Causal Inference Framework for Ocean Microbial Community Responses to Warmer Temperature

(Minh) Viet Tran
Helmholtz Munich, Ludwig-Maximilians-Universität München,
Munich Center for Machine Learning (MCML), Munich
`viet.tran@helmholtz-munich.de`

Christian L. Müller
Helmholtz Munich, Ludwig-Maximilians-Universität München,
Munich Center for Machine Learning (MCML), Munich
Flatiron Institute, New York
`christian.mueller@helmholtz-munich.de`

Abstract

Understanding how ocean temperature influences microbial communities is critical for forecasting ecological responses to climate change. This paper outlines a framework for applying the Rubin Causal Model (RCM) to observational microbiome sequencing data using matching methods. The aim is to estimate the causal effect of temperature on microbial taxa, while controlling for confounding environmental variables. This methodology offers a transparent, interpretable, and simulation-free way to extract causal signals from complex ecological datasets.

1 Introduction

Ocean microbes form the base of the marine food web and play a foundational role in global biogeochemical cycles, such as carbon fixation, and thus in climate regulation. Changing conditions, particularly the increase in sea temperature, poses a threat to the stability and function of this vital ecosystem. Numerous observational studies have documented a strong correlation between temperature and shifts in microbial community structures and diversity [1, 2, 3].

Recent global-scale surveys, such as Tara Oceans [1, 4, 5] and the GRUMP database [6], begin to provide microbial sequencing and environmental covariate data at a scale that is amenable to statistical modeling. Common techniques include correlational analysis [1], eco-dynamic modeling [2], joint taxon-environment models [7], and generative latent variational models [8]. Methods that determine potentially causal drivers of microbial patterns in the global ocean are thus far largely elusive.

Here, we propose a potential outcomes framework to estimate the causal effects of environmental variables on microbial taxa using observational data. We apply different matching strategies to construct a quasi-experimental design, allowing the estimation of the effect of temperature increases on microbial community composition.

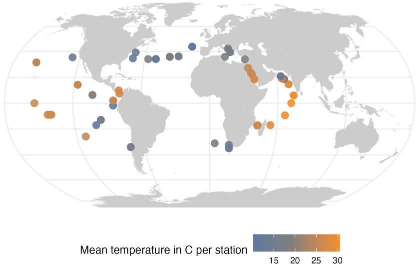


Figure 1: Global sampling stations from the Tara Oceans project, colored by mean sea surface temperature (°C)

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2025.

2 Related work

Our proposal employs matching [9] to construct a quasi-experimental design [10] by balancing covariates between treatment and control groups. This approach is part of a larger toolkit, including inverse probability of treatment weighting [11, 12], machine learning methods such as Causal Forests [13] for estimating heterogeneous effects, and double machine learning [14]. While similar methods have been successfully applied to other complex biological systems, such as environmental exposure effects on human gut microbiota [15] or ecology [16], their application to marine ecology remains underexplored. Our study builds on this foundation to formally estimate the effect of warming temperatures on ocean microbial communities.

3 Methodology

3.1 Causal framework

We adopt the potential outcomes framework [17] to formalize the problem of estimating the causal effect of temperature on microbial communities from observational data. For each sampling unit i, let $T_i \in \{0, 1\}$ be a binary treatment indicator, where $T_i = 1$ denotes exposure to high temperature and $T_i = 0$ exposure to low temperature. Let Y_i be the observed outcome of interest. We define two potential outcomes for each unit: $Y_i(1)$, the outcome that would be observed under treatment, and $Y_i(0)$, the outcome under control. The individual treatment effect is $\tau_i = Y_i(1) - Y_i(0)$. Since we can only observe one potential outcome for each unit, we focus on estimating the average treatment effect on the treated (ATT) assuming the standard *Stable unit treatment value* and *ignorability assumption*:

$$\text{ATT} = E[\tau_i | T_i = 1] = E[Y_i(1) - Y_i(0) | T_i = 1] \, .$$

3.2 Data and covariate balancing via matching

The metagenomic-derived 16S ribosomal RNA gene tags used in this study were sourced from the global Tara Oceans project [1] [1]. The sequencing reads were clustered into operational taxonomic units (OTUs) at a 97$\%$ similarity threshold, and taxonomic assignments were made using the SILVA database (v115) [18]. The dataset was filtered to include only bacterial taxa and to exclude samples with missing values for the covariates of interest.

