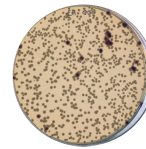
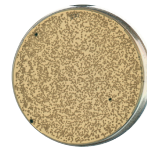
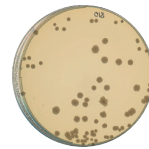
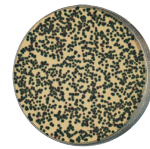
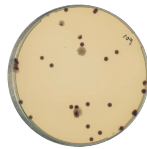
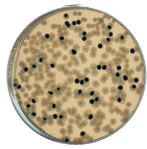


Serratia proteomaculans

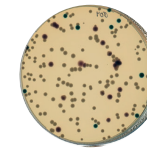
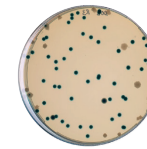
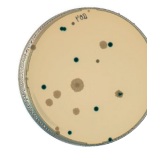
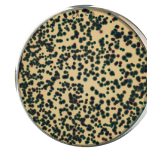
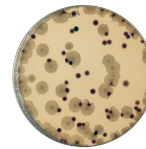
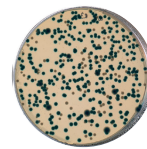
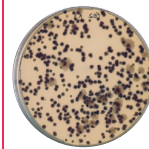
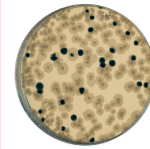
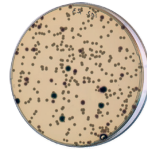
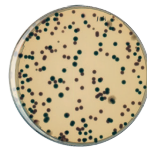


Lelliota amnigena

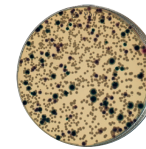
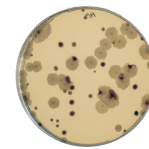
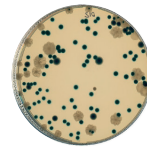
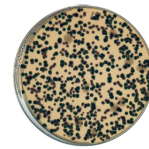
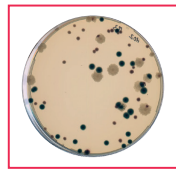
PAIRS



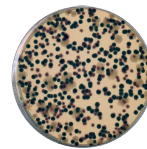
TRIOS



FOUR



FIVE



 Communities missing one specie

Figure S18: **Agar plate images of all five-species combinations after the 10-day experiment.** The 31 agar plates each contain a different species combination, including monocultures, pairs, trios, four-species, and five-species communities. Red squares indicate the only two final communities in which one species was missing after 10 days.

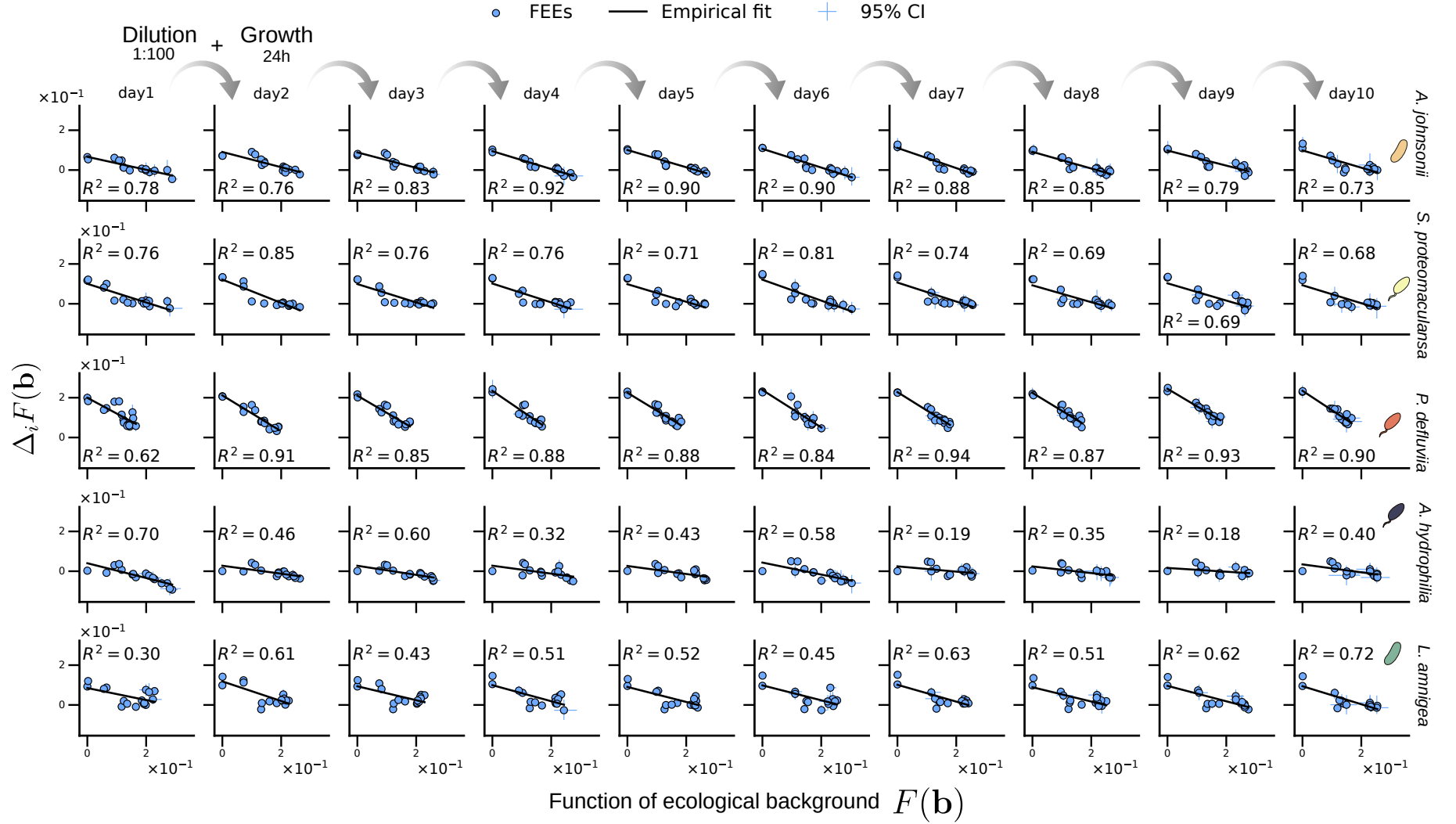


Figure S19: Functional Effect Equations remain predictable during serial dilution. Functional Effect Equations (FEEs) are shown for each of the ten daily passages of the serial-dilution experiment. Columns correspond to passages, and rows to focal species. In each panel, the functional effect of adding species i , $\Delta_i F(\mathbf{x})$, is plotted against the function of the corresponding background community, $F(\mathbf{x})$. Points show measured functional effects, error bars denote 95% bootstrap confidence intervals, and solid black lines show empirical linear fits. The coefficient of determination, R_i^2 , is reported in each panel. Species names are indicated on the right.

