



Metropolitan Water Reclamation District of Greater Chicago

Welcome to the April Edition
of the 2026 M&R Seminar Series

NOTES FOR SEMINAR ATTENDEES

- Remote attendees' microphones are muted at entry to minimize background noise.
For attendees in the auditorium, please silence your phones.
- A question and answer (Q/A) session will follow the presentation.
- For remote attendees, please use “**Chat**” only to type questions for the presenter. For other issues, please send emails to MnRseminars@mwrld.org.
For attendees in the auditorium, please raise your hand and wait for the microphone to ask a verbal question during the Q/A session.
- The presentation slides will be posted on the MWRD website after the seminar.
- This seminar has been approved by the Engineering Society of Illinois (ESI) for one PDH and approved by the IEPA for one TCH. Certificates will be issued only to participants who attend the entire presentation. For PDH certificate seekers, **completing a brief course evaluation and submitting it are required.**

FRANCIS L. DE LOS REYES II, PH.D., BCEEM, WEF FELLOW, TED FELLOW

Glenn E. and Phyllis J. Futrell Distinguished Professor of Civil, Construction,
and Environmental Engineering and Alumni Distinguished Undergraduate Professor,
North Carolina State University, Raleigh, North Carolina



Dr. de los Reyes received a B.S. in Agricultural Engineering from the University of the Philippines-Los Baños, M.S. in Civil Engineering from Iowa State University, and Ph.D. in Environmental Engineering from the University of Illinois at Urbana-Champaign. In 2020, he was elected to the National Academy of Science and Technology, Philippines (NAST PHL) as Corresponding Member in recognition of his outstanding scientific contributions at the interface of microbial ecology/biotechnology and environmental engineering, and for his continuous collaboration in the Philippines for the advancement of science and technology.

He was a Balik Scientist of the Philippine government in 2008. Dr. de los Reyes was a Board Member (2011-2013), Chair of the Board (2012-2013), Vice President (2013), and President (2014) of the Philippine-American Academy of Science and Engineering (PAASE). He was the recipient of the Severino and Paz Koh Lectureship in Engineering from PAASE (2013).



Exploring the use of machine learning to connect microbial community assembly to water resource recovery facility performance

Savanna Smith, Ph.D.

Francis L. de los Reyes III, Ph.D., BCEEM, F. WEF

Glenn E. and Phyllis J. Futrell Distinguished Professor of Environmental
Engineering

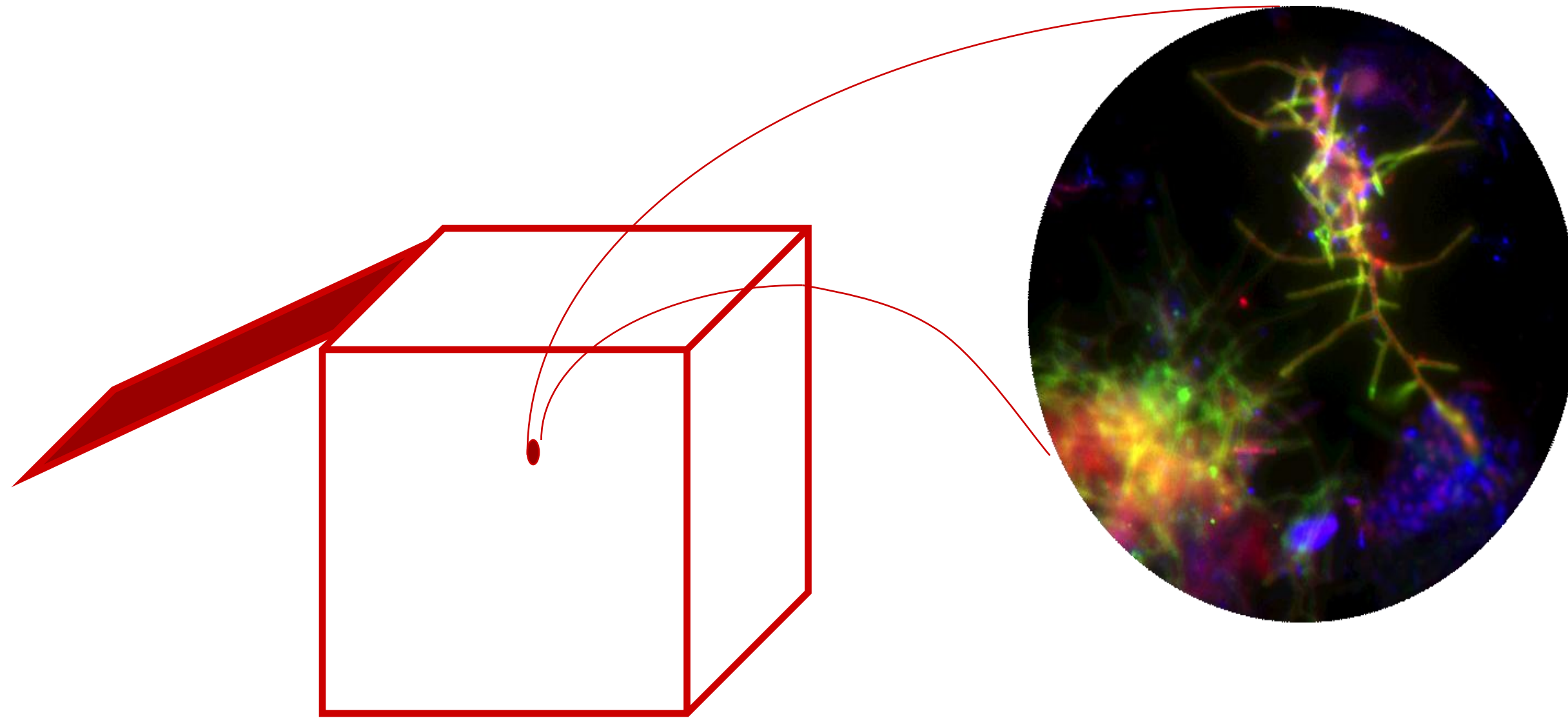
Alumni Distinguished Undergraduate Professor

University Faculty Scholar

North Carolina State University

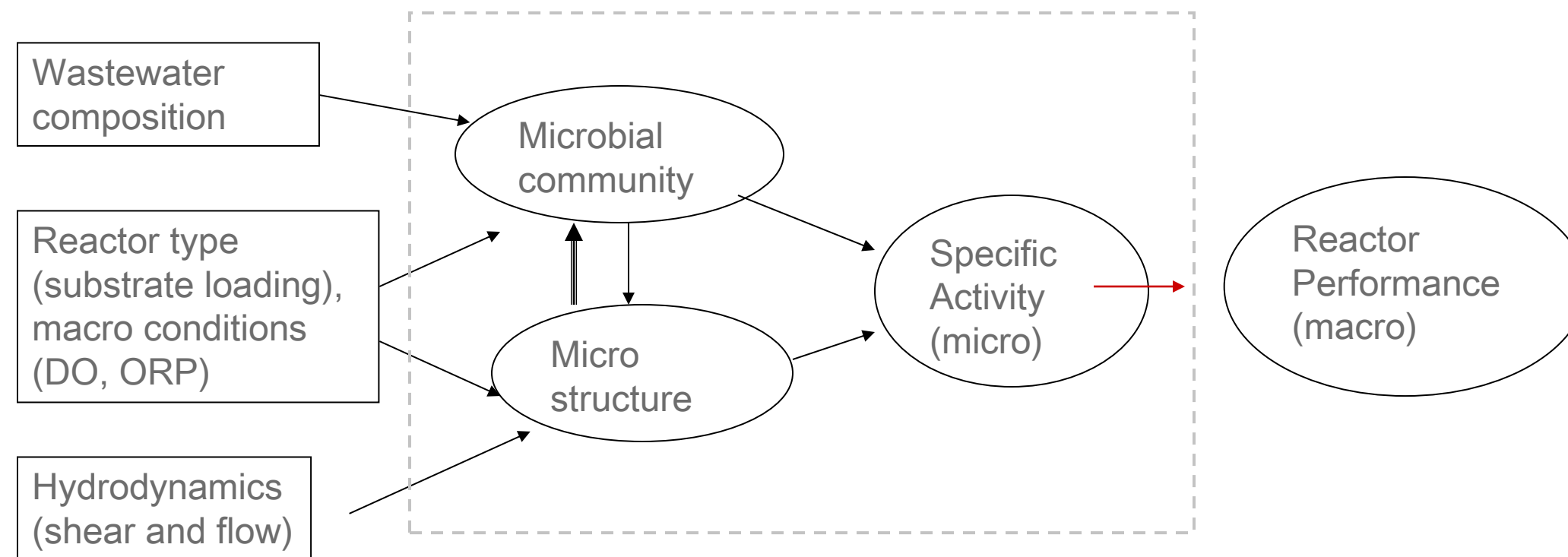


Microbial Populations → Inside the environmental “Black Box”



Who's there? How many? What are they doing (causative)?
Can we influence them?

‘Structure-Function’ Analysis

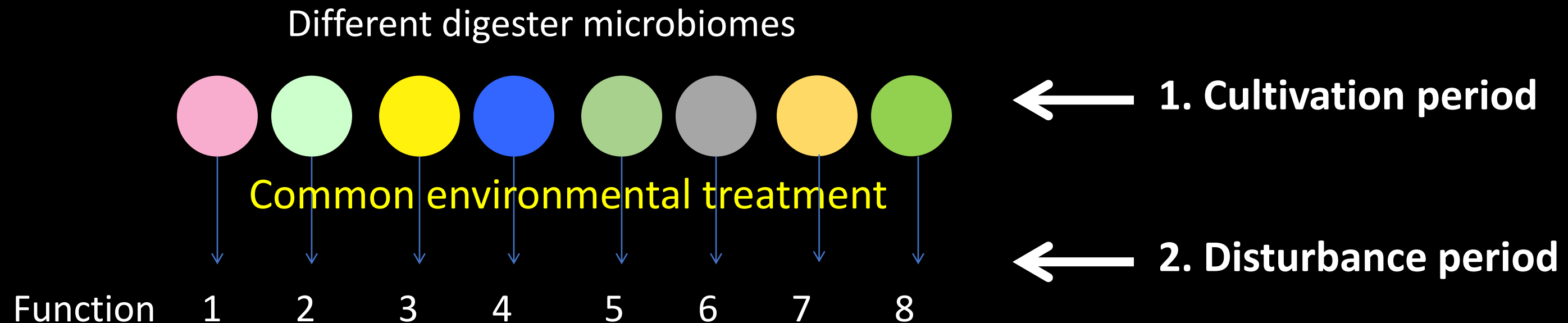


Environments: **activated sludge, anaerobic reactors, landfills, soils, sediments, groundwater, graywater, rumen, pit latrines, mudsprings**

Experiment I: Common garden experiment

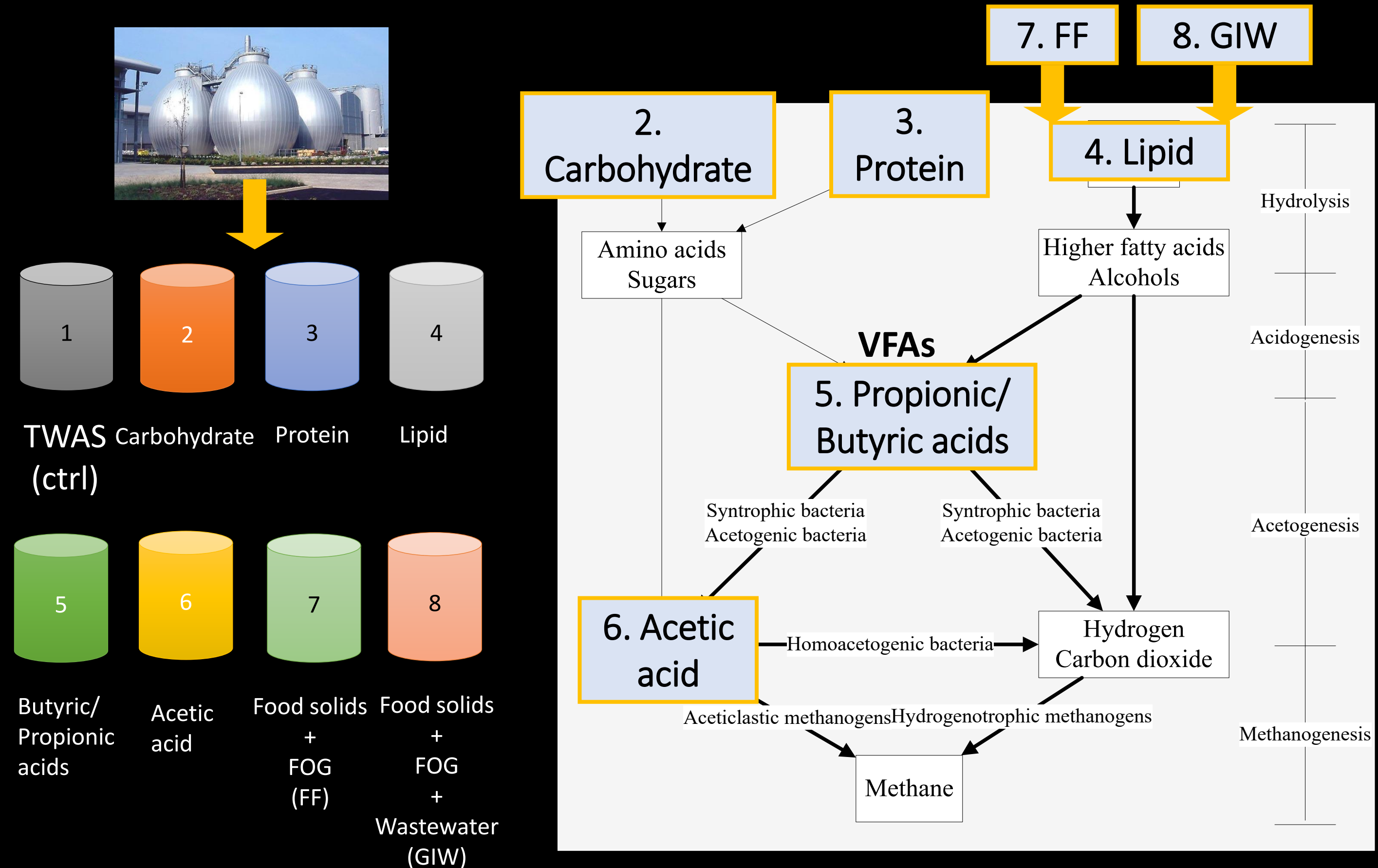
From the discipline of community ecology (de los Reyes 2015; Reed and Martiny, 2007)

- Evaluate the causal relationship between community composition, function, and environmental disturbances.
- Directly attribute differences in the function to differences in the composition.



1. Cultivation period

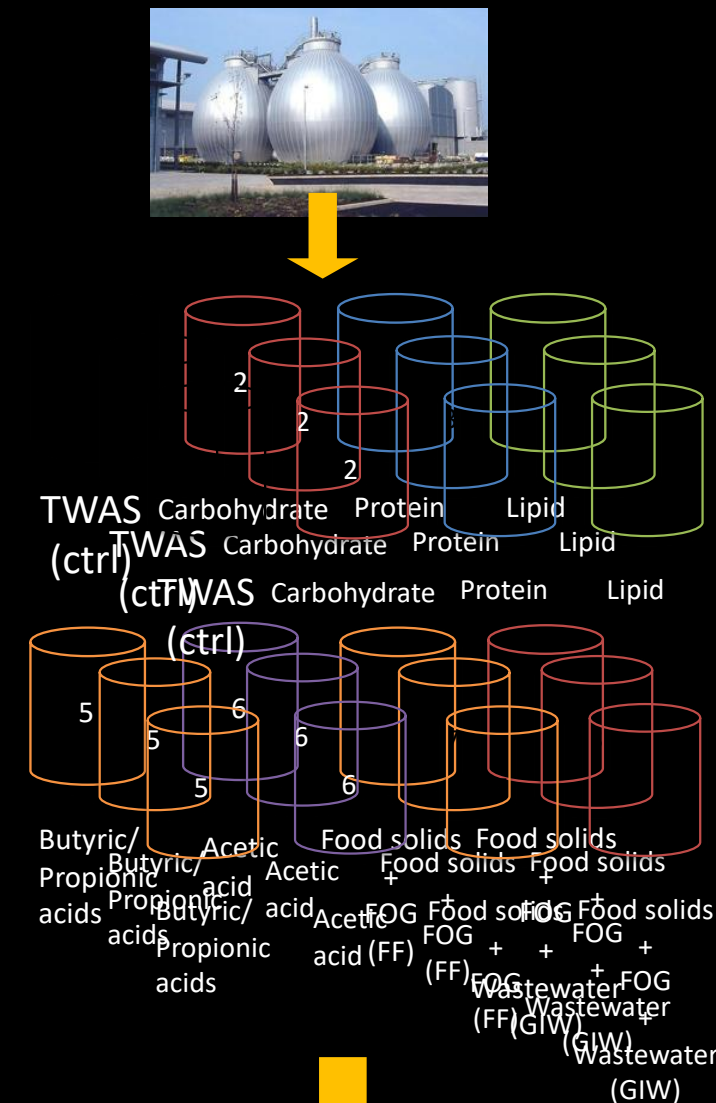
Using substrate enrichment to develop 8 different digester microbiomes



2. Disturbance period

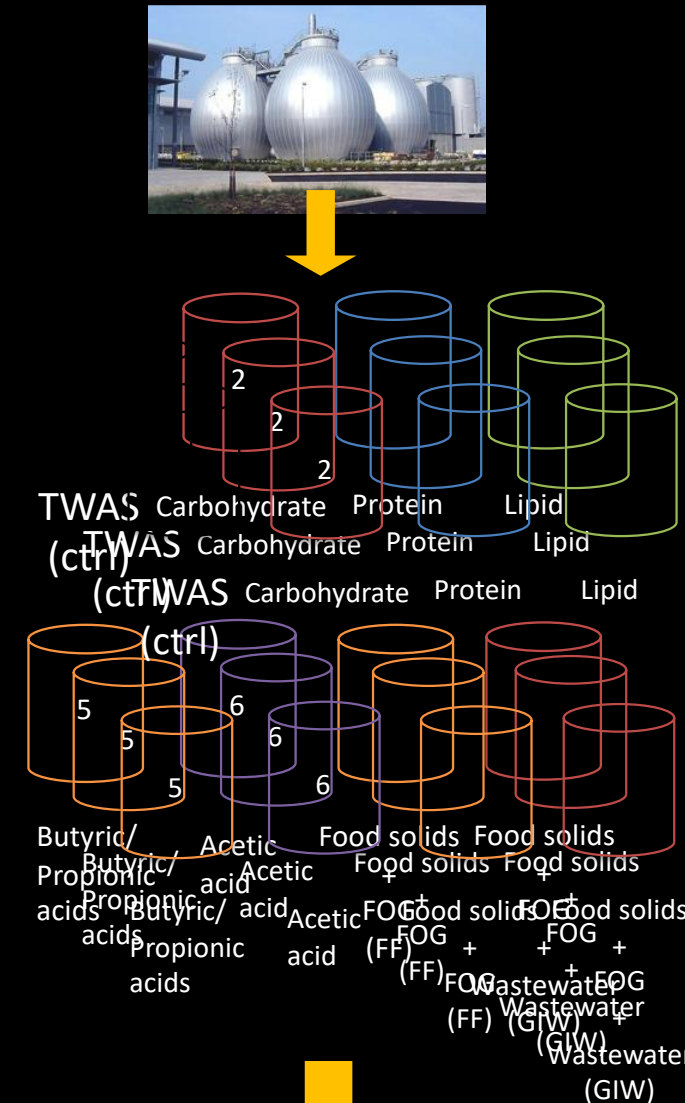
Using substrate enrichment to develop 8 different digester microbiomes

28-30% COD
(TWAS 70% COD
+
substrate 30%
COD)



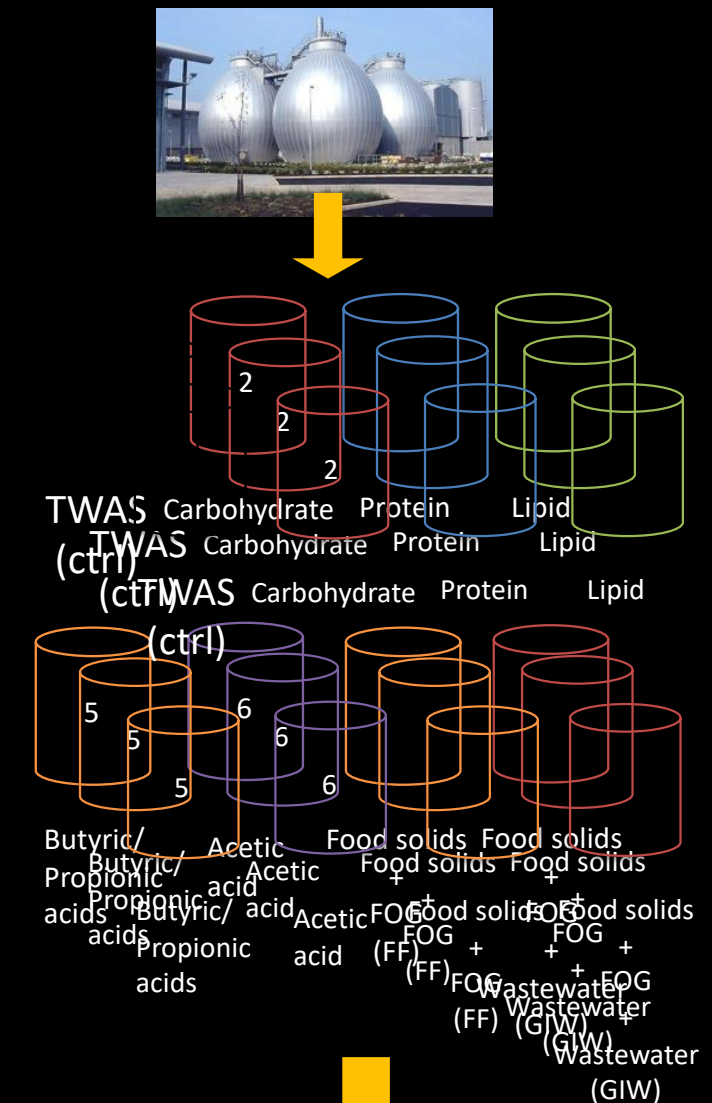
Low-GIW disturbance

63% COD
(TWAS 37% COD +
GIW 63% COD)