The treatment variable T was defined by partitioning samples into terciles based on sea surface temperature. The upper tercile ($> 25°C$) constitutes the treatment group and the lower tercile ($< 19°C$) the control. Key confounding variables, including salinity, nutrient concentrations (phosphate, nitrate), and water column depth, were identified as covariates (X) for matching, resulting in 72 (out of 139) samples suitable for matching.

To mitigate selection bias from the observational data, we use two matching methods to create a balanced comparison between the treatment and control groups, thereby mimicking a randomized experiment: (i) Propensity Score Matching (PSM) where we estimate the probability of being in the high-temperature group based on covariates X and match treated units to control units with similar propensity scores [19]; and (ii) Bipartite Matching (BM) where we construct a bipartite graph of treated and control units, creating edges only between units with sufficiently small distances in their covariate vectors, and then find the maximum number of matched pairs [15, 20].

3.3 Estimating average treatment effect on the treated

We estimate the ATT of two quantities of interest by comparing the outcomes of the matched groups. Firstly, we assess the treatment effect on the whole microbial community, as measured by (i) observed richness (the cardinality of the set of unique taxa) and (ii) alpha-diversity via the Shannon entropy [21]. Secondly, we identify individual taxa affected by the treatment via compositionally-aware differential abundance (DA) testing. We employ LinDA [22], a state-of-the-art DA method that fits a robust linear model for each taxon to test for a significant association with the treatment variable while controlling for multiple testing [23].

---

[1] Data were retrieved from http://ocean-microbiome.embl.de/companion.html

4 Results

4.1 Matching strategies result in more balanced covariates

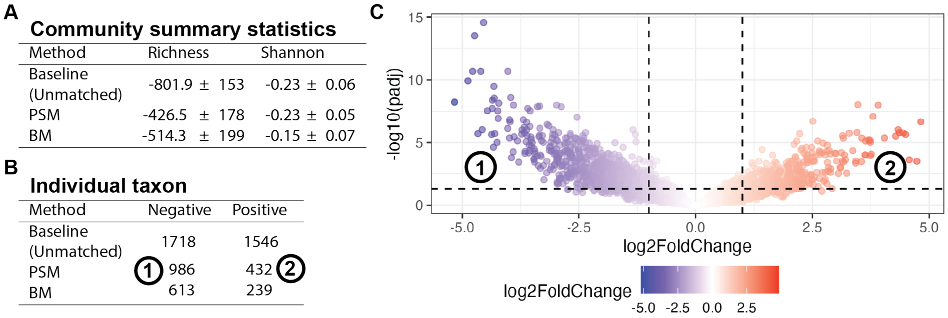
Prior to matching, there were substantial imbalances in nutrient levels (see Table 1 and Supplementary Fig. 4). After PSM and BM, most differences were substantially reduced, thus approximating the conditions of a randomized experiment.

Table 1: Standardized mean difference of covariates before and after matching

		Standardized Mean Difference					
	N	Salinity	Phosphate	Nitrate + Nitrite	SRF (Depth)	DCM (Depth)	MES (Depth)
Unmatched	72	0.059	0.665	0.693	0.884	0.296	0.843
PSM	36	0.216	0.023	0.118	0.000	0.000	0.000
BM	32	0.163	0.018	0.003	0.000	0.000	0.000

4.2 Analysis suggests a negative effect of higher temperature on diversity and a shift in the bacterial composition

We first estimated the ATT of high temperature on microbial community composition, consistently showing that higher temperatures had a negative effect on microbial richness and diversity (Fig. 2 (A)). For example, we estimated a reduction of approximately 427 unique taxa using PSM, indicating that warming sea surface temperatures lead to a loss of microbial richness.



