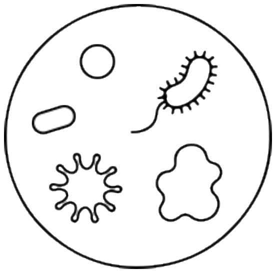
Causal Inference Framework for Ocean Microbial Community Responses to Warmer Temperature

(Minh) Viet Tran, Christian L. Müller

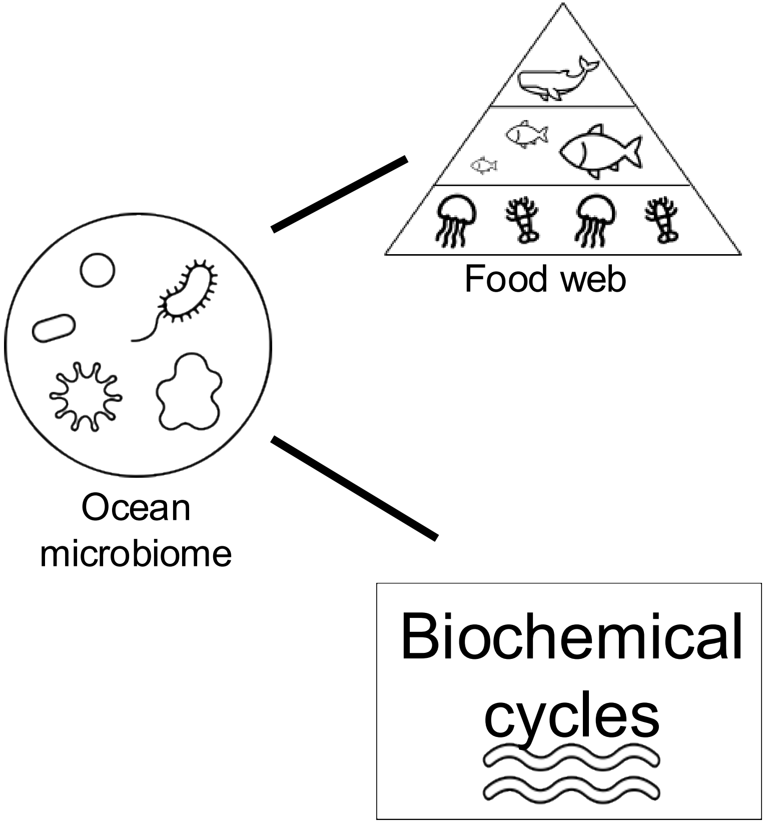
03.11.2025

The ocean microbiome plays a central role in climate regulation

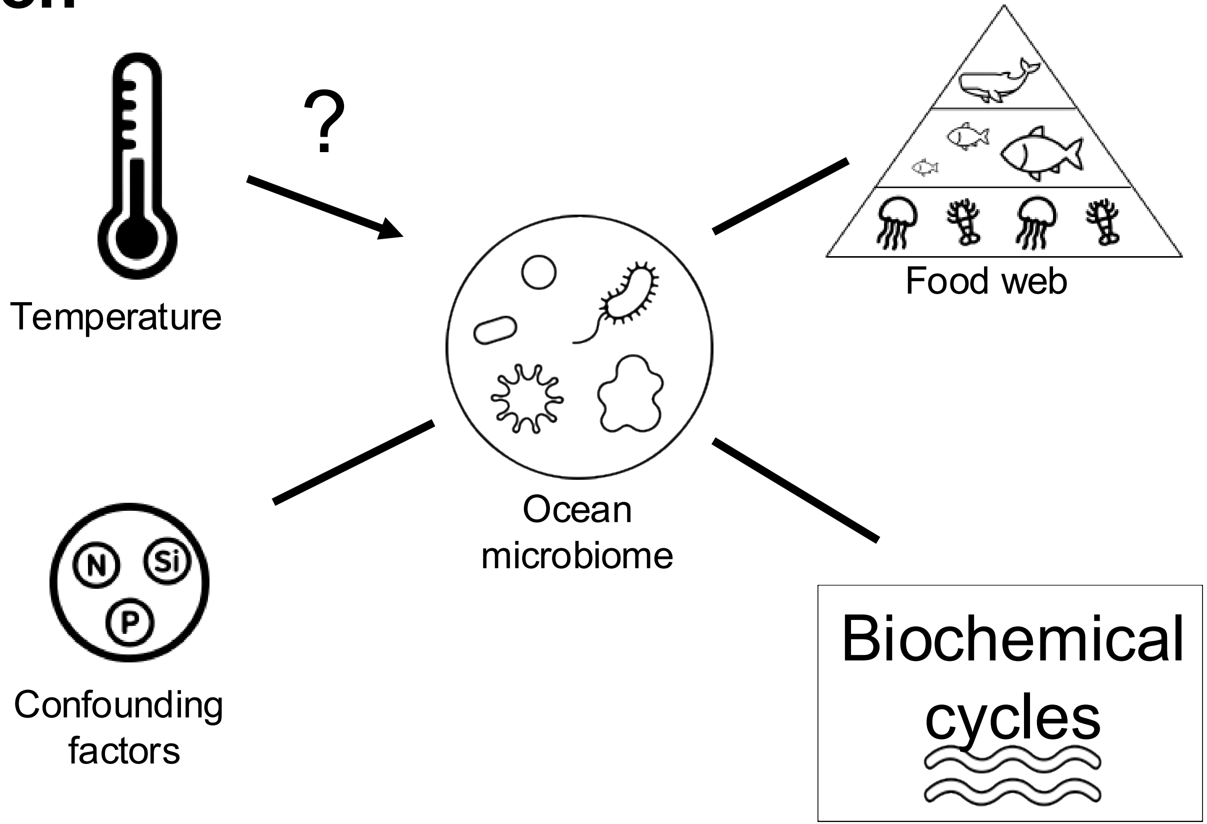


Ocean
microbiome

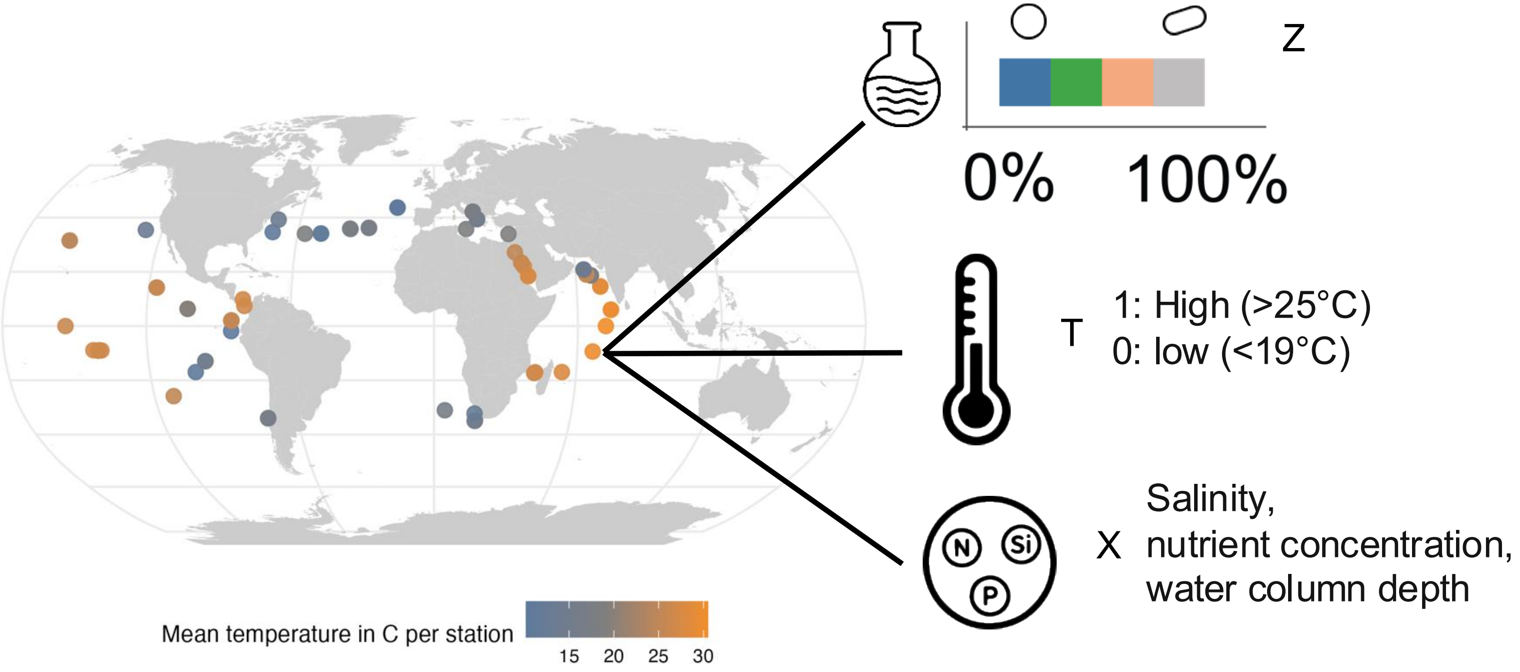
The ocean microbiome plays a central role in climate regulation



The ocean microbiome plays a central role in climate regulation

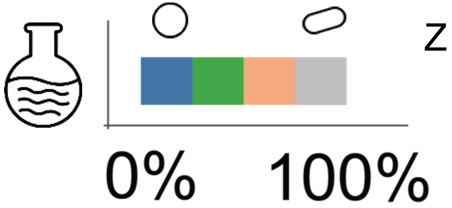


Global ocean survey enable us to study those microorganisms in more detail



Shinichi Sunagawa, Luis Pedro Coelho, Samuel Chaffron, Jens Roat Kultima, Karine Labadie, Guillem Salazar, Bardya Djahanschiri, Georg Zeller, Daniel R Mende, Adriana Alberti, et al. Structure and function of the global ocean microbiome. Science, 348(6237):1261359, 2015.

