



HAL
open science

Challenges in microbial interaction inference from time series

Felix Roy

► **To cite this version:**

Felix Roy. Challenges in microbial interaction inference from time series. Theomodive 2025 - Réunion annuelle du Groupe de Travail - Théorie et modélisation de la biodiversité, Dec 2025, Bordeaux, France. ⟨hal-05559195⟩

HAL Id: hal-05559195

<https://inria.hal.science/hal-05559195v1>

Submitted on 19 Mar 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

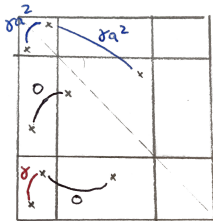
L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Challenges in microbial interaction inference from time series

Felix Roy (postdoc with Simon Labarthe)



INRAE

Inria

Takeaways

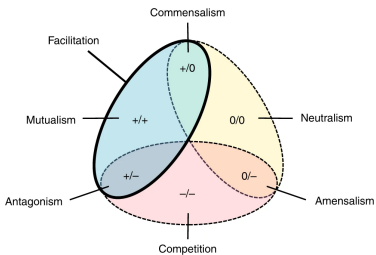
- ▶ **Standard procedure:** When inferring ecological interactions from time series, the procedure is to use the Lotka-Volterra model, and a specific algorithm "gradient matching" (GM).
- ▶ **Issue:** Even though it is highly efficient numerically, precise understanding of the inference is lacking. Furthermore, the precision of series reconstruction is not always correlated to the precision of interaction inference, which raises the issue of practical identifiability.
- ▶ **Contribution:** By using an intermediate model, we are able to study better the parameter estimation. We find that GM presents some limitations, and study ways to circumvent them.
- ▶ **Application:** We present two different experimental setups and data, that are representative of microbial population series.

- 1 Context
- 2 Inference algorithms
- 3 Inference results
- 4 Experimental datasets
- 5 Ongoing work

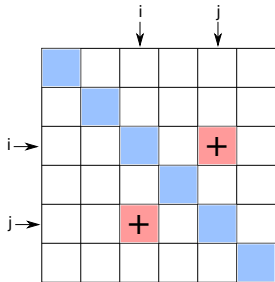
Theoretical setup

- ▶ **Data:** Given S species, T timesteps, and R replicates, the population time series are $X_{it}^r = x_i^r(t)$. Dimensions: (S, T, R) .
- ▶ **Goal:** Find the interaction matrix M_{ij} that would generate the time series through a Lotka-Volterra model. Dimensions: (S, S) .
- ▶ **Application:** Determine interactions between microbiota from synthetic communities experiments, studying grapevine mildew antagonists, or bean seeds microbiota.

Interaction matrix encodes ecological interaction



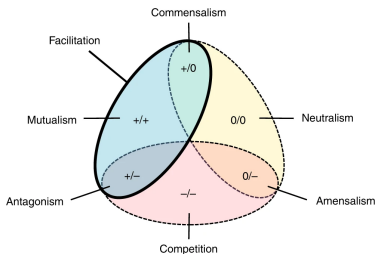
Ecology and evolution of facilitation among symbionts [Nature comm. - 2018]



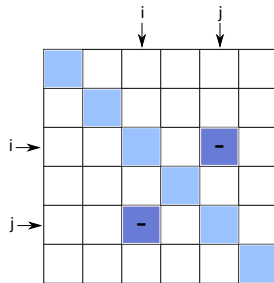
Mutualism

Ex: Plant-pollinators

Interaction matrix encodes ecological interaction



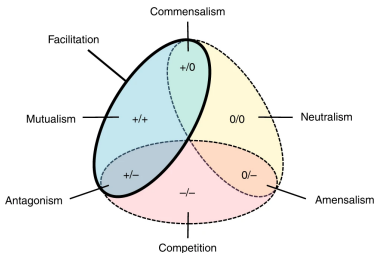
Ecology and evolution of facilitation among symbionts [Nature comm. - 2018]



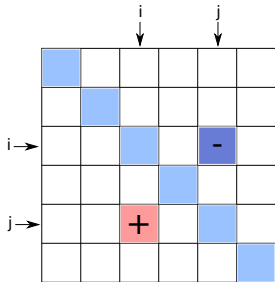
Competition

Ex: Plants compete for sunlight and nutrients

Interaction matrix encodes ecological interaction



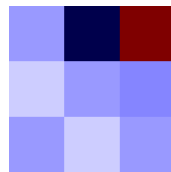
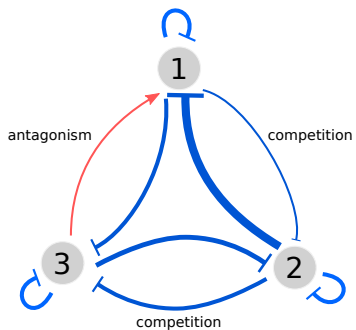
Ecology and evolution of facilitation among symbionts [Nature comm. - 2018]



Antagonism

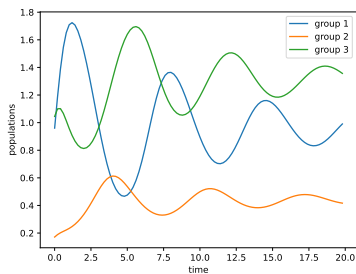
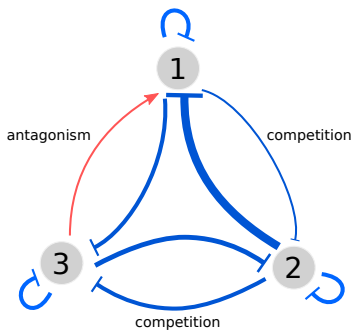
Ex: Foxes-rabbits

Toy example: 3 species community



Interaction matrix

Toy example: 3 species community



Time evolution

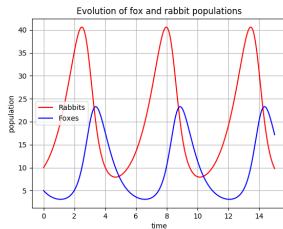
Lotka-Volterra model

$$\forall i \in \mathbb{N}_S, \quad \frac{dx_i}{dt} = x_i (\mu_i + (M \cdot \vec{x})_i)$$

- ▶ $x_i(t)$: absolute abundance of species i
- ▶ μ_i : intrinsic growth rate of species i
- ▶ M : interaction matrix

Goal:

$$X \rightarrow (\vec{\mu}, M)$$
$$(S, T, R) \rightarrow (S, S + 1)$$



- 1 Context
- 2 Inference algorithms**
- 3 Inference results
- 4 Experimental datasets
- 5 Ongoing work

Gradient matching (GM) and series reconstruction (REC)

Gradient matching

- ▶ Estimate the time derivative from the data

$$G_{it}^{\log} \sim \frac{d \ln x_i(t)}{dt}$$

- ▶ Linear regression

$$G = \mathcal{L}_X(M)$$

Series reconstruction

- ▶ Start from an initial guess for interactions
- ▶ Run the dynamics
- ▶ Compute the error to the real data
- ▶ Minimize the error (least square)

Both methods can be implemented with regularization (L1 or L2).