A

Community summary statistics

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

B

Individual taxon

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

Figure 2: (A) Effect of higher temperature on diversity statistics. (B) The number of significantly ($\alpha = 0.05$) positive or negatively affected taxa. (C) Volcano plot showing taxon log-fold changes on the x-axis and (log-scaled) adjusted p-values on the y-axis after PSM.

We next investigated how high temperatures affected differential abundance of microbial taxa (Fig. 2 (B)). Our analysis revealed potentially widespread causal effects on microbial taxa (see Fig. 2 (C)). For example, after PSM, we estimated 986 taxa to have significant negative log-fold changes at high temperatures and 432 taxa to have significant positive log-fold changes. Our analysis also confirmed prior observations [2] that, among the set of differentially abundant taxa, higher temperatures favor microbes with slower growth rates (see Supplementary Fig. 5).

5 Discussion

We have presented a causal inference framework, grounded in Rubin's potential outcomes model, to estimate causal effects of temperature on ocean microbial communities from large-scale microbiome survey data. Using data from Tara Oceans [5], we exemplified our proposal by applying (i) two different matching to construct a quasi-experimental design after balancing key environmental covariates and (ii) by investigating the effect of temperature on microbial community composition and individual taxon abundances.

The framework indicates an overall negative effect of increased temperature on microbial richness and diversity, sharpening results from the (unmatched) observational analysis. While the analysis on (unmatched) observational data suggested a nearly balanced differential abundance response, our causal framework suggests a predominantly negative impact on species abundances, particularly for fast-growing bacteria [2].

Our work highlights the need for causal methodologies to disentangle drivers in complex global environmental systems. While our current analysis is limited by the low sample size, the proposed framework is readily scalable and amenable to future large-scale ocean survey data analysis.

Acknowledgments and Disclosure of Funding

This research was supported by the Simons Collaboration on Computational Biogeochemical Modeling of Marine Ecosystems (CBIOMES) through grant 986803 (awarded to CLM) and core funding from the Munich Center for Machine Learning (MCML).

References

[1] 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.

[2] 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.

[3] Jingdi Li, Kieran A Bates, Kim L Hoang, Tobias E Hector, Sarah CL Knowles, and Kayla C King. Experimental temperatures shape host microbiome diversity and composition. *Global Change Biology*, 29(1):41–56, 2023.

[4] Lucas Paoli, Hans-Joachim Ruscheweyh, Clarissa C Forneris, Florian Hubrich, Satria Kautsar, Agneya Bhushan, Alessandro Lotti, Quentin Clayssen, Guillem Salazar, Alessio Milanese, et al. Biosynthetic potential of the global ocean microbiome. *Nature*, 607(7917):111–118, 2022.

[5] Shinichi Sunagawa, Silvia G Acinas, Peer Bork, Chris Bowler, Damien Eveillard, Gabriel Gorsky, Lionel Guidi, Daniele Iudicone, Eric Karsenti, et al. Tara oceans: towards global ocean ecosystems biology. *Nature Reviews Microbiology*, 18(8):428–445, 2020.

[6] Jesse McNichol, Nathan LR Williams, Yubin Raut, Craig Carlson, Elisa R Halewood, Kendra Turk-Kubo, Jonathan P Zehr, Andrew P Rees, Glen Tarran, Mary R Gradoville, et al. Characterizing organisms from three domains of life with universal primers from throughout the global ocean. *Scientific Data*, 12(1):1078, 2025.

[7] Yuqing Yang, Ning Chen, and Ting Chen. Inference of environmental factor-microbe and microbe-microbe associations from metagenomic data using a hierarchical bayesian statistical model. *Cell systems*, 4(1):129–137, 2017.

[8] Aditya Mishra, Jesse McNichol, Jed Fuhrman, David Blei, and Christian L Müller. Variational inference for microbiome survey data with application to global ocean data. *ISME communications*, 5(1):ycaf062, 2025.