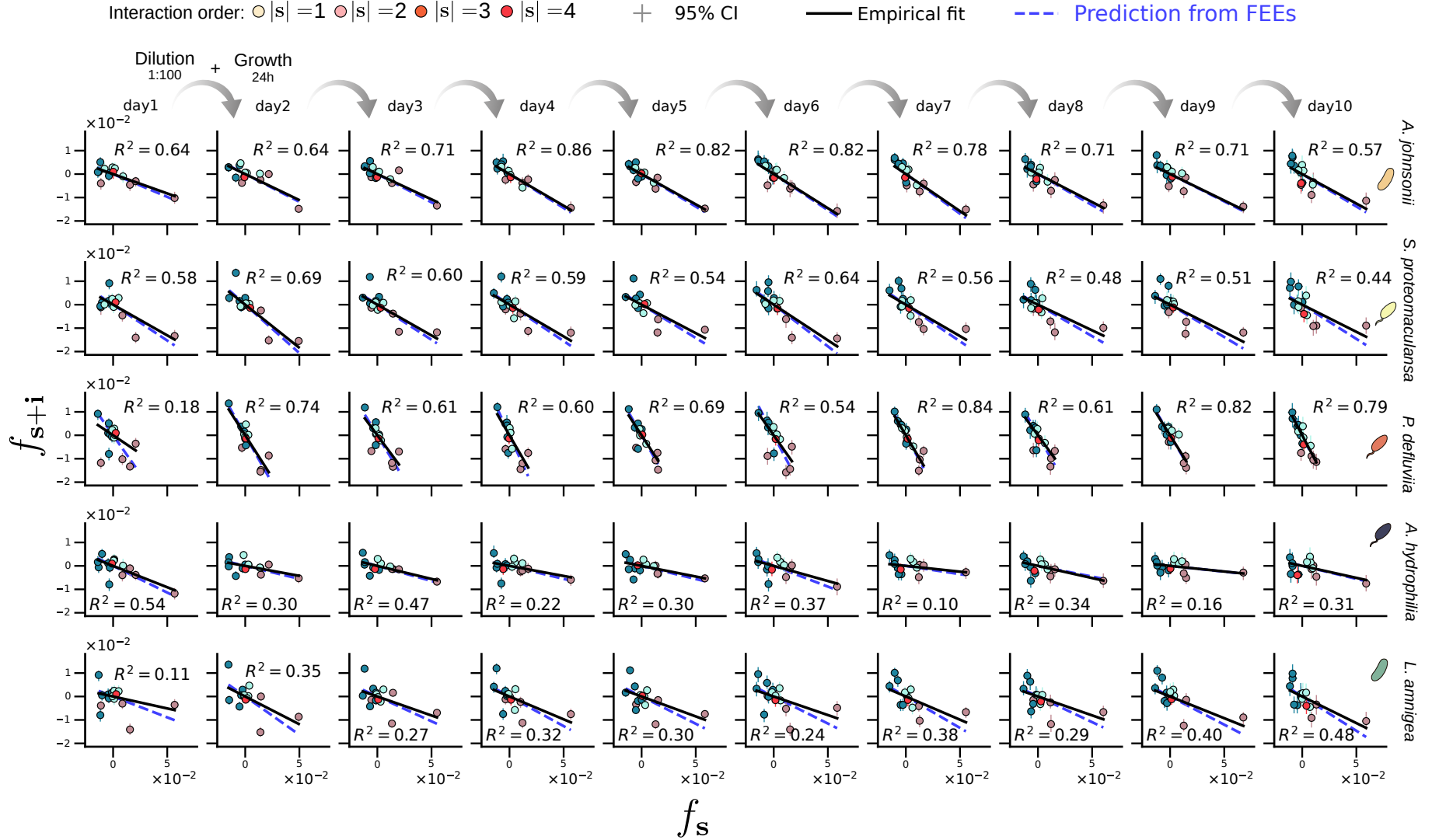


Figure S20: **WH interaction coefficients retain simple cross-order relationships during serial dilution.** Cross-order relationships among WH interaction coefficients are shown for each of the ten daily passages of the serial-dilution experiment. Columns correspond to passages, and rows to focal species. In each panel, WH coefficients involving focal species i , f_{s+i} , are plotted against the corresponding coefficients one order below, f_s , for subsets s not containing species i . Points are colored by interaction order, and error bars denote 95% bootstrap confidence intervals. Solid black lines show empirical linear fits, whereas dashed blue lines indicate the slopes predicted from the corresponding fitted FEE slope b_i , using $c_i = b_i/(2 + b_i)$ as given by Eq. (15). The coefficient of determination of the interaction-space fit, R_i^2 , is reported in each panel. Species names are indicated on the right.

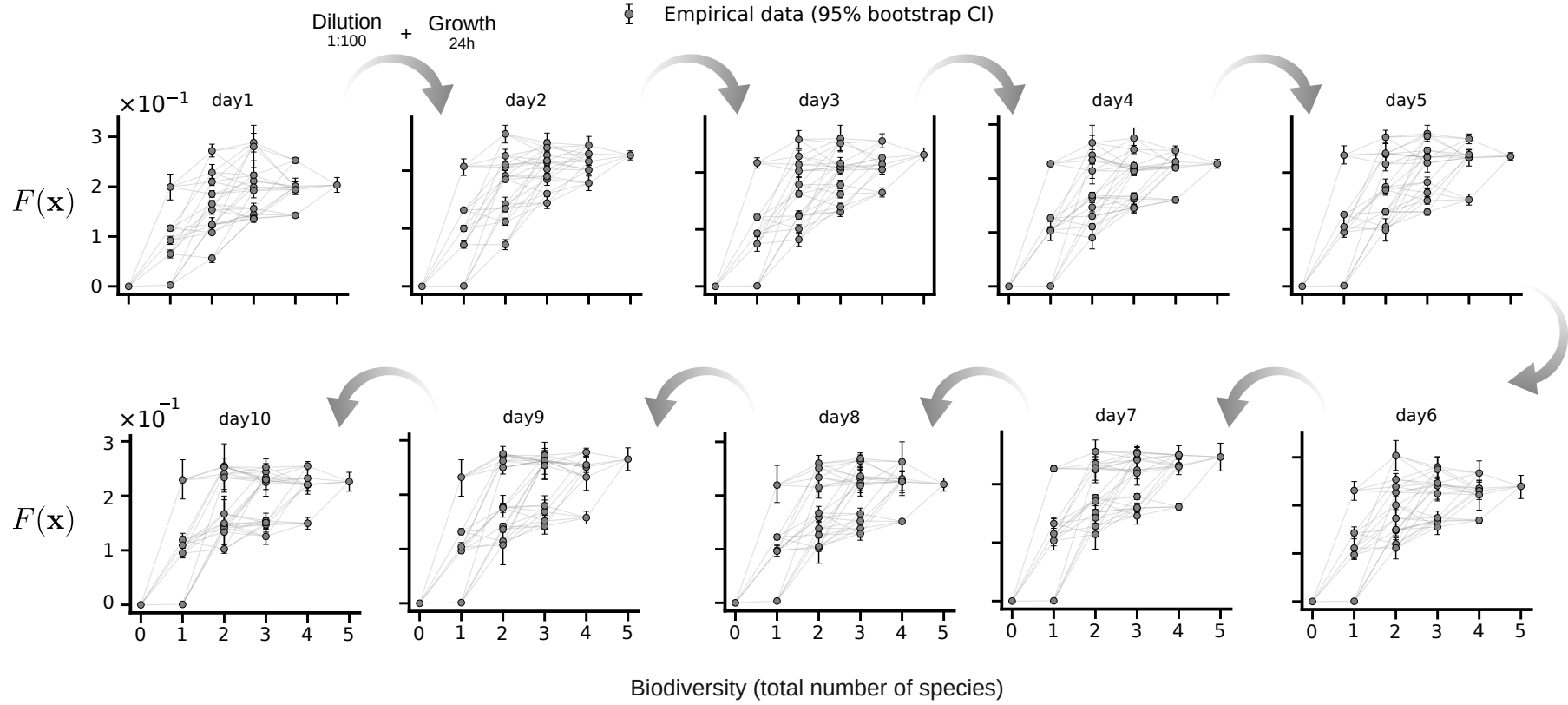


Figure S21: **Biomass landscapes projected onto community biodiversity during serial dilution.** For each of the ten daily passages, the biomass community-function landscape was projected onto community biodiversity, defined as the number of species present in the initial assemblage. Points show measured biomass values, $F(\mathbf{x})$, and error bars denote 95% bootstrap confidence intervals. Light grey lines connect communities that differ by the addition or removal of a single species, preserving the adjacency structure of the community-composition hypercube. The relationship between biomass and biodiversity remains broadly conserved across passages, indicating that the coarse organization of the biomass landscape persists during population dynamics.