Mid-GIW disturbance

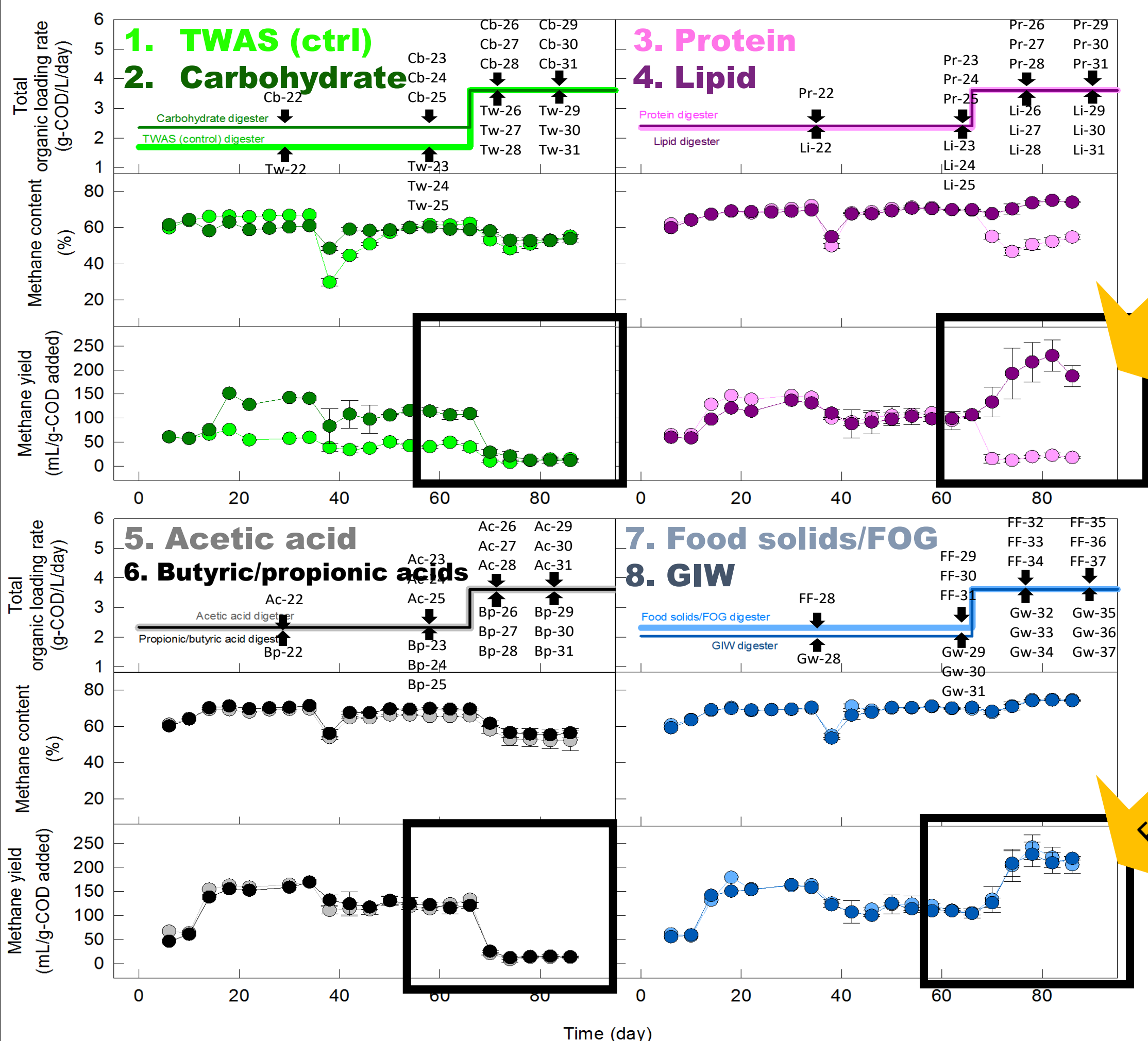
78% COD



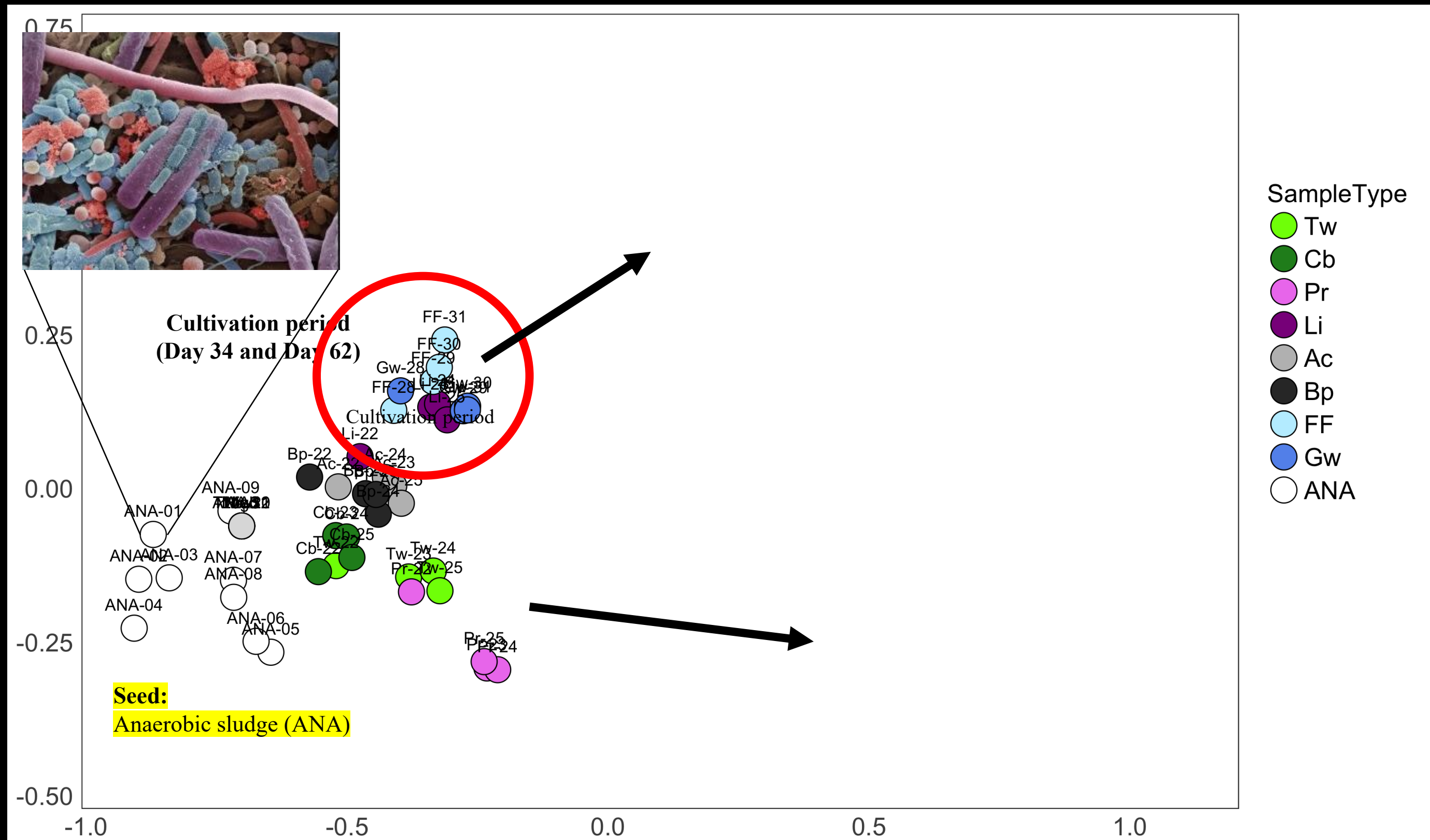
High-GIW disturbance

95% COD

Digester performance at low-GIW disturbance



Experiment I: Community analysis

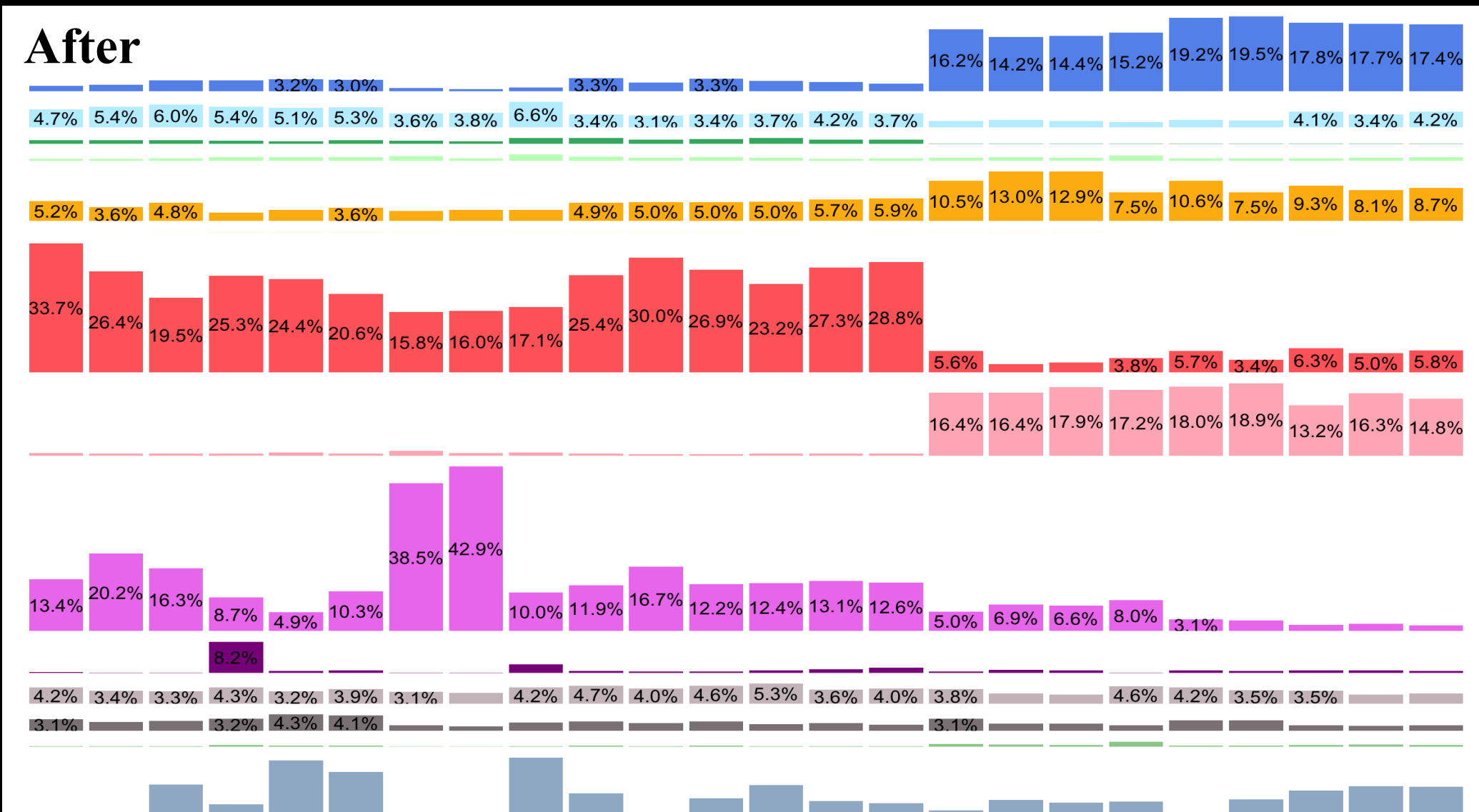


- Despite identical seed and environment, microbial community dissimilarity due to substrate history (**feed selection** and **intensity of preshocks**) determined response to overloading GIW stresses



Methanosaeta

Syntrophomonas



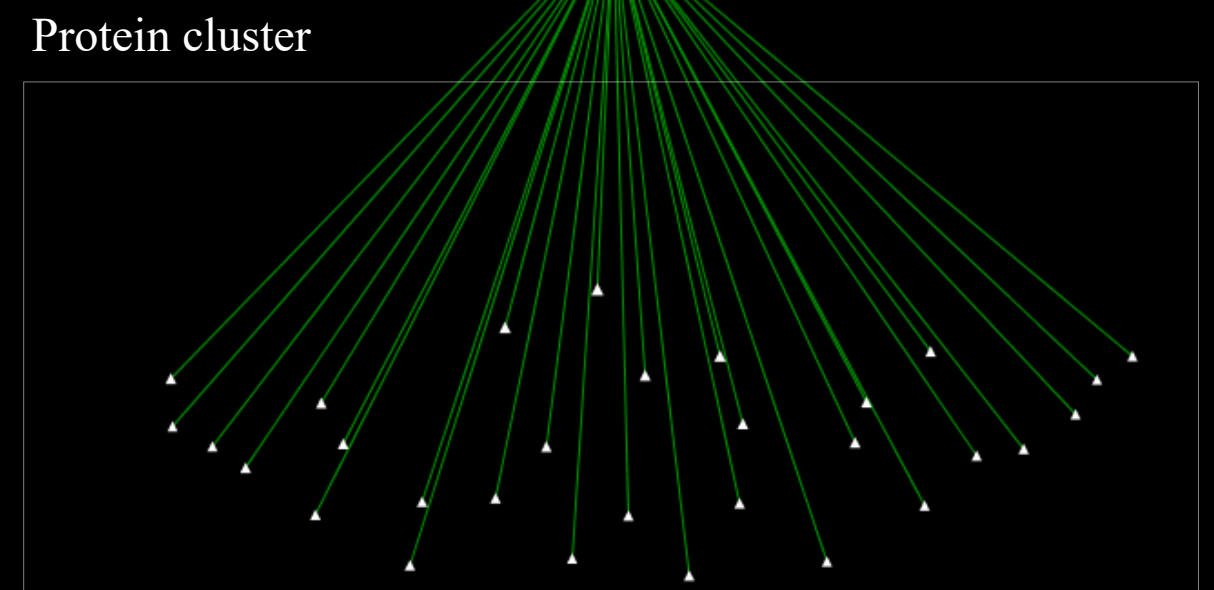
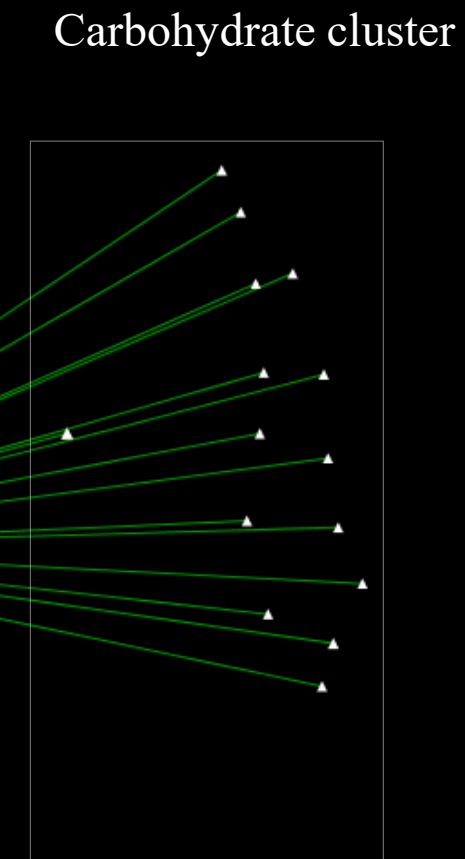
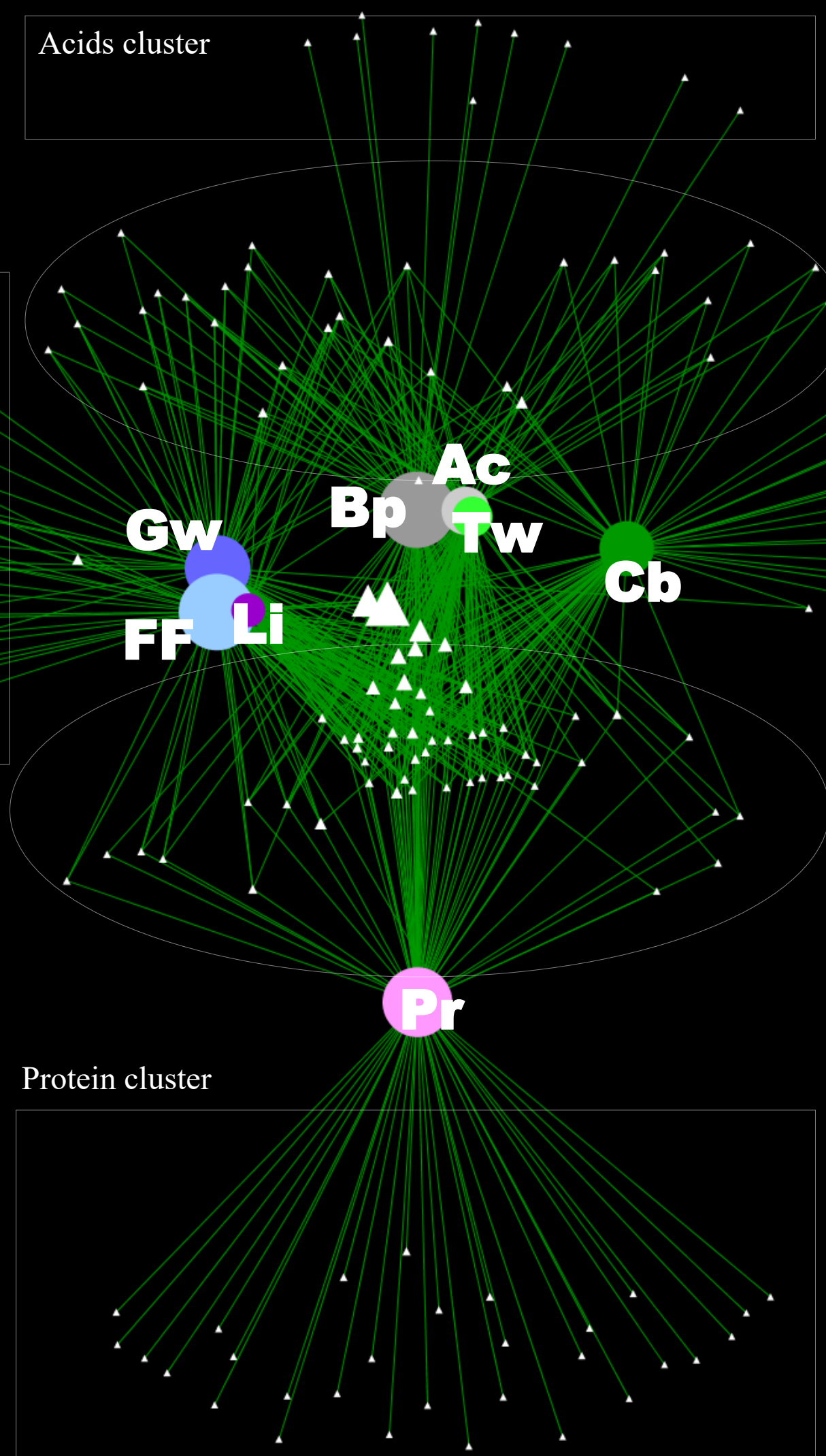
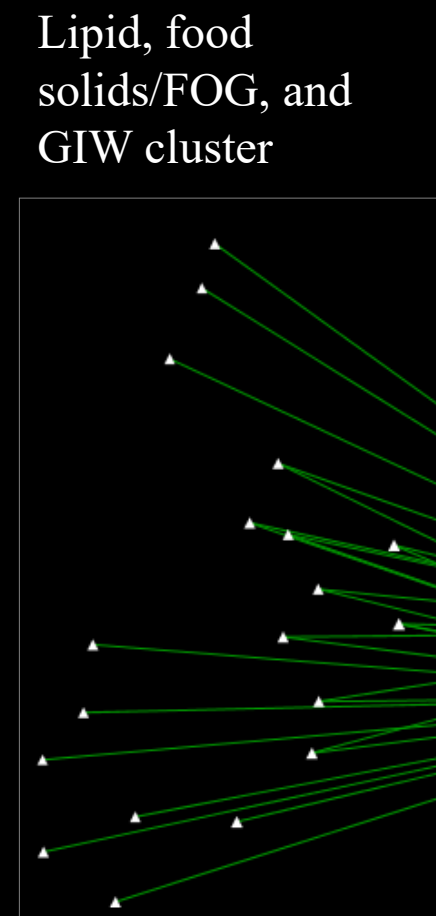
Methanosaeta