[9] Elizabeth A Stuart. Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 25(1):1, 2010.

[10] Marie-Abele C Bind and Donald B Rubin. Bridging observational studies and randomized experiments by embedding the former in the latter. *Statistical methods in medical research*, 28(7):1958–1978, 2019.

[11] James M Robins, Miguel Angel Hernan, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.

[12] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[13] Susan Athey and Stefan Wager. Estimating treatment effects with causal forests: An application. *Observational studies*, 5(2):37–51, 2019.

[14] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

[15] 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.

[16] Kaitlin Kimmel, Laura E Dee, Meghan L Avolio, and Paul J Ferraro. Causal assumptions and causal inference in ecological experiments. *Trends in Ecology & Evolution*, 36(12):1141–1152, 2021.

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

[18] Christian Quast, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, and Frank Oliver Glöckner. The silva ribosomal rna gene database project: improved data processing and web-based tools. *Nucleic acids research*, 41(D1):D590–D596, 2012.

[19] 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.

[20] Silvio Micali and Vijay V Vazirani. An o (v| v| c| e|) algoithm for finding maximum matching in general graphs. In *21st Annual symposium on foundations of computer science (Sfcs 1980)*, pages 17–27. IEEE, 1980.

[21] Claude E Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[22] 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.

[23] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

[24] Gabor Csardi and Tamas Nepusz. The igraph software. *Complex syst*, 1695:1–9, 2006.

[25] Daniel Ho, Kosuke Imai, Gary King, and Elizabeth A Stuart. Matchit: nonparametric preprocessing for parametric causal inference. *Journal of statistical software*, 42:1–28, 2011.

6 Supplements

6.1 Matching

We employ two matching methods to create groups with balanced covariate distributions: (I) Propensity score matching (PSM) : (1) estimate the propensity score, $e(\mathbf{X}_i) = P(T_i = 1|\mathbf{X}_i)$, using a logistic regression. (2) Perform 1:1 nearest neighbor matching on the logit of the propensity scores with a caliper of 0.1 standard deviations, without replacement. Maximum bipartite matching [15, 20]: Reframe as finding the maximum pairs in a bipartite graph. The two sets of nodes represent the treatment and control groups, and the edge exist between a treated unit i and a control unit j if their distance in covariates (see Supplement 6.4) is small enough.

6.2 Estimating the ATT

To estimate the Average treatment effect on the treated (ATT), we employ a regression the matched samples. We then estimate the ATT by applying OLS to the following model on the matched data:

$$Y_i = \alpha + \tau_{\text{ATT}} T_i + \varepsilon_i$$

Here, Y_i is the outcome, T_i is the treatment indicator, ε_i is the error term, α denotes the intercept and the coefficient τ_{ATT} is our estimate of the treatment effect.

Let Z^{nxp} be the matrix with the abundances, with samples as the rows and microbes as columns. As the outcomes Y_i we use two summary statistics:

- The observed richness for a given sample i: $\sum_{j=1}^{p} \mathbb{I}(Z_{ij} > 0)$, where $\mathbb{I}(\cdot)$ is the indicator function, which is equal to 1 if the condition $Z_{ij} > 0$ is true, and 0 otherwise.
- The alpha diversity (as measured by the Shannon entropy) for a given sample i: $-\sum_{j=1}^{p} p_{ij} \ln(p_{ij})$ with p_{ij} denoting the count of a single microbe divided by the total count of all microbes in that sample (the row sum) $p_{ij} = \frac{Z_{ij}}{\sum_{k=1}^{p} Z_{ik}}$

Since the matrix Z is compositional, we used LinDA as a method to estimate the ATT to correct for the constraint of the data. Then every microbe is tested individually.