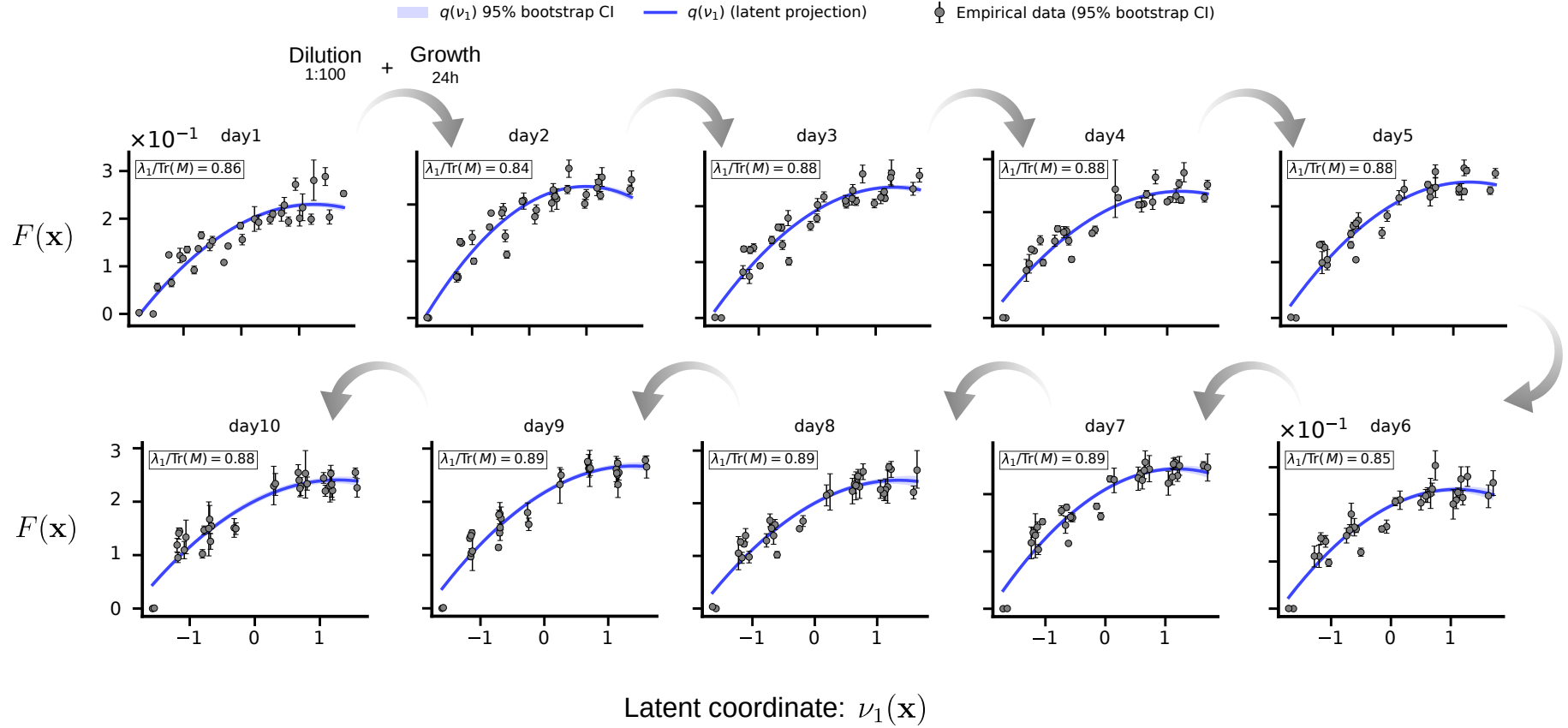


Figure S22: **Biomass landscapes retain a stable projection onto the leading collective coordinate during serial dilution.** For each of the ten daily passages, the biomass community-function landscape was projected onto its leading latent coordinate, $\nu_1(\mathbf{x}) = \mathbf{u}_1^\top \mathbf{x}$. Points show measured biomass values, $F(\mathbf{x})$, and error bars denote 95% bootstrap confidence intervals. Solid blue lines show the latent-coordinate prediction, $q(\nu_1)$, and shaded blue regions denote the corresponding 95% bootstrap confidence intervals. The relationship between biomass and $\nu_1(\mathbf{x})$ remains stable across passages, indicating that the dominant collective axis organizing biomass production is preserved during population dynamics.

