Quang P. Nguyen, Ph.D.
Department of Biomedical Data Science
Geisel School of Medicine at Dartmouth College

April 6th, 2023

Dear Editors,

I am writing on behalf of my co-authors to submit an original research article titled "Evaluating trait-based sets for taxonomic enrichment analysis applied to human microbiome data sets" for consideration by BMC Bioinformatics.

We confirm this work is original and has not been submitted or published elsewhere. All authors have reviewed and approved the submission of this manuscript.

One of the major challenges in understanding microbiome taxonomic abundance data is the ability to interpret microbes that were identified to be significantly associated with a disease outcome, environmental condition, or host phenotype. This issue is commonly faced in various other high-throughput genomic data sets such as RNA sequencing. Gene set testing can help alleviate this challenge by focusing on cellular processes whose effect is the combination of a collection of genes. More importantly, this allows researchers to incorporate *apriori* biological knowledge into statistical analysis, thereby generating mechanistically meaningful hypotheses. Various researchers have begun to adapt these approaches to microbiome datasets.

However, there is limited research on curating and evaluating the annotation databases, such as MSigDB, that are required to apply these methods. Here in this manuscript, we constructed taxonomic sets using curated databases about microbial traits at the genus and species level, sourced from prior experimental studies. Traits represent biochemical and metabolic phenotypes (such as sulfate reduction) that can serve as annotations for potential ecosystem function. This sets our approach apart from other databases such as Disbiome or BugSigDB, which focus on generating microbe sets relating to pathogenicity and disease status. We assessed the overall coverage of our database in a typical microbiome data set and evaluated the utility of using set-based features in disease prediction. This manuscript would provide a novel approach to enrichment analysis in the microbiome space that would be of interest to readers of your journal.

We do not have any competing interests to declare. Please address all correspondence concerning this manuscript to me at quangpmnguyen@gmail.com

Thank you so much for your consideration.

Sincerely,

Quang P. Nguyen, Ph.D.