Bibliography

▶ Introduction of the method

Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota
[PLoS Computational Biology - 2013 - Stein, Bucci, Toussaint *et al.*]

▶ Pitfalls numerical study

Inferring Human Microbial Dynamics from Temporal Metagenomics Data: Pitfalls and Lessons
[BioEssays - 2017 - Cao, Gibson, Bashan and Liu]

▶ Theoretical identifiability, for a few species ($S = 2$)

Structural identifiability of the generalized Lotka-Volterra model for microbiome studies
[Royal Society Open Science - 2021 - Remien, Eckwright and Ridenhour]

▶ Bayesian setup

Learning Ecosystem-Scale Dynamics from Microbiome Data with MDSINE2
[Nature Microbiology - 2025 - Gibson, Kim, Acharya *et al.*]

Replicator model (Lotka-Volterra rescaled)

Series reconstruction is hard to test with Lotka-Volterra, because the model has the tendency to explode 🤖

There is an exact mapping to rescaled variables: the **Replicator model**.

$$\forall i \in \mathbb{N}_{S+1}, \quad \frac{dn_i}{dt} = n_i ((A \cdot \vec{n})_i - \phi(t))$$

- ▶ $n_i(t)$: relative abundance of species i
- ▶ A : interaction matrix
- ▶ $\phi(t)$: Lagrange multiplier

The variables are **relative** abundances: $\forall t, \quad \sum_{i=1}^{S+1} n_i(t) = 1$

Pros and cons

Algorithm	Gradient matching	Series reconstruction
Speed	●	●
Initialization	●	●
Understanding	●	●

- 1 Context
- 2 Inference algorithms
- 3 Inference results**
- 4 Experimental datasets
- 5 Ongoing work

Gradient matching (GM) issues

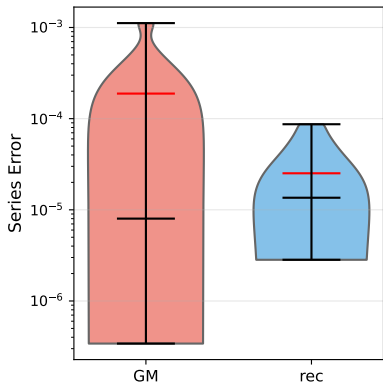
- ▶ **Precision:** Exact interaction reconstruction is only achieved with perfect Euler scheme
- ▶ **Sampling:** Increasing time sampling only helps up to a plateau. Afterwards, only providing different initial conditions helps improving the inference.
- ▶ **Regularization:** Without regularization, it fails completely.
- ▶ **Error/loss correlation:** Not obvious.

Benchmark comparison

Protocol

- ▶ $S = 10, T = 100, R = 1$
- ▶ Sample $N = 10$ interaction matrices from a normal distribution
- ▶ Simulate dynamics with the Replicator model
- ▶ Run GM and REC ($n_{\text{init}} = 10$), varying regularization
- ▶ Compare best results for each sample and algorithm

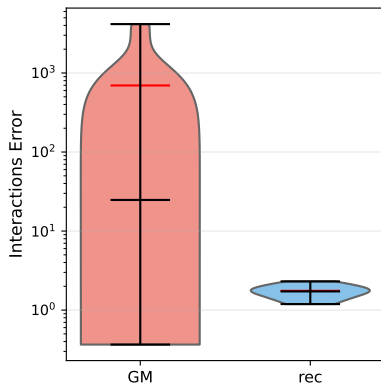
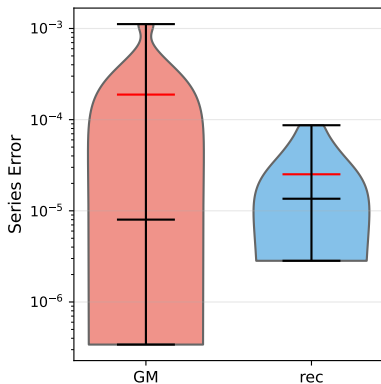
Benchmark comparison



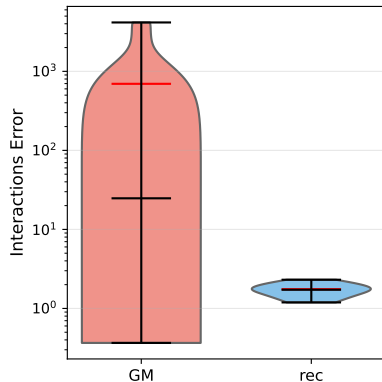
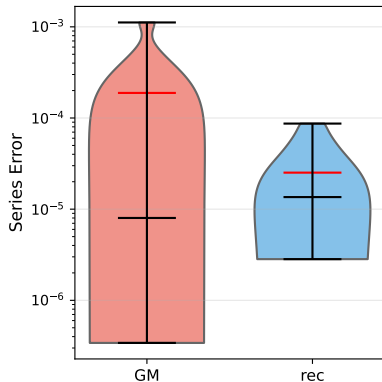
Protocol

- ▶ $S = 10, T = 100, R = 1$
- ▶ Sample $N = 10$ interaction matrices from a normal distribution
- ▶ Simulate dynamics with the Replicator model
- ▶ Run GM and REC ($n_{\text{init}} = 10$), varying regularization
- ▶ Compare best results for each sample and algorithm

Benchmark comparison



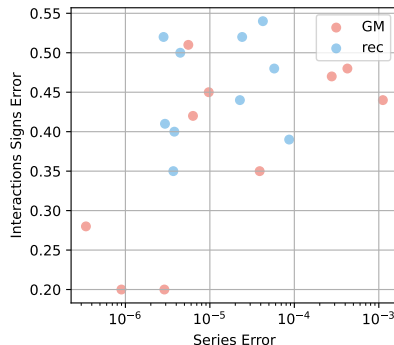
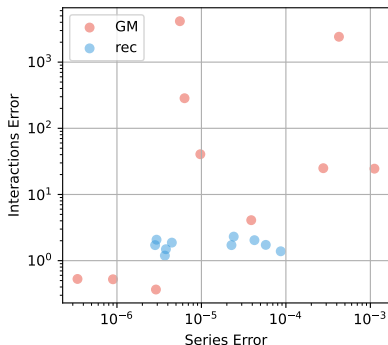
Benchmark comparison



- ▶ GM performs more randomly than REC, and is typically worse.