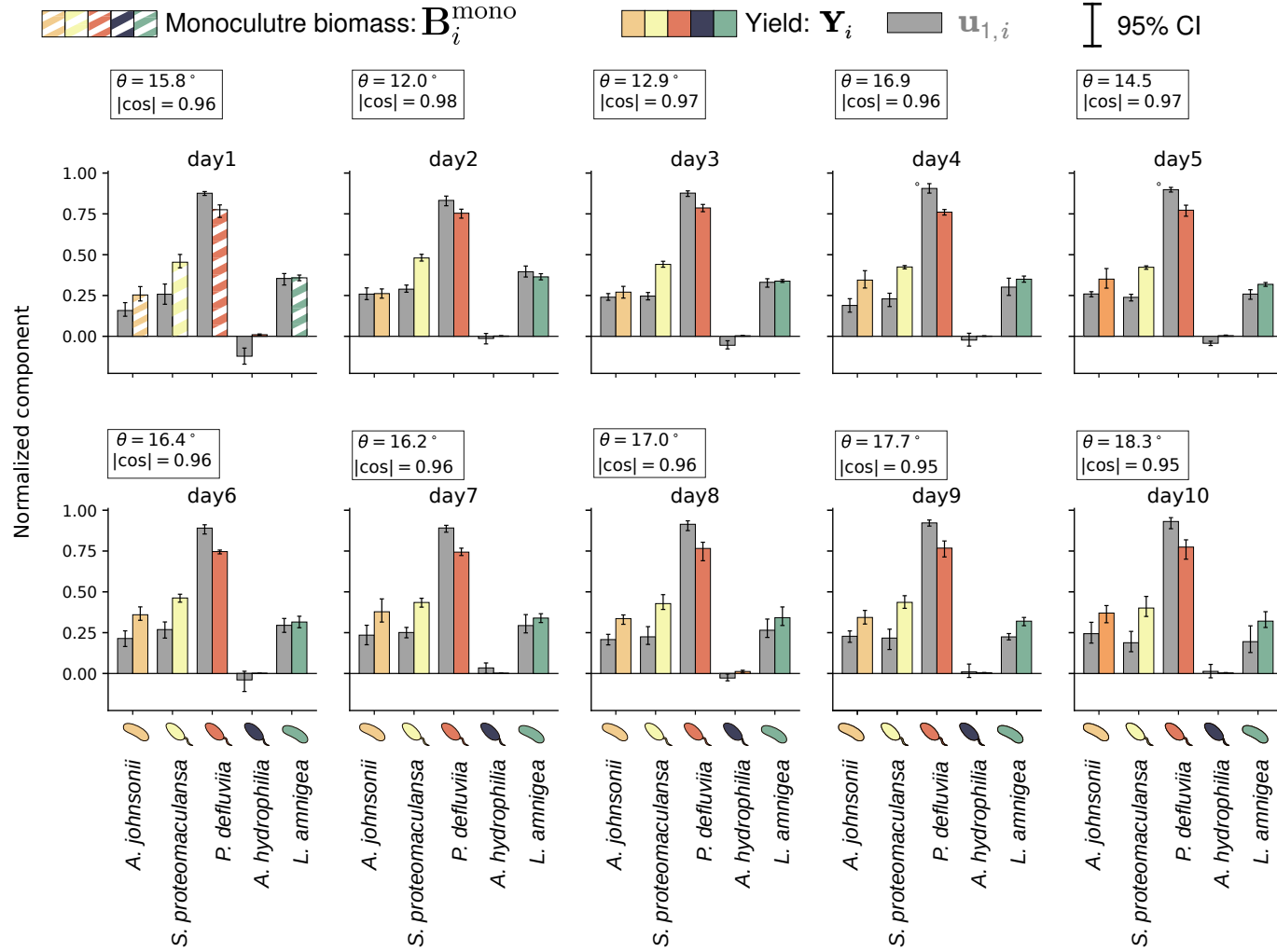


Figure S23: **The leading productivity mode remains aligned with species-level yield estimates during serial dilution.** For each passage, bars compare the normalized components of the leading collective mode, $\mathbf{u}_1$, with the corresponding normalized species-level productivity vector. On day 1, species productivity is approximated by monoculture biomass, $\mathbf{Y} \approx \mathbf{B}_{\text{mono}}$, because the yield estimator requires measurements from two consecutive passages. For days 2–10, species yields are estimated from monoculture biomass measured in consecutive passages using Eq. (1). Error bars denote 95% bootstrap confidence intervals. Insets report the sign-invariant angle, $\theta = \arccos(|\mathbf{u}_1^\top \mathbf{Y}|)$, and the corresponding absolute cosine similarity, $|\mathbf{u}_1^\top \mathbf{Y}|$, between the leading collective mode and the productivity vector used for each passage. The consistently high similarity indicates that the dominant collective mode remains aligned with species-specific productivity throughout the serial-dilution experiment.

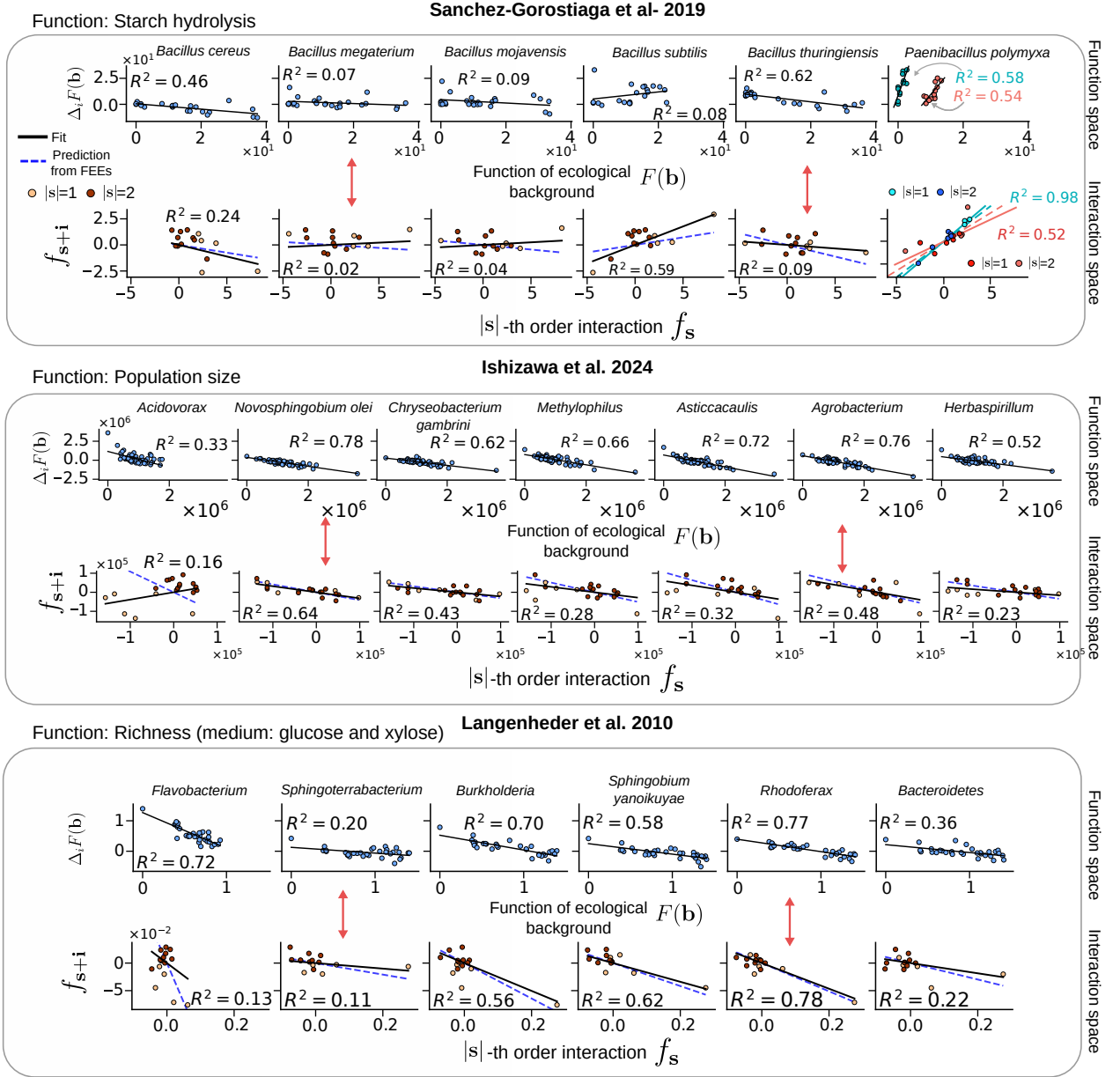


Figure S24: Functional Effect Equations predict cross-order dependencies across published community-function landscapes. Top, biomass data from Sanchez-Gorostiaga et al. (2019) [20]. Middle, population-size data from Ishizawa et al. (2024) [15]. Bottom, richness data from Langenheder et al. (2010) [17]. Each column corresponds to one focal taxon. Upper rows show the functional effect, $\Delta_i F(\mathbf{b})$, as a function of background function, $F(\mathbf{b})$. Lower rows show f_{s+i} against the corresponding lower-order coefficient f_s . Point colors indicate interaction order $|s|$. Black lines show empirical linear fits. Blue dashed lines show FEE predictions. Values report R^2 . For *Paenibacillus polymyxa* in Sanchez-Gorostiaga et al. (2019), the FEE separates into two branches depending on whether *Bacillus thuringiensis* is present or absent in the ecological background. Only first- and second-order relationships are shown because higher-order coefficients are difficult to estimate reliably without replication [3].

