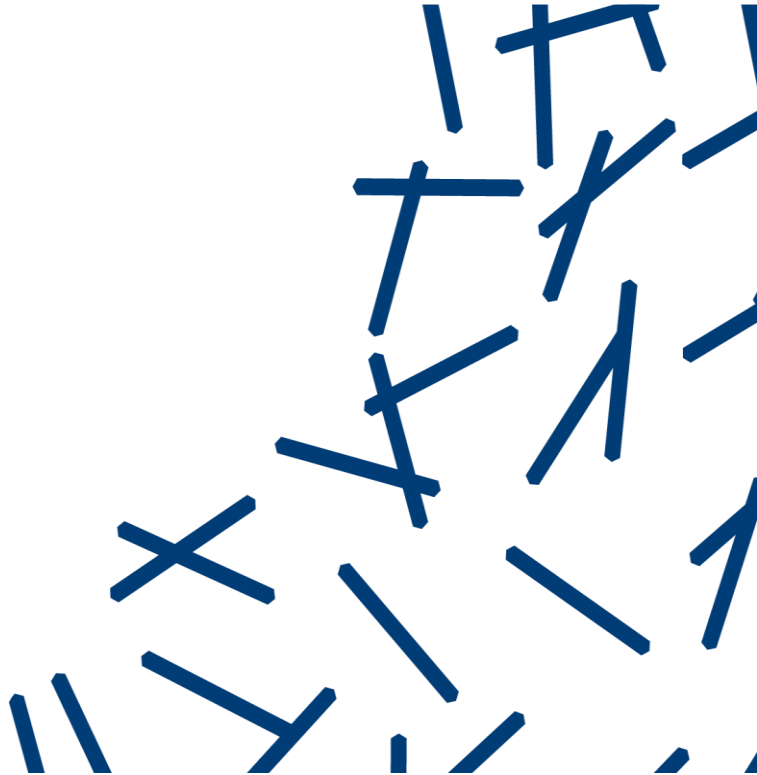
Niche dimensionality governs microbial community structure

Purushottam Dixit
Biomedical Engineering



Yale ENGINEERING

Acknowledgments

Current group

- Jeremy Barrios (Physics)
- Brooke Emison (BME)
- **Karthik Srinivasan (BME)**
- Lucy Sullivan (BME)

Alumni

- Andrew Goetz (Genentech)
- Hoda Akl (Generate: Biomedicine)
- Mayar Shahin
- Bridgette Gifford (NIH IRTA program)
- Lukas Herron (IPST, UMD)
- Xiaochuan Zhao (Industry)

Collaborators

- **German Plata** (BiomEdit LLC)
- Luca Gerosa (Genentech)

Funding

My Agenda

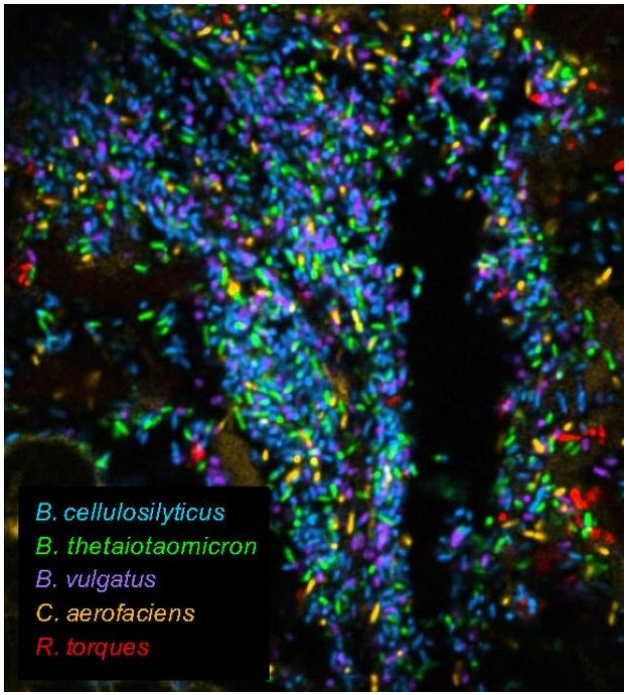
- Question a key assumption in models of microbiomes built by physicists
  - No *spin glasses*

- Advocate for a *low dimensional* latent variable picture

Ecosystems are diverse and we are still grappling with it!

Riparian ecosystem

mouse gut microbiome

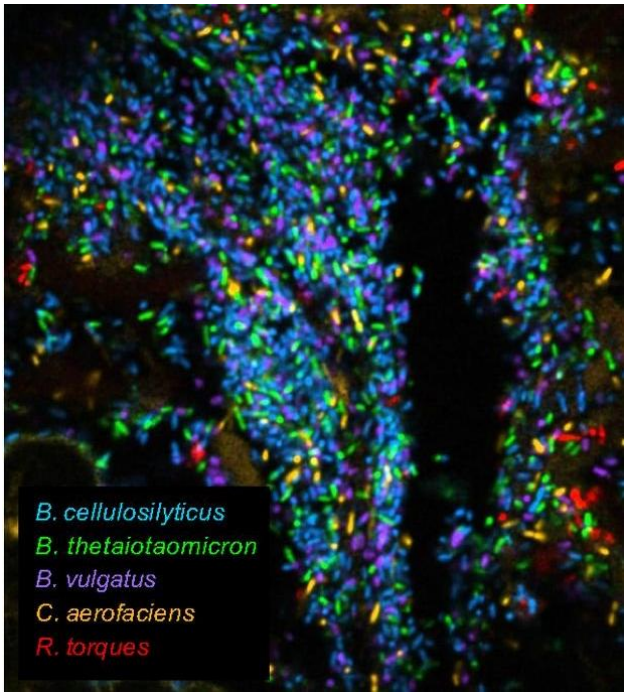


Welch et al. PNAS 2017

Ecosystems are diverse and we are still grappling with it!

Riparian ecosystem

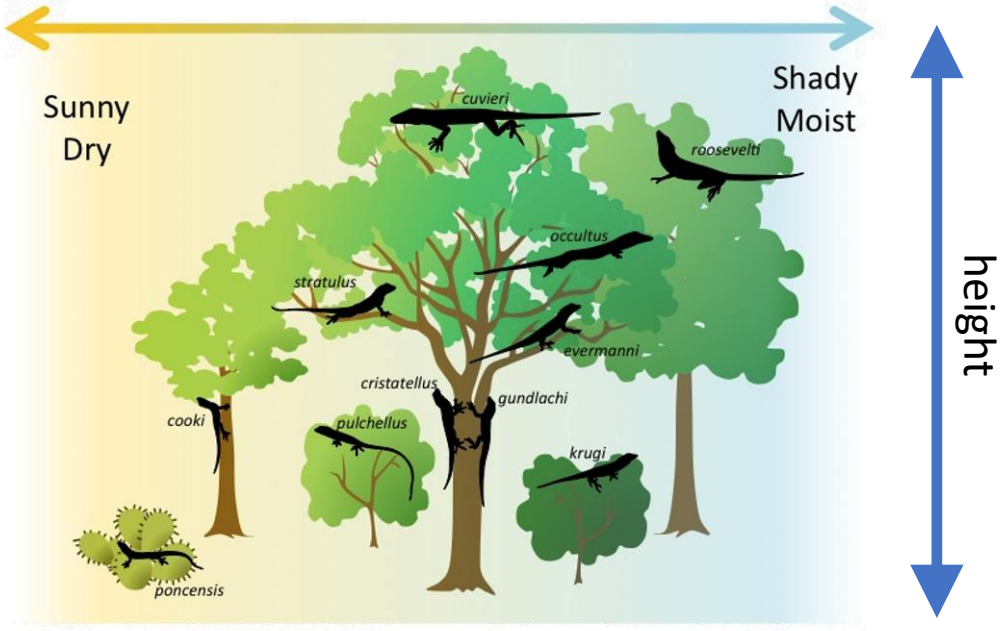
mouse gut microbiome



Welch et al. PNAS 2017

- In most ecosystems, 100s-1000s of species live in similar area at the same time
  - Represent billions of years of evolutionary relationships!

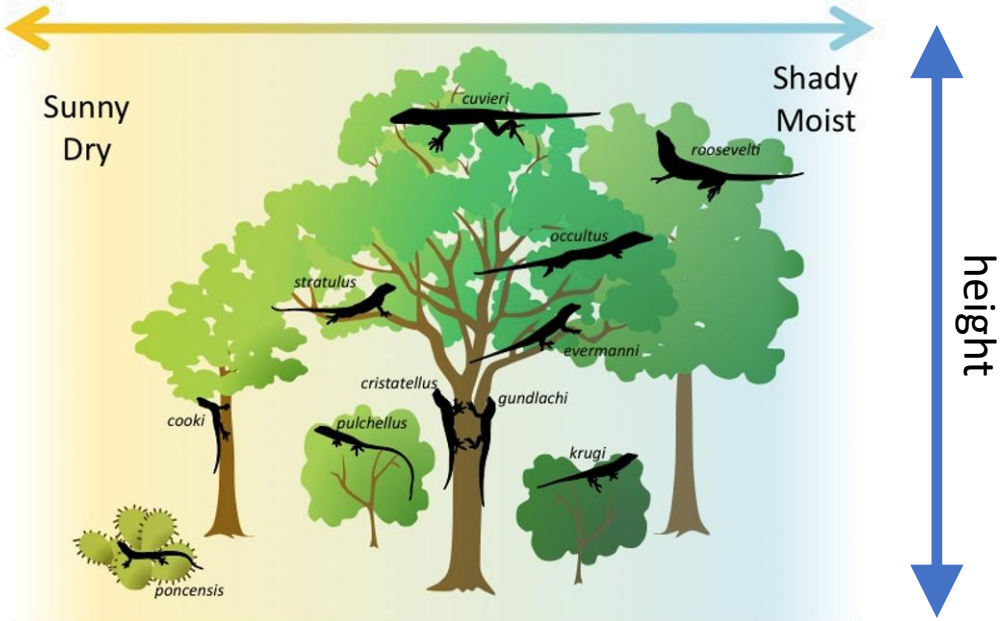
Hutchinson's multidimensional *niche theory* rationalizes species coexistence via niche overlap

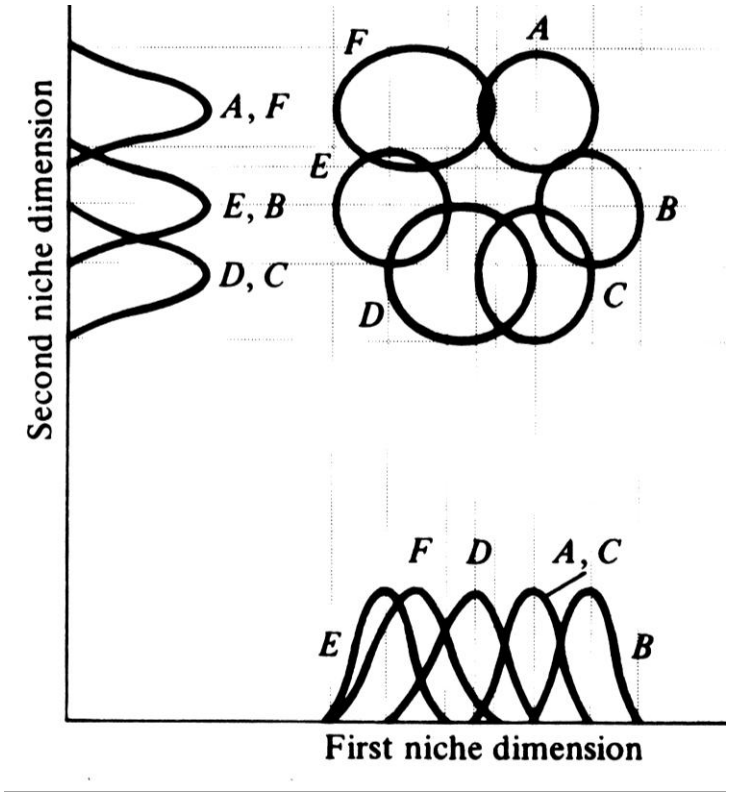


- Species thrive in specific regions of a N-dimensional *niche hypervolume*
  - Coexistence is facilitated by reduced overlap between species

MacArthur, 1984

Hutchinson's multidimensional *niche theory* rationalizes species coexistence via niche overlap



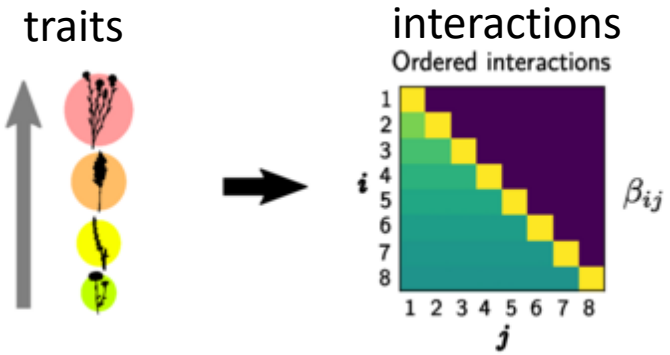


- Species thrive in specific regions of a N-dimensional *niche hypervolume*
  - Coexistence is facilitated by reduced overlap between species
- High niche dimensionality facilitates coexistence

MacArthur, 1984

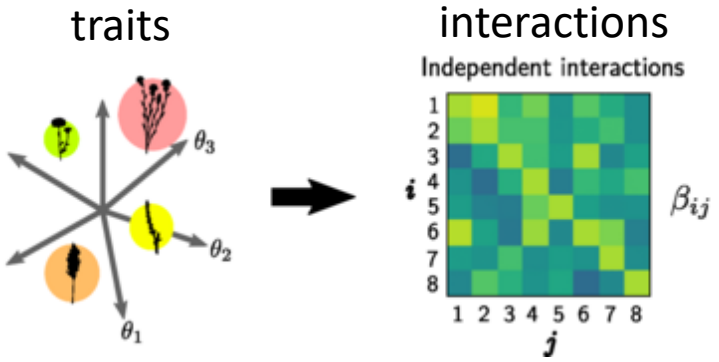
Niche dimensionality η_{D} dictates modes of co-existence