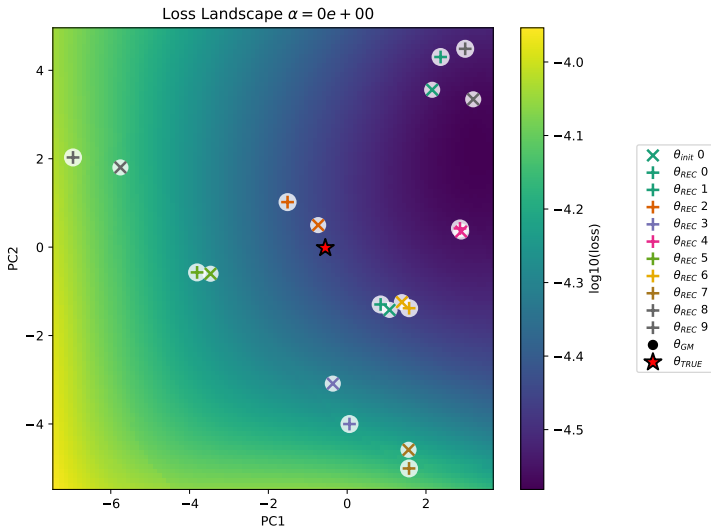
Benchmark error/loss

Error/loss comparison | Best regularization and zero noise

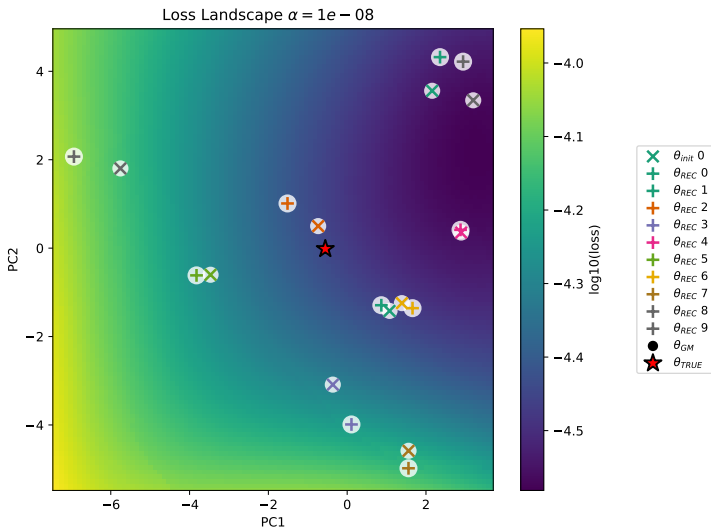


- ▶ The error/loss correlation is not obvious, and the signs are quite wrong. The same conclusions extend to other metrics such as Pearson or Spearman correlations.