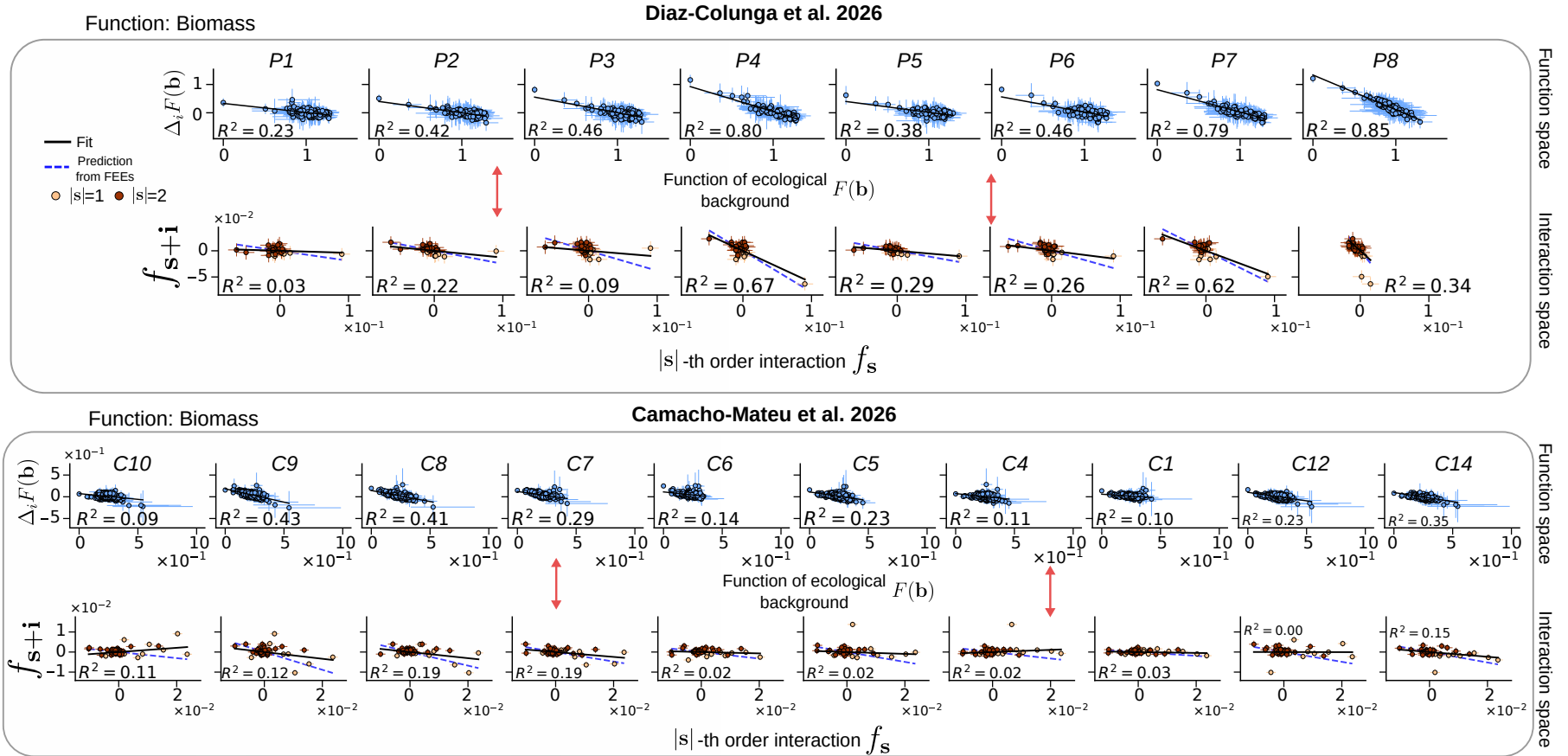


Figure S25: **Functional Effect Equations predict cross-order dependencies among ecological interactions in published biomass landscapes.** Top, Diaz-Colunga et al. (2026) [8]. Bottom, Camacho-Mateu et al. (2026) [3]. Each column corresponds to one focal species. Upper rows show the functional effect, $\Delta_i F(\mathbf{b})$, as a function of background biomass, $F(\mathbf{b})$. Lower rows show interaction coefficients involving the focal species, f_{s+i} , against the corresponding lower-order coefficients, f_s . Point colors indicate interaction order $|s|$. Black lines show empirical linear fits. Blue dashed lines show the slopes predicted from the corresponding FEEs. Values report the coefficient of determination, R^2 . Red arrows highlight representative correspondences between function space and interaction space. Stronger FEEs are generally associated with more pronounced cross-order organization of interaction coefficients.