Tradeoffs in low dimensional ecosystems



niche dimensionality hypothesis
(Harpole, Nature 2007)

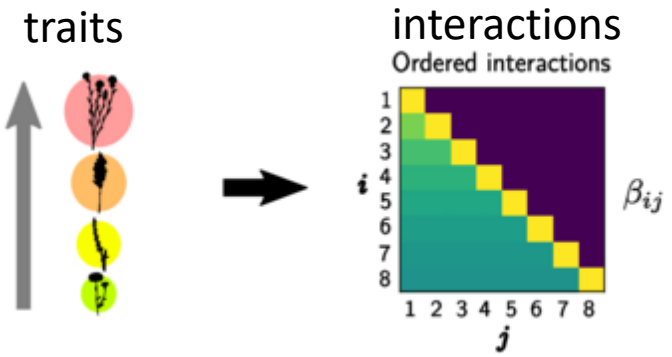
High dimensional competition



Barbier et al. Phys. Rev. X (2021)

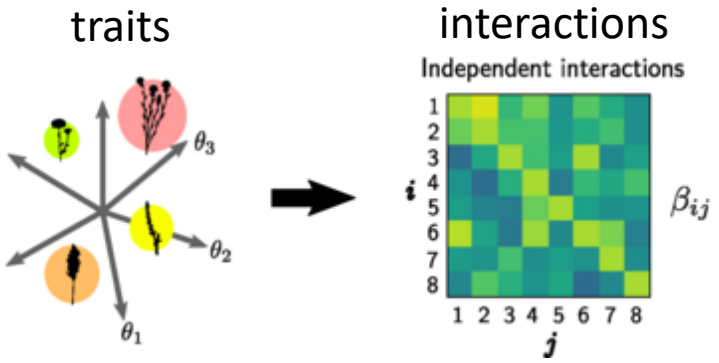
Niche dimensionality η_{D} dictates modes of co-existence

Tradeoffs in low dimensional ecosystems



niche dimensionality hypothesis
(Harpole, Nature 2007)

High dimensional competition

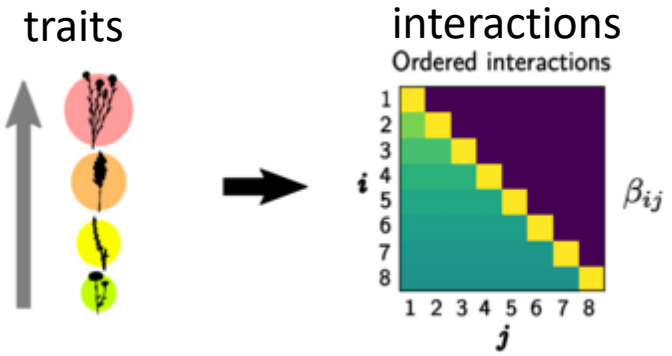


- BUT, estimating η_{D} is difficult: **impossible to catalog** all relevant niche variables
  - organism/environment relationship is heterogeneous
  - environmental dimensions are be correlated

Barbier et al. Phys. Rev. X (2021)

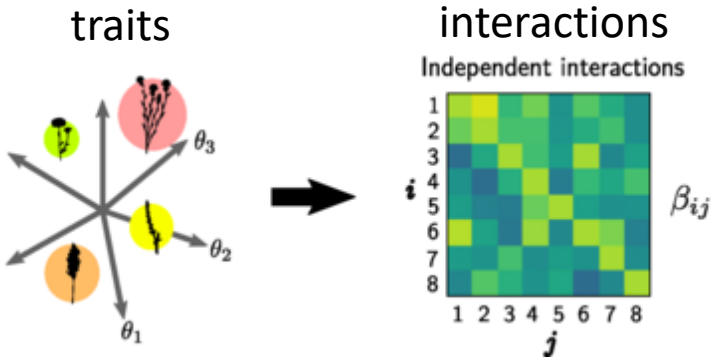
Niche dimensionality η_{D} dictates modes of co-existence

Tradeoffs in low dimensional ecosystems



niche dimensionality hypothesis
(Harpole, Nature 2007)

High dimensional competition



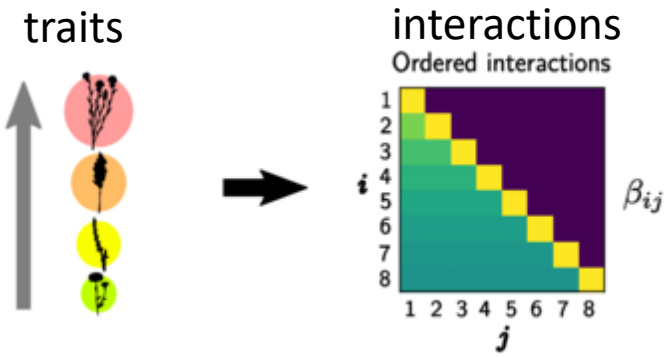
- BUT, estimating η_{D} is difficult: **impossible to catalog** all relevant niche variables
  - organism/environment relationship is heterogeneous
  - environmental dimensions are be correlated

- Ecologists take an effective approach (JE Cohen, 1977 and others):
  - search for a low rank embedding that reproduces ecological patterns

Barbier et al. Phys. Rev. X (2021)

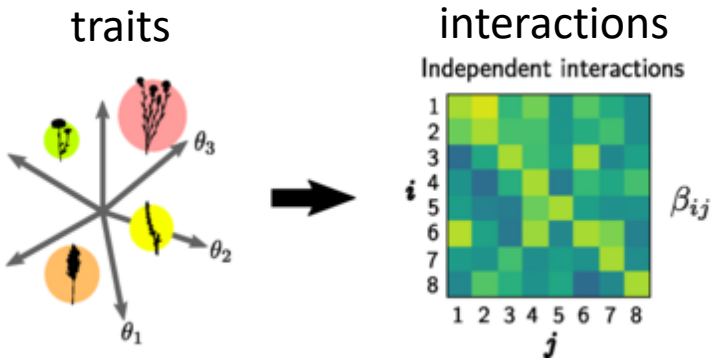
Niche dimensionality η_{D} dictates modes of co-existence

Tradeoffs in low dimensional ecosystems



niche dimensionality hypothesis
(Harpole, Nature 2007)

High dimensional competition



- BUT, estimating η_{D} is difficult: **impossible to catalog** all relevant niche variables
  - organism/environment relationship is heterogeneous
  - environmental dimensions are be correlated

- Ecologists take an effective approach (JE Cohen, 1977 and others):
  - search for a low rank embedding that reproduces ecological patterns
  - **very low η_{D}** ($\eta_{\mathrm{D}} \sim 1 - 10$) explains structure of food webs (Cohen, Allesina, etc.)

Barbier et al. Phys. Rev. X (2021)

Models of microbiomes have invariably assumed $\eta_{\mathrm{D}} \to \infty$

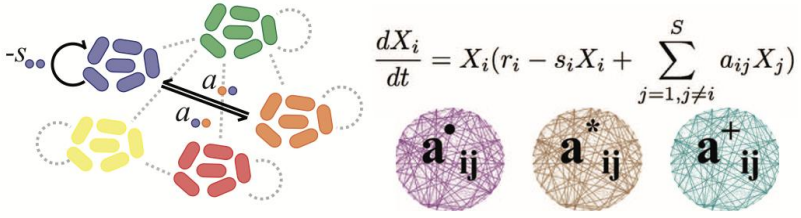
- Niche dimensionality of microbiomes is largely absent from our vocabulary

Models of microbiomes have invariably assumed $\eta_{\mathrm{D}} \to \infty$

- Niche dimensionality of microbiomes is largely absent from our vocabulary
- Theoretical studies often assume high dimensional niches
  - interaction parameters in GLV models are ***iid***

Models of microbiomes have invariably assumed $\eta_D \to \infty$

- Niche dimensionality of microbiomes is largely absent from our vocabulary
- Theoretical studies often assume high dimensional niches
  - interaction parameters in GLV models are ***iid***



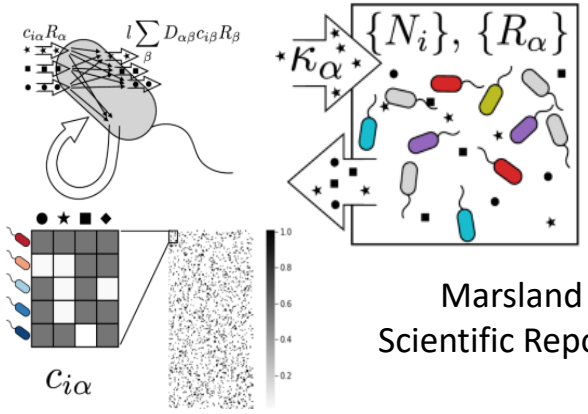
Coyte et al., Science (2015)

Microbiomes Through the Looking Glass: Linking Species Interactions to Dysbiosis through a Disordered Lotka-Volterra Framework

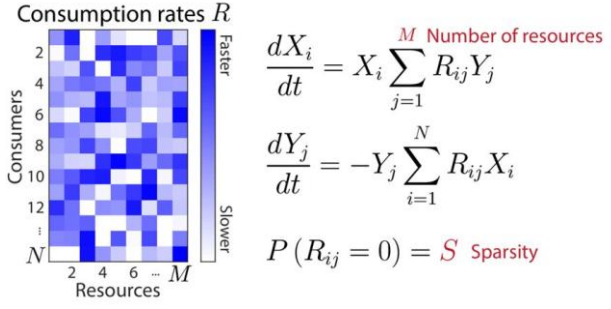
Jacopo Pasqualini, Amos Maritan, Andrea Rinaldo, Sonia Facchin, Edoardo Savarino, Ada Altieri, Samir Suweis

Dipartimento di Fisica "G. Galilei" e INFN sezione di Padova, Università di Padova, Padua, Italy • Dipartimento di Ingegneria Civile, Edile e Ambientale ICEA, University of Padova, Padua, Italy • EPFL, Ecole Polytechnique Fedéralé Lausanne, Lausanne, Switzerland
... show 3 more

Pasqualini et al. eLife, 2025



Marsland et al.,
Scientific Reports (2020)

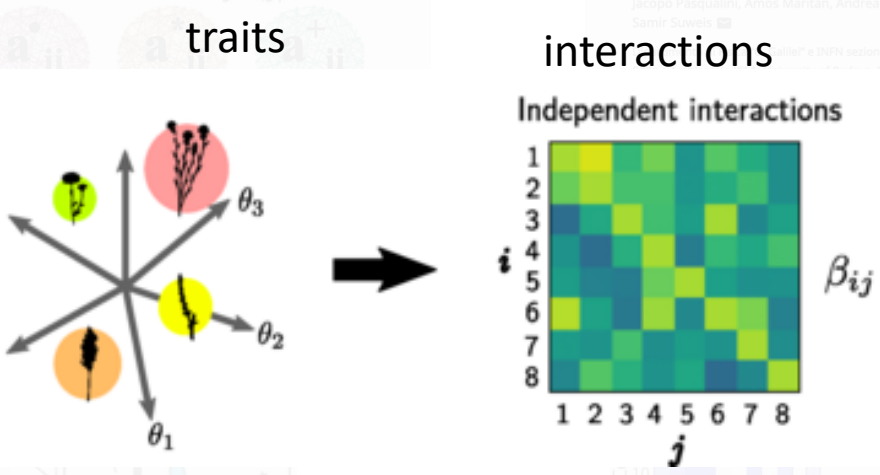


Ho et al., eLife (2022)

Models of microbiomes have invariably assumed $\eta_{\mathrm{D}} \to \infty$

- Niche dimensionality of microbiomes is largely absent from our vocabulary
- Theoretical studies often assume high dimensional niches
  - interaction parameters in GLV models are ***iid***

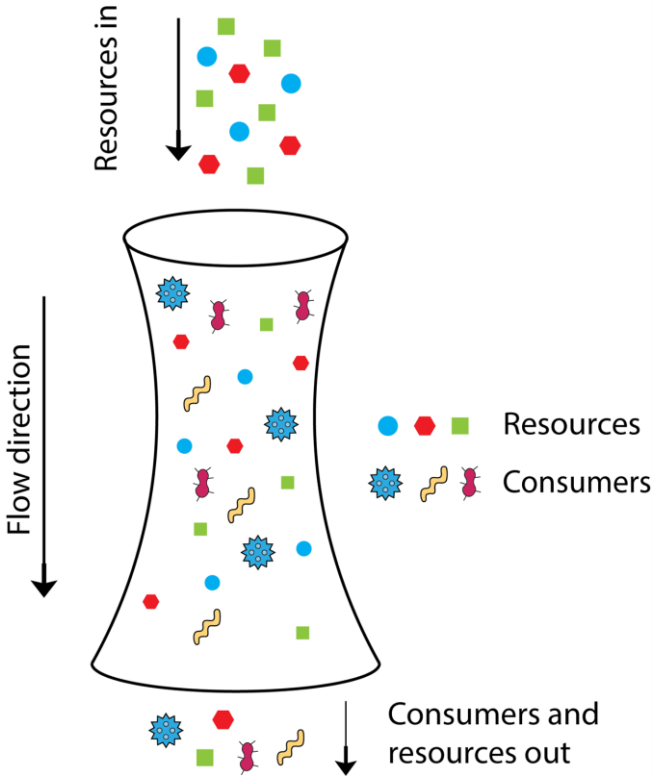
High dimensional competition



Is niche dimensionality relevant for microbiomes?