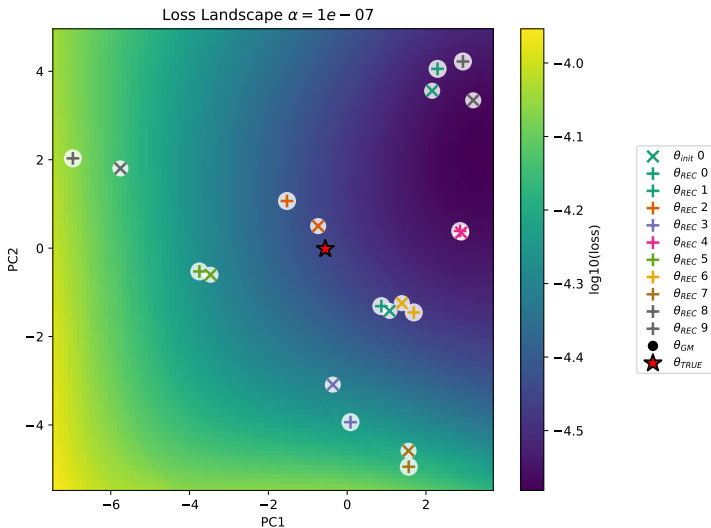
Projected loss landscape new



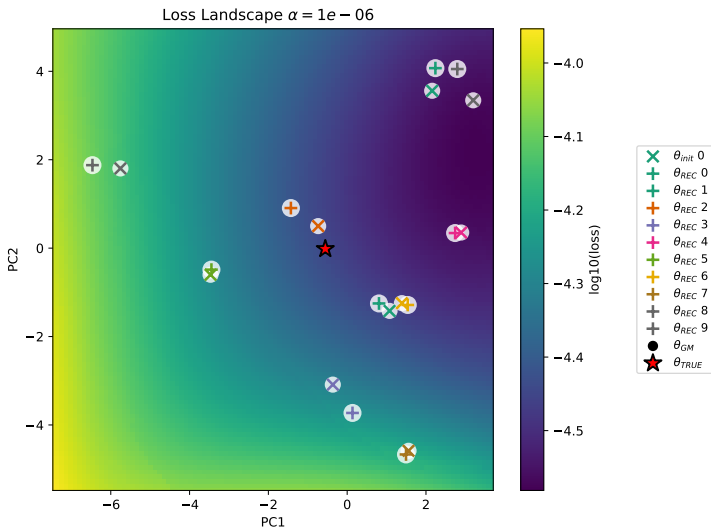
Projected loss landscape new



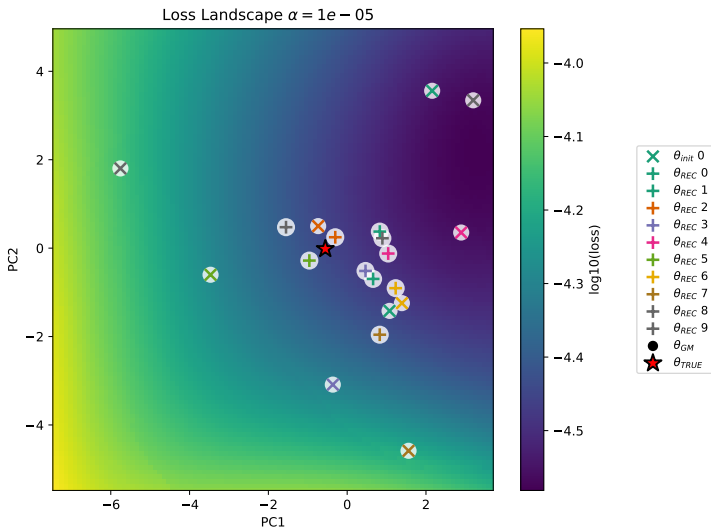
Projected loss landscape new



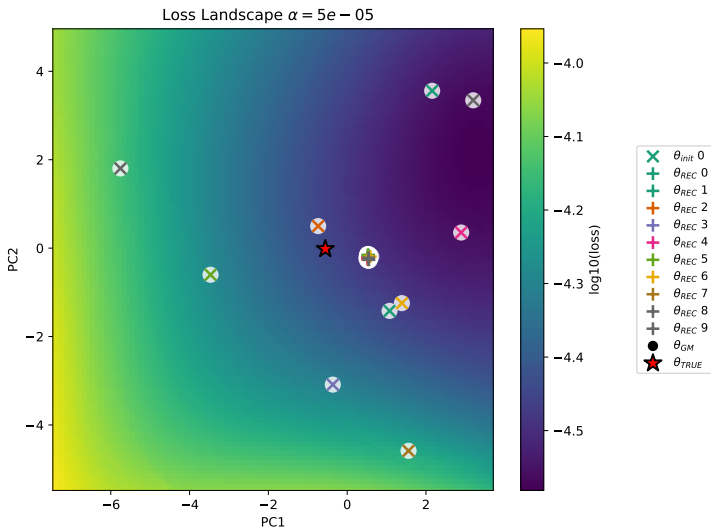
Projected loss landscape new



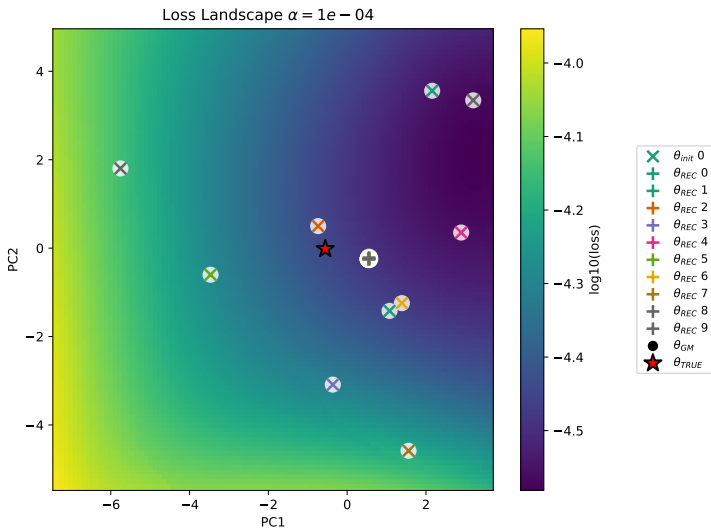
Projected loss landscape new



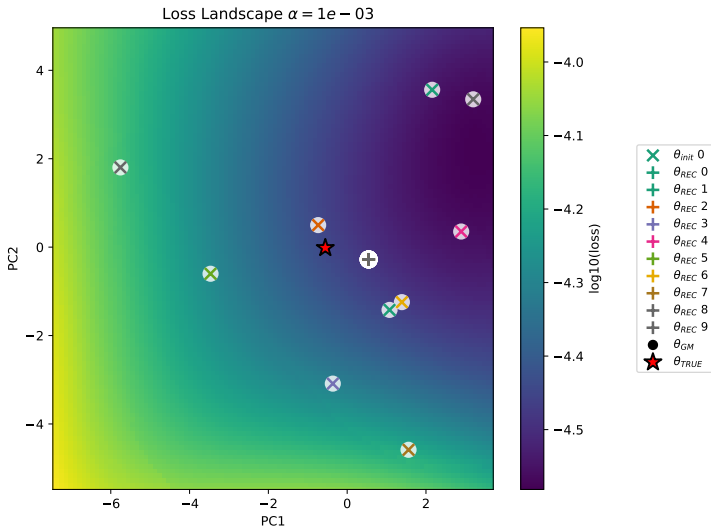
Projected loss landscape new



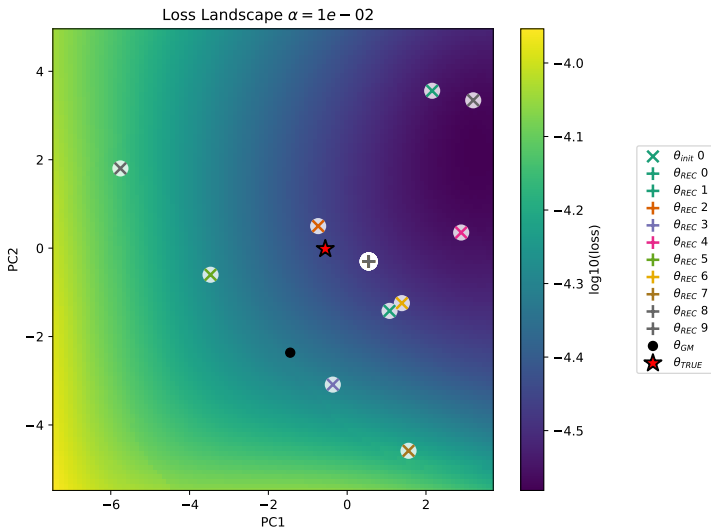
Projected loss landscape new



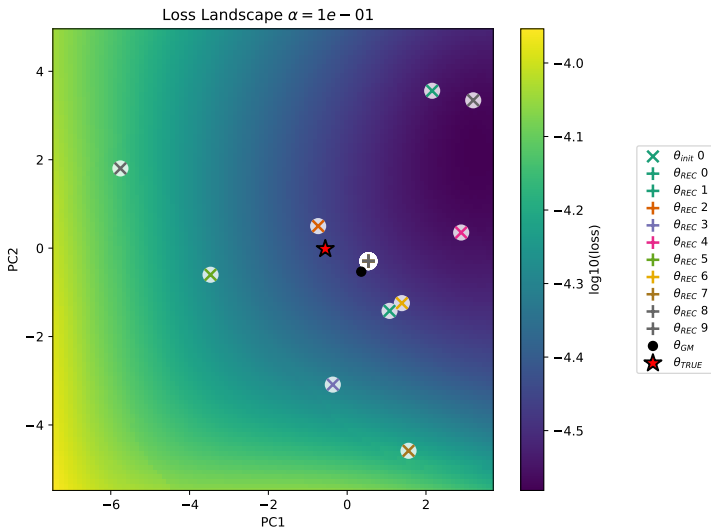
Projected loss landscape new



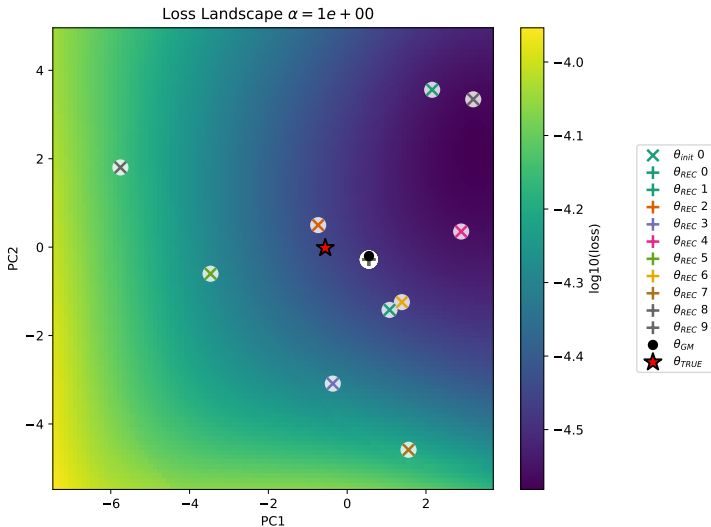
Projected loss landscape new



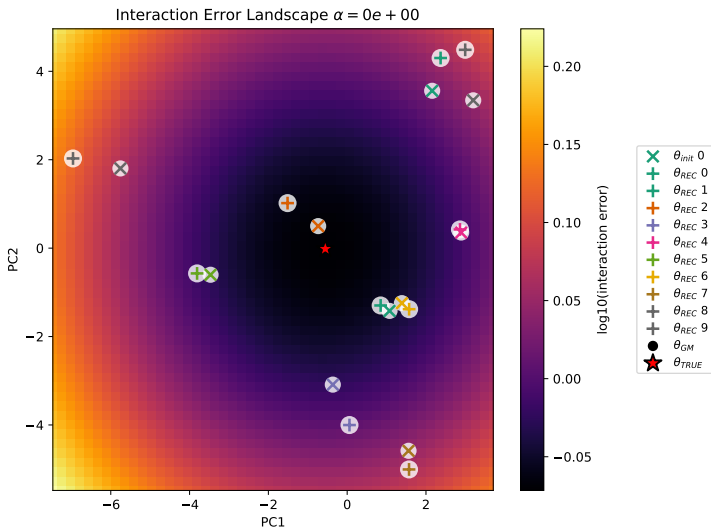
Projected loss landscape new



Projected loss landscape new



Projected loss landscape new



- 1 Context
- 2 Inference algorithms
- 3 Inference results
- 4 Experimental datasets**
- 5 Ongoing work

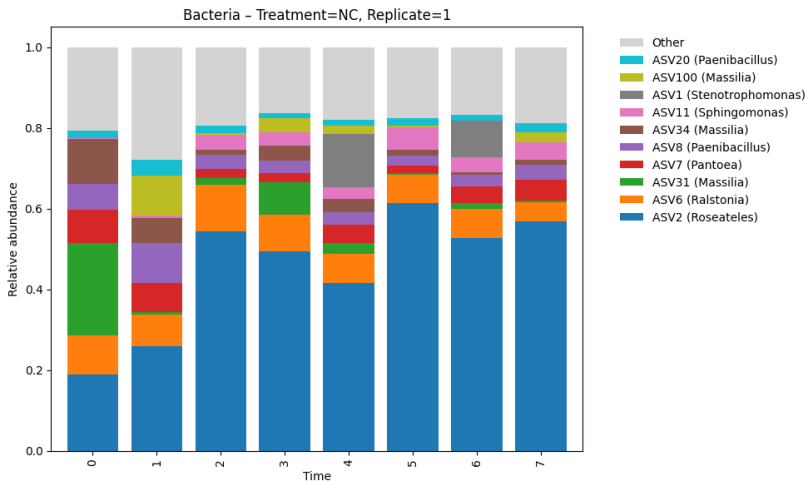
Vitae experiment

- ▶ Study grapevine mildew resistance with synthetic microbial communities inoculations
- ▶ Relative abundances, through metabarcoding
- ▶ High diversity $S \sim 160$