Gould et al. 2018

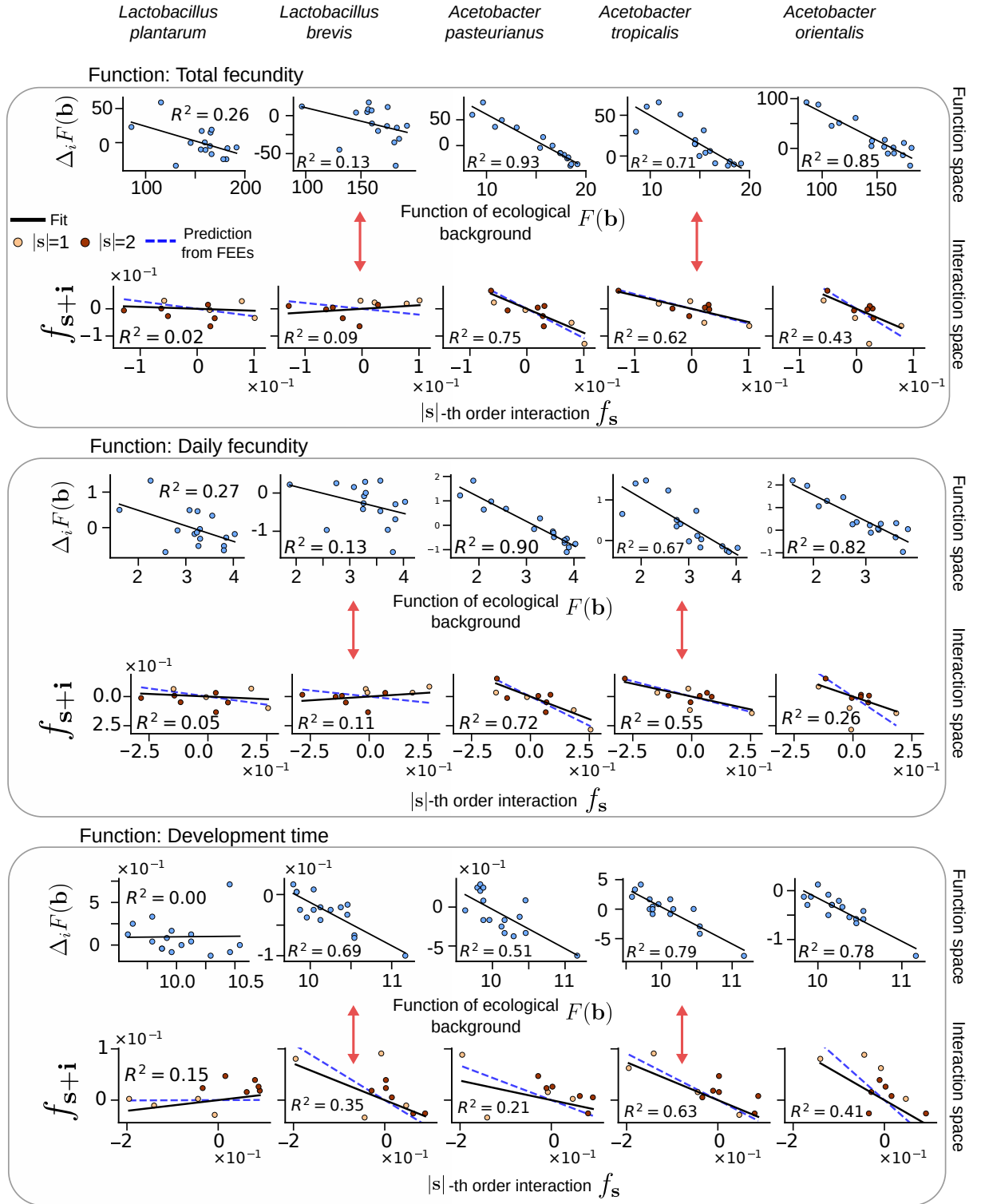


Figure S26: Functional Effect Equations and cross-order relationships among WH coefficients in the *Drosophila* microbiome landscape of Gould et al. (2018). Data are from Gould et al. [11]. Columns correspond to the five bacterial species and rows to total fecundity, daily fecundity, and development time. Upper panels show the functional effect of focal species i , $\Delta_i F(b)$, as a function of the community function in ecological background b . Lower panels compare WH coefficients involving the focal species, f_{s+i} , with the corresponding lower-order coefficients, f_s , where $|s|$ denotes interaction order. Solid black lines show empirical fits, dashed blue lines show slopes predicted from the FEEs, and R^2 values quantify linear fit quality.

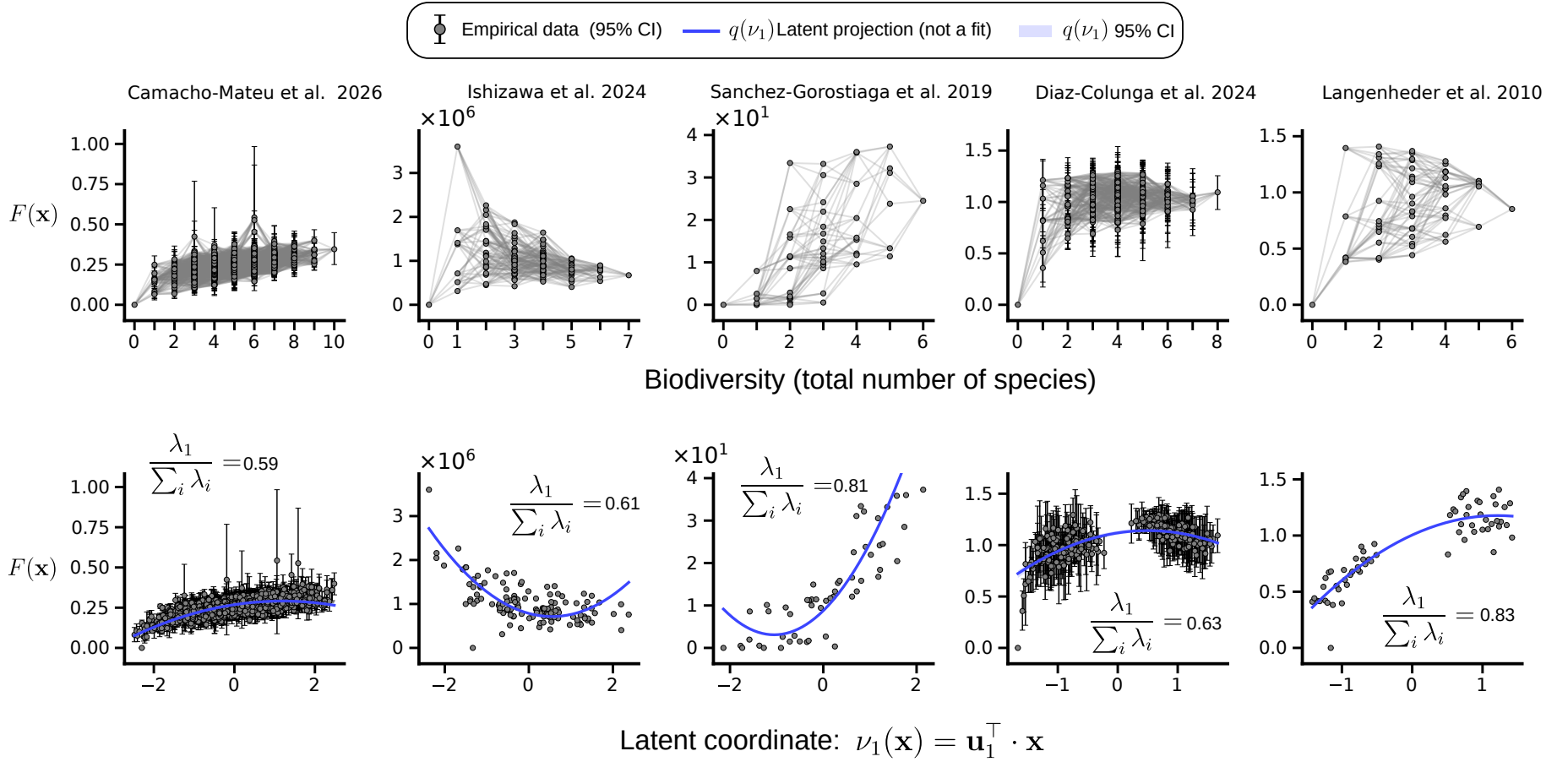


Figure S27: **One-dimensional latent-coordinate projections of previously published community-function landscapes.** Five fully sampled community-function landscapes from previously published datasets were analyzed using the same framework applied to our experimental data (Table T2). Each row corresponds to one dataset. Left panels show the community function, $F(\mathbf{x})$, as a function of biodiversity, whereas right panels show $F(\mathbf{x})$ projected onto the leading latent coordinate, $\nu_1(\mathbf{x}) = \mathbf{u}_1^\top \mathbf{x}$. Black points show empirical measurements, and vertical error bars denote 95% bootstrap confidence intervals when replicate measurements are available. Solid blue lines show the predicted response, $q(\nu_1)$, and shaded blue regions denote the corresponding 95% bootstrap confidence intervals. Labels report the fraction of total spectral weight captured by the leading mode, $\lambda_1 / \text{Tr}(\mathbf{M})$.