Is niche dimensionality relevant for microbiomes?
(**YES**)

Variations in microbiome abundances are often conceptualized using the consumer/resource model



species abundance evolution

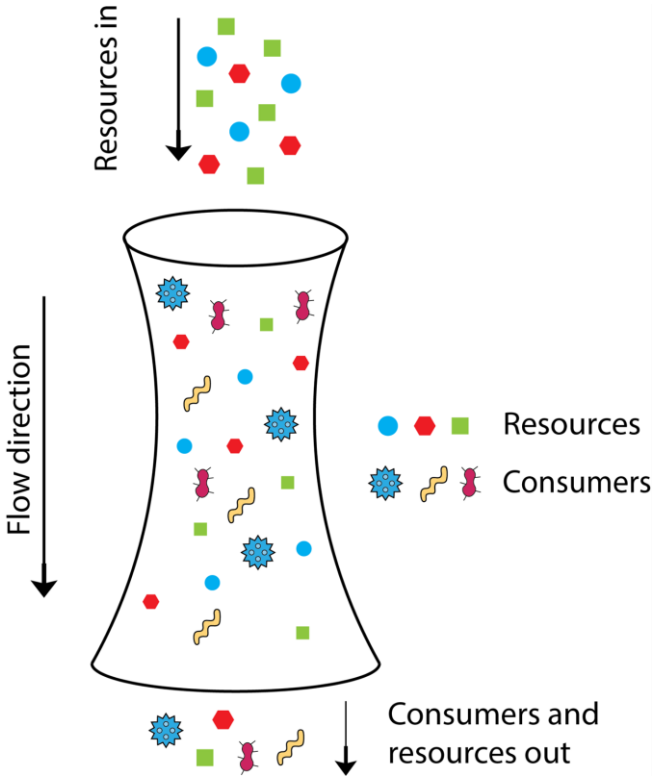
$$\frac{1}{N_o}\frac{dN_o}{dt} = \mu_o(t) = \sum_k r_k \theta_{ko}$$

resource dynamics

$$\frac{dr_k}{dt} = I_k(t) - \sum_o \theta_{ko} N_o r_k$$

Stephanopolous, Aristidou, Nielsen, Metabolic Engineering
Macarthur, PNAS 1970

But the C/R model only captures broad statistical features of microbiomes

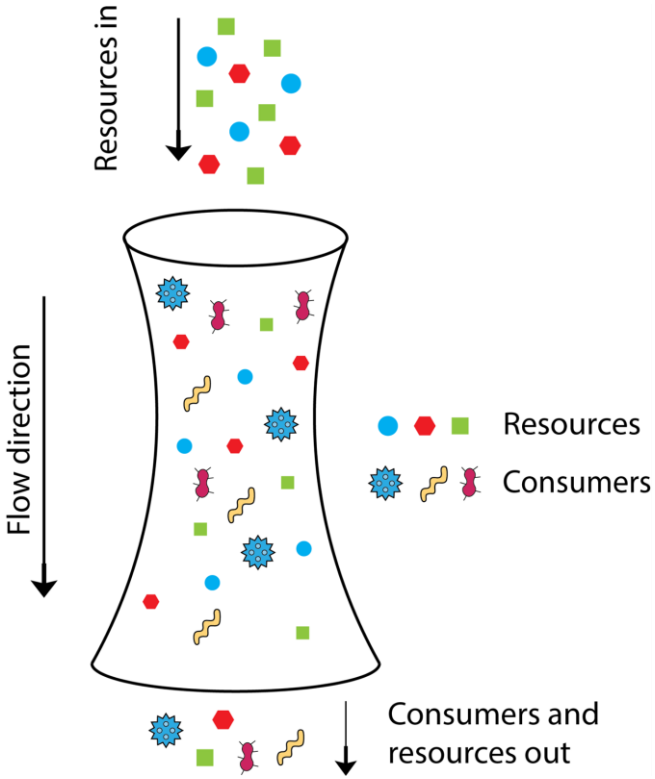


growth rate $$\frac{1}{N_o}\frac{dN_o}{dt} = \mu_o(t) = \sum_k r_k \theta_{ko}$$

resource dynamics $$\frac{dr_k}{dt} = I_k(t) - \sum_o \theta_{ko} N_o r_k$$

- **Practical**: Fitting model to data is essentially impossible
  - Needs absolute abundance measurement
  - Knowledge of consumer preferences, cross-feeding, resource flow rates
  - Dynamics could be chaotic
- **Only statistical predictions**:
  - Most current works use **randomly drawn parameters**
    - Predicts general features such as Taylor's law, species abundance distribution, etc.
  - **Cannot model** specific host-microbiome co-variations

We reformulate the C/R model as a latent histories model



Abundance fluctuations

$$N_o(t) \approx \exp\left(\int_0^t \mu_o(\tau)d\tau\right)$$

Growth rate from resources

$$\mu_o(\tau) = \sum_{k=1}^{K} r_k(\tau)\theta_{ko}$$

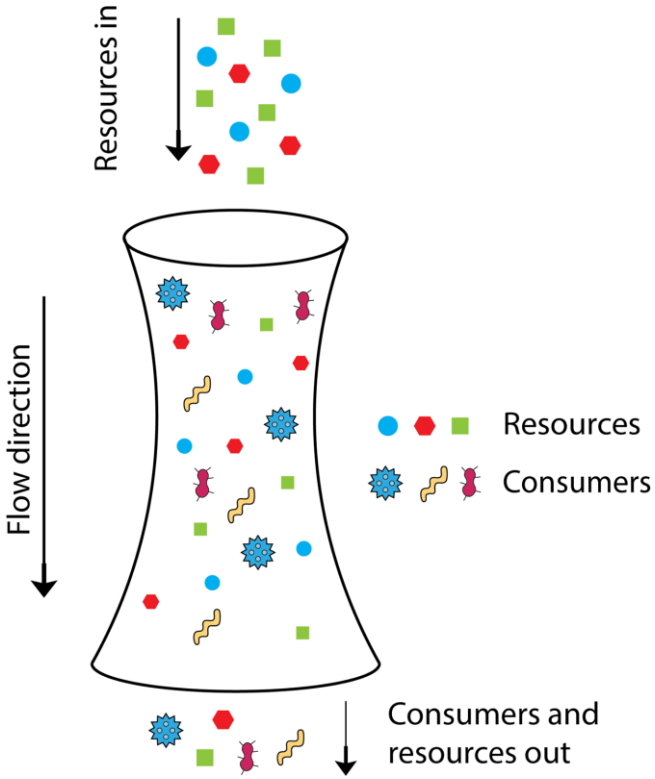
growth rate

$$\frac{1}{N_o}\frac{dN_o}{dt} = \mu_o(t) = \sum_k r_k\theta_{ko}$$

resource dynamics

$$\frac{dr_k}{dt} = I_k(t) - \sum_o \theta_{ko}N_o r_k$$

We reformulate the C/R model as a latent histories model



growth rate $\frac{1}{N_o}\frac{dN_o}{dt} = \mu_o(t) = \sum_k r_k \theta_{ko}$

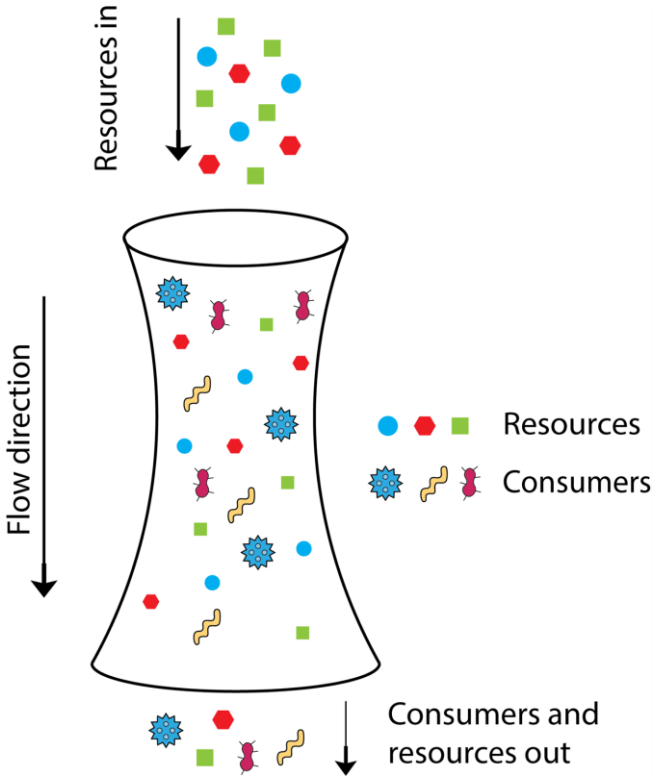
resource dynamics $\frac{dr_k}{dt} = I_k(t) - \sum_i \theta_{ki} N_i r_k$?

Abundance fluctuations $N_o(t) \approx \exp\left(\int_0^t \mu_o(\tau) d\tau\right)$

Growth rate from resources $\mu_o(\tau) = \sum_{k=1}^{K} r_k(\tau)\theta_{ko}$

- Resource dynamics are often unknown
  - Host is not a simple chemical reactor with passive inflow and outflow
- We take a phenomenological approach

We reformulate the C/R model as a latent histories model



Abundance fluctuations

$$N_o(t) \approx \exp\left(\int_0^t \mu_o(\tau)d\tau\right)$$

Growth rate from resources

$$\mu_o(\tau) = \sum_{k=1}^{K} r_k(\tau)\theta_{ko}$$

Define latent histories

$$z_k(t) = -\int_0^t r_k(\tau)d\tau$$

growth rate

$$\frac{1}{N_o}\frac{dN_o}{dt} = \mu_o(t) = \sum_k r_k\theta_{ko}$$

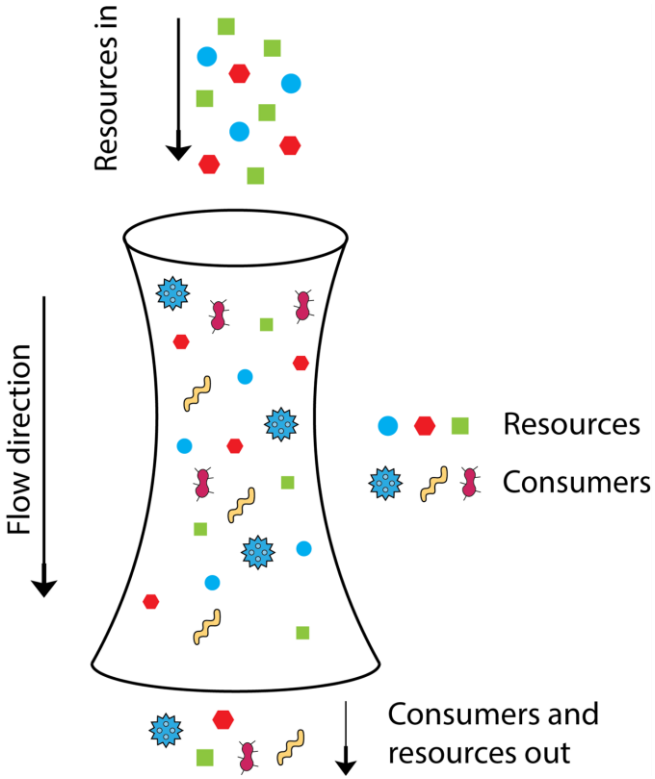
resource dynamics

$$\frac{dr_k}{dt} = I_k(t) - \sum_i \theta_{ki}N_i r_k$$

?

Shahin, Ji, **Dixit**, NPJ Sys Bio
Plata et al., mSystems
Srinivasan et al., in review

We reformulate the C/R model as a latent histories model



growth rate $\frac{1}{N_o}\frac{dN_o}{dt} = \mu_o(t) = \sum_k r_k \theta_{ko}$

resource dynamics $\frac{dr_k}{dt} = I_k(t) - \sum_i \theta_{ki} N_i r_k$?

Abundance fluctuations $N_o(t) \approx \exp\left(\int_0^t \mu_o(\tau)d\tau\right)$

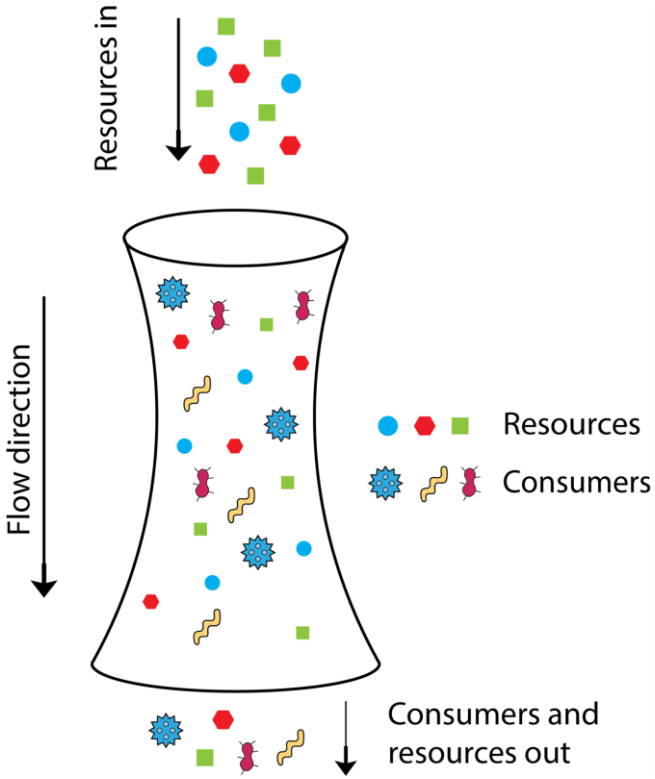
Growth rate from resources $\mu_o(\tau) = \sum_{k=1}^{K} r_k(\tau)\theta_{ko}$

Define latent histories $z_k(t) = -\int_0^t r_k(\tau)d\tau$

Latent histories represent unobserved resource dynamics

Shahin, Ji, **Dixit**, NPJ Sys Bio
Plata et al., mSystems
Srinivasan et al., in review

We reformulate the C/R model as a latent histories model



Abundance fluctuations

$$N_o(t) \approx \exp\left(\int_0^t \mu_o(\tau)d\tau\right)$$

Growth rate from resources

$$\mu_o(\tau) = \sum_{k=1}^{K} r_k(\tau)\theta_{ko}$$

Define latent histories

$$z_k(t) = -\int_0^t r_k(\tau)d\tau$$

Relative abundance x_o

$$x_o = \frac{N_o}{N_{tot}} = \frac{1}{\Omega}\exp\left(-\sum_k z_k\theta_{ko}\right)$$

growth rate

$$\frac{1}{N_o}\frac{dN_o}{dt} = \mu_o(t) = \sum_k r_k\theta_{ko}$$

resource dynamics

$$\frac{dr_k}{dt} = I_k(t) - \sum_i \theta_{ki}N_i r_k$$

?