Porphyromonadaceae

Sporanaerobacter

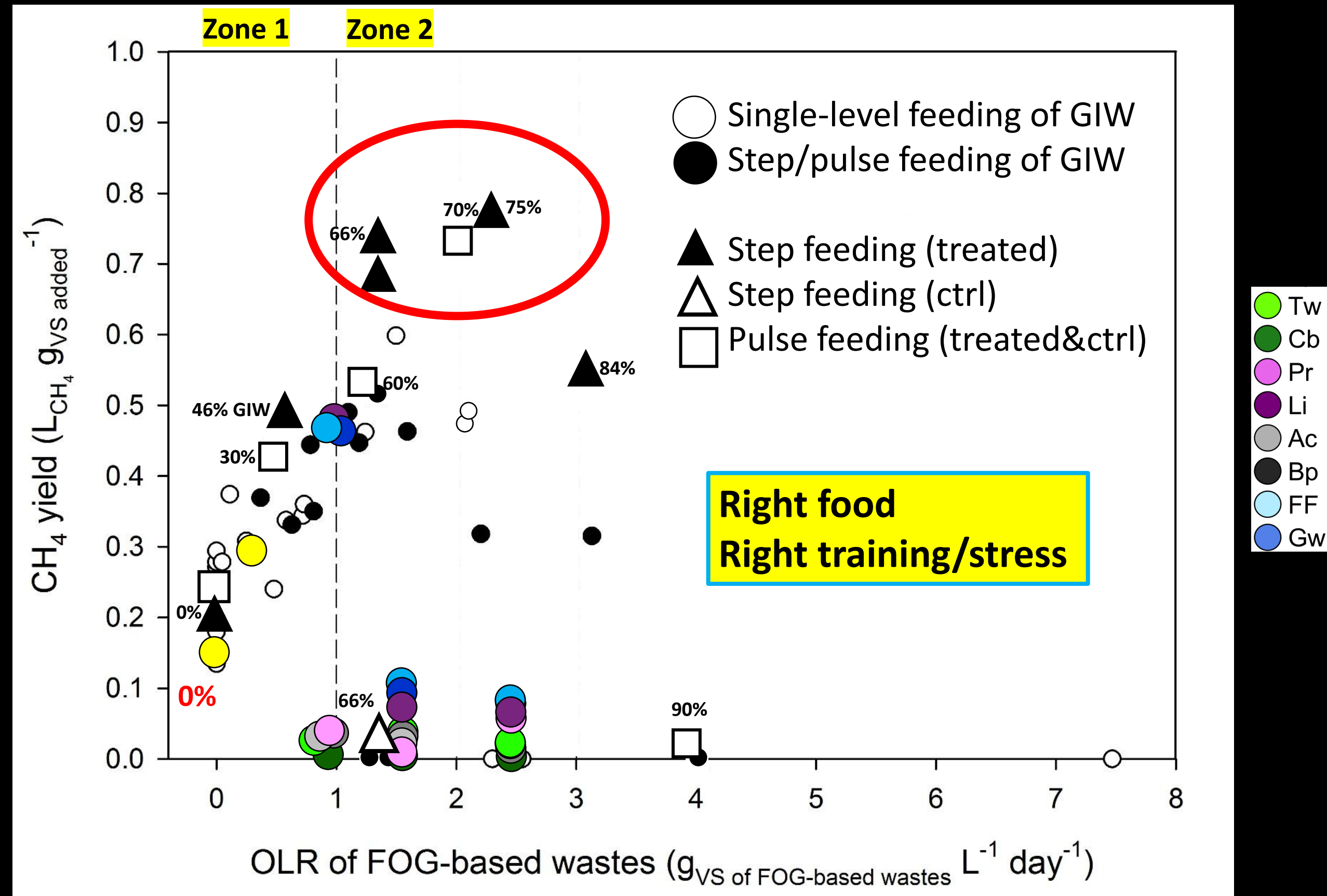
Syntrophomonas

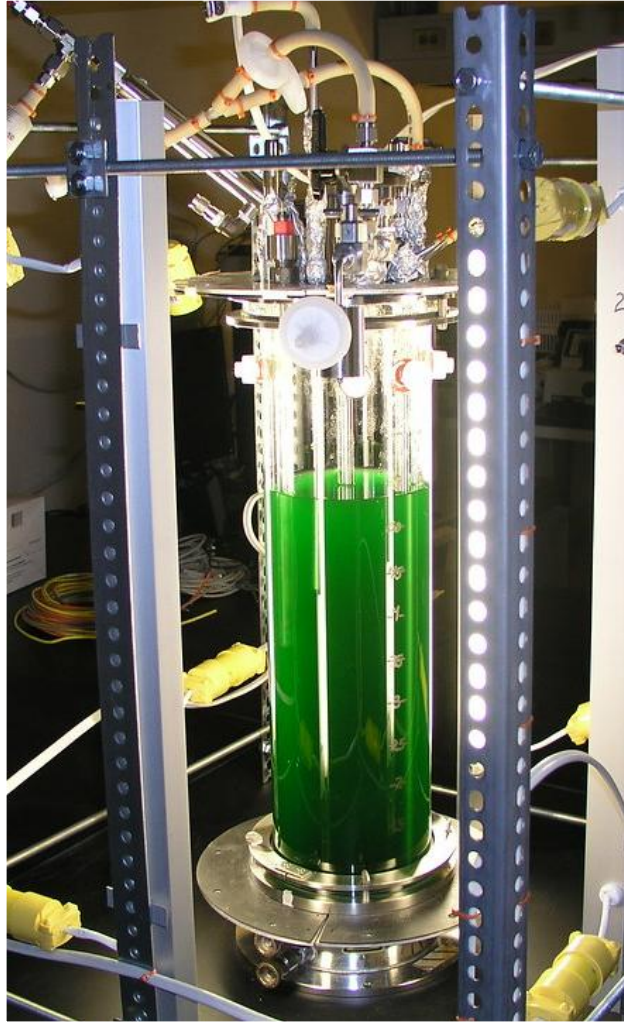
Family *Methanospirillaceae*, bacterial orders Bacteroidales (Bacteroidia) and Rhizobiales (Alphaproteobacteria), families *Saprospiraceae* (Bacteroidia) and *Methylocystaceae* (Alphaproteobacteria), and genera *Paludibacter* (Bacteroidia), *Sporanaerobacter* and *Syntrophomonas* (Clostridia), *Comamonas* (Betaproteobacteria), *Syntrophus* (Deltaproteobacteria), and *Methylocaldum* (Gammaproteobacteria).



TW (TWAS, ctrl)
Cb (Carbohydrate)
Pr (Protein)
Li (Lipid)
Ac (Acetic acid)
Bp (Butyric/propionic acids)
FF (Food solids/FOG)
Gw (GIW)

A "map" to high-performance GIW co-digestion





**Big “IF” (Big Hope):
If we can understand how communities assemble,
we can improve our bioreactors**

Hypothetical: How microbial community assembly can lead to better reactors

These factors are
descriptions of
MCA processes

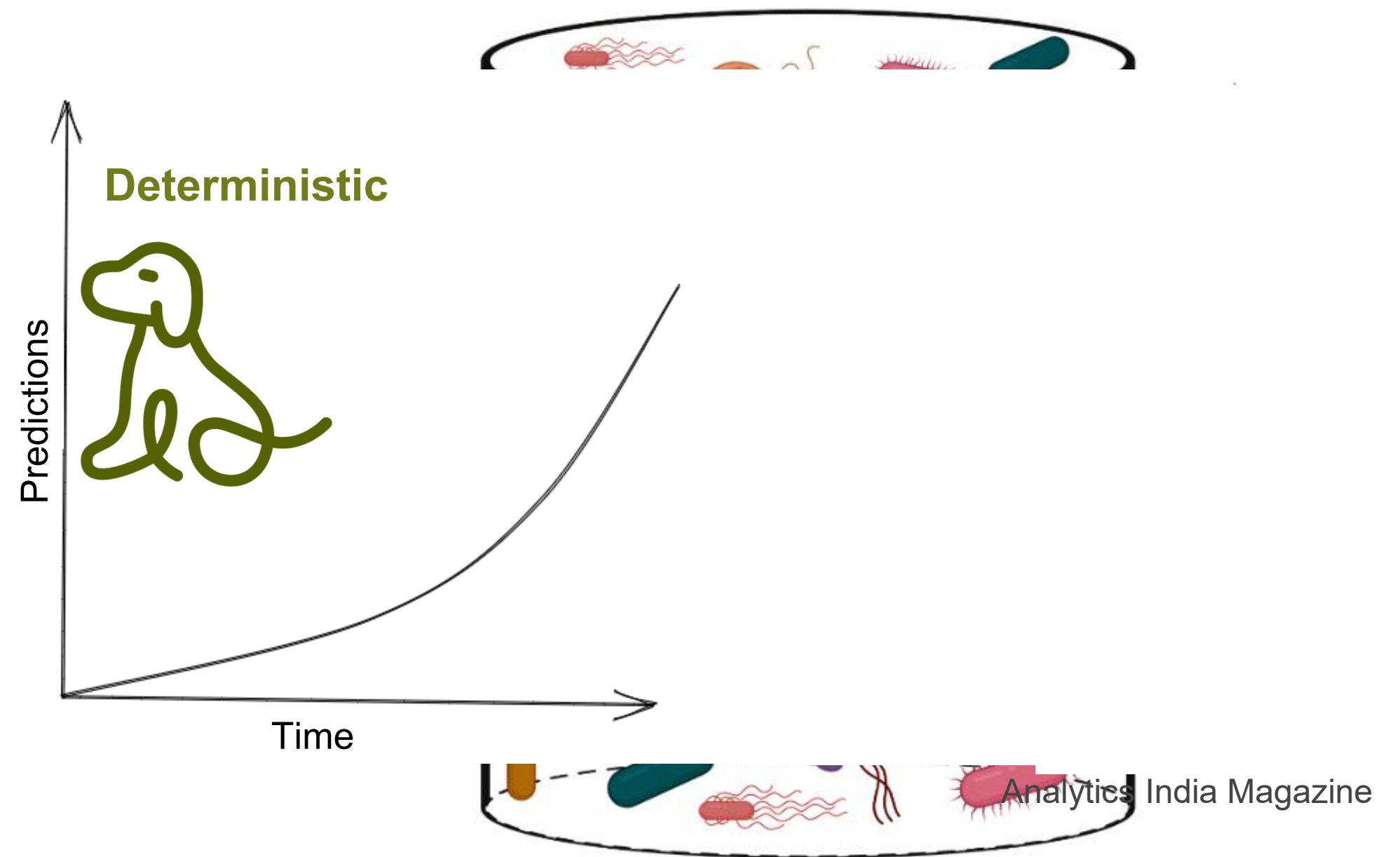
Challenges:

1. Unable to reliably predict bioreactor process upsets or failures
2. Unclear connections between microbial community dynamics and bioreactor performance
3. Rising demands for biological process intensification
4. A deluge of data associated with biological processes converting waste to resources coupled with unclear steps of how to leverage this data to improve process performance

Aim: apply MCA theories and knowledge to improve bioreactor function in engineered biological processes

Microbial community assembly processes describe how microbial communities form

- Each process is **deterministic**, **stochastic**, or both
- Conflicting conclusions about what's important
→ Does dispersal matter?

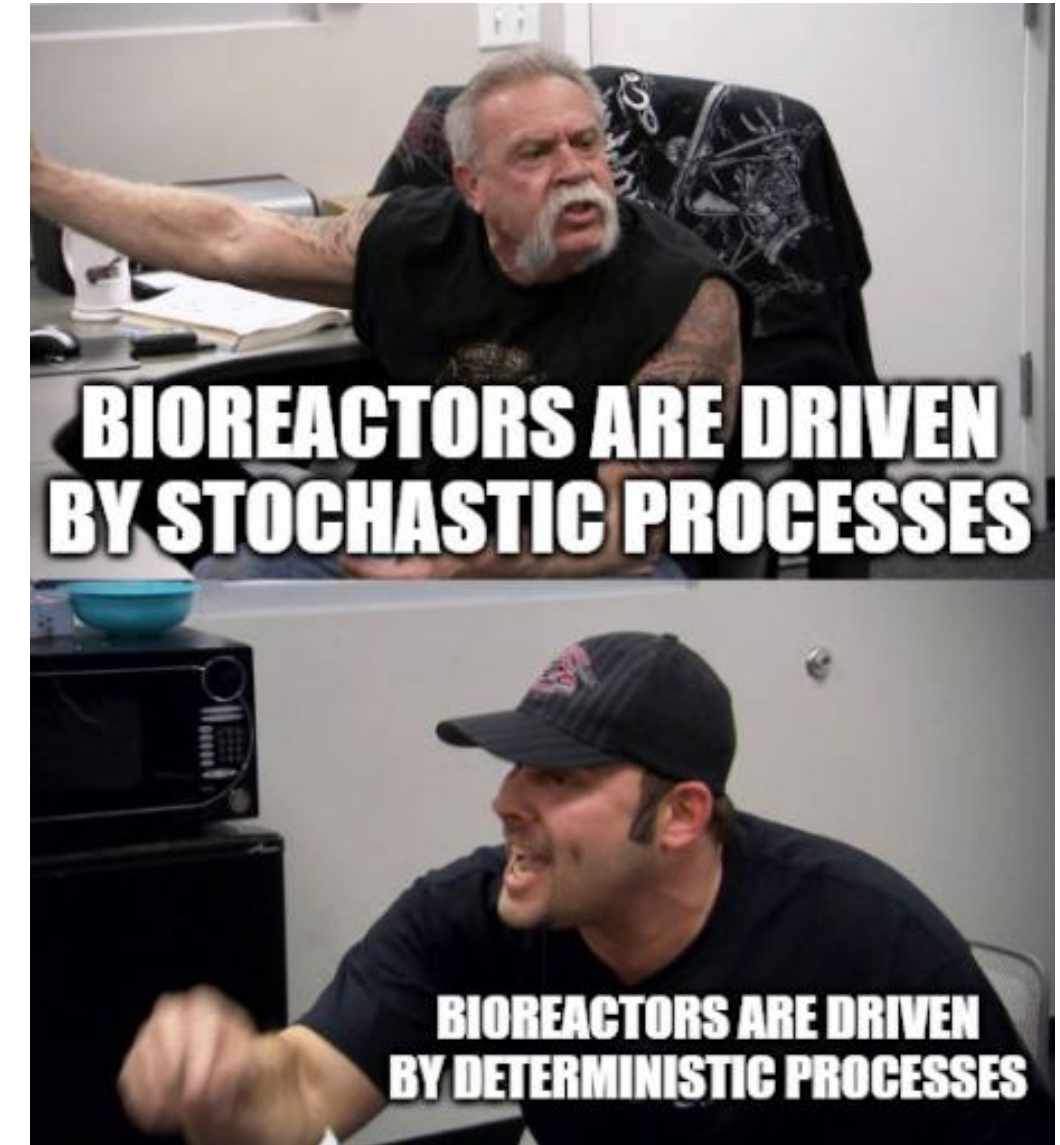


Relevant case studies

- 84% of case studies found both **deterministic** and **stochastic** processes influence MCA of engineered systems
- Disagreement arises from quantification of the relative influence of **deterministic** and **stochastic** processes
 - And from identifying which features affect this balance
- Example:
 - MCA of conventional activated sludge systems depends on the time scale assessed, physical scale of the system and its inputs, and the specific analyses used

Why don't we agree?

- Mismatch of methods used
- Varying system scales, operation, and experimental design
- Extrapolating expectations from non-engineered bioreactors may not be the most accurate
- There may not be generalizable rules or conclusions

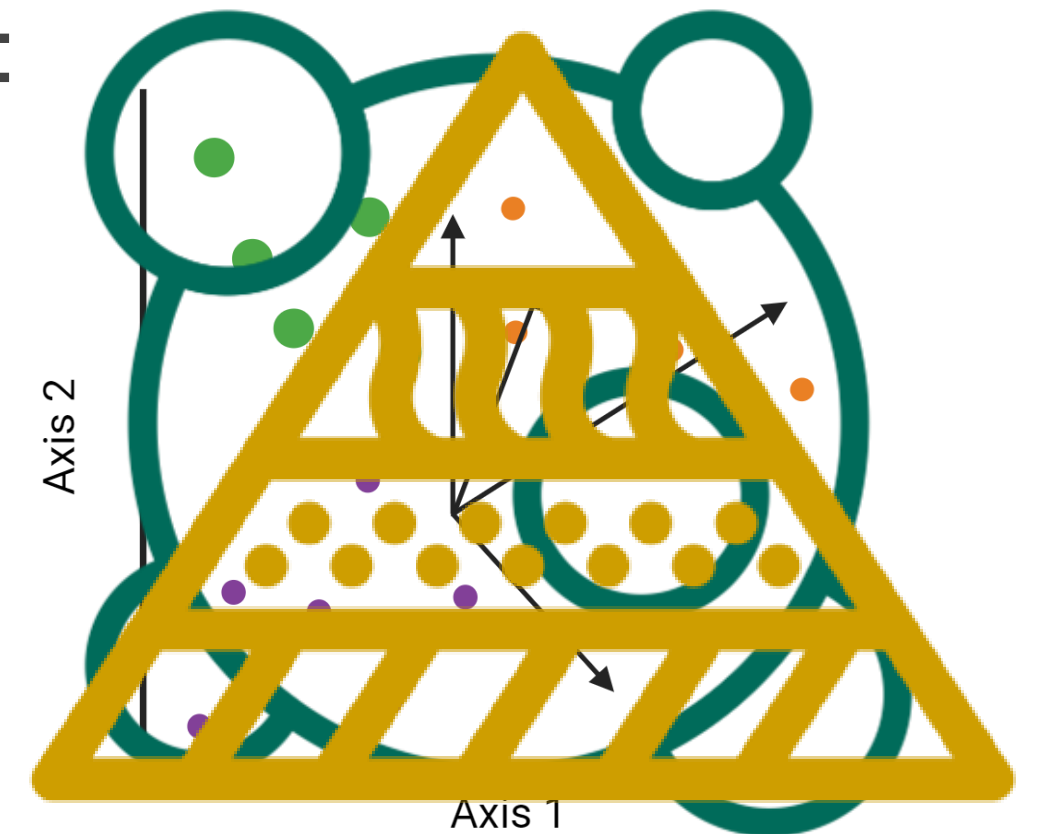


Primary methods to assess MCA processes

1. Multivariate analyses
2. Null modeling
3. Neutral modeling

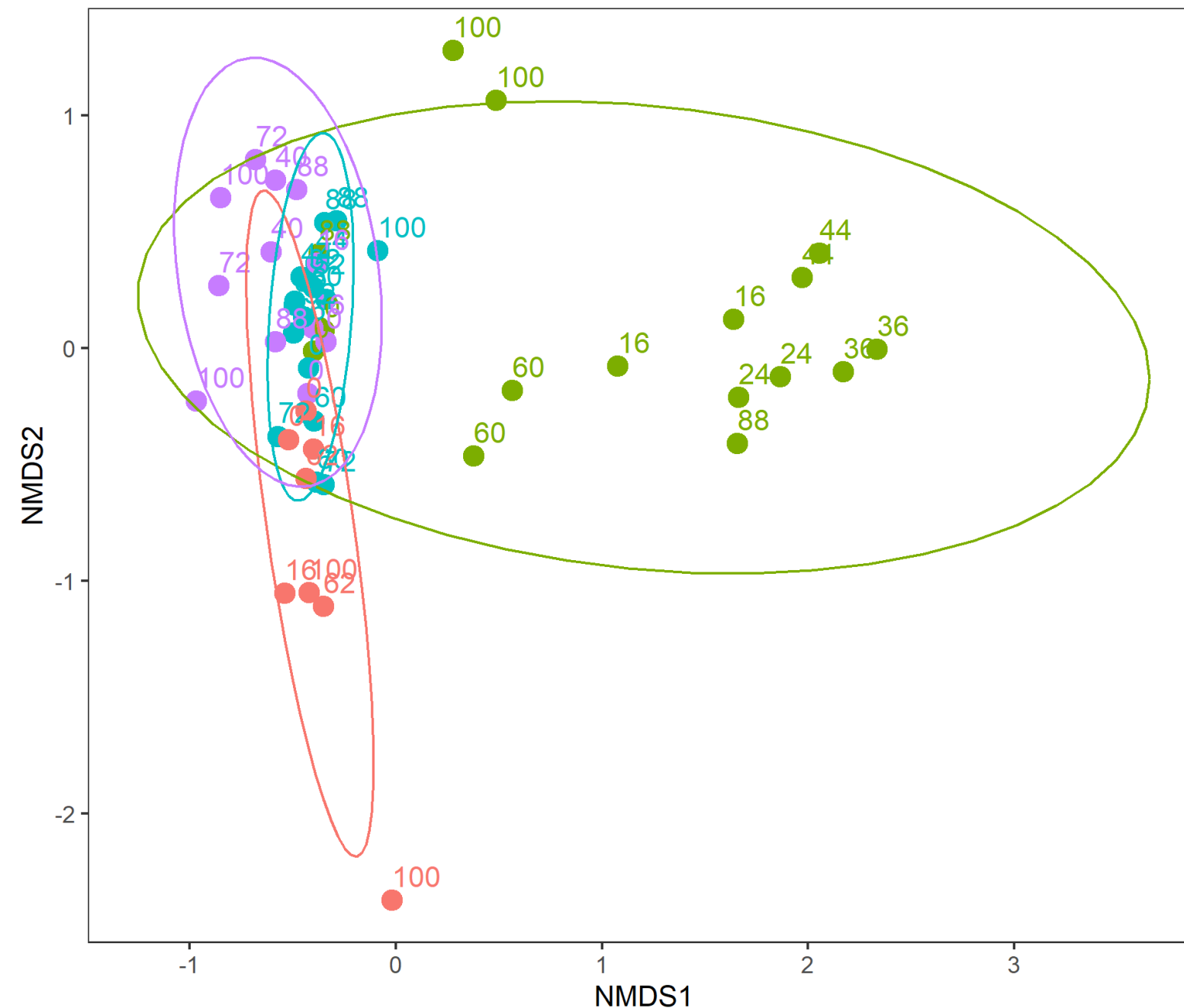
Assembly processes were evaluated using multivariate analyses, null modeling, and neutral modeling

- Completed for whole communities and subsets identified by:
 - Multivariate analyses
 - Metabolism groups from MiDAS 4 database
 - Specialists and generalists (Levins' niche breadth index)