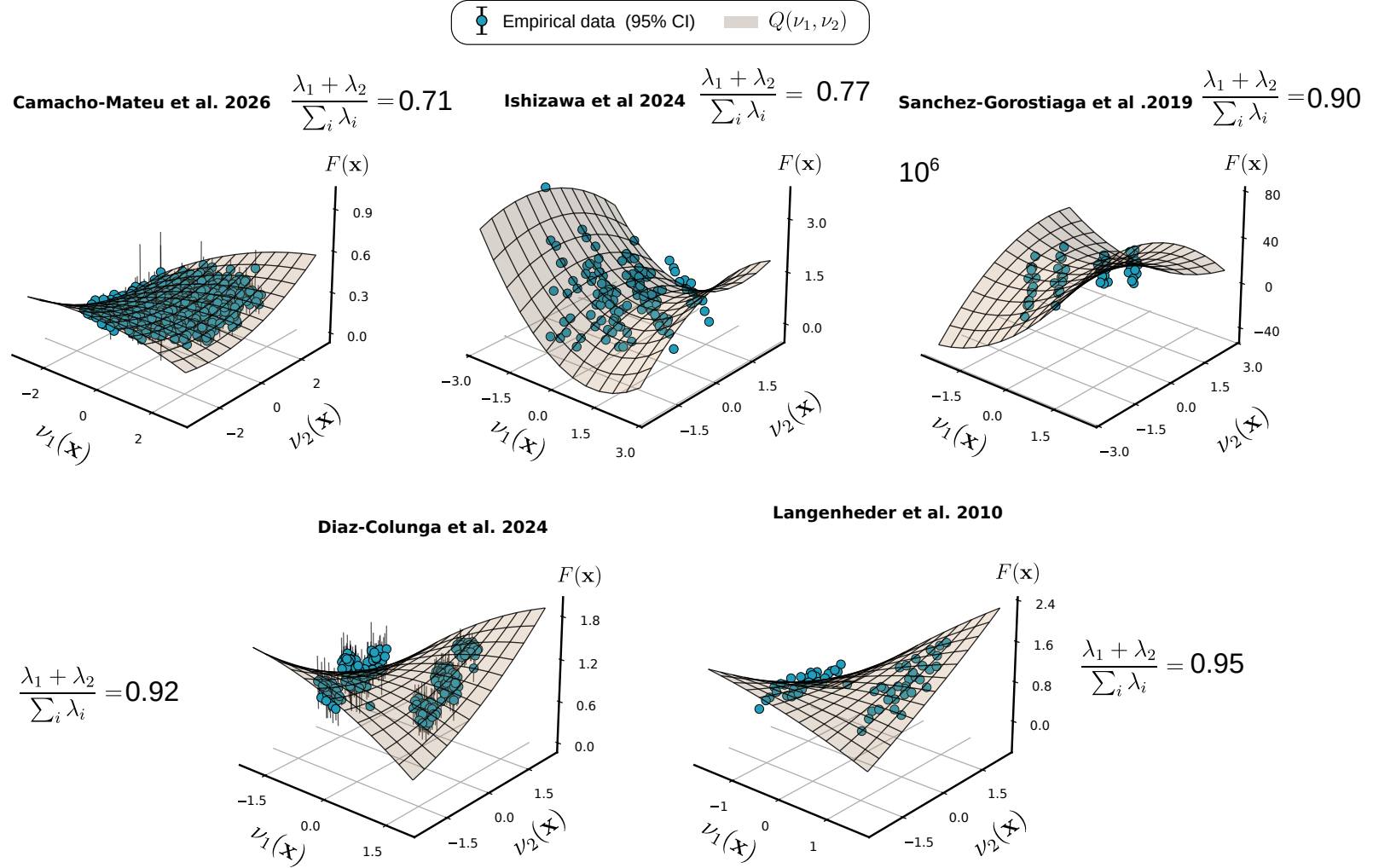


Figure S28: **Two-dimensional latent-coordinate manifolds of previously published community-function landscapes.** The same five datasets shown in Fig. S27 were projected onto their first two latent coordinates, $\nu_1(\mathbf{x}) = \mathbf{u}_1^\top \mathbf{x}$ and $\nu_2(\mathbf{x}) = \mathbf{u}_2^\top \mathbf{x}$. Points show empirical measurements, and vertical error bars denote 95% bootstrap confidence intervals when replicate measurements are available. Surfaces show the predicted quadratic latent manifold, $Q(\nu_1, \nu_2)$. Labels report the fraction of total spectral weight captured by the first two modes, $(\lambda_1 + \lambda_2)/\text{Tr}(\mathbf{M})$.