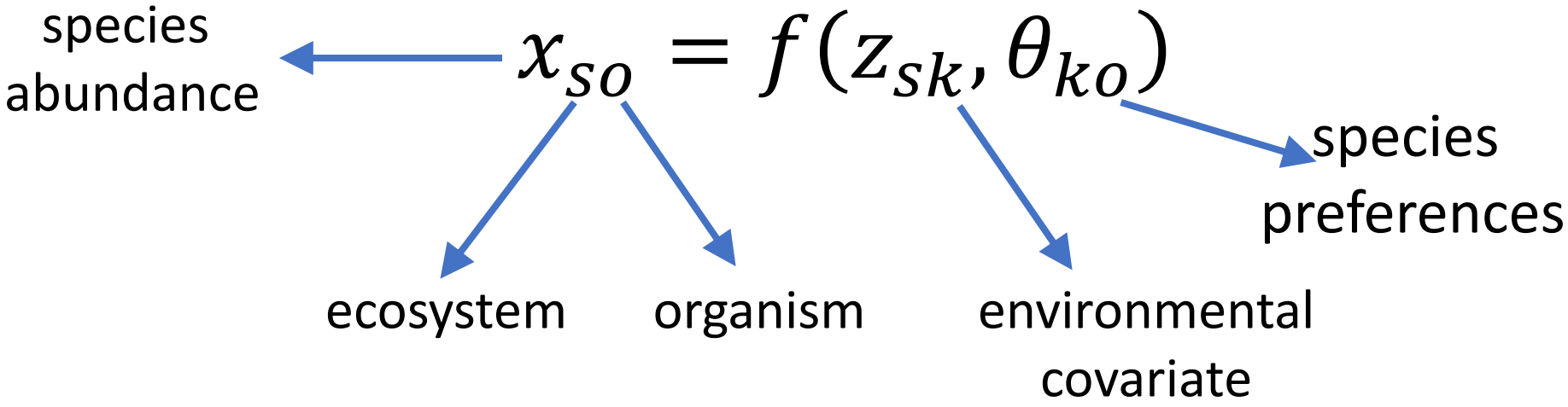
Shahin, Ji, **Dixit**, NPJ Sys Bio
Plata et al., in review, BiorXiv
Srinivasan et al., in preparation

We have converted the CRM into a **J**oint **S**pecies **D**istribution **M**odel (JSDM)

$$x_{so} = f(z_{sk}, \theta_{ko})$$

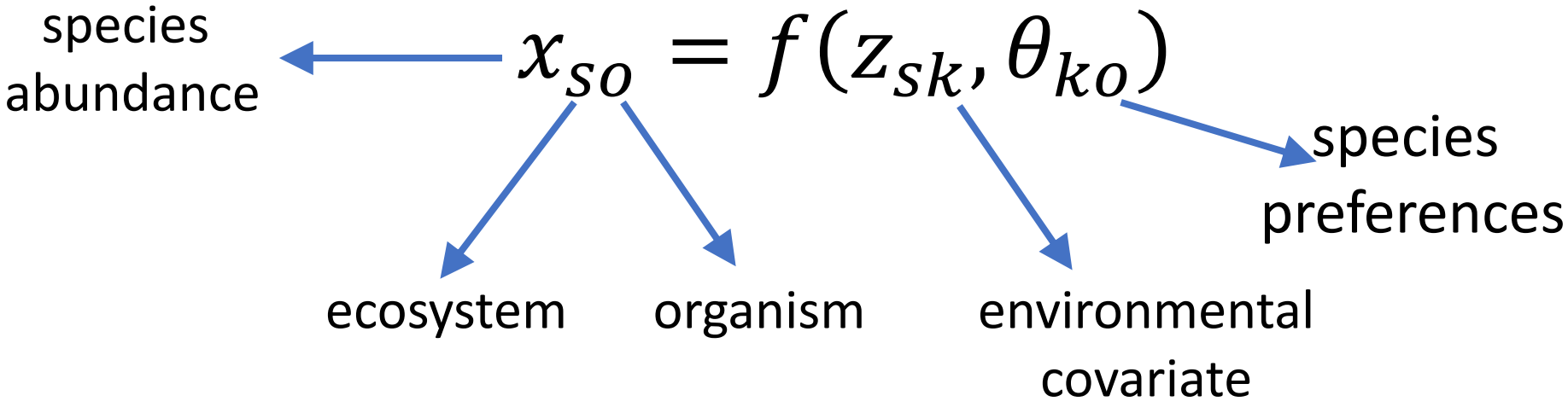
Ovaskainen and Abrego (2020)

JSDMs relate species distribution to environmental factors and estimate niche dimensionality



Ovaskainen and Abrego (2020)

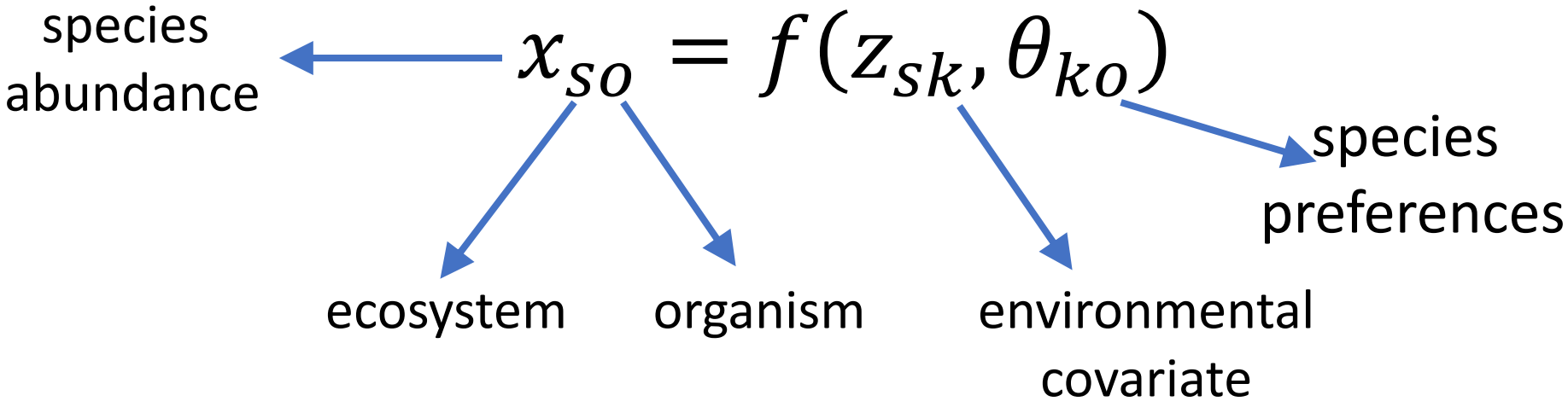
JSDMs relate species distribution to environmental factors and estimate niche dimensionality



- $k \in [1, K]$: *measured* temperature, pH, nutrient concentrations, etc.

Ovaskainen and Abrego (2020)

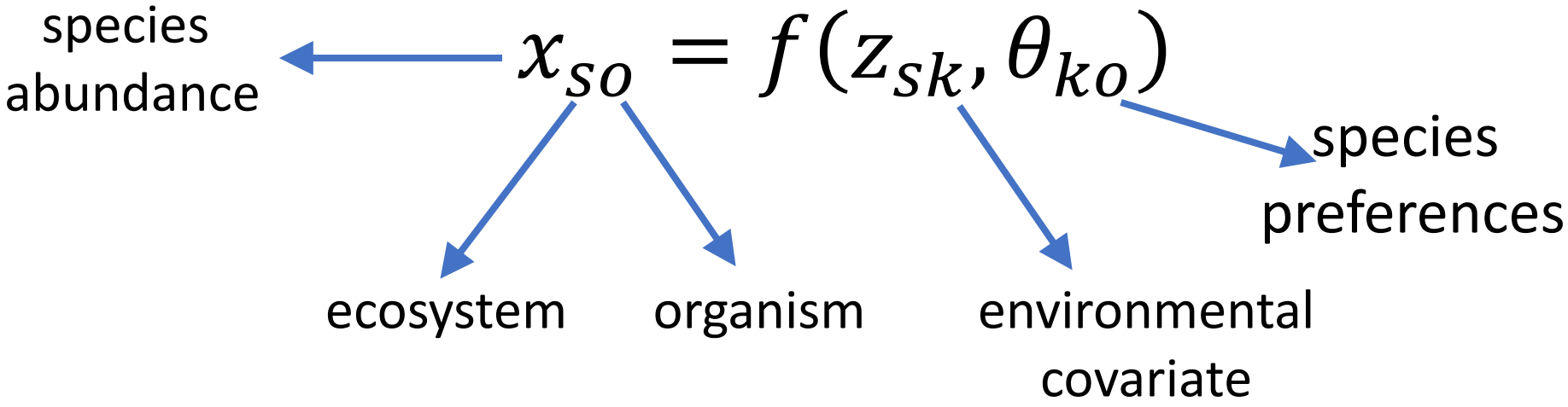
JSDMs relate species distribution to environmental factors and estimate niche dimensionality



- $k \in [1, K]$: *measured* temperature, pH, nutrient concentrations, etc.
  - many factors may be correlated, unmeasured, and unmeasurable
  - others may not be known

Ovaskainen and Abrego (2020)

JSDMs relate species distribution to environmental factors and estimate niche dimensionality

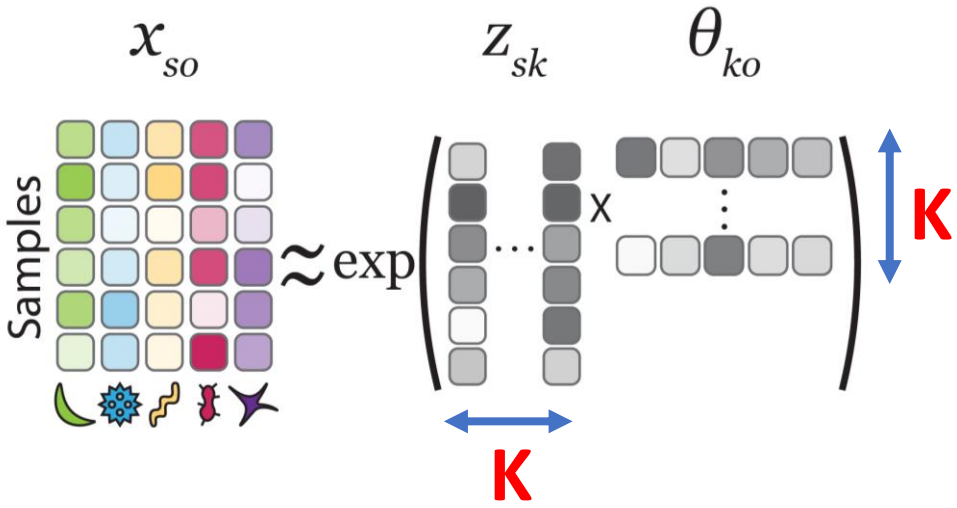


- $k \in [1, K]$: *measured* temperature, pH, nutrient concentrations, etc.
  - many factors may be correlated, unmeasured, and unmeasurable
  - others may not be known

- Latent variable JSDMs capture variation arising from known/unknown covariates

Ovaskainen and Abrego (2020)

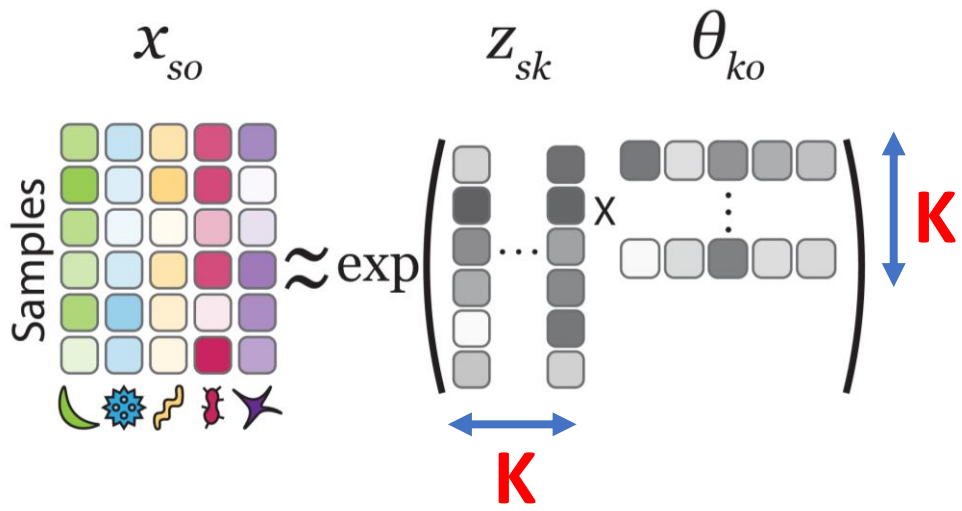
JSDMs relate species distribution to environmental factors and estimate niche dimensionality