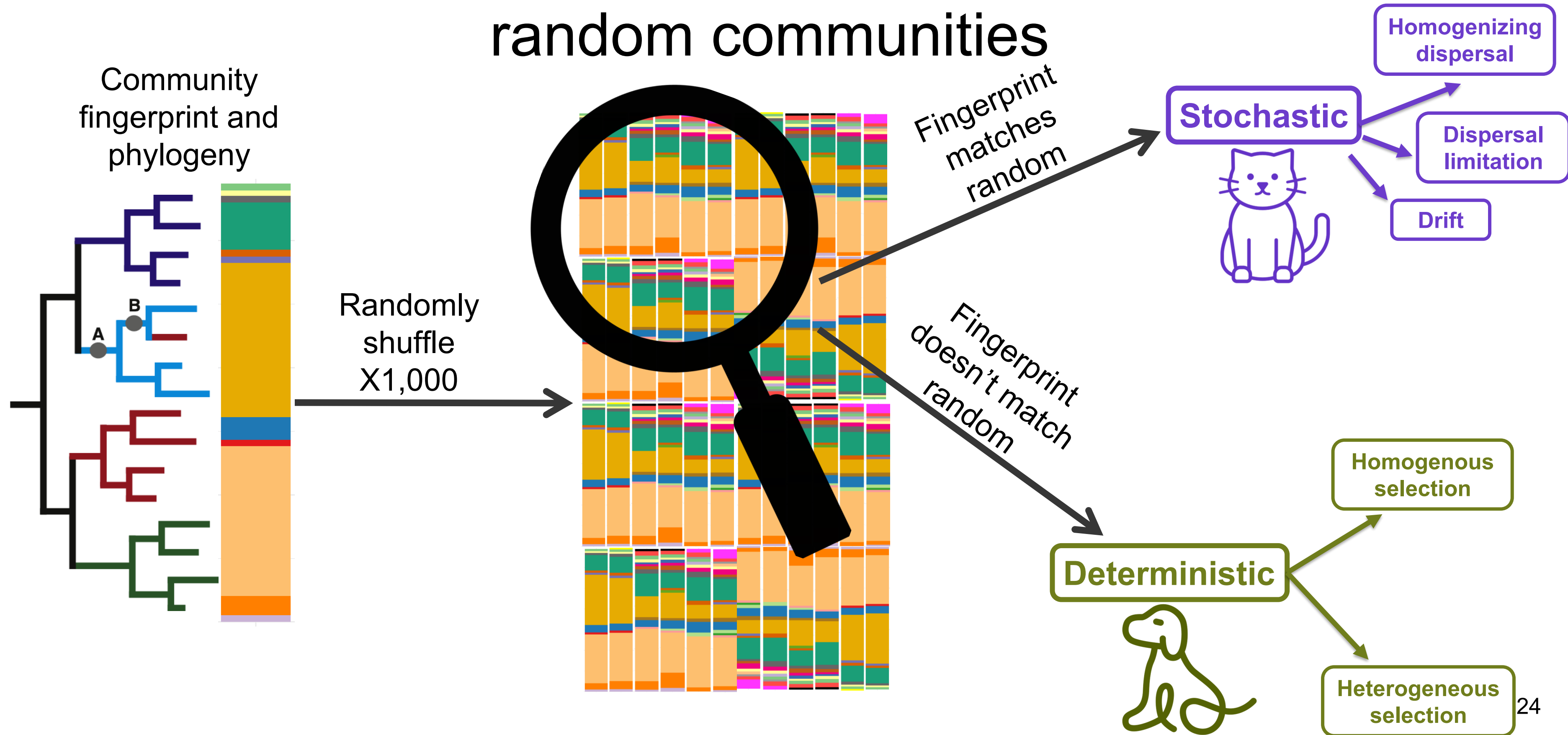


1. Multivariate analyses help us identify relationships and patterns

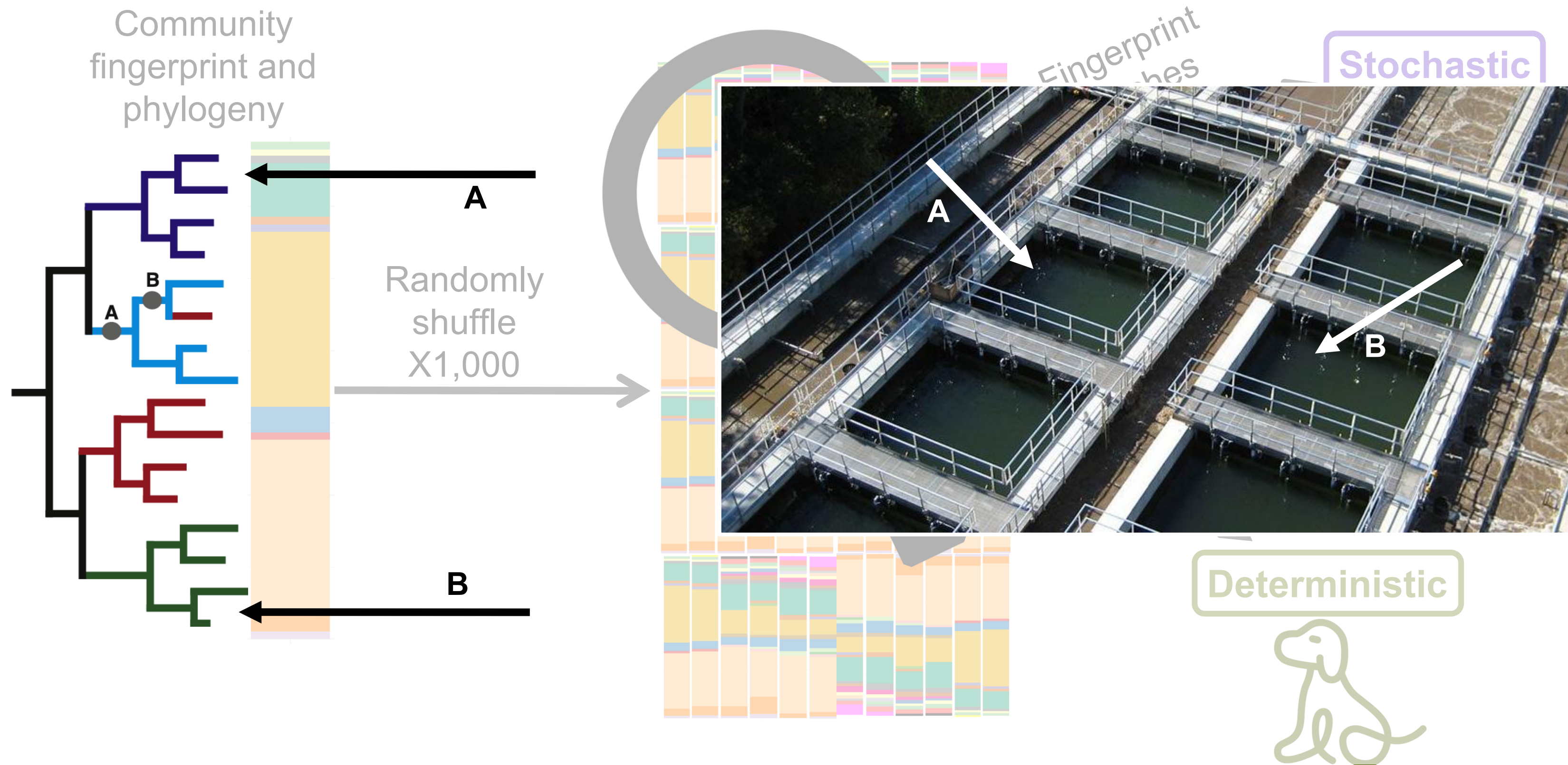
- Summarize high-dimensional data and examine relationships between large sets of variables
- Direct comparison examines microbial community structure differences between and within groups



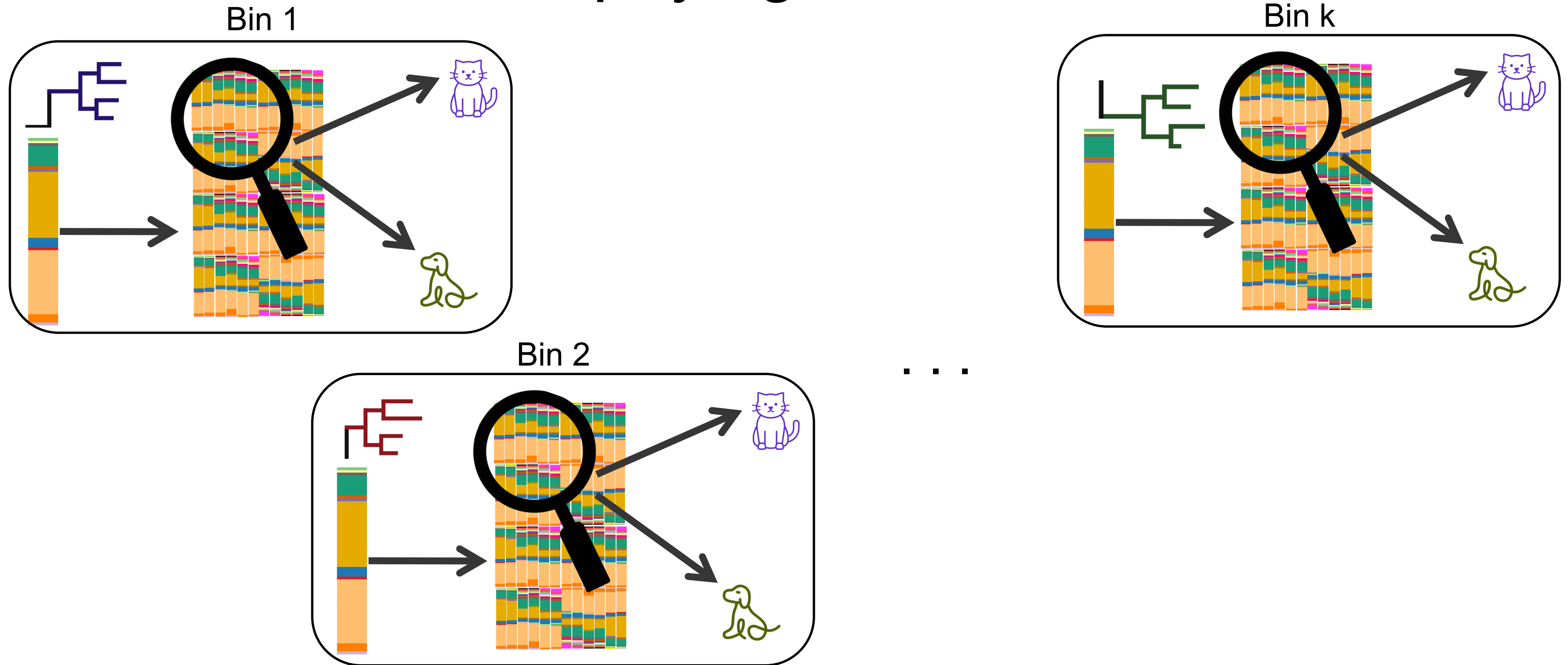
2. Null modeling compares our real community to 1,000 random communities



2. Null modeling compares our real community to 1,000 random communities

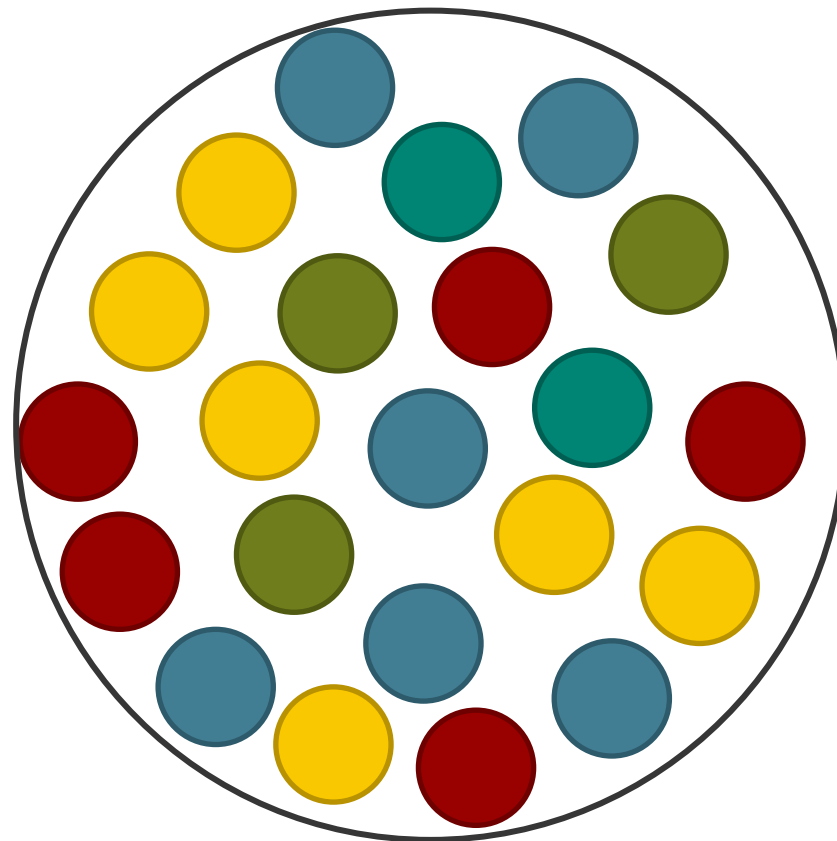


Bin-based null models first subset communities into phylogenetic bins

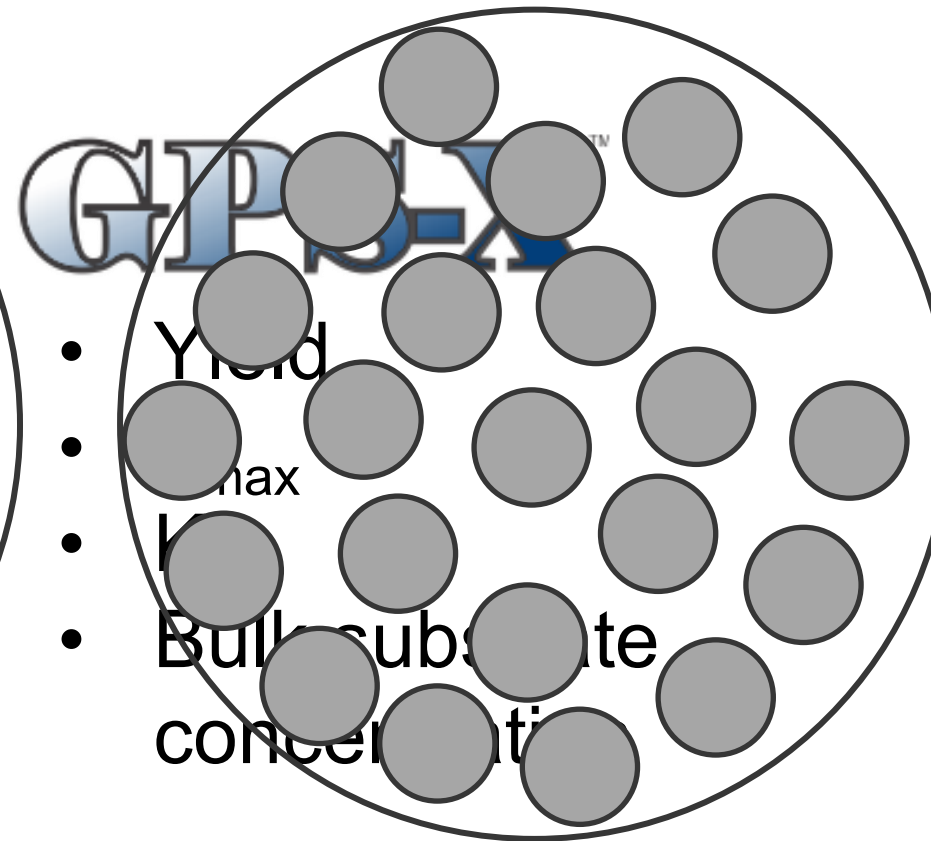


3. Neutral modeling assumes individual bugs are equivalent

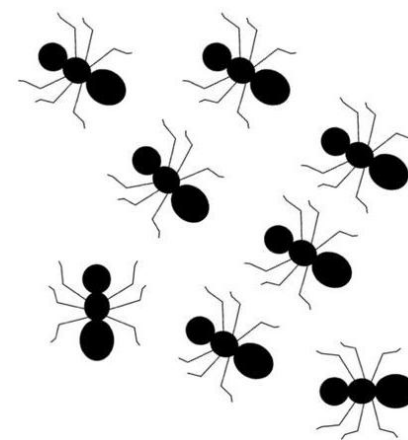
In process modeling,
individuality matters



In neutral modeling,
there is no individuality



- Yield
- K_{max}
- K
- Bulk substrate concentration



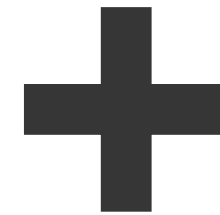
Important parameters:

- Growth rate
- Death rate
- Immigration probability

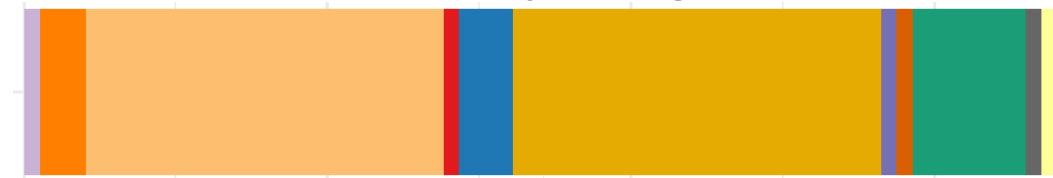
Important parameters + equivalent bugs are put into a differential equation

Important parameters:

- Growth rate
- Death rate
- Immigration probability



Community fingerprint



Stochastic differential equation

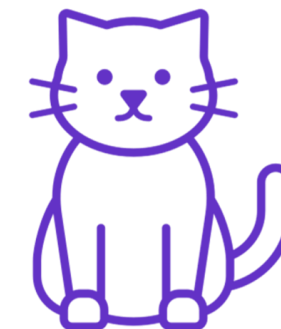
Deterministic



Doesn't match
random

Matches
random

Stochastic



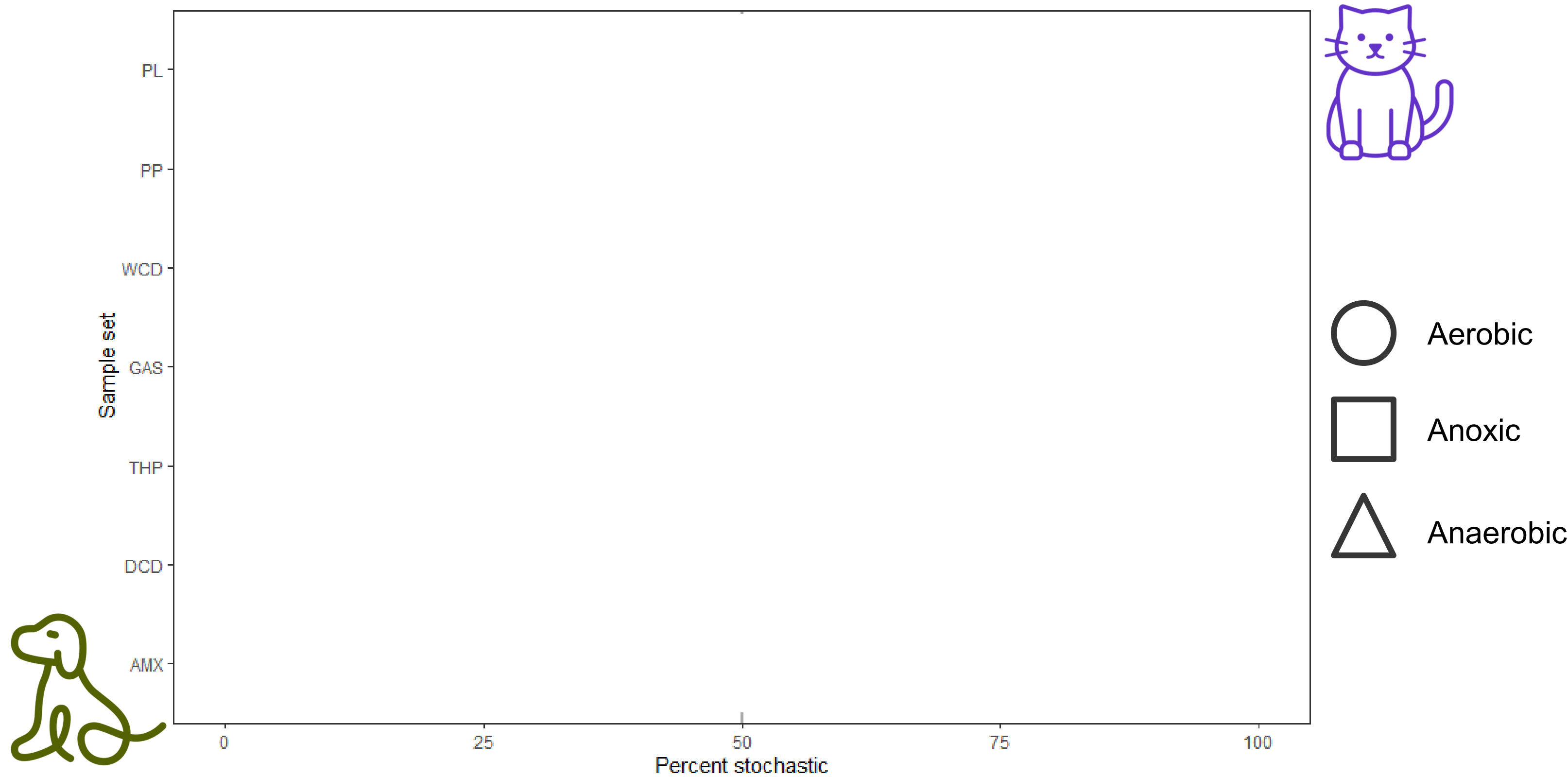
Building on lessons we learned from our previous work

This work:

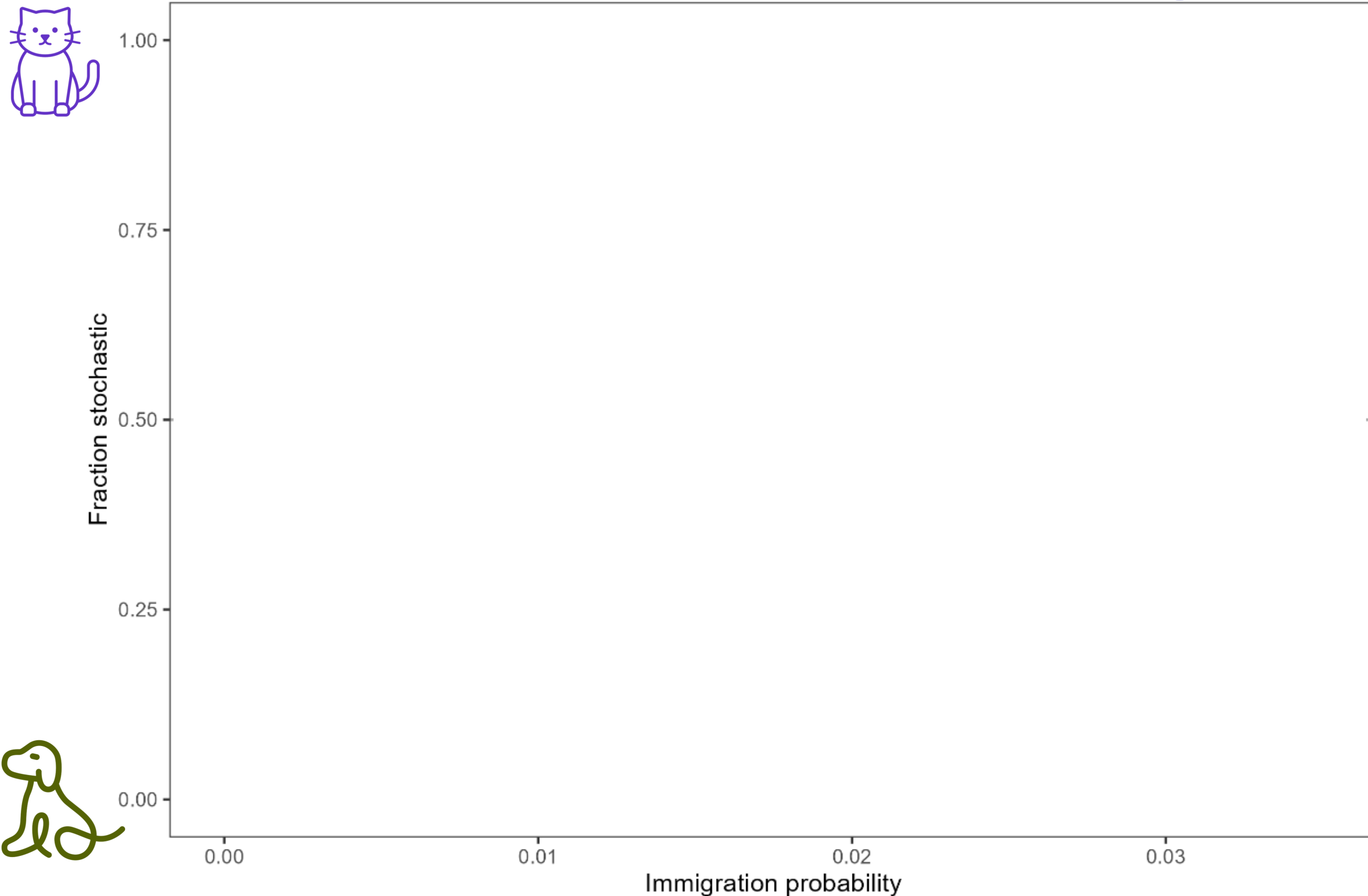
- Proposes and employs a community-wide tool set
- Assess multiple bioreactor systems across a range of process types and scales
- Connects MCA patterns to relevant microbial groups in each system