Outcomes of interest



Let Z be the compositional data matrix with n rows and p species.

Community effects:

- Observed richness: $\sum_{j=1}^{p} \mathbb{I}(Z_{ij} > 0)$
- Alpha diversity with Shannon entropy: $-\sum_{j=1}^{p} Z_{ij} ln(Z_{ij})$

Individual species effect:

For each taxa $Z_{.j}$ we applied LinDA, a robust regression model to test for a significant association with the treatment variable while controlling for multiple testing

Huijuan Zhou, Kejun He, Jun Chen, and Xianyang Zhang. Linda: linear models for differential abundance analysis of microbiome compositional data. Genome biology, 23(1):95, 2022.

Rubin causal model (RCM)

Unit i	$Y_i(1)$	$Y_i(0)$	Individual causal effect τ
1	100	10	90
2	1000	100	900
3	1000	90	910
4	1200	300	900

Guido W Imbens and Donald B Rubin. Causal inference in statistics, social, and biomedical sciences. Cambridge university press, 2015.

Rubin causal model (RCM)

	Treated	Untreated	
Unit i	$Y_i(1)$	$Y_i(0)$	Individual causal effect τ
1	100	10	90
2	1000	100	900
3	1000	90	910
4	1200	300	900

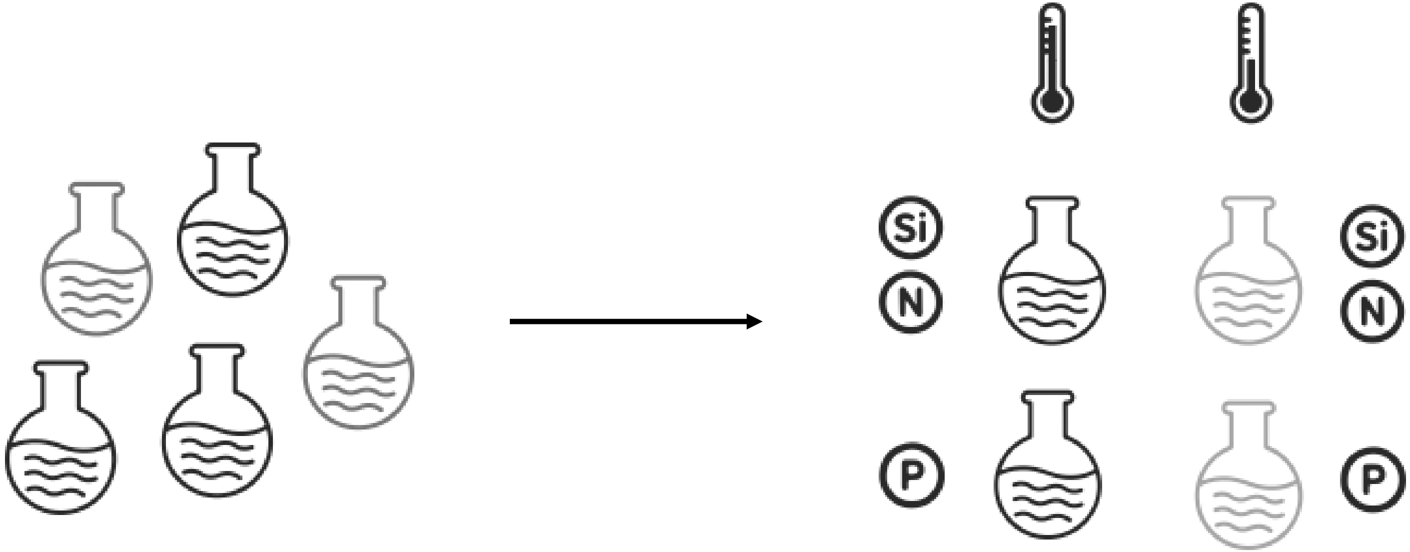
Rubin causal model (RCM)

Rubin causal model frames causal inference as a "missing data" problem: We can only observe one of the potential outcomes

Unit i	$Y_i(1)$	$Y_i(0)$	Individual causal effect τ
1	100	?	?
2	?	100	?
3	1000	?	?
4	?	300	?

