

Integrative analysis across metagenomic taxonomic classifiers

Tanya T. Karagiannis^{1,2}, Ye Chen³, Sarah Bald⁴, Albert Tai⁵⁻⁶, Sofiya Milman⁷, Stacy Andersen⁸, Thomas T. Perls⁸, Daniel Segre^{4,9}, Paola Sebastiani^{1-2,4,6}, Meghan I. Short ^{1,2}

¹ Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, USA
 ² Department of Medicine, Tufts University, Medford, MA, USA
 ³ Clinical Translational and Science Institute, Tufts Medical Center, Boston, MA, USA
 ⁴ Bioinformatics Program, Faculty of Computing and Data Science, Boston University, Boston, MA, USA
 ⁵ Department of Immunology, Tufts University School of Medicine, Boston, MA, USA
 ⁶ Data Intensive Studies Center, Tufts University, Medford, MA, USA
 ⁷ Albert Einstein College of Medicine, Bronx, New York, USA
 ⁸ Department of Medicine, Geriatrics Section, Boston University School of Medicine, Boston, MA, USA.



INTRODUCTION

There are various well-validated taxonomic classifiers for profiling shotgun metagenomics sequencing data, with two popular methods, MetaPhIAn and Kraken, at the forefront of many studies. Despite substantial differences between classification approaches and calls for consensus-based methods, most metagenomic studies rely on a single taxonomic classifier. To compare inferences from multiple taxonomic classifiers, we performed an analysis using MetaPhIAn4 and Kraken2 in parallel and examined diversity trends and species relative abundance associations with age in two studies of extreme human longevity. We used a consensus-based and a novel meta-analytic approach to compare and integrate findings from both taxonomic classifiers.

METHODS





Complementary approaches for taxonomic classification

- We processed datasets using KneadData² (v0.12.0), and taxonomic classification was performed using both MetaPhIAn4³ (v.4.1) and Kraken2⁴ (v. 2.1.2) via an in-house metagenomic pipeline (available at <u>https://github.com/Integrative-Longevity-Omics/MGS_pipeline</u>).
- We used Bracken⁵ (v.2.9) to estimate the Kraken2 taxonomic relative abundances at the species level.
 Association of microbiome diversity with age
- We calculated normalized alpha diversity⁶ of the taxonomic relative abundances at each taxonomic level using both classifiers and examined alpha diversity changes with age using linear regression.
- We performed principal coordinate analysis of the Bray-Curtis dissimilarities of sample profiles to visualize the similarities and differences between samples, testing association with age using PERMANOVA.

Taxon-specific differential abundance with age

Figure 2. Age associations with beta diversity are concordant across taxonomic classifiers. (Top) Principal coordinate analysis plots displaying the Bray-Curtis dissimilarities between samples with age across taxonomic levels when using both classifiers in each cohort, with age association evaluated via PERMANOVA with significance threshold p < 0.05. (Bottom) Procrustes analysis performed on Bray-Curtis dissimilarities of samples from both classification methods within each cohort, with lines connecting the same samples. Procrustian randomization (Monte Carlo) test to evaluate whether the concordance between the distances based on the taxonomic classifiers is greater than expected due to chance, with significance threshold p < 0.05.



 We performed differential abundance analysis at the species level to model species log-transformed relative abundances with age utilizing linear regression with generalized estimating equations (GEE) to account for within-family correlations, adjusting for sex and education.

Comparison/Integration of downstream results

- We performed Pearson correlation and Procrustes analysis of diversity metrics to compare across methods.
- We introduce and used a correlated meta-analyses approach⁷ to integrate results across both classifier methods and adjust the level of statistical significance based on individual studies.



RESULTS

Figure 3. Differential abundance analyses using both classifier methods capture more age-associated taxa. (Top) Volcano plot of the differentially abundant species with age in each cohort via linear regression and using correlated meta-analysis approach to generate combined p-values and combined effect estimates for species identified by both classifiers within a cohort. (Bottom) Upset plots containing all species significantly associated with age via individual tests or correlated meta-analysis approach. The vertical bar heights show the number of species associated in the Bracken profiles, MetaPhIAn4 (Metaphlan) profiles, both, or neither based on individual tests.

CONCLUSIONS

- While many results were consistent across the two classifiers, we found classifier-specific inferences that would be lost when using one classifier alone.
- Both classifiers captured similar age-associated changes in diversity in two cohorts, with classifier-specific differences driving variability in alpha diversity of species.
- Correlated meta-analysis of differential abundance results identified 17 taxa robustly associated with age across classifiers and cohorts.
- We highlight the utility of integrating results from multiple classifiers when performing downstream analysis of metagenomics data with phenotypes of interest.

ACKNOWLEDGEMENTS

Figure 1. Alpha diversity displays similar changes with age at higher taxonomic levels and vary at lower taxonomic levels across classifiers. (Top) Scatterplots of normalized alpha diversity for each sample with age, comparing across taxonomic levels when using both classifiers in each cohort. We employed linear regression models to evaluate the association with age with significance threshold p < 0.05. (Bottom) Scatterplots comparing the normalized alpha diversity scores of samples based on classification method, evaluated via Pearson correlation analysis with significance threshold p < 0.05.

This research is supported by NIH-NIA: UH2AG064704.



- 1. Xu, Qian, et al. "Metagenomic and metabolomic remodeling in nonagenarians and centenarians and its association with genetic and socioeconomic factors." *Nature Aging* 2.5 (2022): 438-452.
- 2. Mclver, Lauren J., et al. "bioBakery: a meta'omic analysis environment." *Bioinformatics* 34.7 (2018): 1235-1237.
- 3. Blanco-Míguez, Aitor, et al. "Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhIAn 4." Nature Biotechnology 41.11 (2023): 1633-1644.
- 4. Wood, Derrick E., et al. "Improved metagenomic analysis with Kraken 2." Genome biology 20 (2019): 1-13.
- 5. Lu, Jennifer, et al. "Metagenome analysis using the Kraken software suite." Nature protocols 17.12 (2022): 2815-2839.
- 6. Karagiannis, Tanya T., et al. "Cell type diversity statistic: an entropy-based metric to compare overall cell type composition across samples." Frontiers in Genetics 13 (2022): 855076.
- 7. Reed, E., and Sebastiani, P. "A Simple Strategy for Identifying Conserved Features across Non-independent Omics Studies." bioRxiv. (2024).

RESEARCH POSTER PRESENTATION TEMPLATE © 2019 WWW.PosterPresentations.com