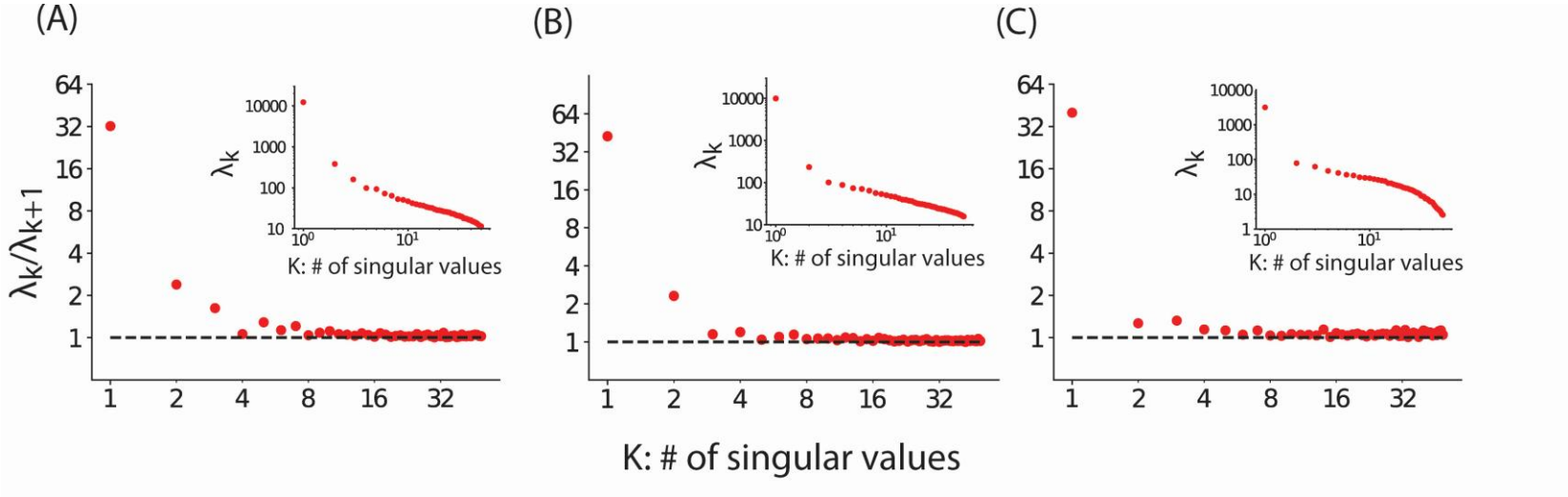


- How many Ks do we need to accurately model the data?
- Singular value decomposition suggests $K = o(1)$ may be enough!

Fisher et al. PLoS Comp Bio (2017), Shahin et al., npj Sys Bio (2023), Plata et al. mSystems (2025)

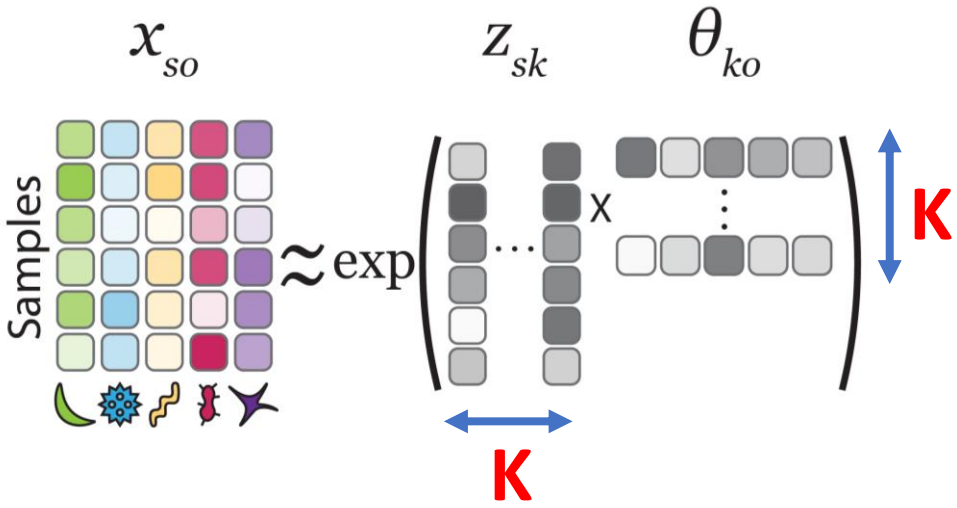
JSDMs relate species distribution to environmental factors and estimate niche dimensionality

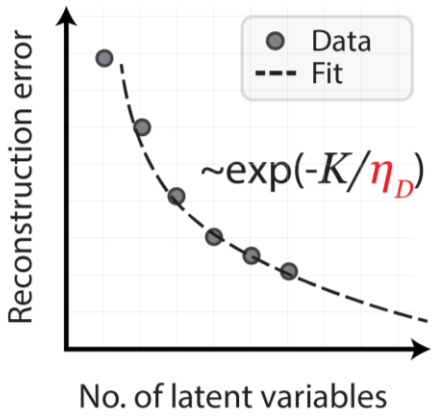




Fisher et al. PLoS Comp Bio (2017), Shahin et al., npj Sys Bio (2023), Plata et al. mSystems (2025)

JSDMs relate species distribution to environmental factors and estimate niche dimensionality

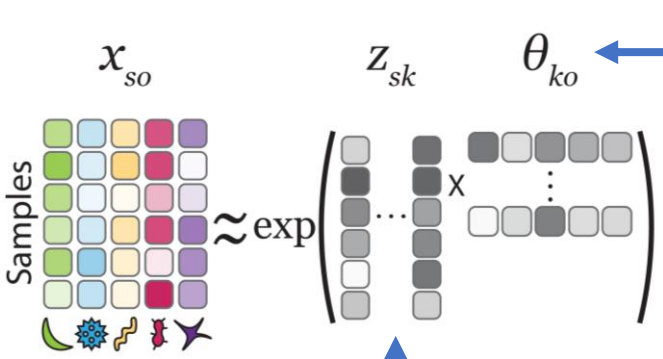




- niche dimensionality ($\boldsymbol{\eta_D}$): effective environmental dimensionality that explains composition

Fisher et al. PLoS Comp Bio (2017), Shahin et al., npj Sys Bio (2023), Plata et al. mSystems (2025)

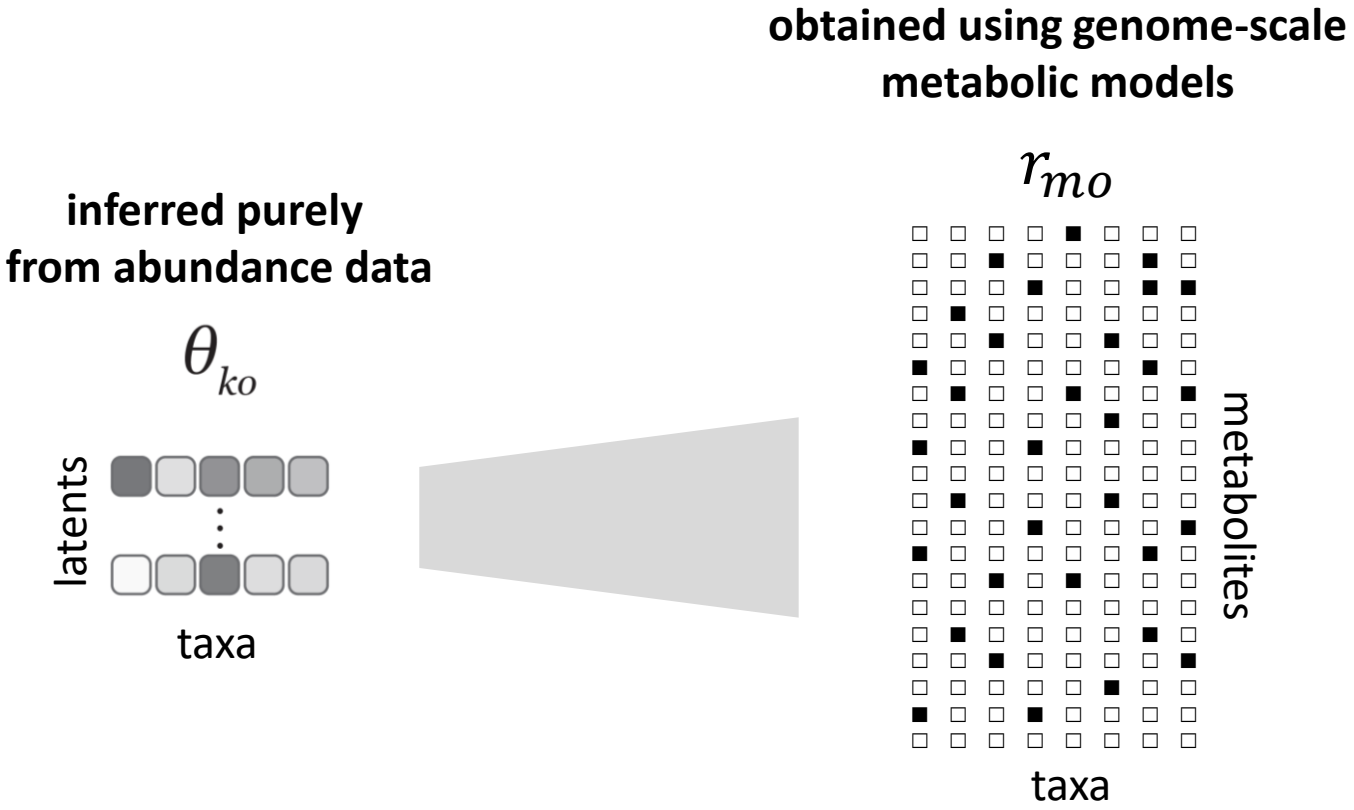
Inferred latent space has ecologically interpretable features



encodes species-specific information

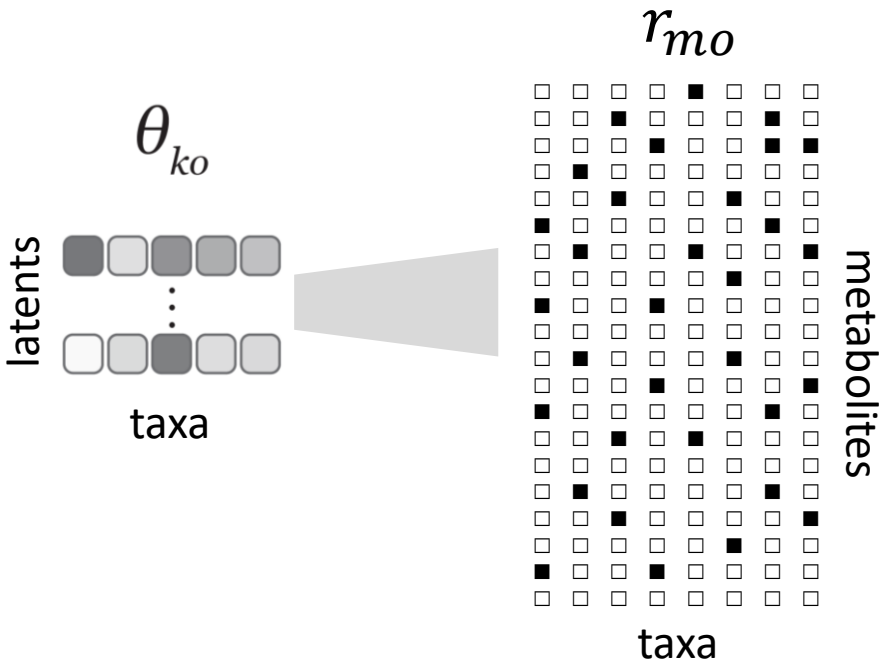
encodes environmental information

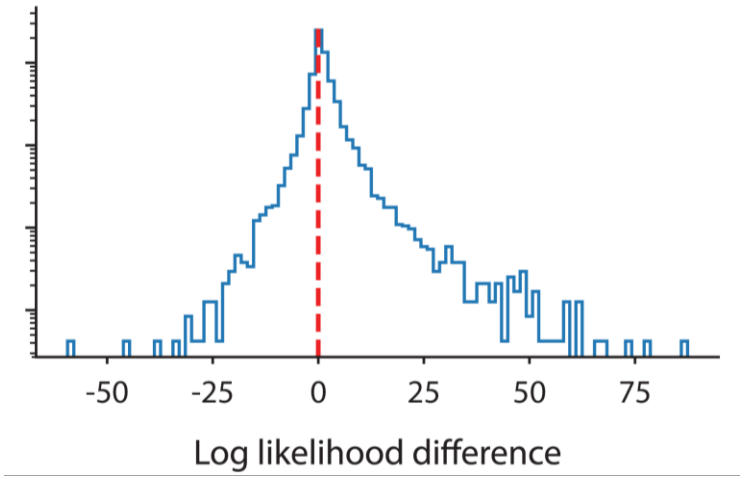
Species specific features predict metabolic traits of taxa



$$p(r_{mo} = 1) \propto \exp(-\sum_{k} \theta_{ko}\Lambda_{km})$$

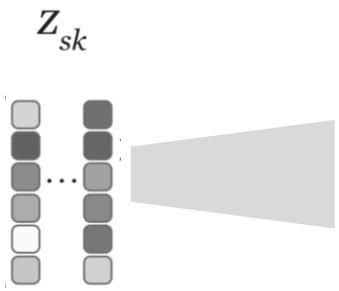
Species specific features predict metabolic traits of taxa



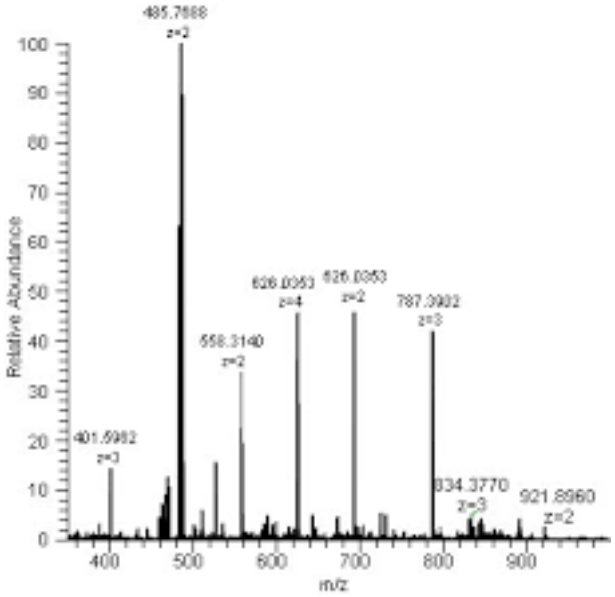


- Latent taxon features inferred from abundances fit species traits better than expected by random chance