If we assume stable unit treatment value, positivity, and ignorability assumption: $ATT = E[Y_i(1) - Y_i(0)|T_i = 1]$.

We want to find comparable units



- Propensity score matching (PSM) 1:1 with caliper
- Bipartite matching (BM)

Paul R Rosenbaum and Donald B Rubin. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. The American Statistician, 39(1):33–38, 1985.

Alice J Sommer, Annette Peters, Martina Rommel, Josef Cyrys, Harald Grallert, Dirk Haller, Christian L Müller, and Marie-Abèle C Bind. A randomization-based causal inference framework for uncovering environmental exposure effects on human gut microbiota. PLoS computational biology, 18(5):e1010044, 2022.

Matching results in better covariate balance

We evaluated the covariance balancing with the standardized mean difference $(\overline{X}_t - \overline{X}_c)/(\sqrt{s_t^2} + \sqrt{s_c^2})/2$

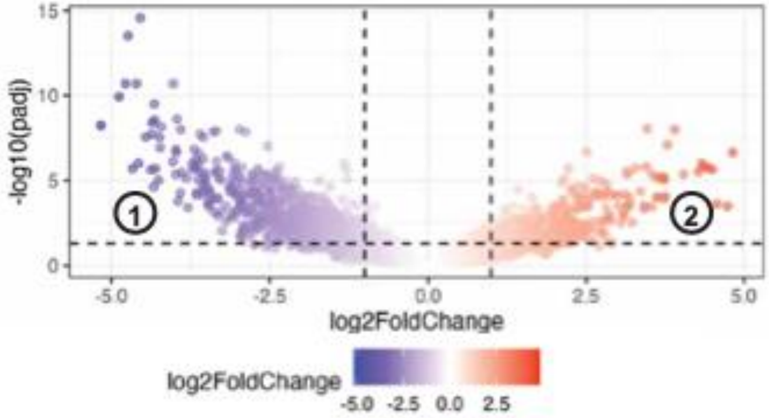
	Unmatched (N = 72, p = 7213)	PSM (N = 36, p = 6594)	BM (N = 32, p = 6547)
Salinity	<u>0.059</u>	0.216	0.163
Phosphate	0.665	0.023	<u>0.018</u>
Nitrate + Nitrite	0.693	0.118	<u>0.003</u>
Depth (SRF)	0.884	<u>0.00</u>	<u>0.00</u>
Depth (DCM)	0.296	<u>0.00</u>	<u>0.00</u>
Depth (MES)	0.843	<u>0.00</u>	<u>0.00</u>

Negative effect on diversity

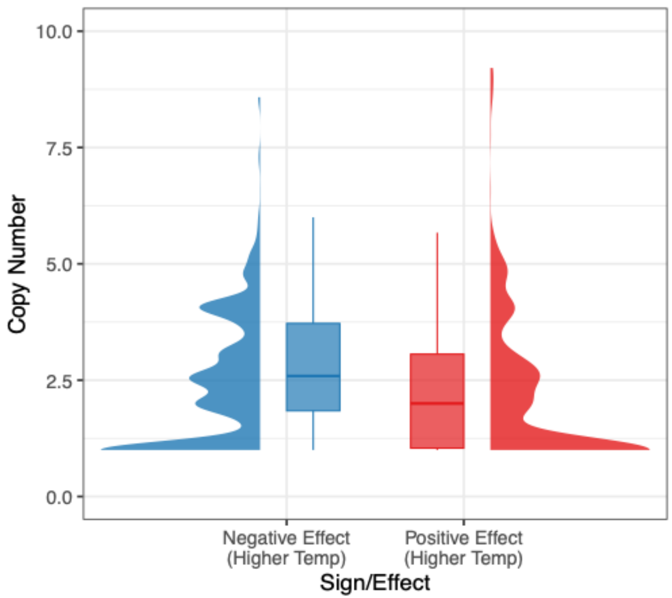
Method	Richness	Shannon
Baseline (Unmatched)	-801.9 ± 153	-0.23 ± 0.06
PSM	-426.5 ± 178	-0.23 ± 0.05
BM	-514.3 ± 199	-0.15 ± 0.07

Effect on individual taxa

Method	Negative	Positive
Baseline (Unmatched)	1718	1546
PSM	1 986	432 2
BM	613	239



Warmer temperature favors slower-growing bacteria



Clare I Abreu, Martina Dal Bello, Carina Bunse, Jarone Pinhassi, and Jeff Gore. Warmer temperatures favor slower-growing bacteria in natural marine communities. Science Advances, 9(19):eade8352, 2023.

Summary and outlook

- Causal inference framework to estimate the causal effect on temperature on ocean microbial communities from large scale surveys
- Extending the idea to continuous treatment variables and a larger dataset
- Conducting a more formal sensitivity analysis