Seven sample sets were used

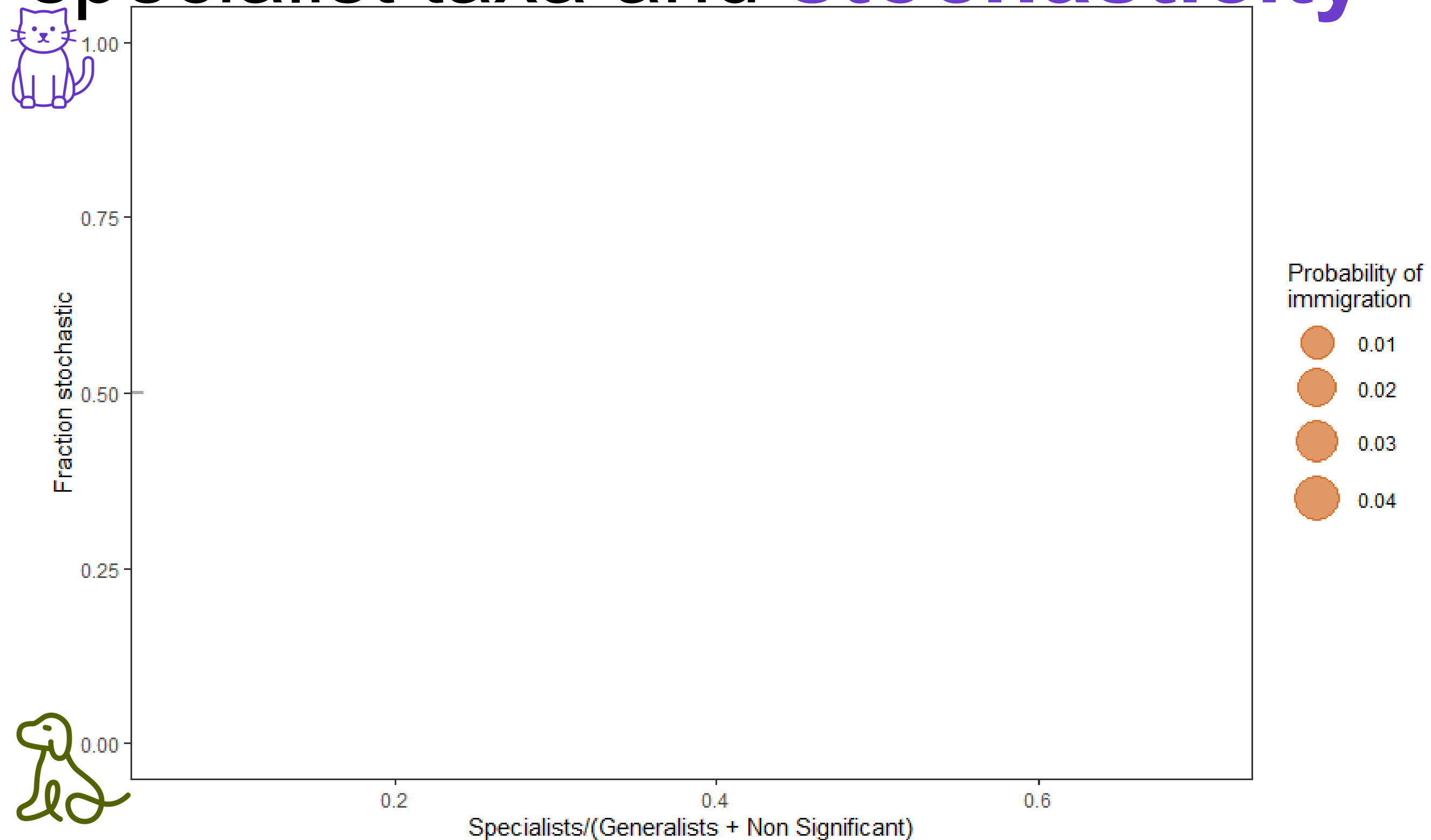
Neutral modeling indicated pit latrines were the most **stochastic** while anammox samples were the most **deterministic**



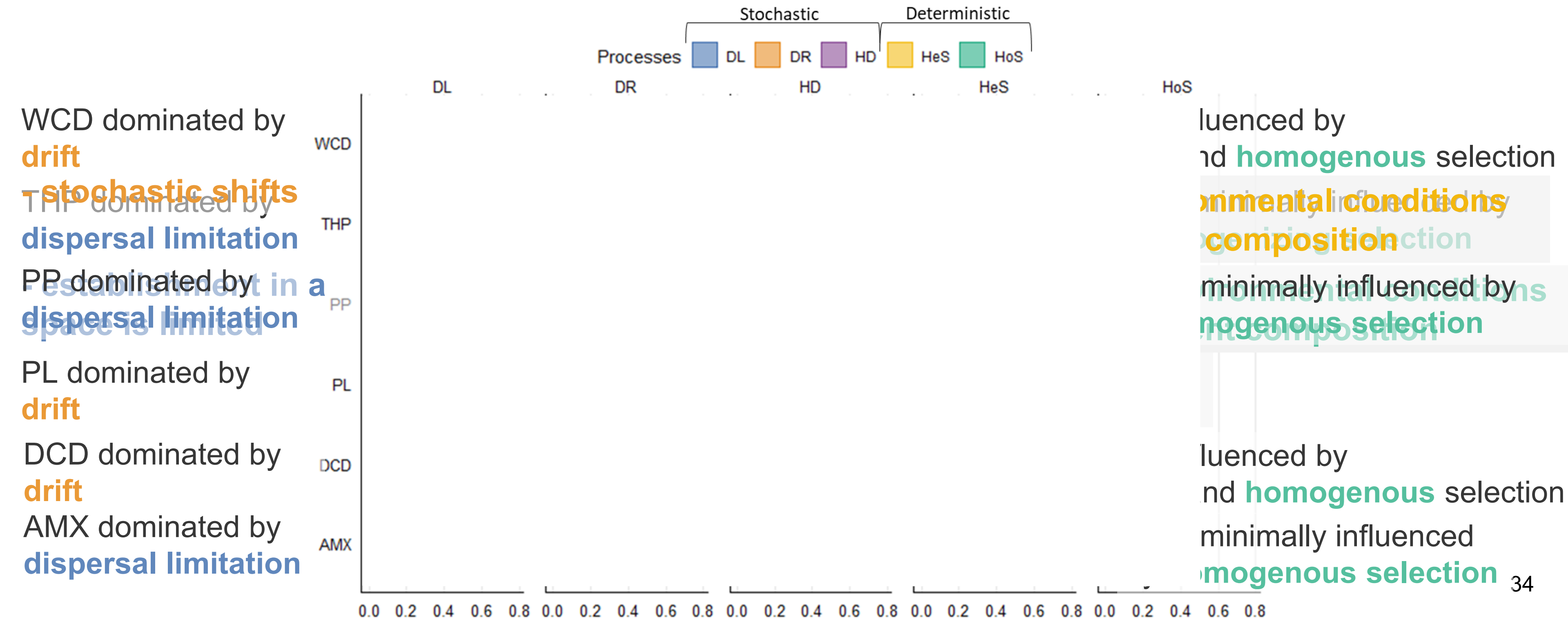
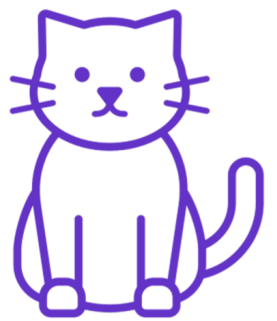
Relationship between immigration and stochasticity is greater for **specialists** than for **generalists**



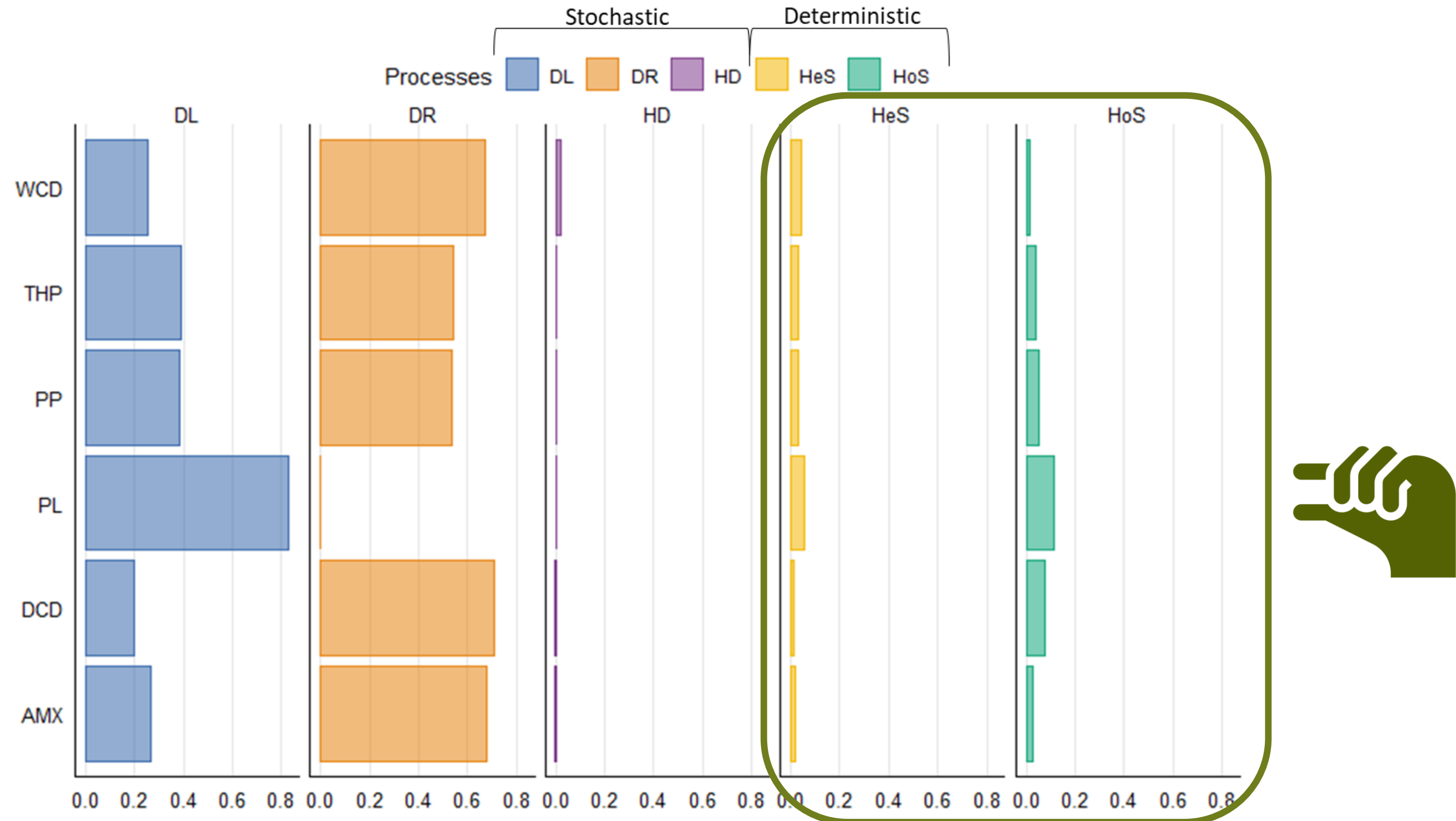
Positive relationship between number of specialist taxa and **stochasticity**



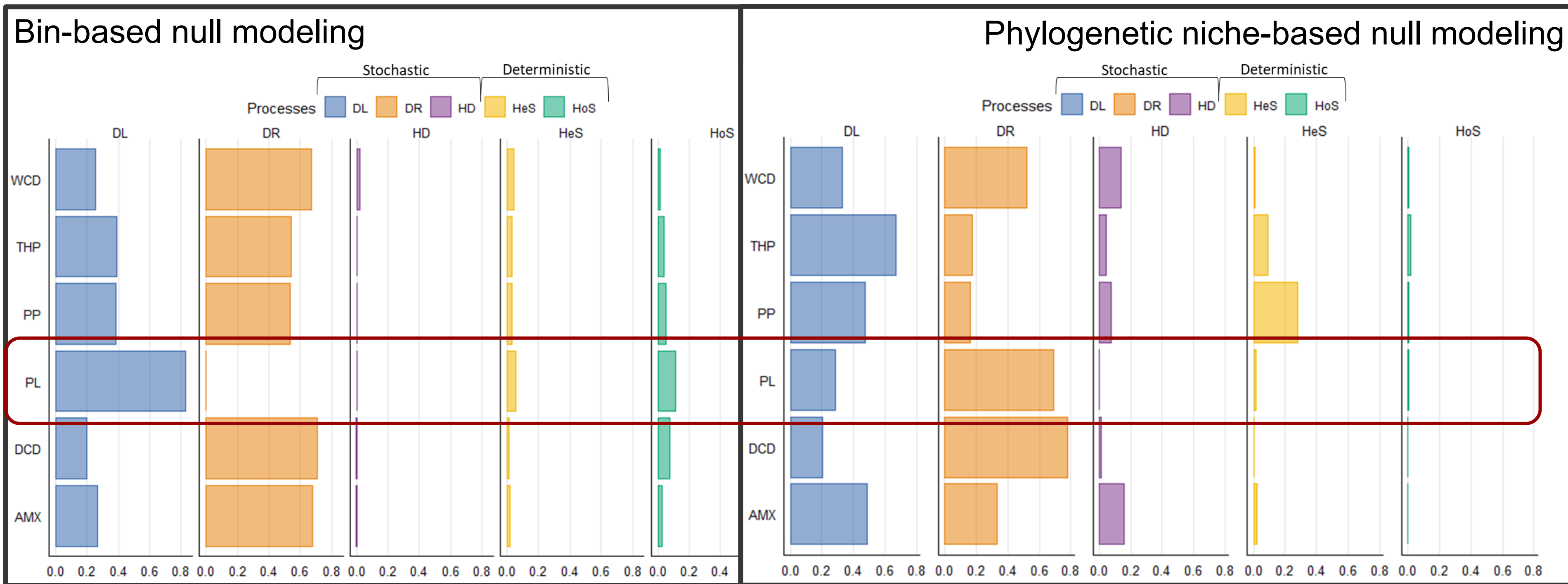
Null modeling indicated all experiments dominated by **stochastic** processes



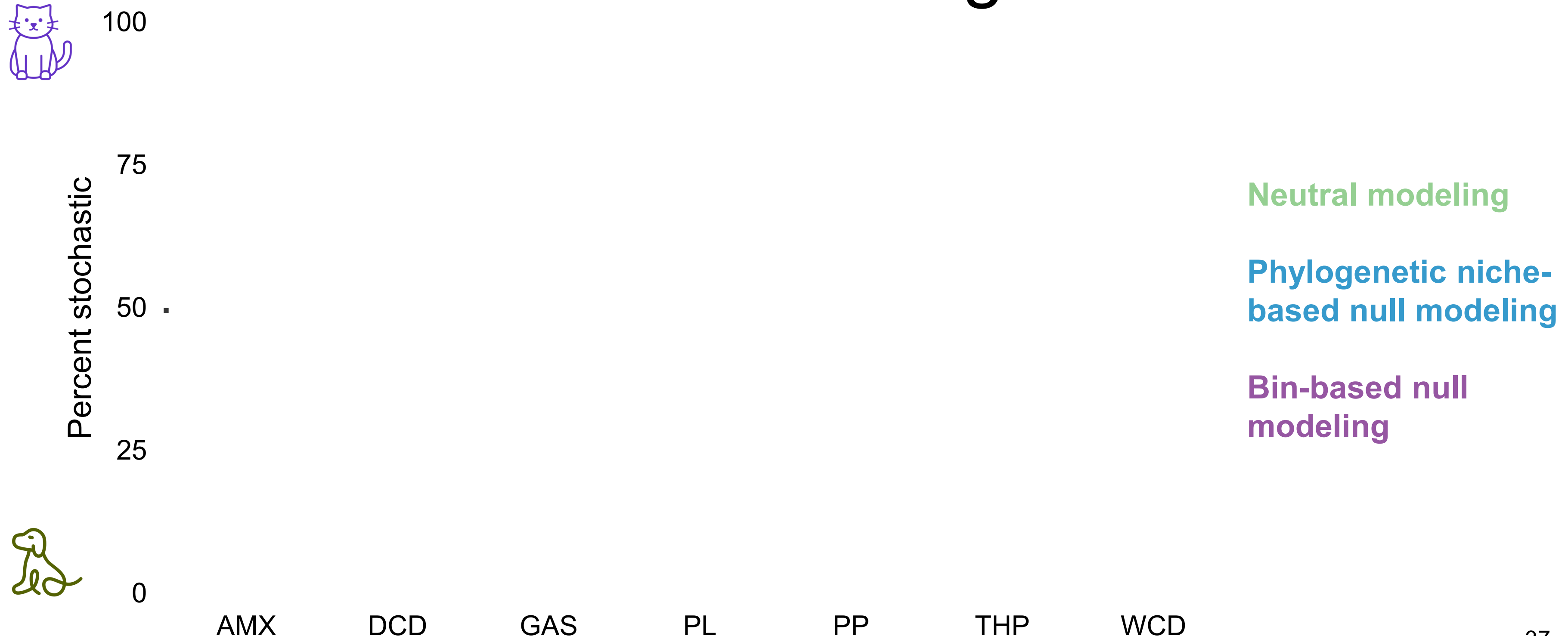
Bin-based null modeling indicated almost no influence by **deterministic** processes



MCA processes varied between null modeling approaches



MCA processes varied between null and neutral modeling



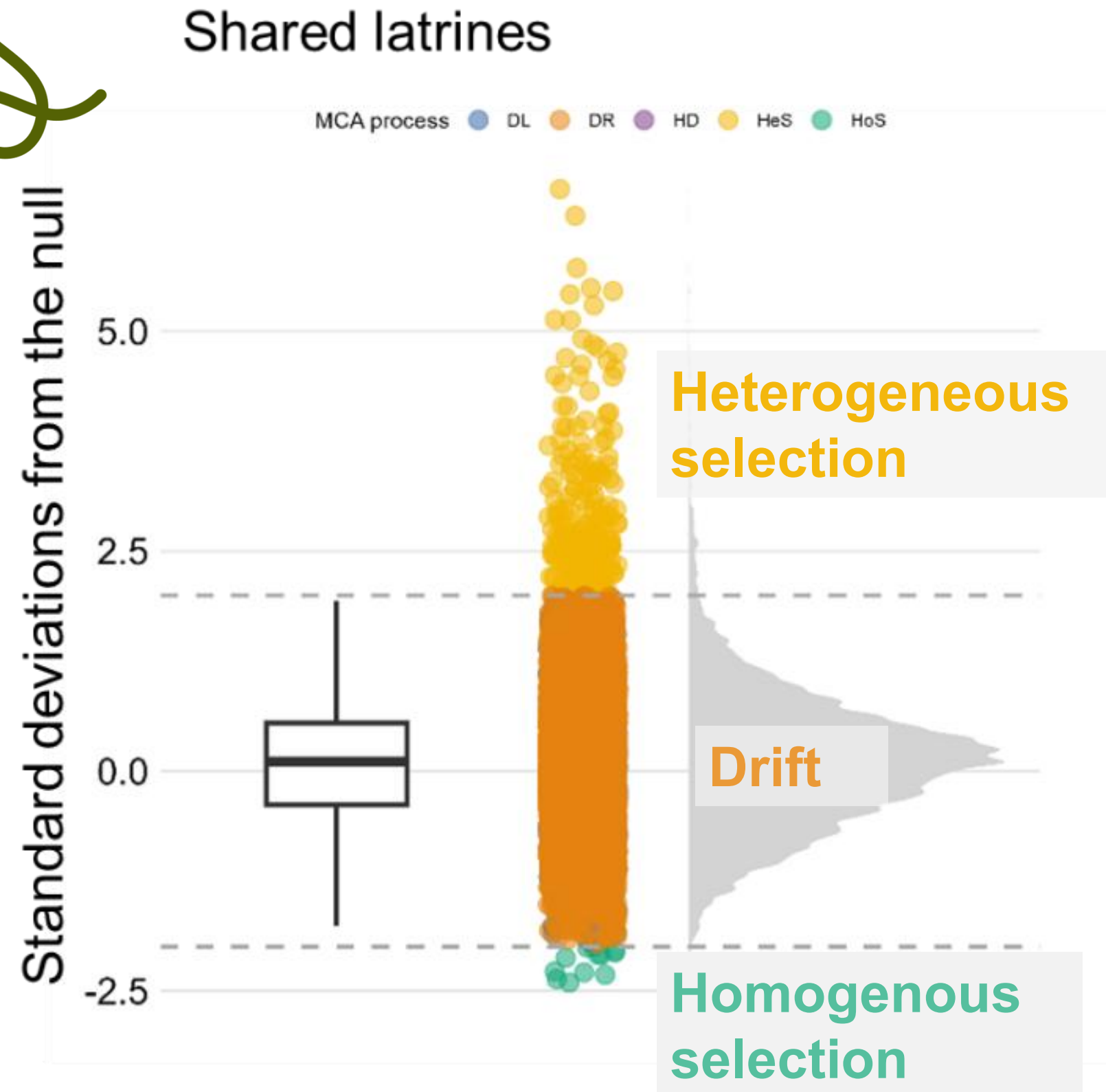
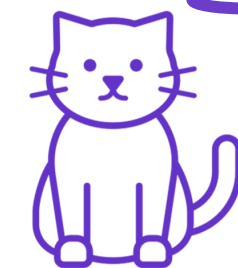
Strategic subsetting revealed
differences in MCA patterns