Gould et al. 2018

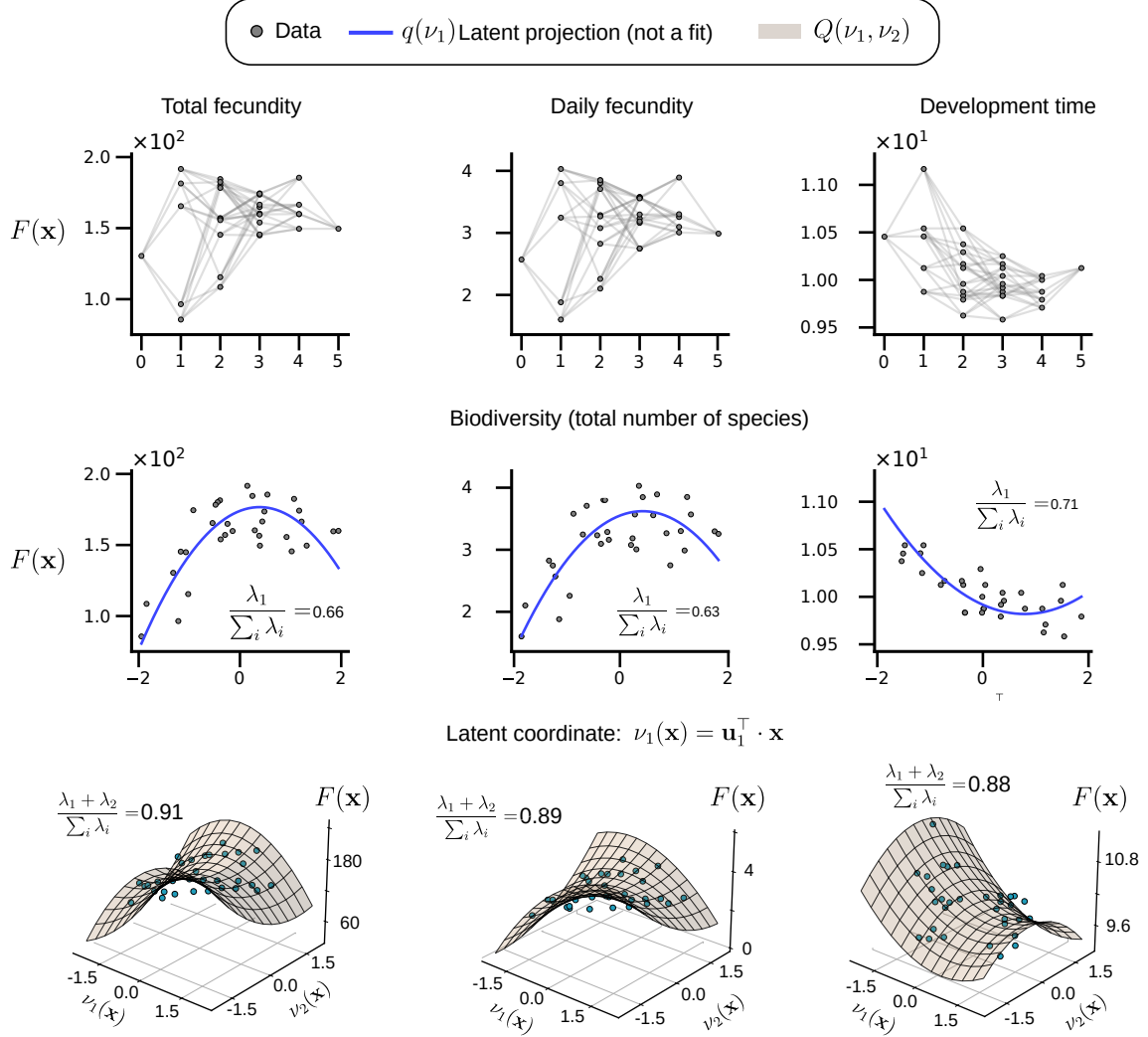


Figure S29: **Low-dimensional latent projections of host fitness landscapes in the *Drosophila* microbiome dataset of Gould et al. (2018).** Data are from Gould et al. [11]. Columns correspond to total fecundity, daily fecundity, and development time. Top row, measured community function $F(\mathbf{x})$ plotted against bacterial biodiversity, defined as the total number of species in each community. Grey points denote observed community functions, and gray lines connect communities that differ by the addition or removal of a single bacterial species. Middle row, the same landscapes projected onto the leading latent coordinate, $\nu_1(\mathbf{x}) = \mathbf{u}_1^\top \mathbf{x}$, where $\mathbf{u}_1$ is the dominant eigenvector of the second-moment matrix of marginal effects. Blue curves show the one-dimensional latent projection $q(\nu_1)$ obtained from the second-order WH representation; these curves are determined by the spectral projection and are not fitted directly to the plotted data. Insets report the fraction of total spectral weight captured by the leading mode, $\lambda_1 / \text{Tr}(\mathbf{M})$. Bottom row, two-dimensional representations of each landscape in the coordinates (ν_1, ν_2) . Cyan points show the measured functions, and the surfaces show the quadratic latent projection $Q(\nu_1, \nu_2)$. The reported values $(\lambda_1 + \lambda_2) / \text{Tr}(\mathbf{M})$ indicate the fraction of spectral weight captured by the first two collective modes. Despite the weak organization of the landscapes by species biodiversity, one or two ecological collective modes capture most of the coordinated variation in marginal effects and provide a compact representation of all three host traits.

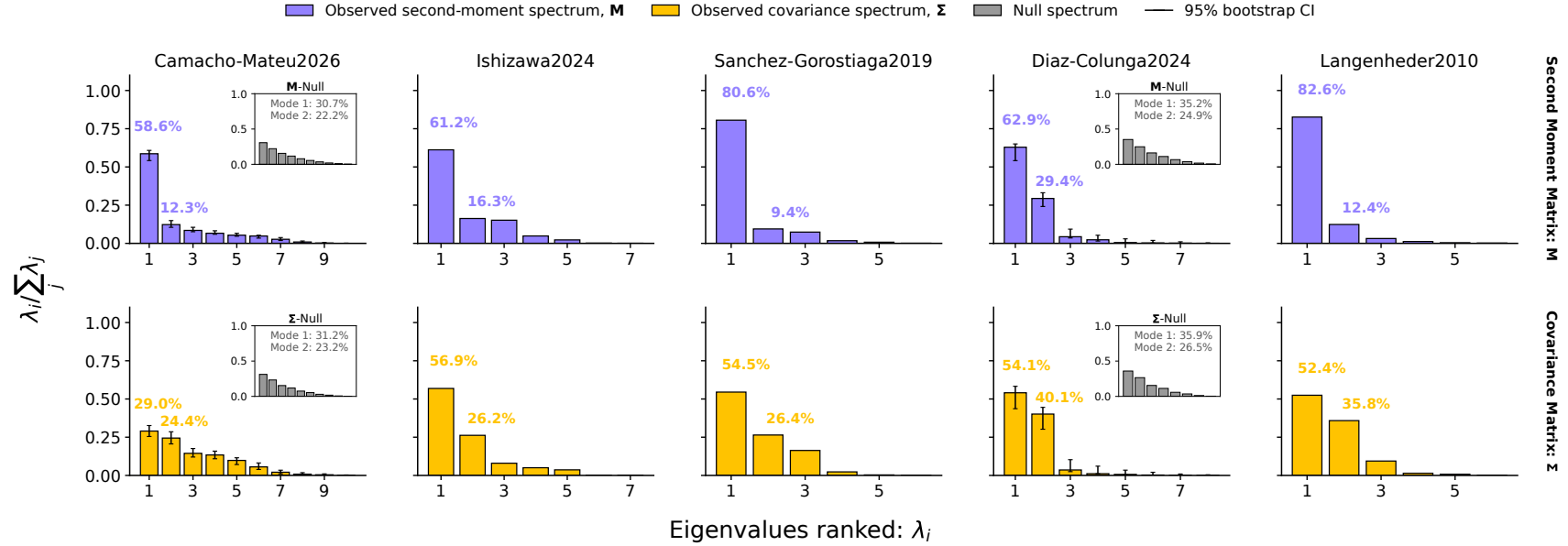


Figure S30: **Second-moment and covariance eigenspectra of previously published community-function landscapes.** Five fully sampled community-function landscapes from previously published datasets were analyzed using the same spectral framework applied to our experimental data (Table T2). Columns correspond to datasets. Top panels show the normalized eigenspectrum of the second-moment matrix of marginal effects, $\mathbf{M}$, whereas bottom panels show the normalized eigenspectrum of the centered covariance matrix, Σ . Bars report $\lambda_i / \sum_j \lambda_j$, with eigenvalues ranked in decreasing order; error bars denote 95% bootstrap confidence intervals when replicate measurements are available. Labels above the first two bars indicate the fractions of total spectral weight captured by the leading two modes. Insets show the corresponding mean null eigenspectra and report the spectral weight of the first two null modes. Across datasets, the spectra of $\mathbf{M}$ are strongly concentrated in one or two leading modes. The spectra of Σ are also concentrated relative to the null expectation, indicating that the background-dependent variation of marginal effects is itself organized along a small number of collective directions.

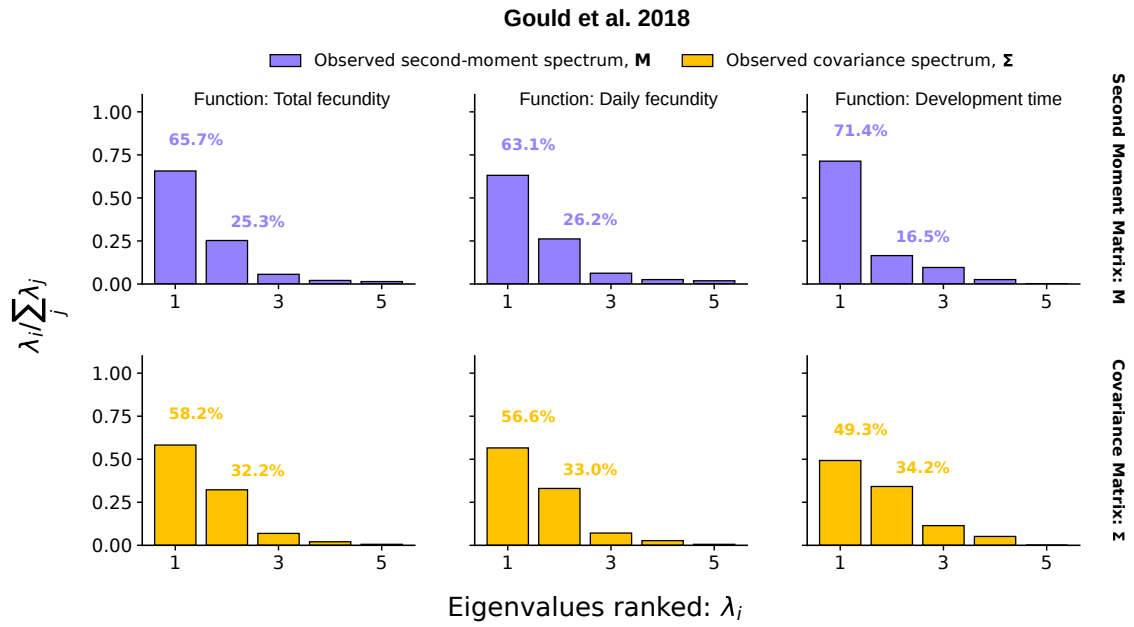


Figure S31: Second-moment and covariance eigenspectra of the *Drosophila* microbiome landscape of Gould et al. (2018). Data are from Gould et al. [11].