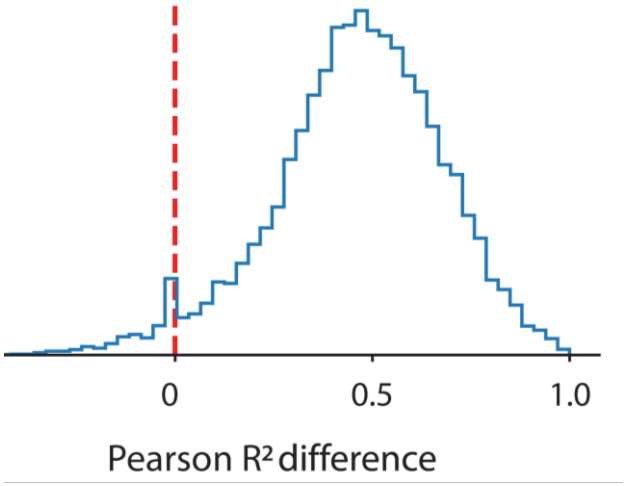
Environment-specific latent embeddings capture metabolic environment



environmental embeddings



fecal metabolite abundances



- Latent embeddings inferred from abundances fit environmental data better than expected by random chance

The inferred latent embeddings correspond to real ecological variables

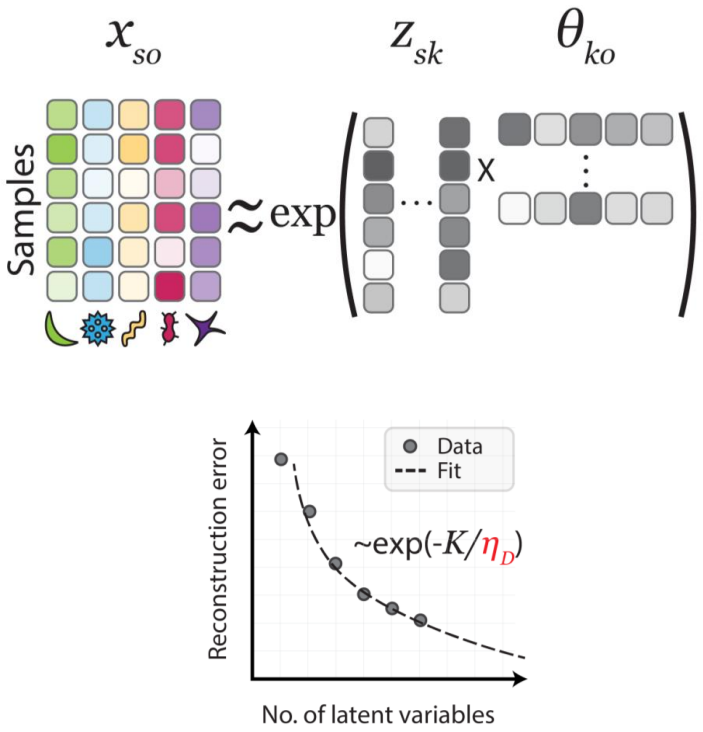
We use a large compendium of human microbiome samples

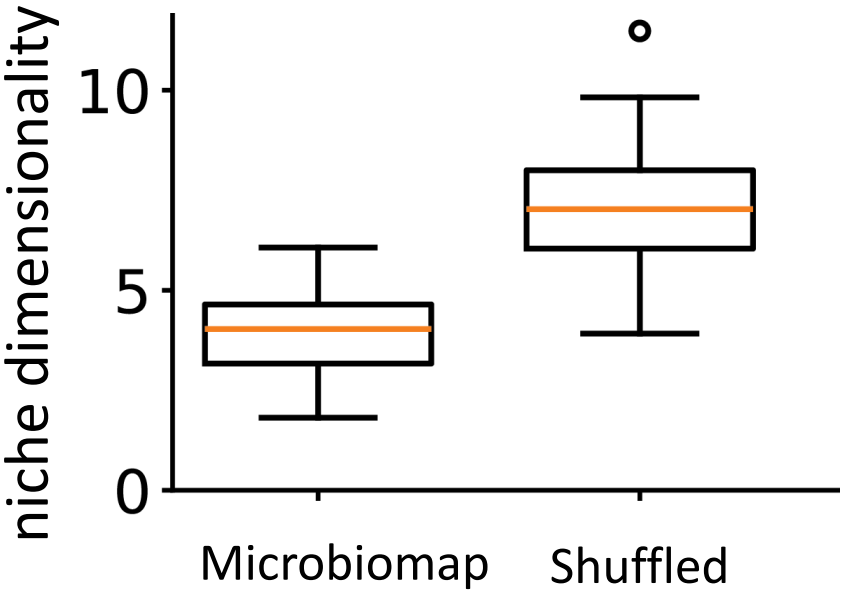
- Microbiomap contains $\sim 1.7 \times 10^5$ samples scattered across ~ 400 independent studies
  - Abdill et al. Cell 2025

- Taxon are represented at the Amplicon Sequence Variant (ASV) level
  - $\sim$genus

- Taxons are uniformly processed, each data set contains the same labels

Estimated η_D is very low and quite variable





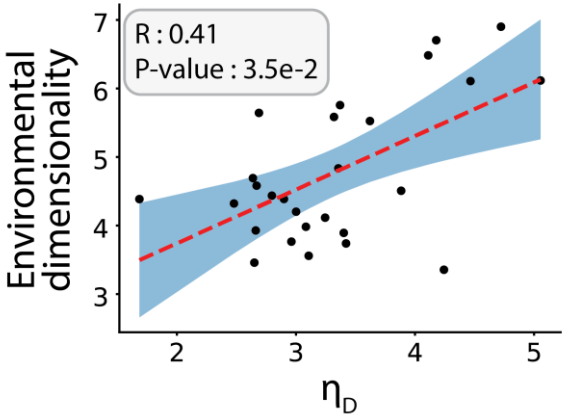
- Estimated niche dimensionality varies across microbiome studies
  - Median η_D: 4 , coefficient of variation: 30%
- Significantly smaller compared to shuffled datasets
  - Median η_D: 7, p-value: 10^{-49}

Abdill et al. Cell (2025)

η_D correlates with environmental complexity, competition, and diversity

Muller et al., npj Biofilms and Microbiomes, 2022, Abdill et al. Cell, 2025

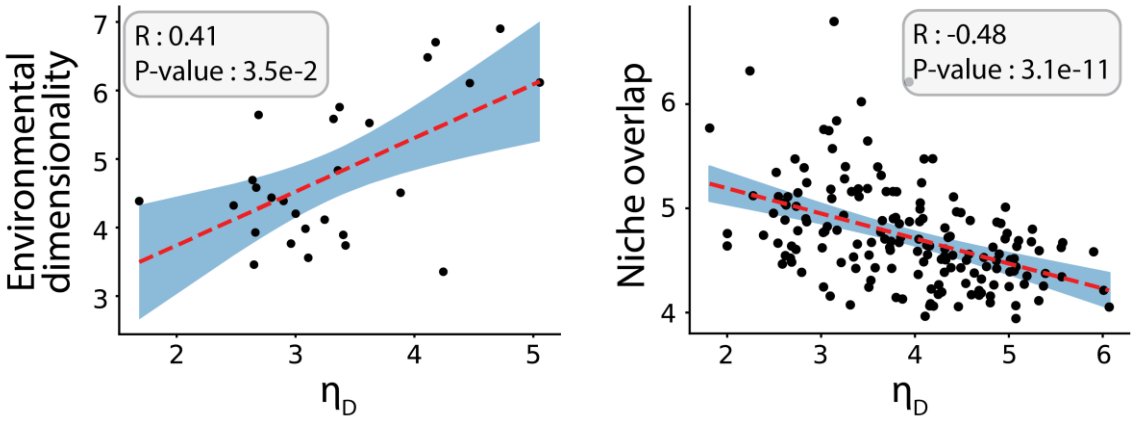
η_D correlates with environmental complexity, competition, and diversity



- Inferred η_D correlates with the dimensionality of the metabolic environment
  - validates the abundance-only approach to estimate niche dimensionality

Muller et al., npj Biofilms and Microbiomes, 2022, Abdill et al. Cell, 2025

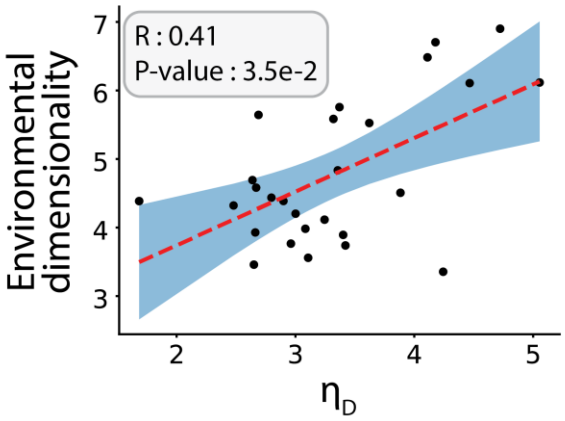
η_D correlates with environmental complexity, competition, and diversity

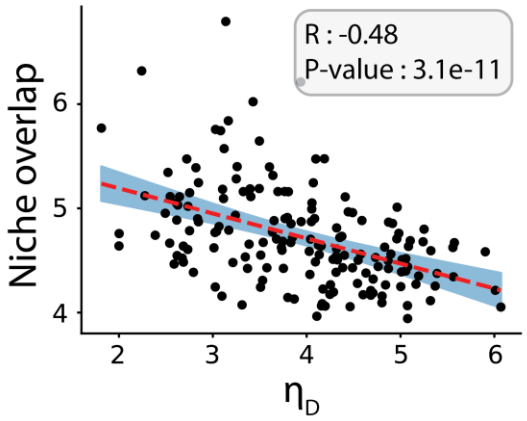


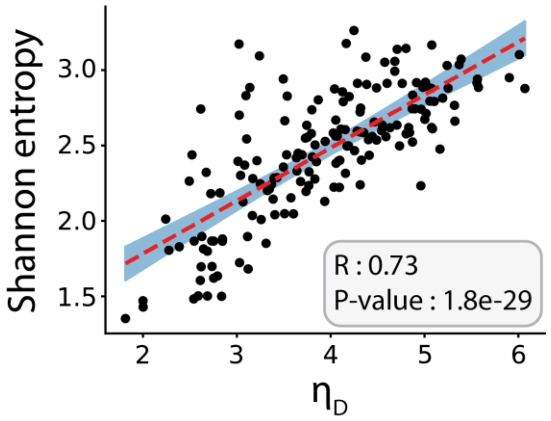
- Inferred η_D correlates with the dimensionality of the metabolic environment
  - validates the abundance-only approach to estimate niche dimensionality
- Niche dimensionality correlates with niche overlap
  - niche overlap computed using flux balance analysis (models from Ines Thiel lab)

Muller et al., npj Biofilms and Microbiomes, 2022, Abdill et al. Cell, 2025

η_D correlates with environmental complexity, competition, and diversity



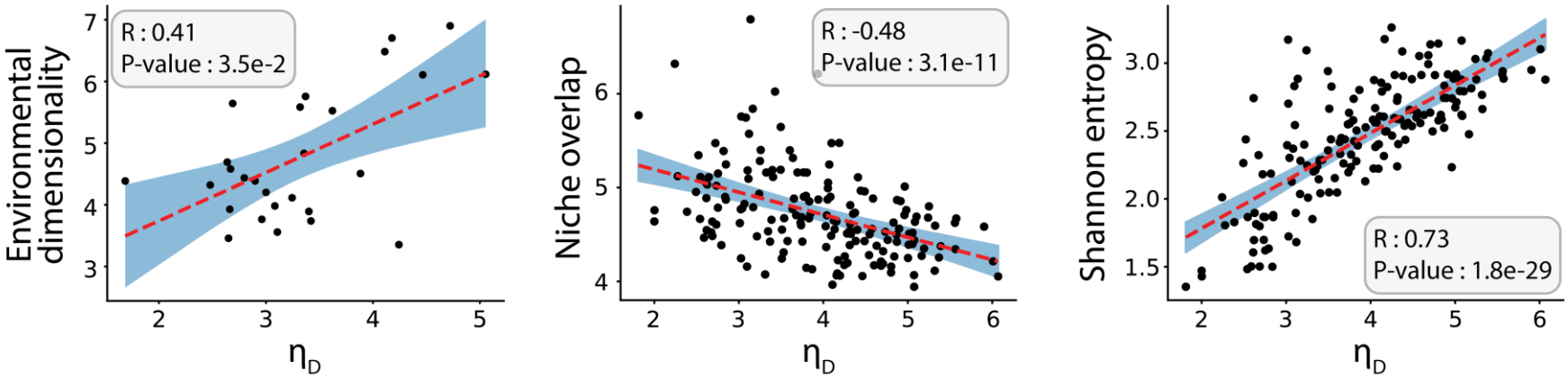




- Inferred η_D correlates with the dimensionality of the metabolic environment
  - validates the abundance-only approach to estimate niche dimensionality
- Niche dimensionality correlates with niche overlap and biodiversity
  - niche overlap computed using flux balance analysis (models from Ines Thiel lab)