- ▶ Low time sampling $T = 8, R = 4$



Aarti Jaswa

Vitae: Time series



Vitae: Difficulties

- ▶ Low resolution data compared to the number of ASVs

$$TR \ll S$$

- ▶ Needs to reduce the studied diversity: functional clustering, taxonomic clustering, or most abundant only
- ▶ The relative abundances are on distinct scales

Millidrop experiment

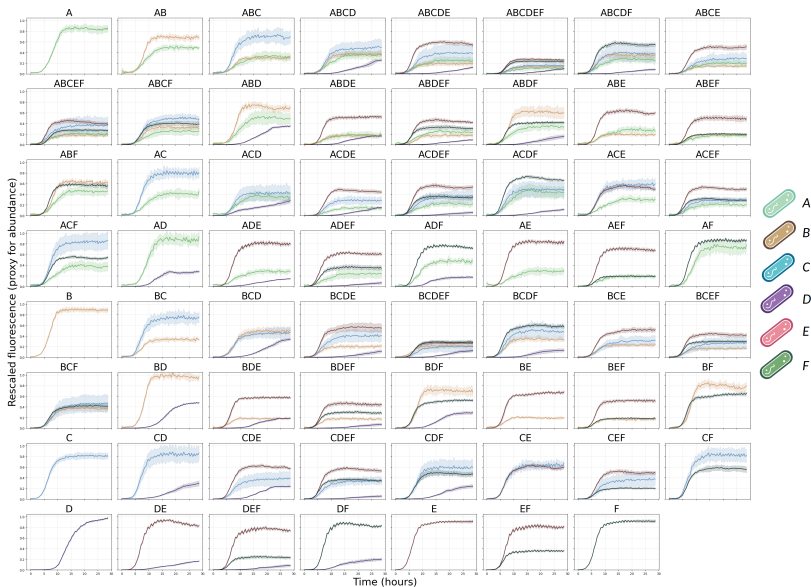
- ▶ Combinatorial study of beans microbes
- ▶ Absolute abundances, through fluorescent marking
- ▶ Low diversity $S \sim 6$
- ▶ High time sampling $T \sim 200, R = 10$



Kenji Maurice



Millidrop: Time series



- 1 Context
- 2 Inference algorithms
- 3 Inference results
- 4 Experimental datasets
- 5 Ongoing work

Ongoing work

Strengthen the study with:

- ▶ Distinct ground-truth sampling structures
- ▶ (S, T, R) dependence
- ▶ Data noise

Acknowledgment

▶ VITAE experiment

- ▶ SAVE team, INRAE Bordeaux
- ▶ Manon Chargy
- ▶ Paola Fournier
- ▶ Corinne Vacher

▶ Millidrop experiment

- ▶ EmerSys Team, INRAE - IRHS Angers
- ▶ Florian Lamouche
- ▶ Samuel Bourdon
- ▶ Elodie Lemmens
- ▶ Brice Marolleau
- ▶ Manuel Blouin
- ▶ Marie Simonin