Assembly patterns in pit latrines varied by latrine use



Deterministic

Stochastic

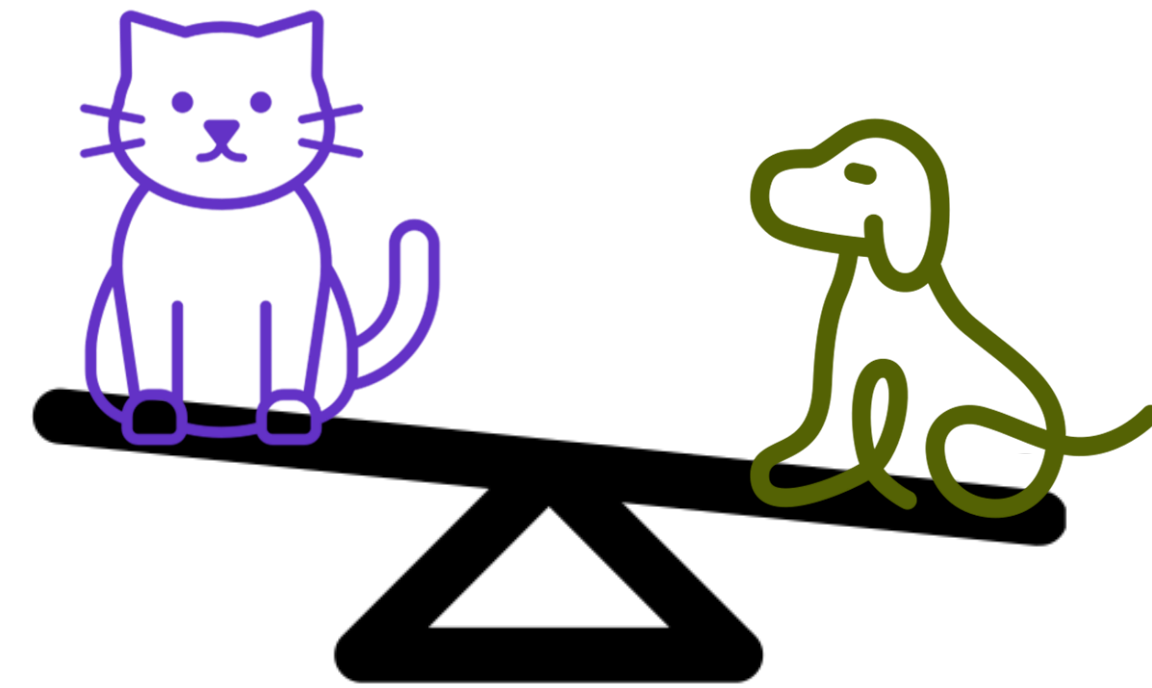


Dispersal limitation

Drift

Community assembly processes vary with bioreactor characteristics, to what extent and in which situations remains to be known

- MCA patterns varied by the specific MCA method applied
 - Null modeling assumes phylogenetic conservatism
 - Neutral modeling assumes taxa individuality
- Most MCA patterns couldn't be explained by general rules
 - Anaerobic generalists assembled more deterministically than specialists
 - No clear patterns by electron acceptor conditions



Caution should be used when applying MCA methods

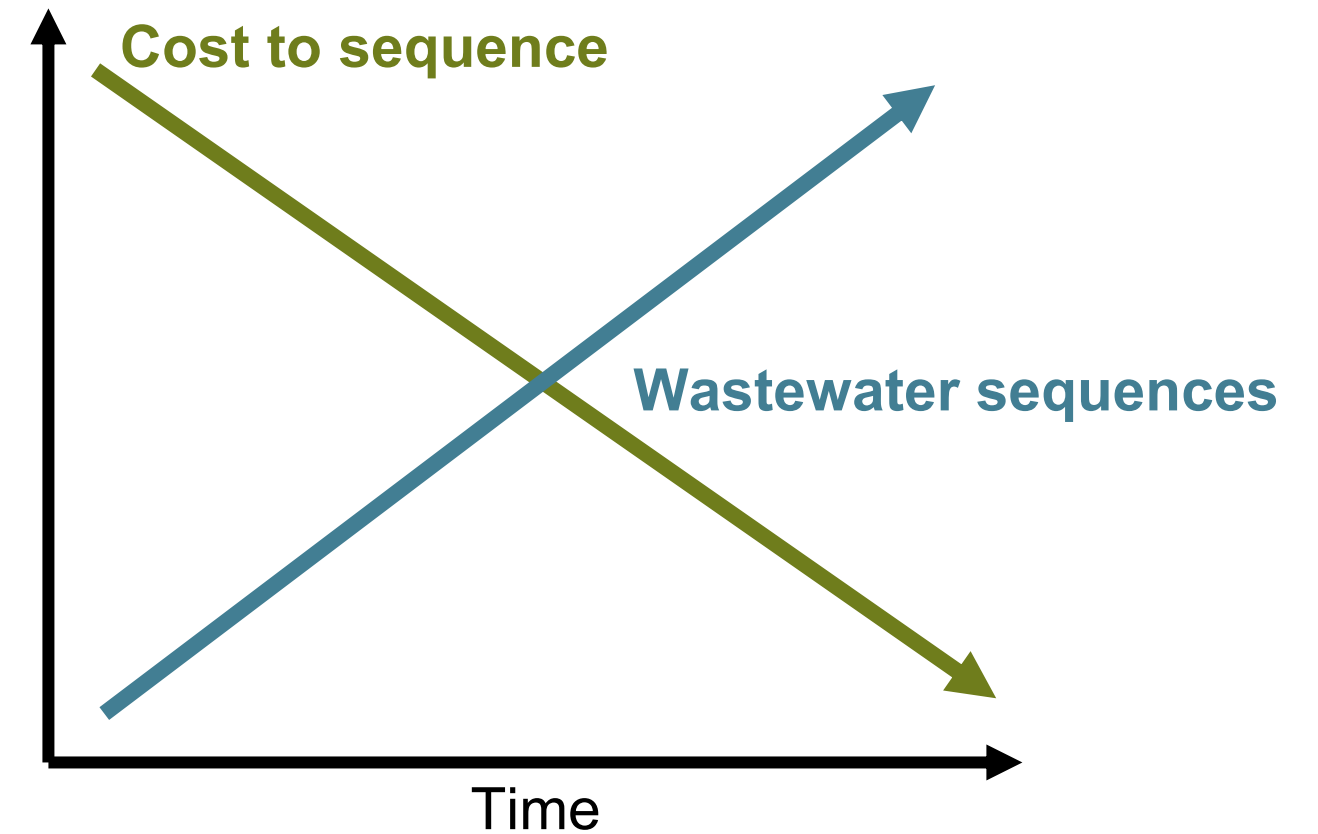


Why combine MCA and ML?

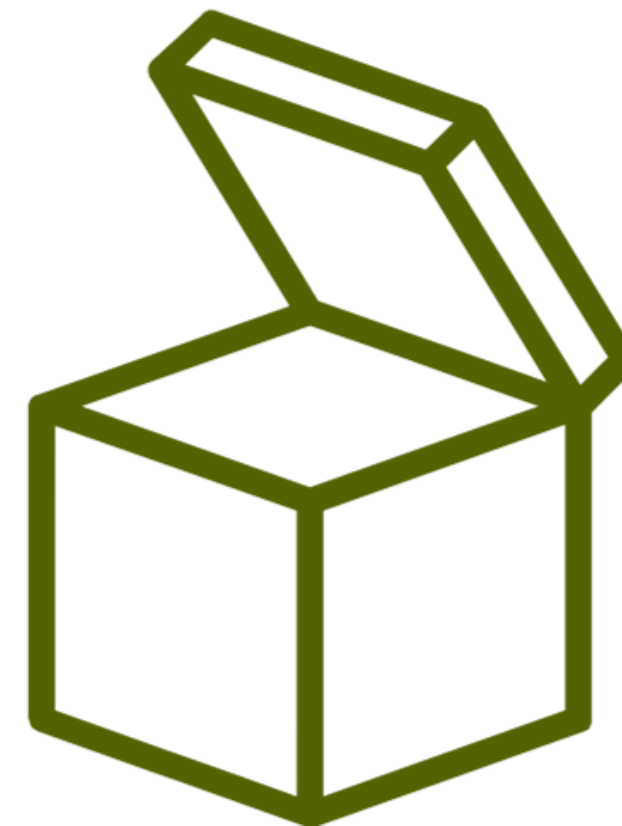
→ Unclear connections between microbial communities and process performance

Microbial communities are really complex

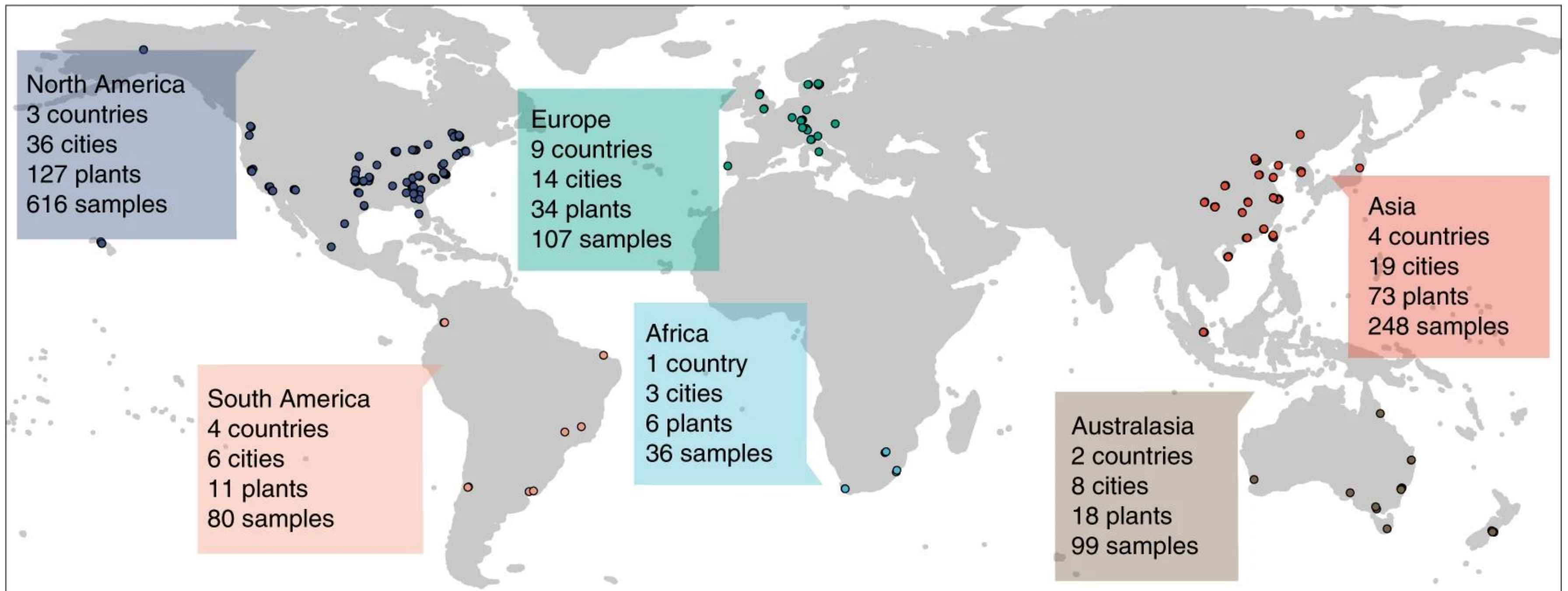
→ Machine learning can help mitigate these challenges



Ecological constraints



~1,200 activated sludge samples from 269 WRRFs were used



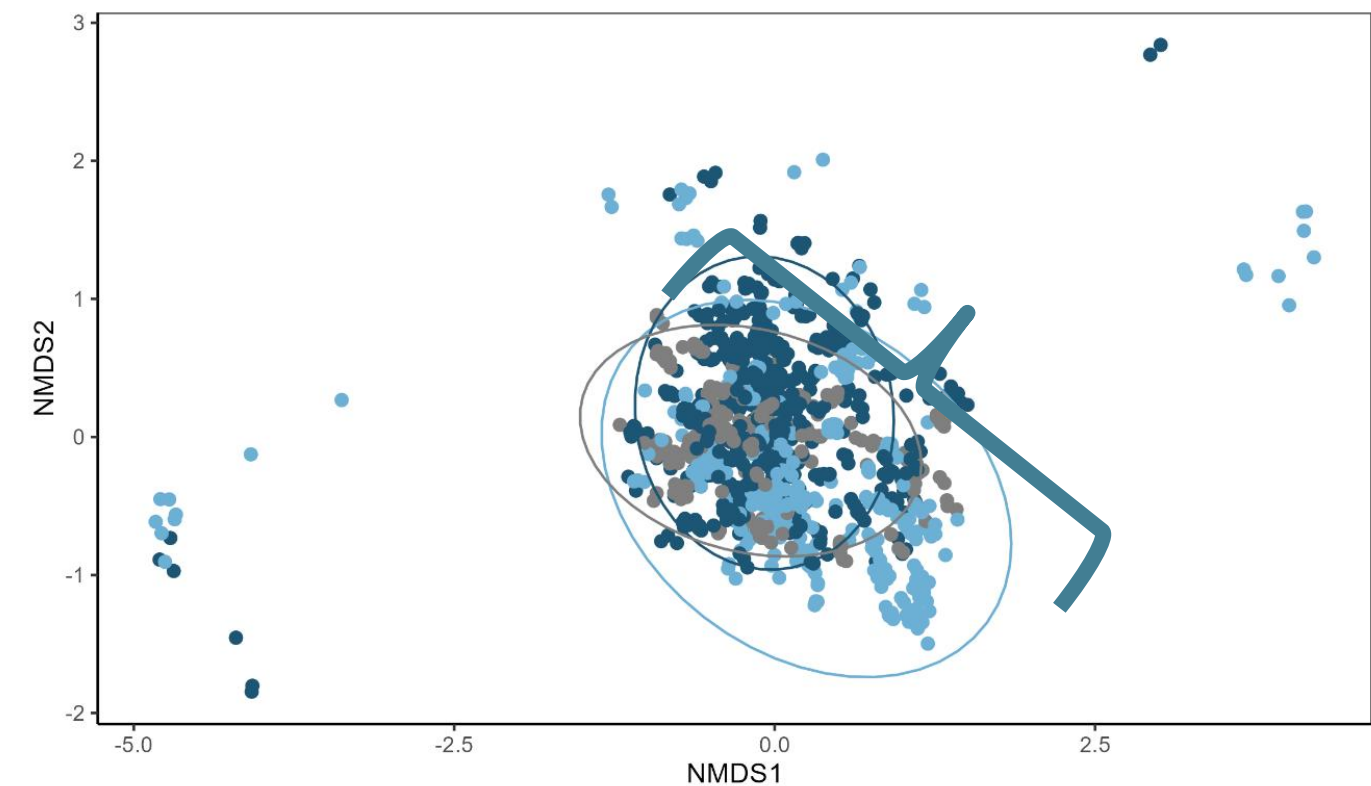
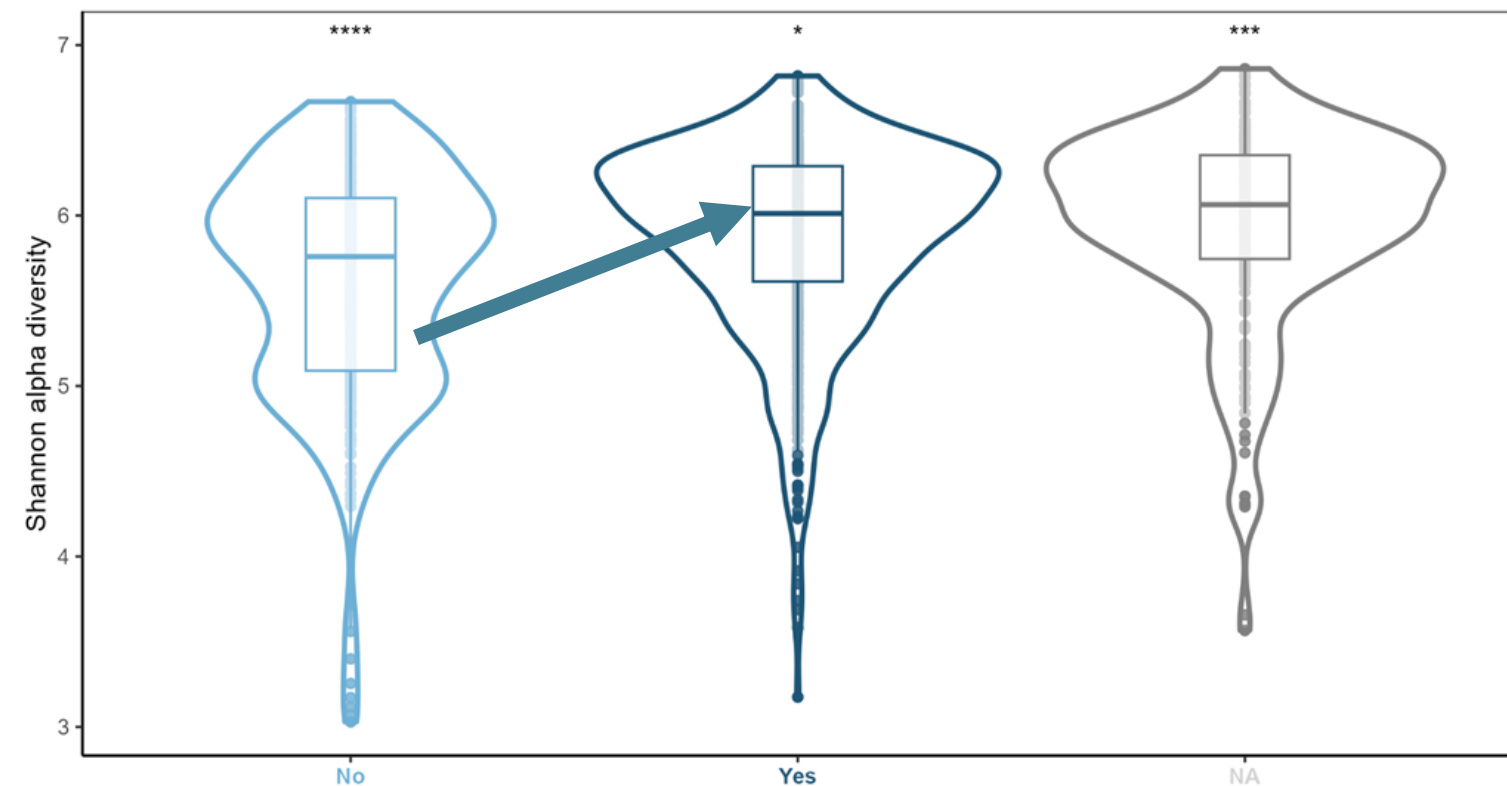
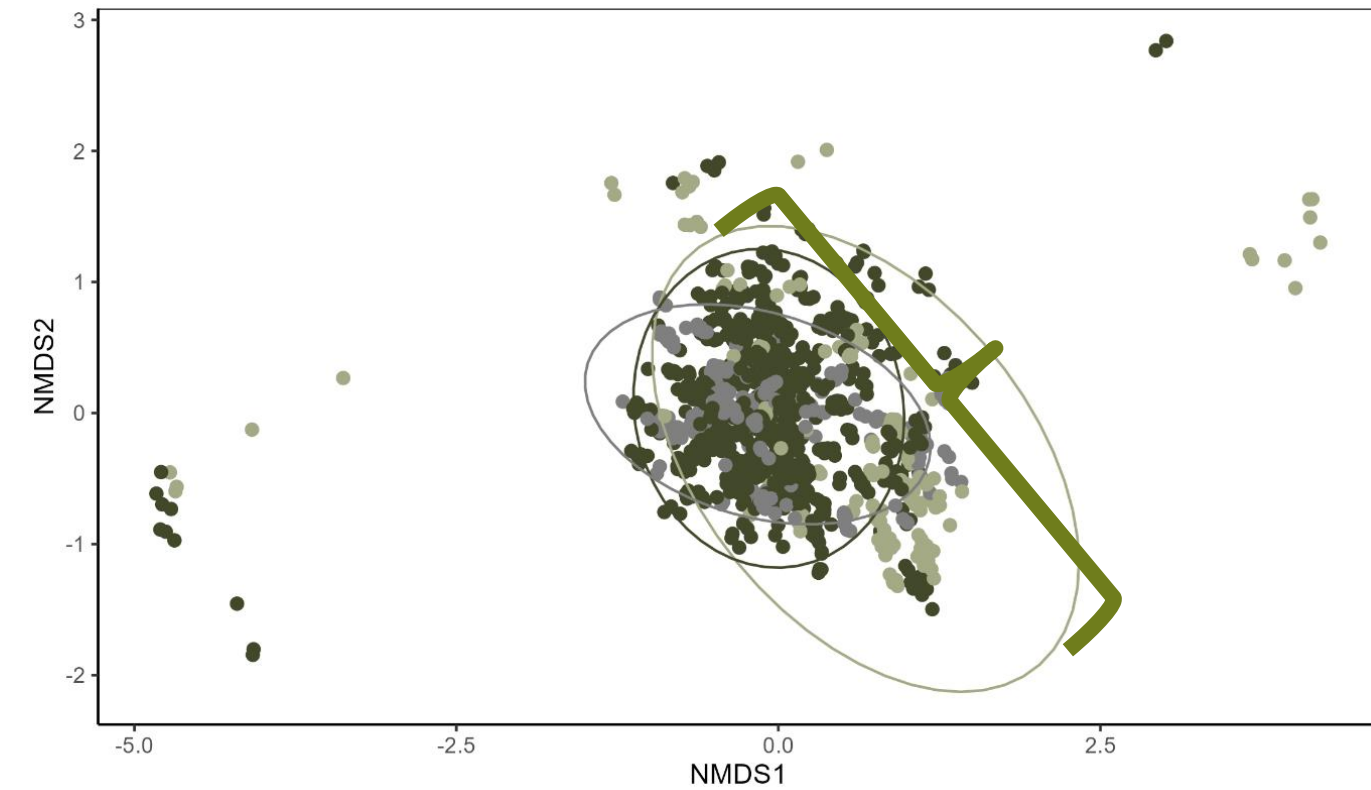
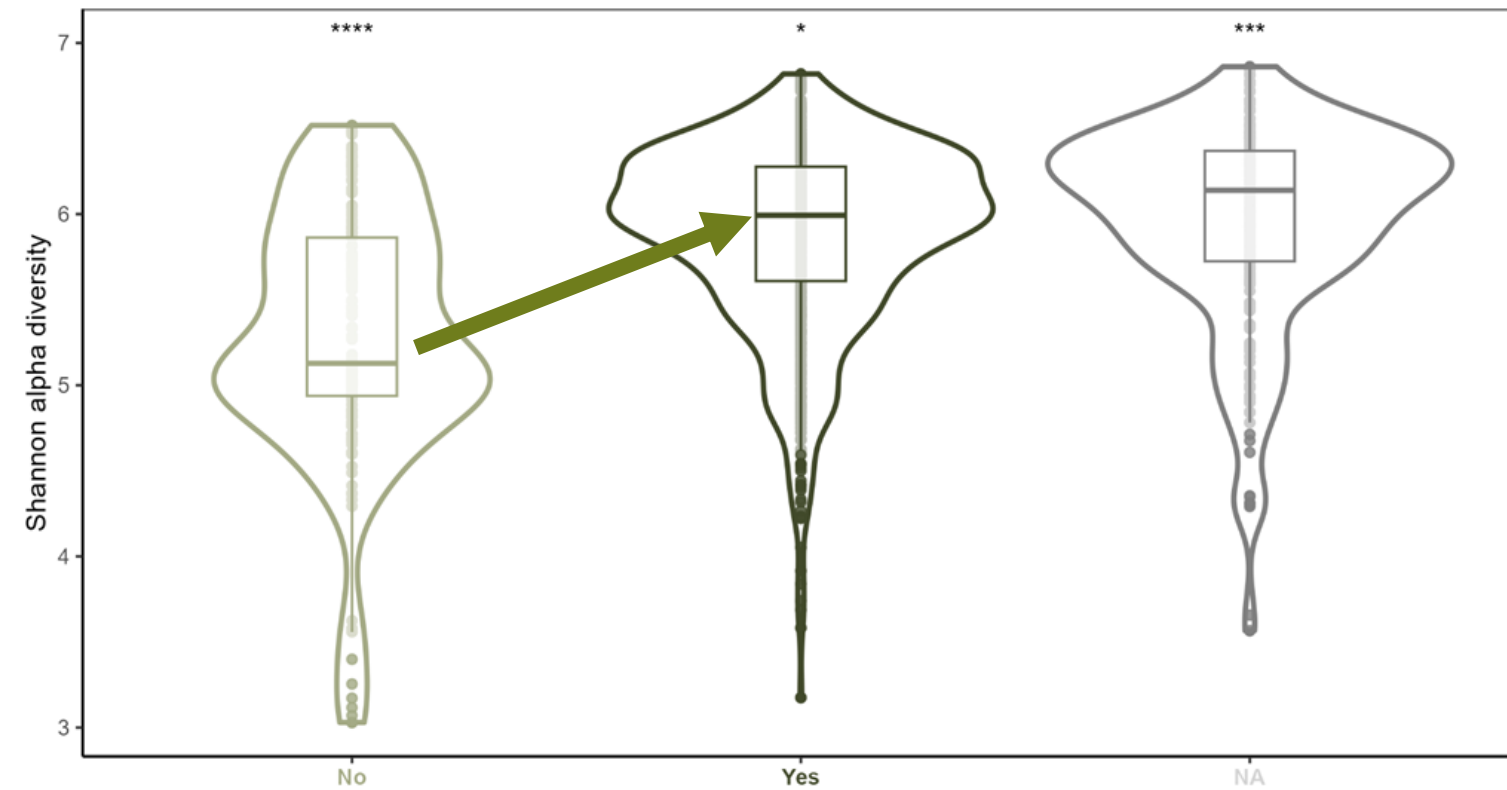
Metadata