Muller et al., npj Biofilms and Microbiomes, 2022, Abdill et al. Cell, 2025

η_D correlates with environmental complexity, competition, and diversity

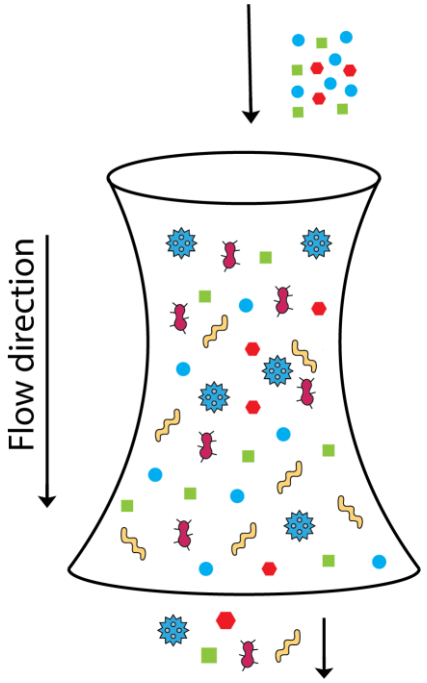


- Inferred η_D correlates with the dimensionality of the metabolic environment
  - validates the abundance-only approach to estimate niche dimensionality
- Niche dimensionality correlates with niche overlap and biodiversity
  - niche overlap computed using flux balance analysis (models from Ines Thiel lab)
  - consistent with the macroecological *niche dimensionality hypothesis*
    - Harpole et al. 2007, Tilman, 1982, MacArthur, 1984

Muller et al., npj Biofilms and Microbiomes, 2022, Abdill et al. Cell, 2025

Diversity metrics are not enough to explain niche dimensionality

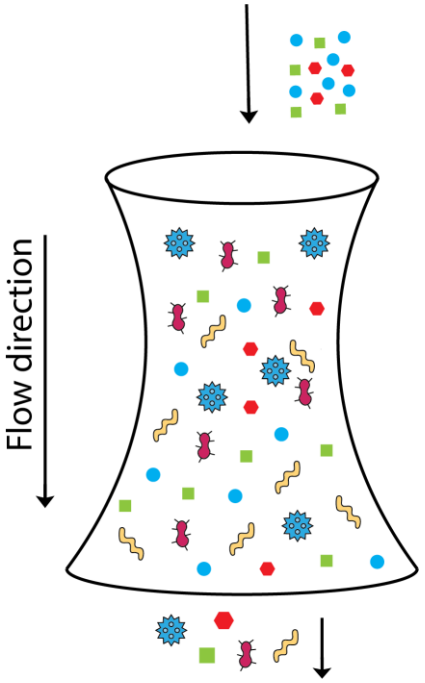
- We use a random consumer/resource model to investigate variation in niche dimensionality



Diversity metrics are not enough to explain niche dimensionality

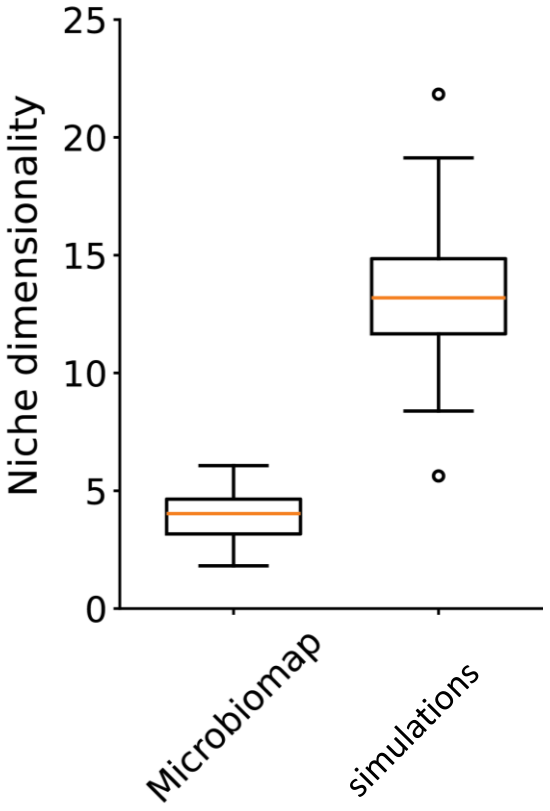
- We use a random consumer/resource model to investigate variation in niche dimensionality

- Previous studies have often used reproduction of diversity metrics to assess realism of simulated communities



Diversity metrics are not enough to explain niche dimensionality

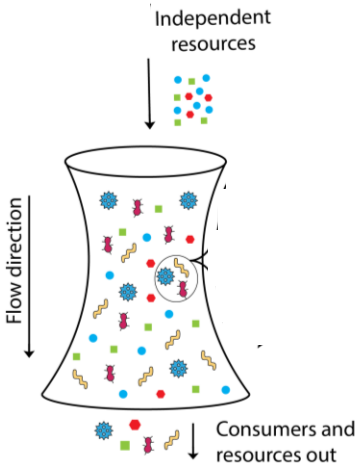
- We use a random consumer/resource model to investigate variation in niche dimensionality
- Previous studies have often used reproduction of diversity metrics to assess realism of simulated communities
- *In silico* communities that reproduce α and β diversity are very high dimensional
  - median η_{D}: 13 compared to median η_{D}: 4 in data



Abdill et al. Cell, 2025

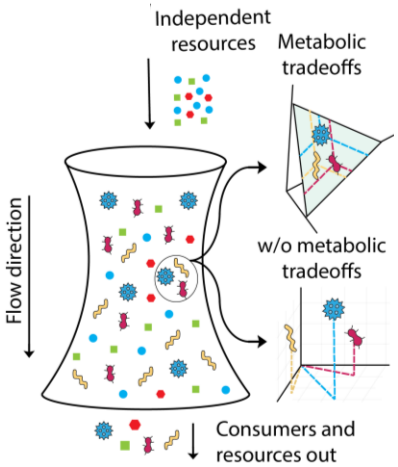
Tradeoffs explains low η_D of microbiomes

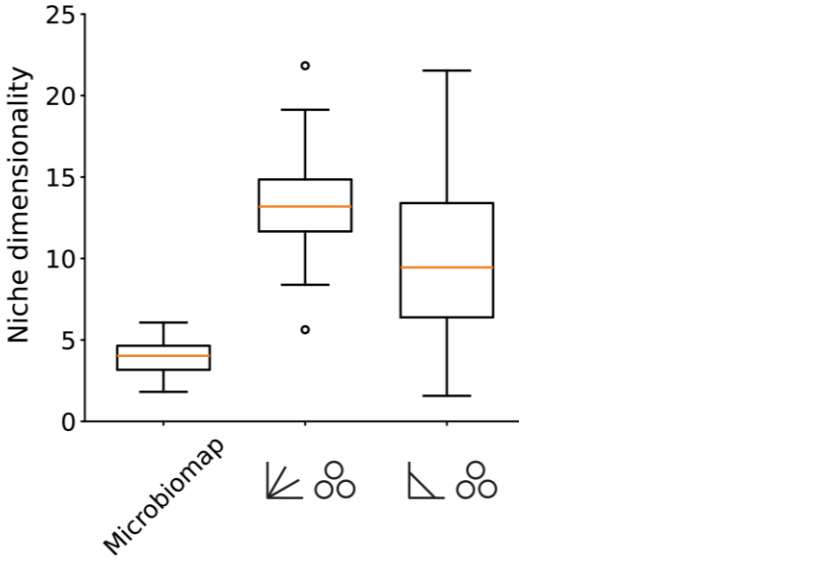
- Species-intrinsic metabolic tradeoffs allow species to exist in low dimensional niches
  - Harpole et al., 2007, Posfai et al. 2017



Tradeoffs explains low η_D of microbiomes

- Species-intrinsic metabolic tradeoffs allow species to exist in low dimensional niches
  - Harpole et al., 2007, Posfai et al. 2017
- Imposition of metabolic tradeoffs does reduce niche dimensionality, but still higher than data

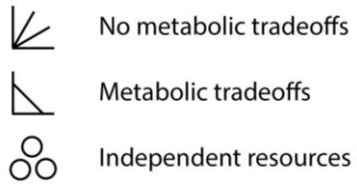


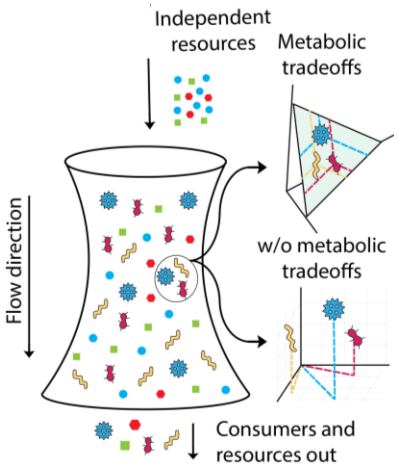


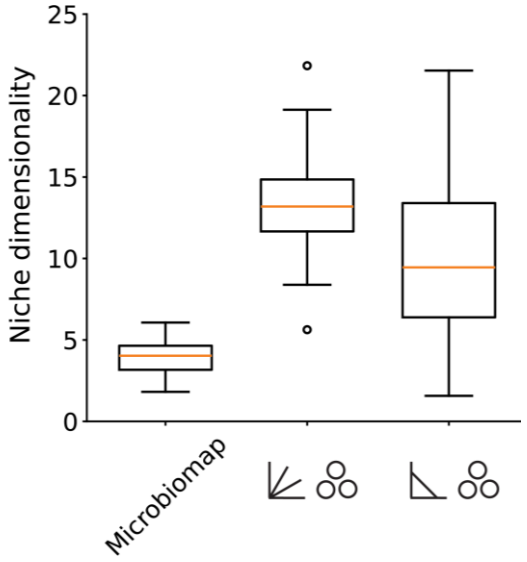
Tradeoffs explains low η_D of microbiomes

- Species-intrinsic metabolic tradeoffs allow species to exist in low dimensional niches
  - Harpole et al., 2007, Posfai et al. 2017

- An overlooked tradeoff is species-extrinsic: hosts consume nutrients in food groups with roughly constant total intake
  - An example tradeoff: high starch <=> low amino acids



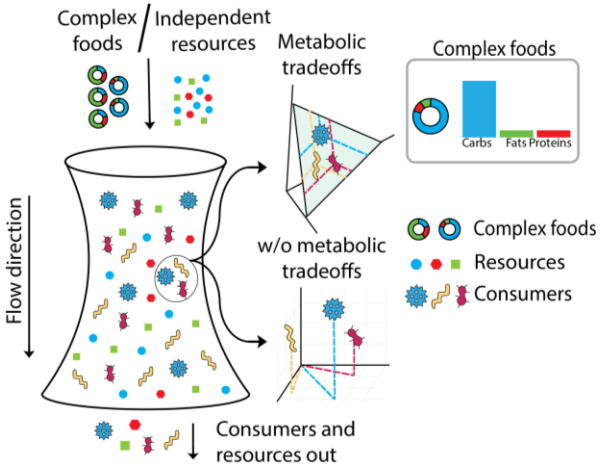


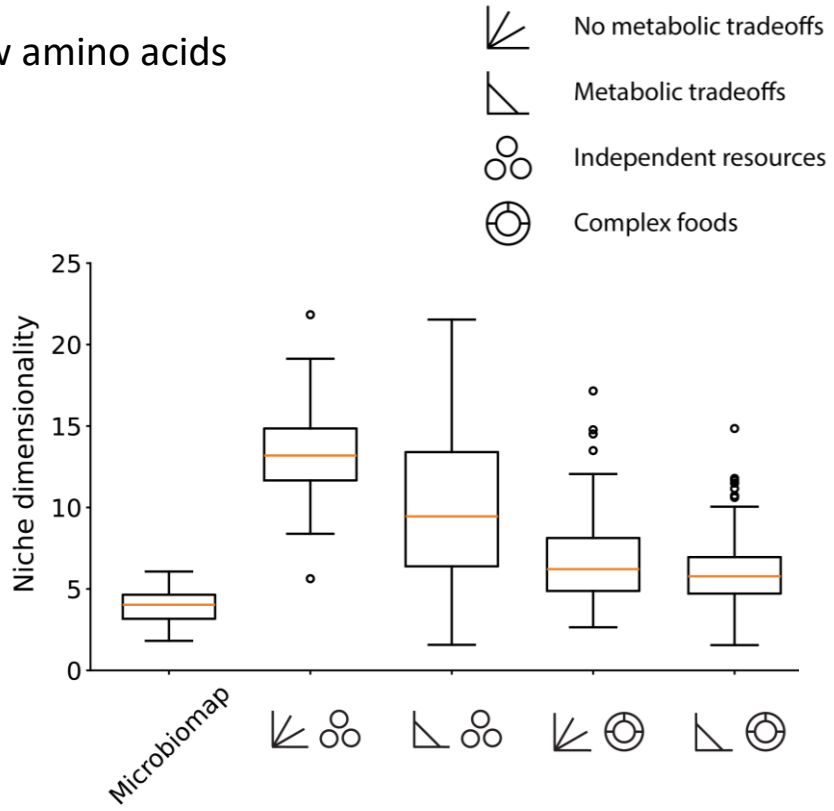


Tradeoffs explains low η_D of microbiomes

- Species-intrinsic metabolic tradeoffs allow species to exist in low dimensional niches
  - Harpole et al., 2007, Posfai et al. 2017

- An overlooked tradeoff is species-extrinsic: hosts consume nutrients in food groups with roughly constant total intake
  - An example tradeoff: high starch <=> low amino acids