Acknowledgment

▶ VITAE experiment

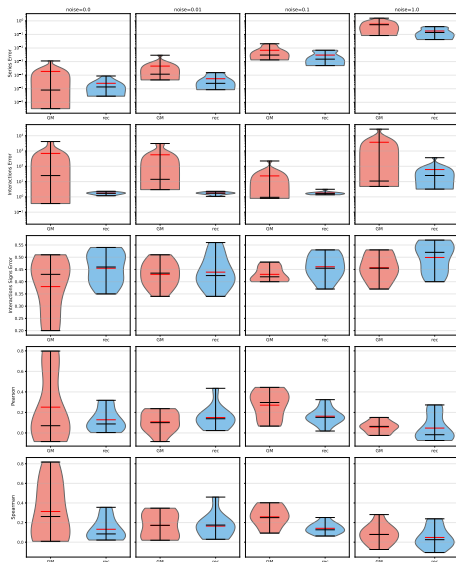
- ▶ SAVE team, INRAE Bordeaux
- ▶ Manon Chargy
- ▶ Paola Fournier
- ▶ Corinne Vacher

▶ Millidrop experiment

- ▶ EmerSys Team, INRAE - IRHS Angers
- ▶ Florian Lamouche
- ▶ Samuel Bourdon
- ▶ Elodie Lemmens
- ▶ Brice Marolleau
- ▶ Manuel Blouin
- ▶ Marie Simonin

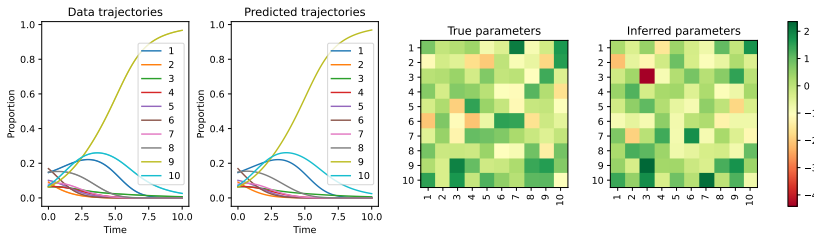
Thank you for your attention

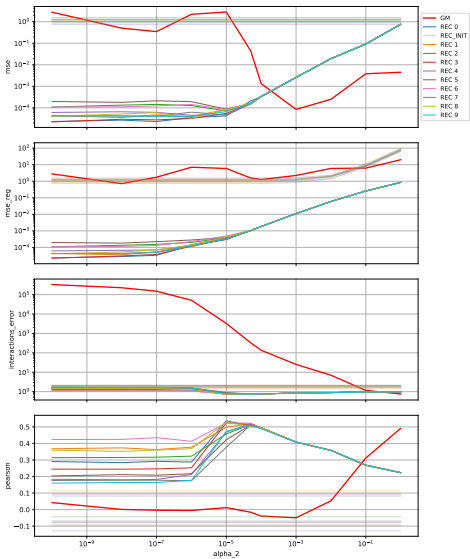
- 6 Inference results
- 7 Loss landscape
- 8 Vitae experiment
- 9 Millidrop experiment
- 10 Picture sources



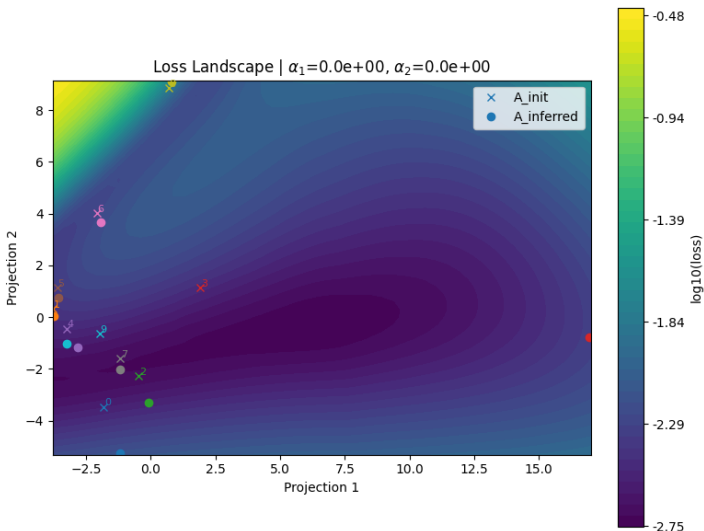
- 6 Inference results
- 7 Loss landscape**
- 8 Vitae experiment
- 9 Millidrop experiment
- 10 Picture sources

