

December 1, 2024

Dear Editor,

We submit our manuscript entitled “**MeLSI: Metric Learning for Statistical Inference in Microbiome Community Composition Analysis**” for consideration as a Methods and Protocols article in mSystems.

Current microbiome beta diversity analysis relies on fixed distance metrics (Bray-Curtis, Euclidean, UniFrac) that treat all taxa uniformly. This approach cannot adapt to dataset-specific signal structure and may miss subtle but biologically meaningful differences. We present MeLSI, a machine learning framework that learns data-adaptive distance metrics while maintaining rigorous statistical inference through permutation testing.

MeLSI integrates ensemble metric learning with PERMANOVA-based hypothesis testing. The method provides interpretable feature importance weights that identify which taxa drive group separation, addressing a key limitation of black-box machine learning approaches. Comprehensive validation demonstrates proper Type I error control, competitive statistical power across synthetic benchmarks, and superior performance on real microbiome datasets (Atlas1006, DietSwap).

The method is implemented as an open-source R package with comprehensive documentation and reproducible validation code (DOI: 10.5281/zenodo.1771484). We believe this work will be of broad interest to the mSystems readership and addresses an important methodological gap in microbiome analysis.

All authors have approved this submission. We declare no conflicts of interest. This work was supported by NIH/NIA grant R56AG079586.

Sincerely,

Nathan Bresette

Aaron C. Ericsson

Carter Woods

Ai-Ling Lin

Corresponding author: ai-ling.lin@health.missouri.edu