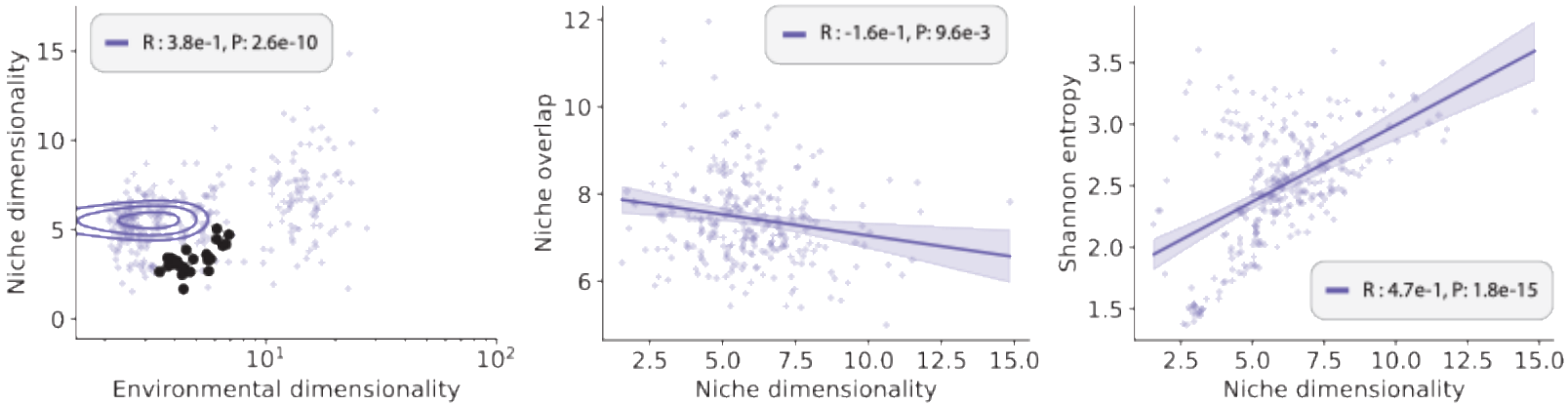


Tradeoffs explains how η_D relates with community structure

- Tradeoffs reproduce relationship between community structure and niche dimensionality

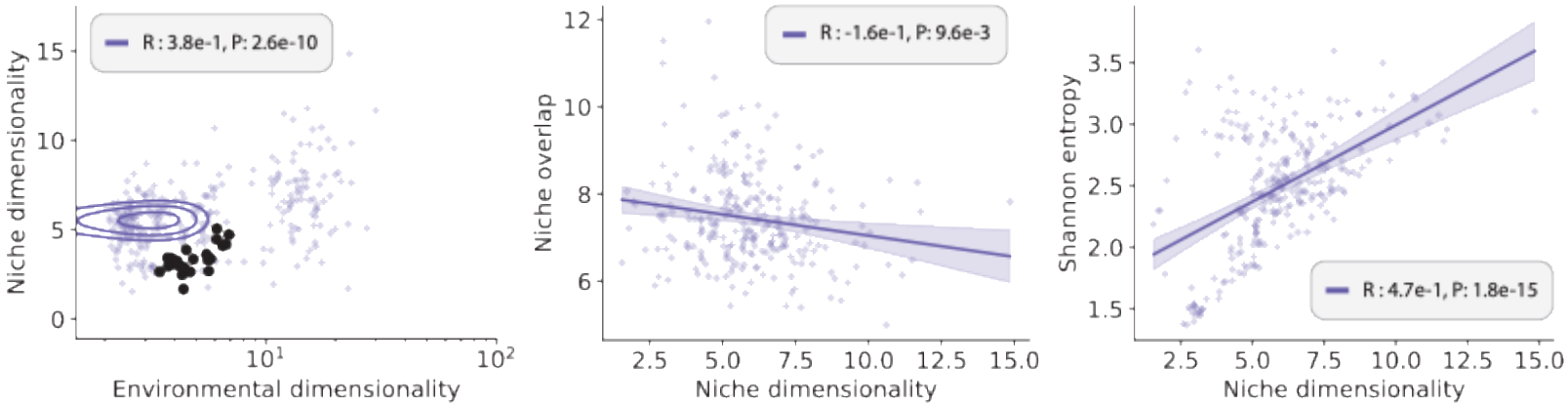
Tradeoffs explains how η_D relates with community structure

- Tradeoffs reproduce relationship between community structure and niche dimensionality



Tradeoffs explains how η_D relates with community structure

- Tradeoffs reproduce relationship between community structure and niche dimensionality



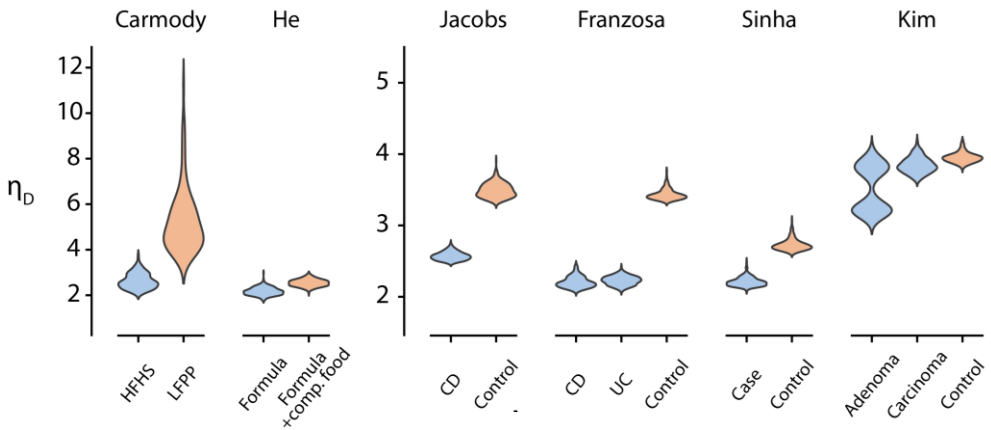
- Both species-intrinsic metabolic tradeoffs and extrinsic environmental tradeoffs are needed to capture variation in niche dimensionality and community structure

η_D correlates with diet and markers of stress

- Complex diets increase niche dimensionality
- Colorectal cancer and IBD lower dimensionality

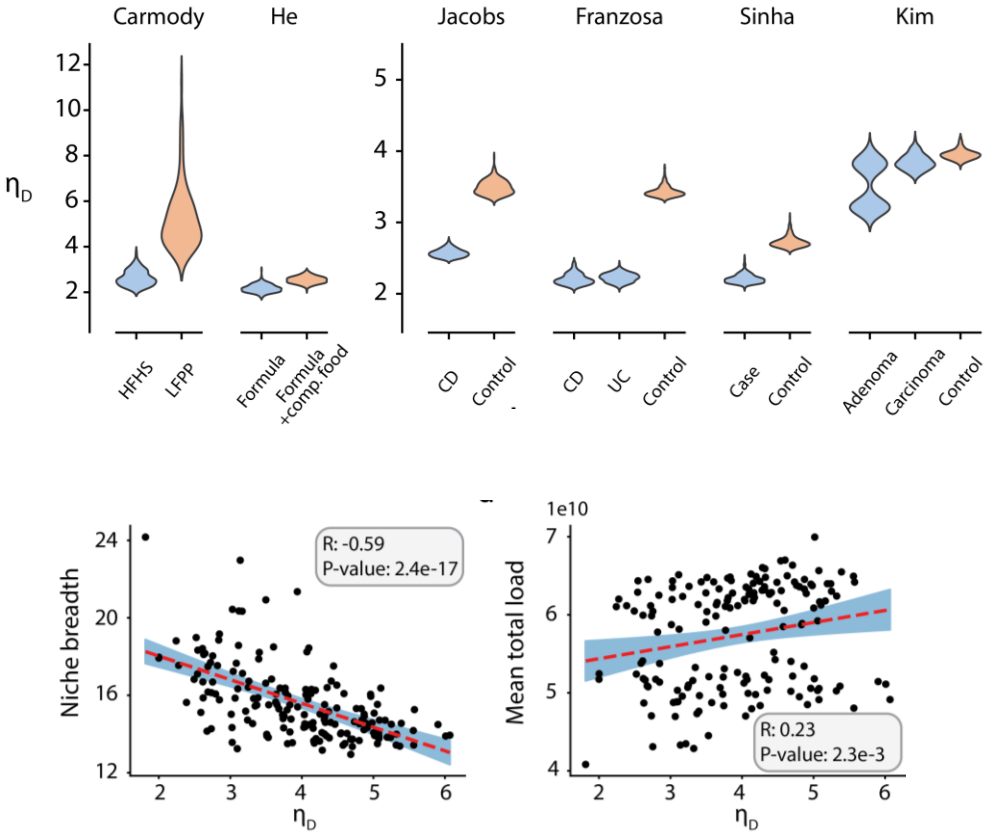
η_D correlates with diet and markers of stress

- Complex diets increase niche dimensionality
- Colorectal cancer and IBD lower dimensionality



η_D correlates with diet and markers of stress

- Complex diets increase niche dimensionality
- Colorectal cancer and IBD lower dimensionality
- Dimensionality correlates with broad markers of stress
  - Prevalence of *generalists* (Veseli et al., eLife, 2025)
  - Total microbial load (Vandeputte et al., Nature, 2017)



Summary

- **Niche dimensionality is a driver for microbiome community structure**
  - Correlates with environment complexity, competition, diversity, generalists, …
  - Low dimensionality arises from tradeoffs
    - Extrinsic tradeoffs are important

Summary

- **Niche dimensionality is a driver for microbiome community structure**
  - Correlates with environment complexity, competition, diversity, generalists, …
  - Low dimensionality arises from tradeoffs
    - Extrinsic tradeoffs are important

- We need to rethink the “spin glass” picture with iid interactions
  - **Dimensionality $\boldsymbol{\eta_D = o(1)}$ NOT $\boldsymbol{\eta_D \to \infty}$**
    - Mizrahi et al. ISME J, 2025 finds hierarchically structured interactions

Summary

- **Niche dimensionality is a driver for microbiome community structure**
  - Correlates with environment complexity, competition, diversity, generalists, …
  - Low dimensionality arises from tradeoffs
    - Extrinsic tradeoffs are important

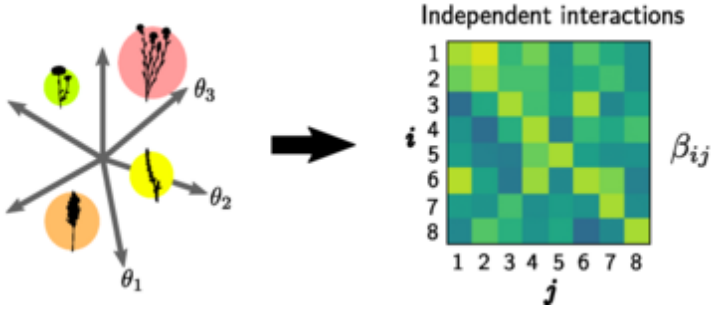
- We need to rethink the "spin glass" picture with iid interactions
  - **Dimensionality $\boldsymbol{\eta_D = o(1)}$ NOT $\boldsymbol{\eta_D \to \infty}$**
    - Mizrahi et al. ISME J, 2025 finds hierarchically structured interactions

- Community models must include low dimensional niches
  - Shahin et al. npj Sys Bio, 2023, Plata et al. mSystems, 2025

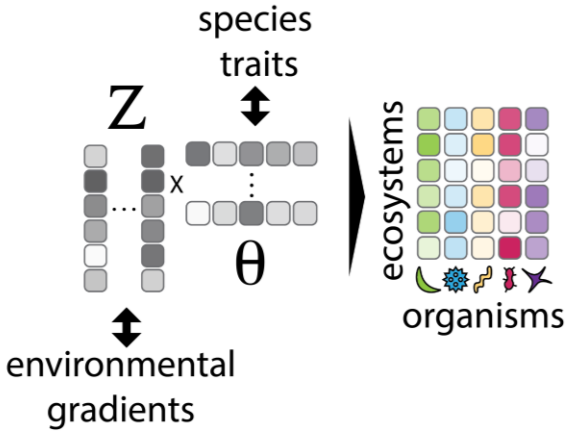
Summary

- **Niche dimensionality is a driver for microbiome community structure**
  - Correlates with environment complexity, competition, diversity, generalists, …
  - Low dimensionality arises from tradeoffs
    - Extrinsic tradeoffs are important

high dimensional
independent interactions



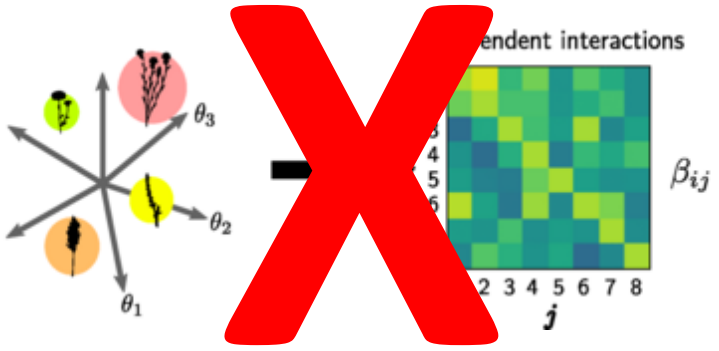
low dimensional
latent space



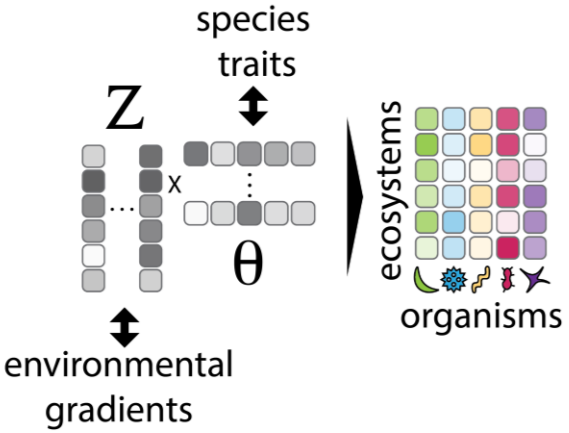
Summary

- **Niche dimensionality is a driver for microbiome community structure**
  - Correlates with environment complexity, competition, diversity, generalists, …
  - Low dimensionality arises from tradeoffs
    - Extrinsic tradeoffs are important

high dimensional
independent interactions



low dimensional
latent space



Thank you!

- Srinivasan, Plata, and Dixit, *Niche dimensionality drives microbial community structure, in review*

- Shahin et al., *EMBED: Essential MicroBiomE Dynamics, a dimensionality reduction approach for longitudinal microbiome studies*, npj Sys. Bio. 2023

- Plata et al., *Designing host-associated microbiomes using the consumer/resource model*, mSystems 2025

BEZOS EARTH FUND

Rumen Digital Twin