6.3 Data preprocessing

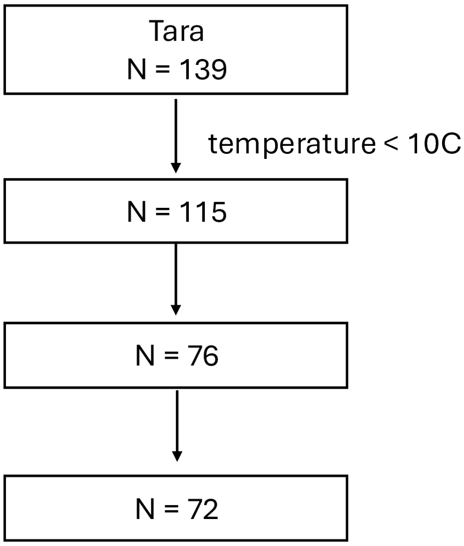


Figure 3: Preprocessing steps: First filter samples with less than 10 degree C, then get the samples for the upper and lower tercile, and finally filter out the two samples with missing values in the matching variables.

6.4 Matching algorithm

Following [15] we employ the matching as follows: Subject i from the observed treatment group is matched to a subject j in the control group if the covariate vector X_i is similar enough to X_j. We define the difference metric as for each variable k of interest.

$$D(X_i, X_j) = \begin{cases} 0, & \text{if } |X_i - X_j| < \delta_k \text{ for } k = 1, ..., K \\ \infty, & \text{otherwise} \end{cases}$$

(1) Create a bipartite network, such that a unit is a node either as a treatment or control node. (2) Construct edges between the unit if the difference metric is smaller than ∞. (3) Using the igraph R packages [24] to find the maximum bipartite matching. With the following δ_k:

Table 2: Threshold used for BM

Variable	**Maximum allowed difference** δ_k
Salinity	1.8
Phosphate	0.8
Nitrate_Nitrite	1
Depth_category	Same category

For propensity score matching we used the package matchit [25].

6.5 Distribution after matching

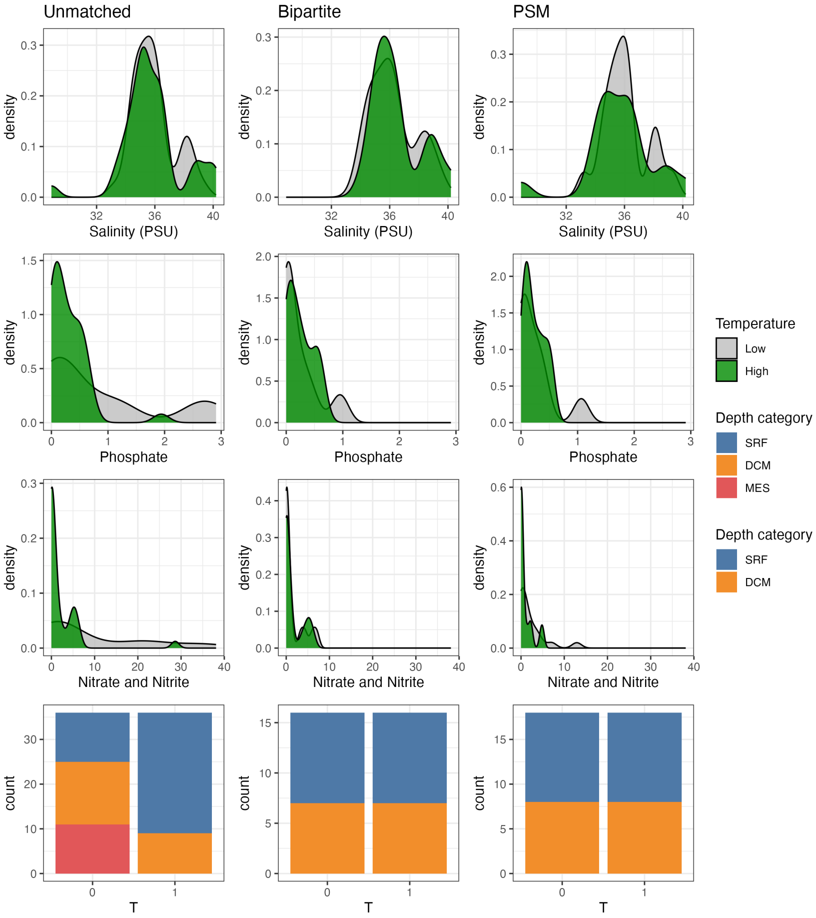


Figure 4: Covariate distributions before (unmatched) and after (Bipartite (BM), PSM) matching, visually confirming improved balance between high and low-temperature groups.