Sequencing
data

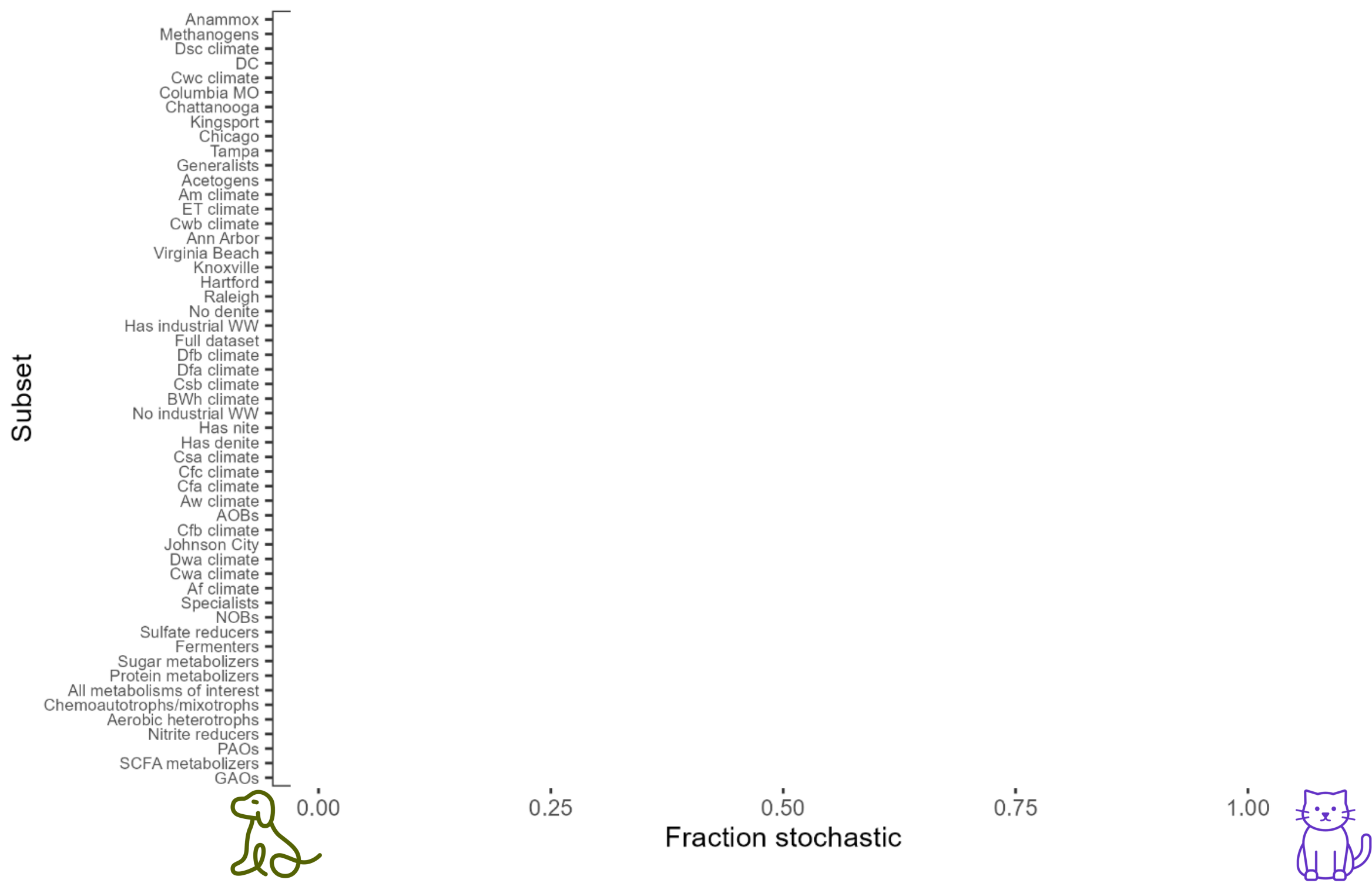


Results

Differences in alpha and beta diversity seen among **nitrification** and **denitrification** in WRRFs

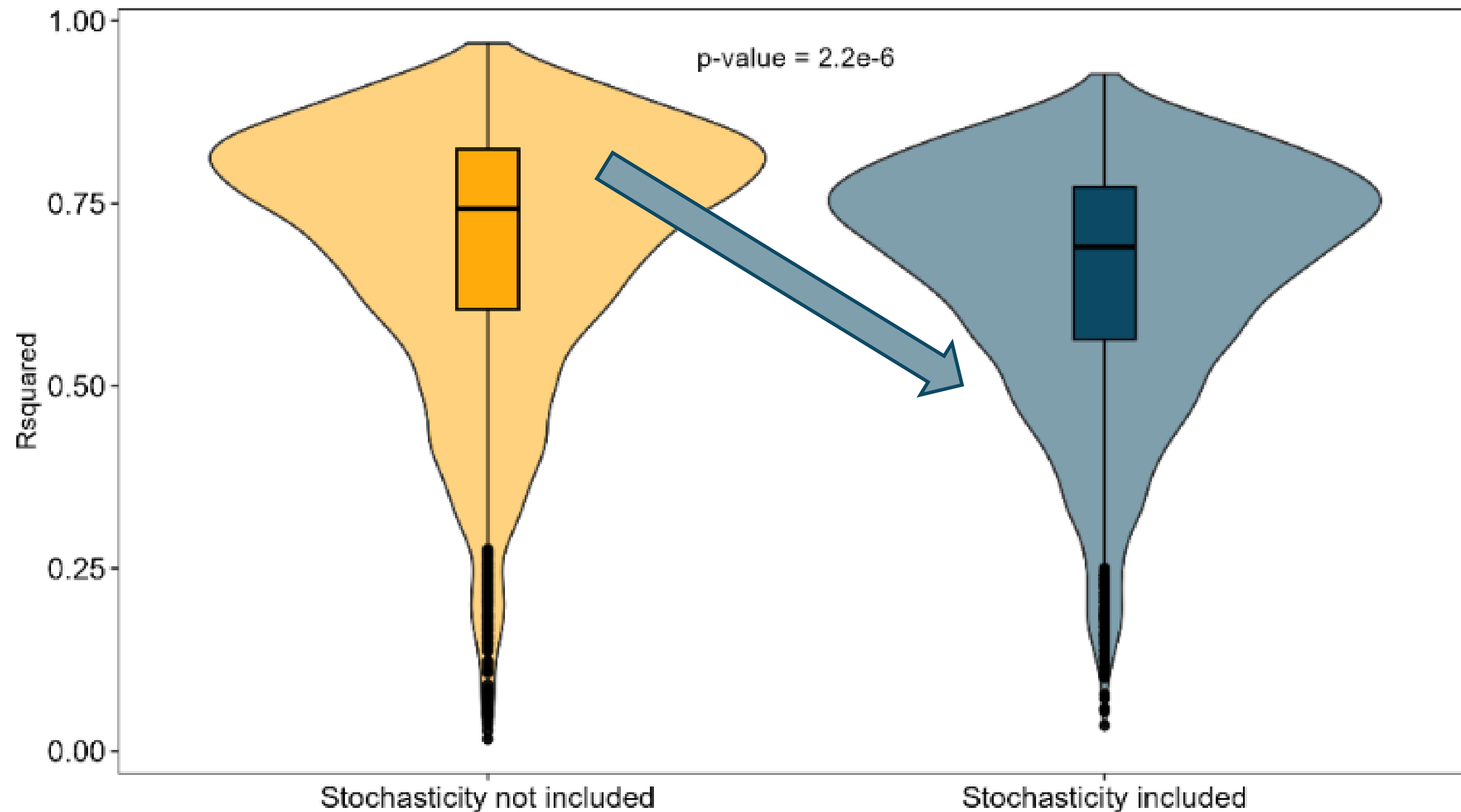


WRRFs assembled 71% **stochastically**, with some groups assembling with varying **stochasticity**

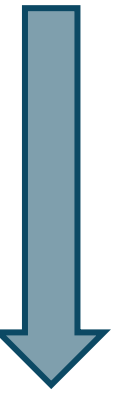


Random forest approach 1

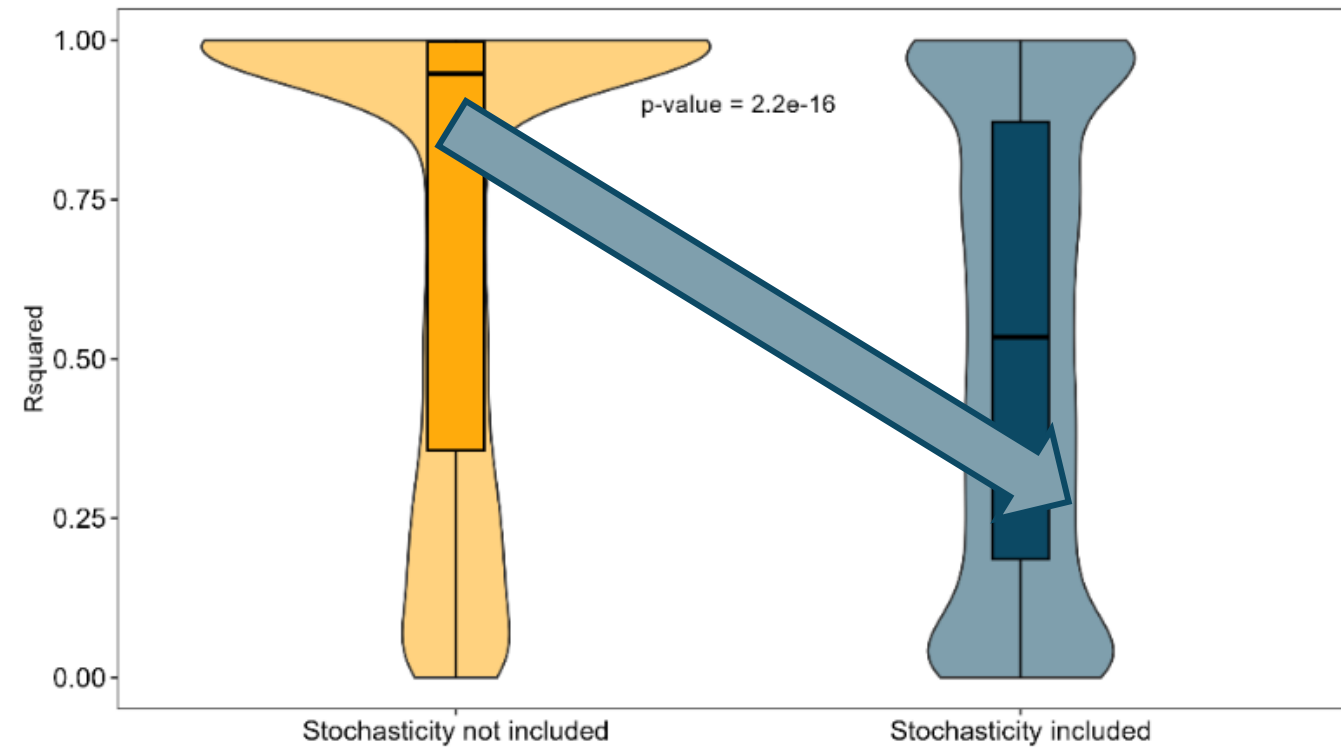
Adding MCA estimate reduced prediction of BOD removal



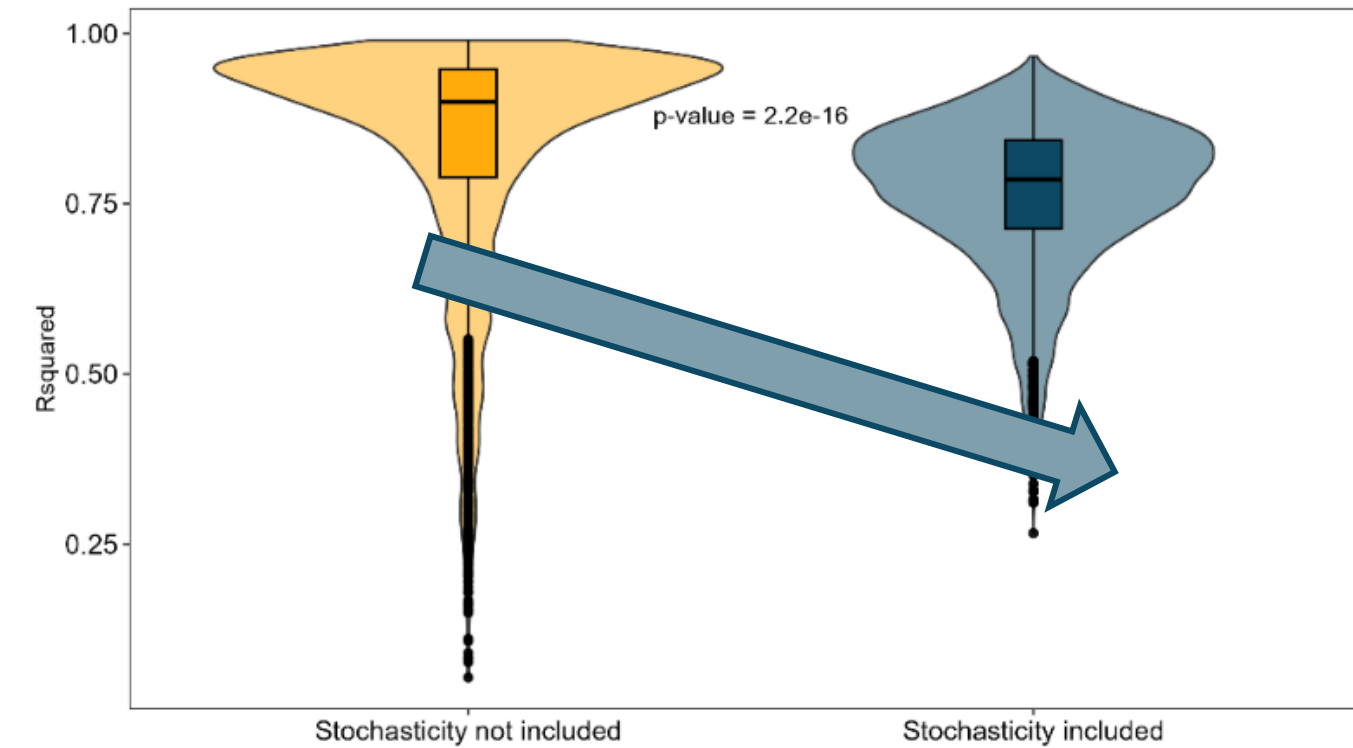
Adding MCA estimate reduced prediction of total N, NH₄-N, and total P removal



