

RESEARCH

Evaluating trait-based sets for taxonomic enrichment analysis in microbiome data sets

Quang P. Nguyen^{1,2}, Anne G. Hoen^{1,2} and H. Robert Frost^{1*}

*Correspondence:

hildreth.r.frost@dartmouth.edu

¹Department of Biomedical Data Science, Dartmouth College, Hanover, NH, USA

Full list of author information is available at the end of the article

Abstract

Background: Set-based pathway analysis is a powerful tool that allows researchers to summarize complex genomic variables in the form of biologically interpretable sets. Since the microbiome is characterized by a high degree of inter-individual variability in taxonomic compositions, applying enrichment methods using on functionally driven taxon sets can increase both the reproducibility and interpretability of microbiome association studies. However, there is still an open question of which knowledge base utilize for set construction. Here, we evaluate microbial trait databases, which aggregates experimentally determined microbial phenotypes, as a potential avenue for meaningful construction of taxon sets.

Methods: Using publicly available microbiome sequencing data sets (both 16S rRNA gene sequencing and whole-genome sequencing), we assessed these trait-based sets on three aspects: first, do they cover the diversity of microbes obtained from a typical data set; second, do they confer additional predictive power on disease prediction tasks when assessed against measured pathway abundances and PICRUSt2 prediction; third, for sets that are found to be enriched, are pathways corresponding to the trait also have increased abundances.

Results:

Conclusions:

Keywords: microbiome; enrichment analysis; trait database

Introduction

Methods

Constructing the trait database

We utilized pre-compiled trait databases from previous publications, Madin et al. [1] and Weissman et al [2]

Results

Discussion

Appendix

Text for this section...

Acknowledgements

Text for this section...

Funding

Text for this section...

Abbreviations

Text for this section. . .

Availability of data and materials

Text for this section. . .

Ethics approval and consent to participate

Text for this section. . .

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Text for this section. . .

Authors' contributions

Text for this section. . .

Authors' information

Text for this section. . .

Author details

¹Department of Biomedical Data Science, Dartmouth College, Hanover, NH, USA. ²Department of Epidemiology, Dartmouth College, Hanover, NH, USA.

References

1. Madin, J.S., Nielsen, D.A., Brbic, M., Corkrey, R., Danko, D., Edwards, K., Engqvist, M.K., Fierer, N., Geoghegan, J.L., Gillings, M., *et al.*: A synthesis of bacterial and archaeal phenotypic trait data. *Scientific Data* **7**(1), 1–8 (2020)
2. Weissman, J.L., Dogra, S., Javadi, K., Bolten, S., Flint, R., Davati, C., Beattie, J., Dixit, K., Peesay, T., Awan, S., *et al.*: Exploring the functional composition of the human microbiome using a hand-curated microbial trait database. *BMC bioinformatics* **22**(1), 1–21 (2021)

Figures



Tables

Table 1 Sample table title. This is where the description of the table should go

	B1	B2	B3
A1	0.1	0.2	0.3
A2
A3

Additional Files

Additional file 1 — Sample additional file title
 Additional file descriptions text (including details of how to view the file, if it is in a non-standard format or the file extension). This might refer to a multi-page table or a figure.

Additional file 2 — Sample additional file title
 Additional file descriptions text.