6.6 Specification of the preprocessing before differential abundance testing

For differential abundance testing we performed a 10% prevalence filter resulting in the following table.

Method	n	Taxa after filter
Baseline (Unmatched)	72	7213
PSM	36	6594
BM	32	6547

Table 3: Number of taxa for differential abundance testing

6.7 Distribution slow and fast growing bacteria

Abreu et al. [2] investigated the effect of higher temperature on slow and fast growing bacteria where maximum growth rate is approximated by 16S rRNA copy numbers. Higher copy numbers indicate higher growth. For the PSM data, we associated the set of differentially abundant taxa to their respective copy numbers and compared the resulting copy number distributions (weighted by the relative abundances of the taxa) for both the positively and negatively differentially abundant taxa, respectively. Supplementary Fig. 5 visualizes the resulting smoothed distributions for the species that profit from warmer temperatures (positive DA, red) and those that are harmed by warmer temperatures (negative DA, blue). We observed a distributional shift toward lower mean weighted copy numbers for the species set that profits from higher temperatures, thus confirming Abreu et al.'s analysis.

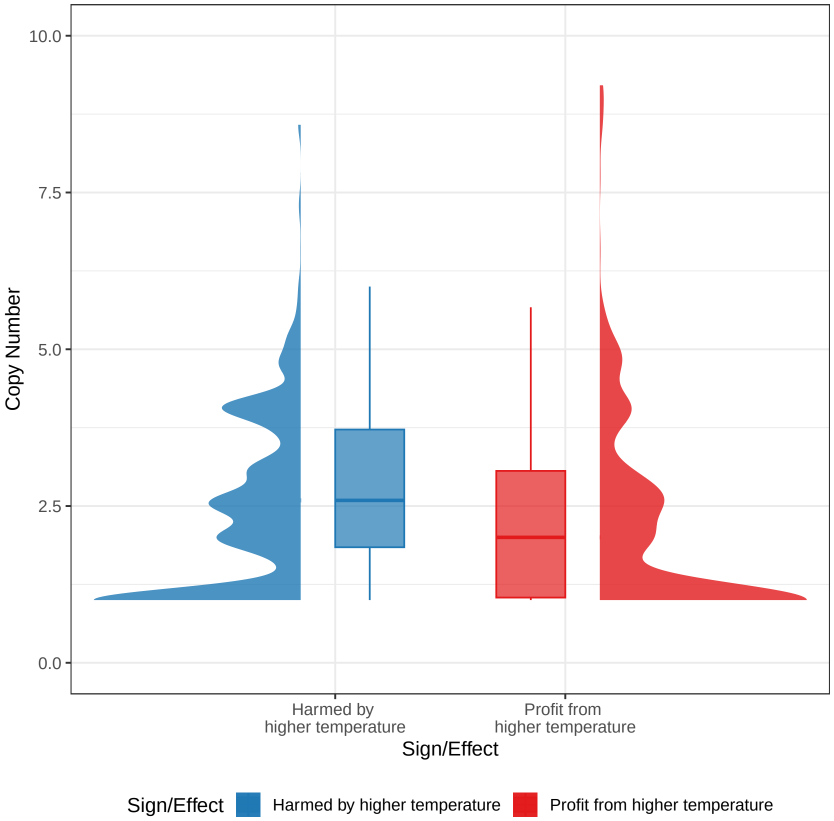


Figure 5: Distribution shift indicates that slow growing bacteria as indicated by low copy number are favored.