ReplicatorInference | MSE=2.215e-05

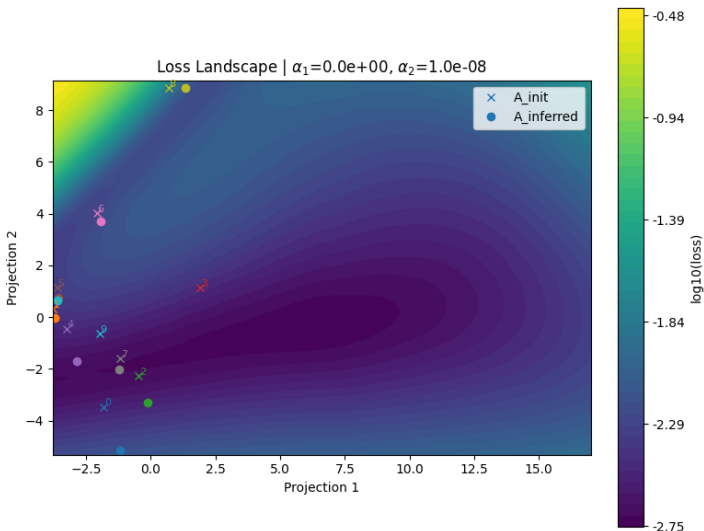




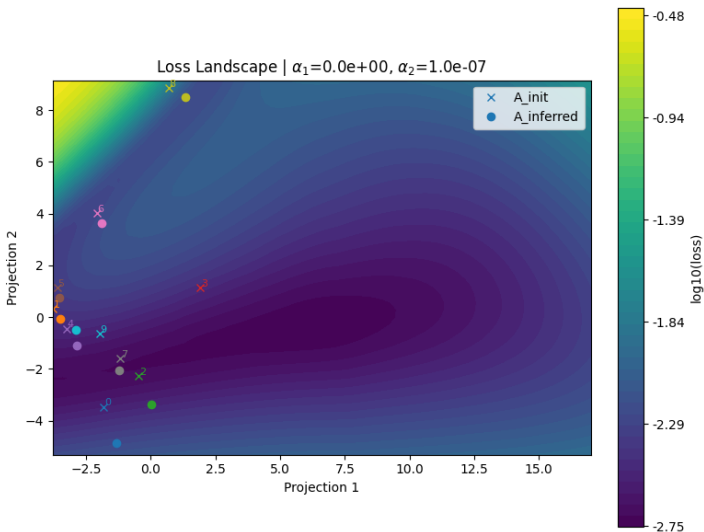
Projected loss landscape



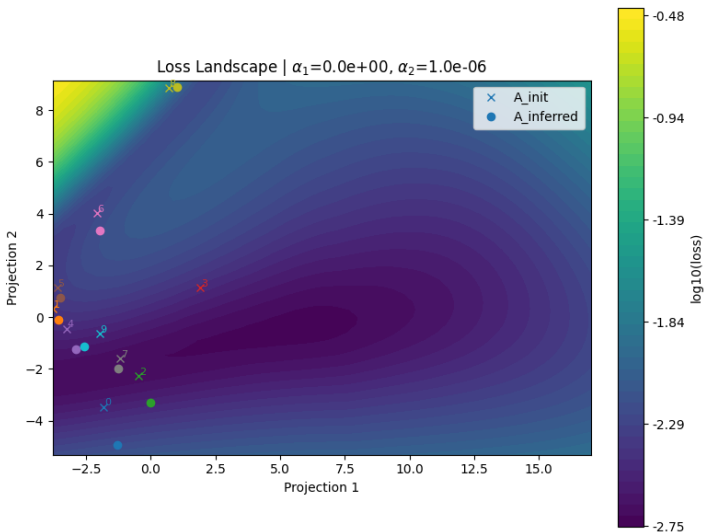
Projected loss landscape



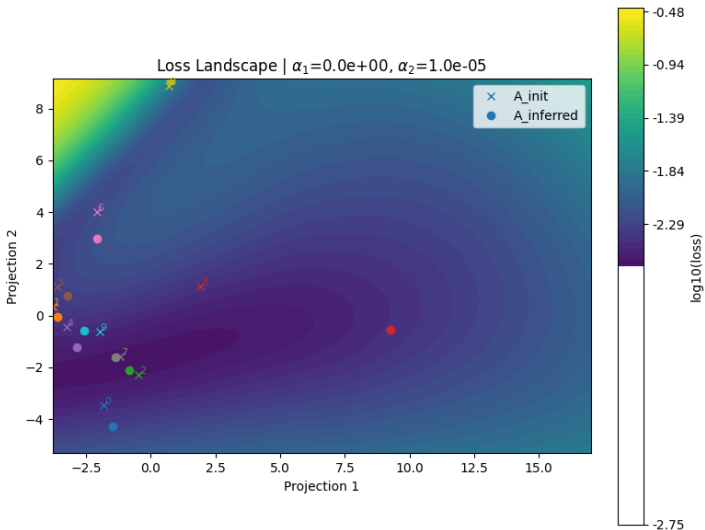
Projected loss landscape



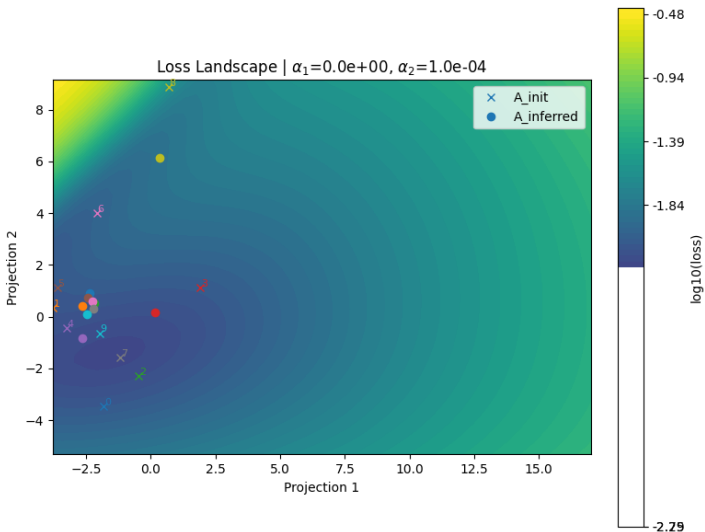
Projected loss landscape



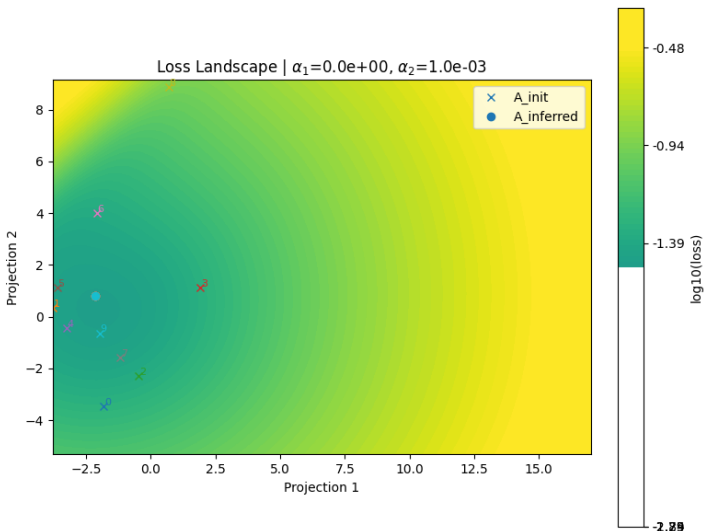
Projected loss landscape



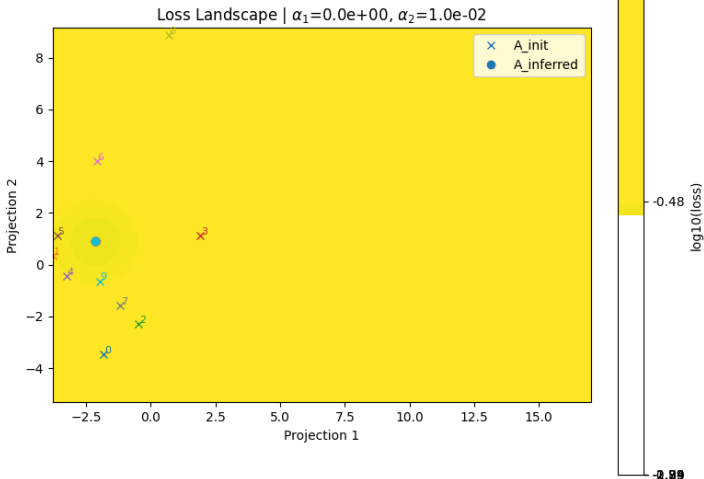
Projected loss landscape



Projected loss landscape

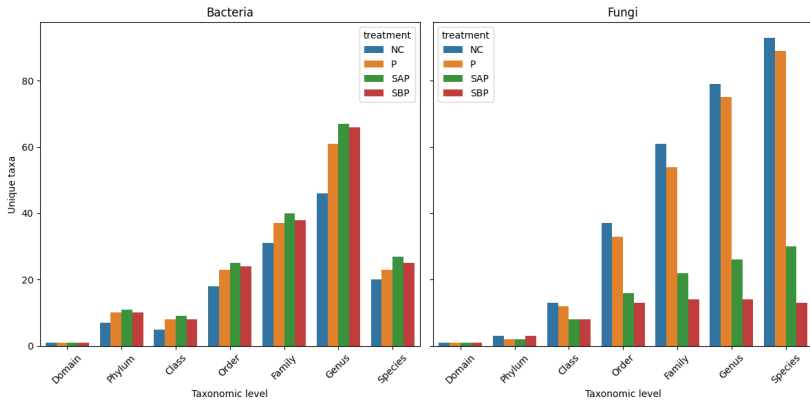


Projected loss landscape



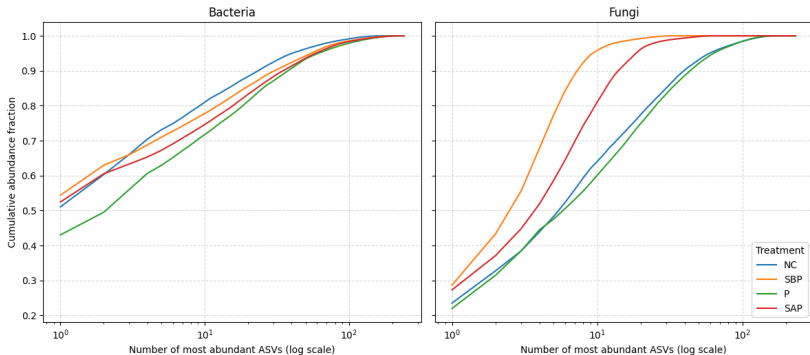
- 6 Inference results
- 7 Loss landscape
- 8 Vitae experiment**
- 9 Millidrop experiment
- 10 Picture sources

VITAE: Data

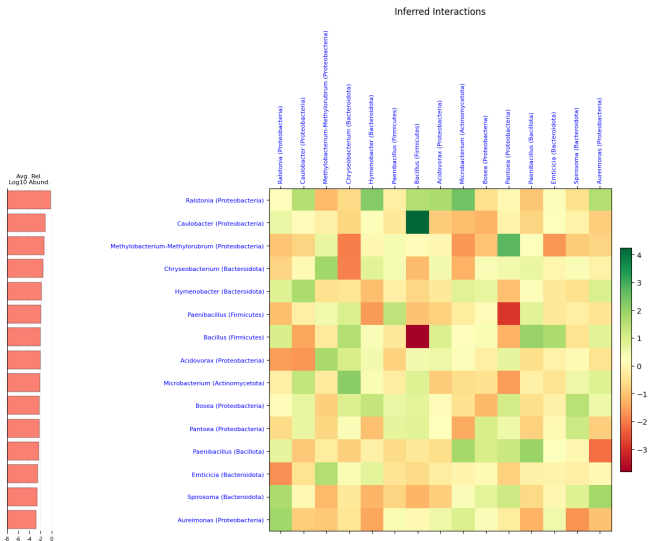