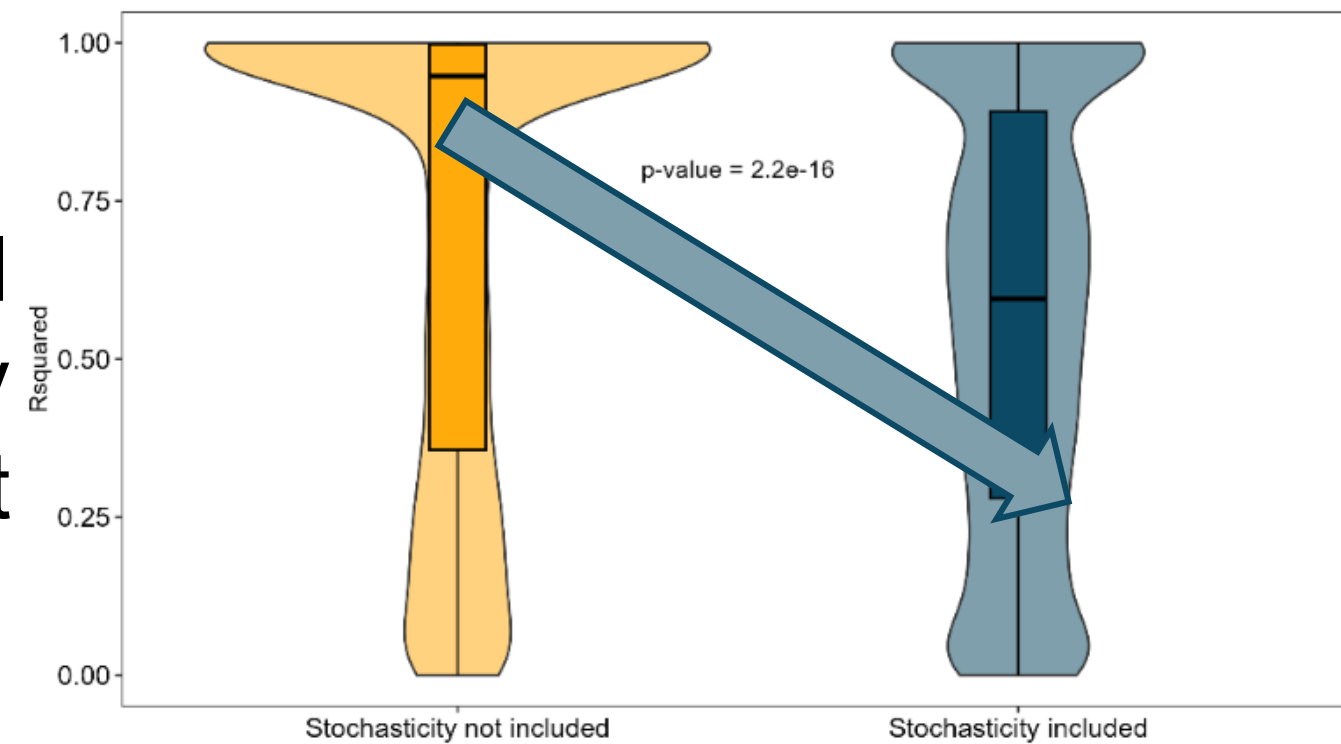
Total N



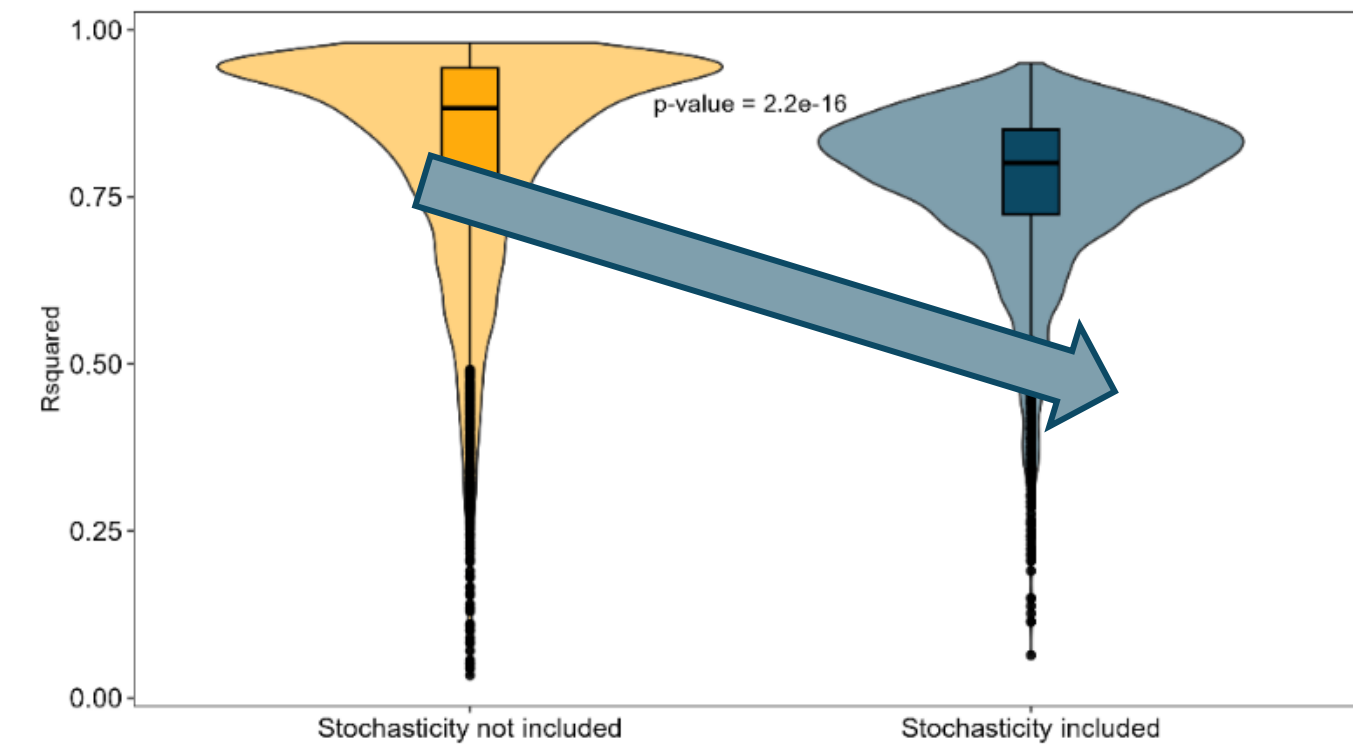
NH₄-N



Total N
community
subset

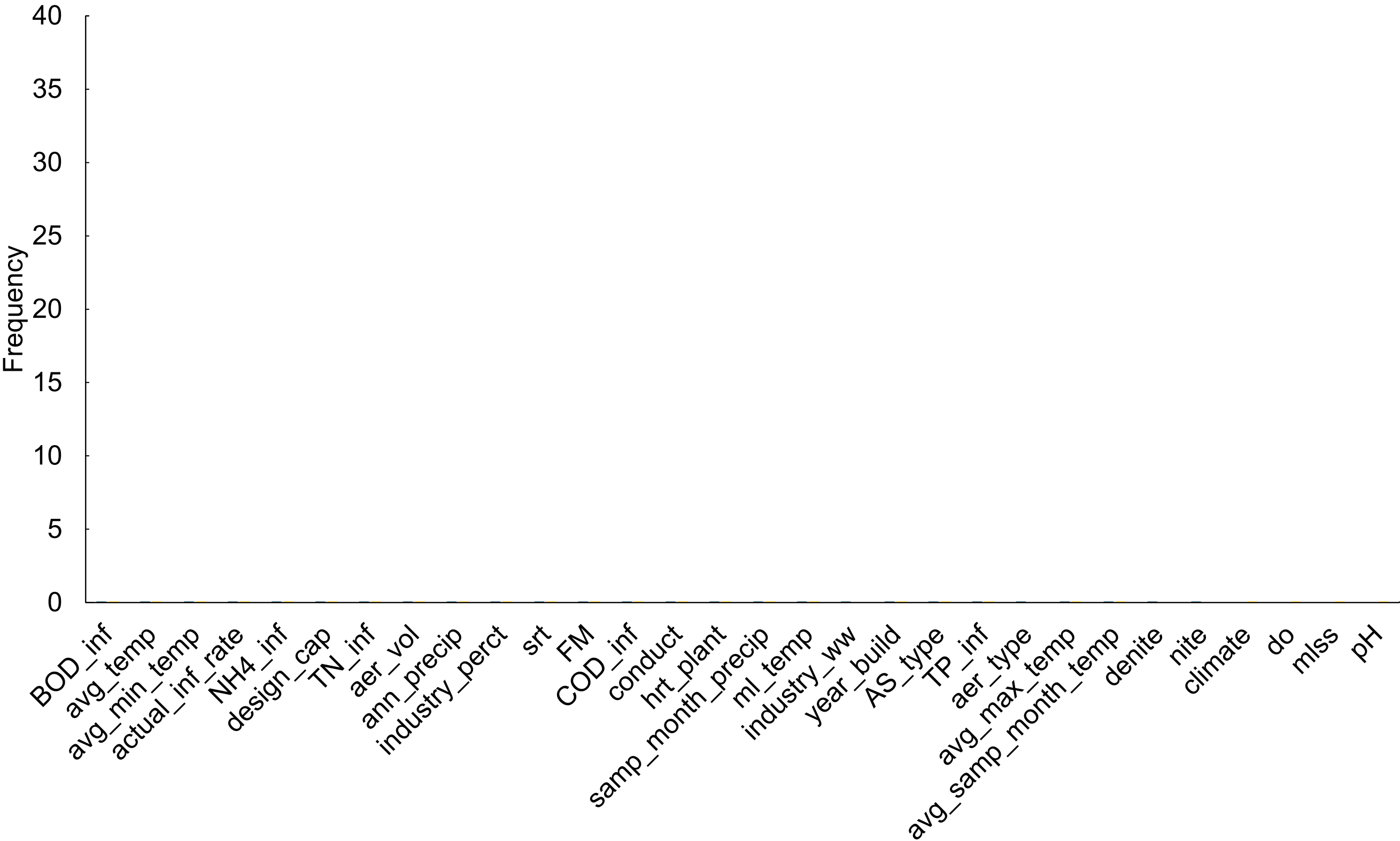


Total P

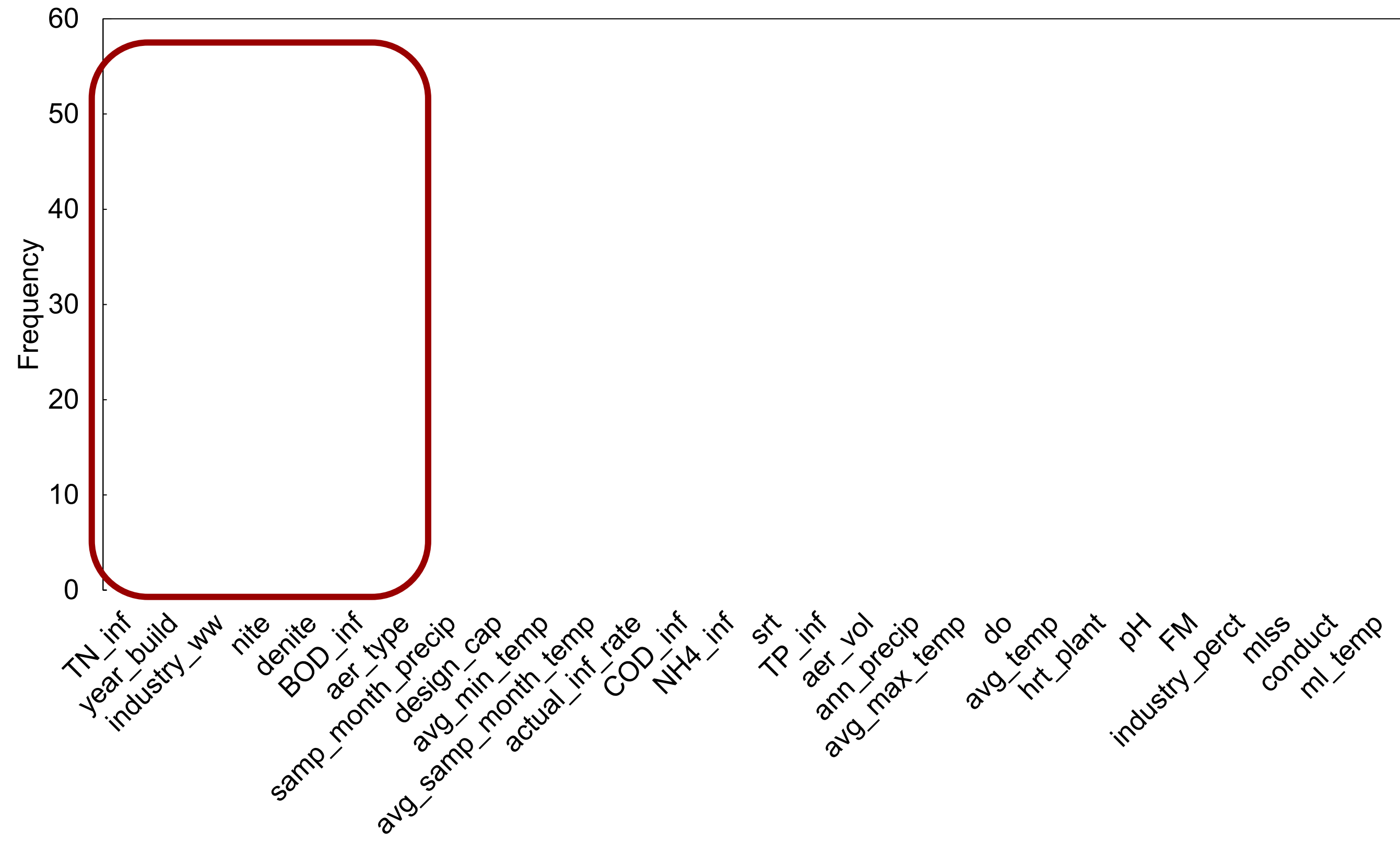


Top models of BOD removal have similar feature frequencies

with and **without** stochasticity

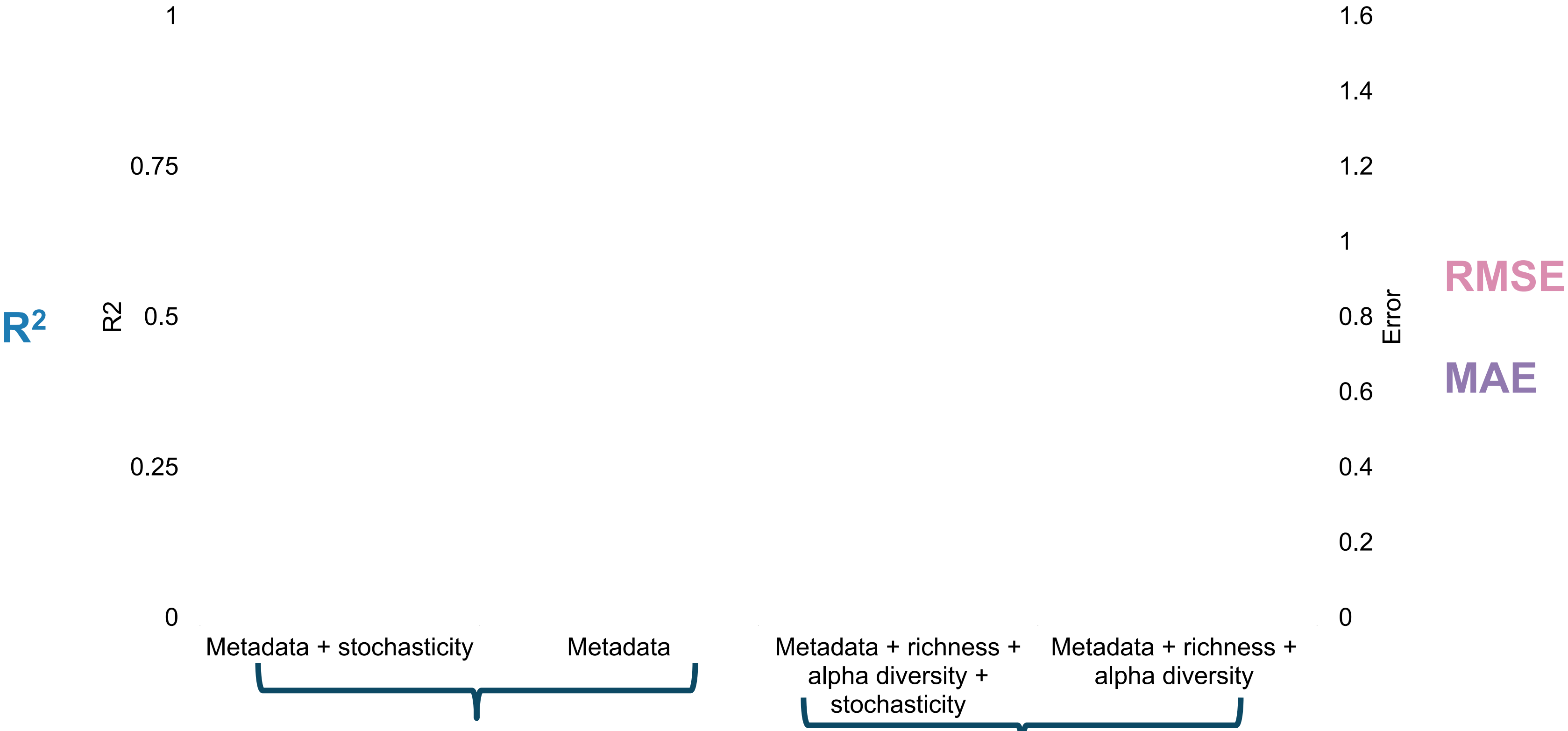


Top models of total N removal have dissimilar feature frequencies **with** and **without** stochasticity

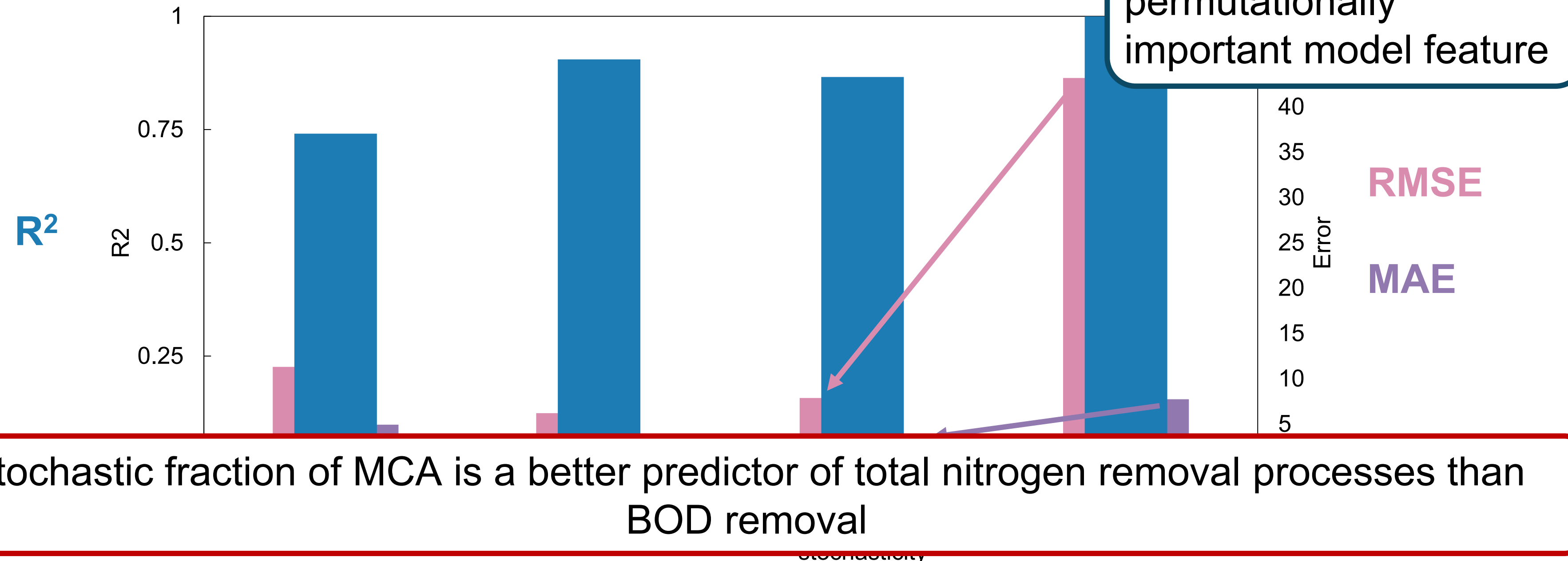


Random forest approach 2

Inclusion of **stochasticity** decreased performance of model predicting BOD removal

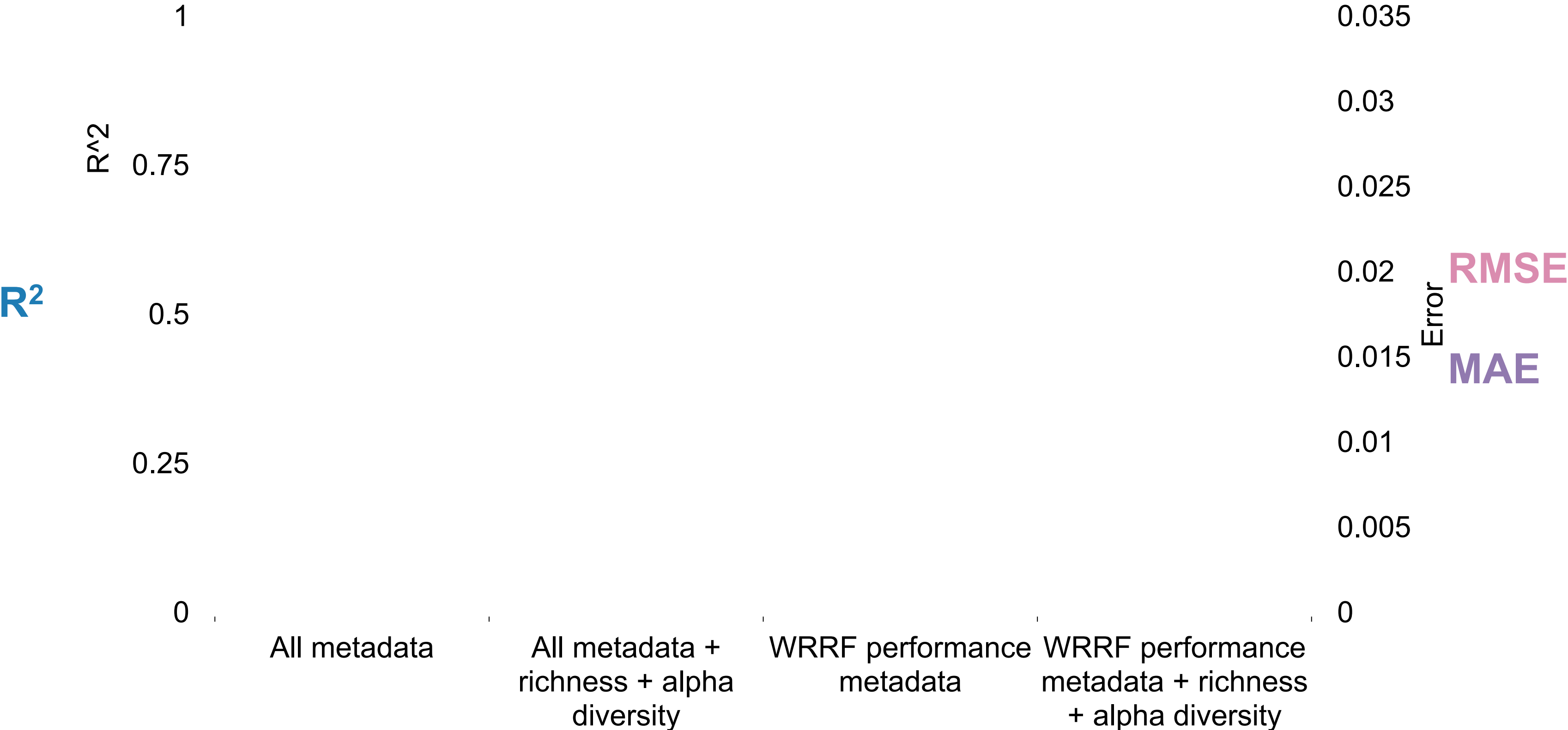


Inclusion of **stochasticity** decreased model errors when predicting total nitrogen removal with richness and alpha diversity



Random forest approach 3:
predicting MCA

Richness and alpha diversity improve prediction of stochastic MCA



Identified future areas of investigation

- but did we gain conclusive evidence of previously unknown phenomena?

- Stochastic fraction of MCA is a better predictor of total nitrogen removal than it is of general BOD, $\text{NH}_4\text{-N}$, and total P removal
- Climate and WRRF design related parameters are important predictors for process performance
- WRRF metadata, process performance, and ecology metrics predict stochastic fraction of MCA

Thank you!



For PDH Certificate seekers,

The link to the new on-line course evaluation form has been posted in the Chat. The link is also available on the District website. The form will **only** be available online until the start of next month's seminar. Please be sure to fill it out and submit promptly.