

A zero-inflated quantile approach (ZINQ) for powerful and robust microbiome data analysis

Wodan Ling

Ni Zhao

Anna M. Plantinga

Lenore Launer

Anthony A. Fodor

Katie A. Meyer

Michael Wu

Video Byte

Keywords: zero-inflated quantile-based approach, type I error control, microbiome differential abundance analysis, heterogeneity, Microbiome, microbiota, zero inflation, differentially abundant, bacterial taxa, statistics, relative abundance, false positive, non-parametric association testing, microbiome profiling, microbes, bacteria, heterogeneous effects, covariate adjustment, logistic regression, quantile regression

Posted Date: October 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-967301/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Studying disease- and treatment-associated changes in bacterial abundance can clarify microbiome-related pathogenic and therapeutic mechanisms. However, the existing tests are not ideal for microbiome data, as they often have high positive rates or lack of power. Most of them also perform poorly when heterogeneous effects are present, and some are ill-suited for more than two groups or covariate adjustment. As an alternative, researchers recently developed a non-parametric zero-inflated quantile approach (ZINQ), which uses a two-part model. In part 1, logistic regression determines whether the variable of interest is associated with taxon presence/absence. In part 2, on selected quantile levels, quantile regression on the non-zero observations tests whether the variable is associated with altered taxon abundance. Finally, the marginal p-values are combined, with a significant aggregate p-value indicating differential abundance. Simulations with real rarefied (a) and CSS-normalized (b) data revealed that ZINQ properly controls type I error and exhibits high power for identification of differentially abundant microbial taxa. The two differential abundance patterns exclusively detected by ZINQ illustrated its ability to identify heterogeneous effects, and the numerical assessment further confirmed ZINQ's power for heterogeneous associations between microbes and clinical variables. Although it may not be ideal for extremely low-frequency taxa or when library size is heavily confounded with the variable of interest, the findings show that ZINQ is a powerful and robust method for microbiome differential abundance analysis.