VITAE: Data

Cumulative abundance fraction vs number of ASVs

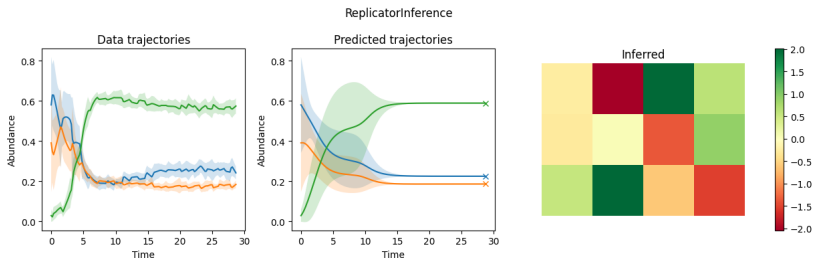


Vitae: First results per domain (bacteria)

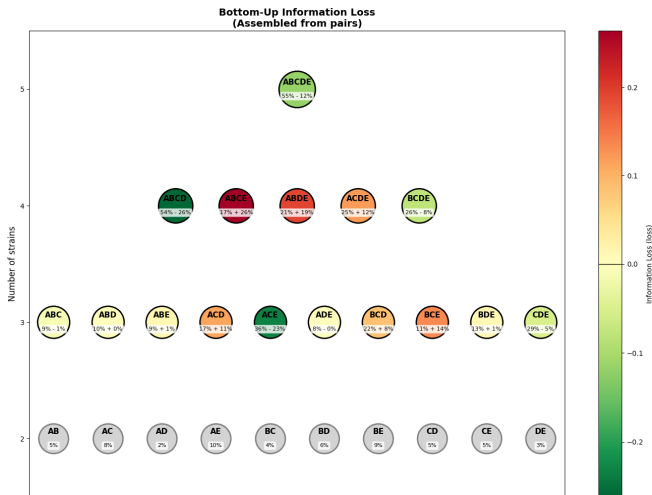


- 6 Inference results
- 7 Loss landscape
- 8 Vitae experiment
- 9 Millidrop experiment**
- 10 Picture sources

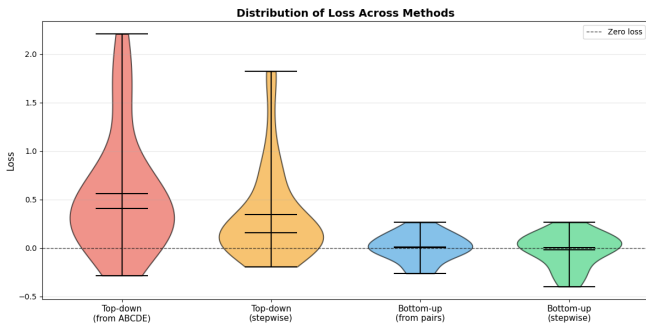
Millidrop: First results on a community



Millidrop: First results on information propagation



Millidrop: First results on information propagation



- 6 Inference results
- 7 Loss landscape
- 8 Vitae experiment
- 9 Millidrop experiment
- 10** Picture sources

Source of the pictures

- ▶ https://fr.wikipedia.org/wiki/Plasmopara_viticola

- ▶ <https://nerdist.com/the-cruelty-of-nature-is-extra-cute-in-this-adult-coloring-book/>