

A phylogenetic isometric log-ratio transforms for analysis of microbiota survey data.

J. D. Silverman¹, A. Washburne², S. Mukherjee², L. A. David³

¹Duke University, NC, USA; *justin.silverman@duke.edu*

²Duke University, NC, USA; *bigalculus@gmail.com*

³Duke University, NC, USA; *sayan@stat.duke.edu*

⁴Duke University, NC, USA; *lawrence.david@duke.edu*

Abstract

The human body is home to a complex ecosystem of microbes that play important roles in human health and disease. Recent technological advances have allowed biomedical researchers to investigate the composition of these communities by quantifying the relative amount of DNA attributable to different biological species in a sample. In addition, models of molecular evolution enable the evolutionary relationships between species to be inferred from DNA sequence data and represented in the form of a binary phylogenetic tree. It has been well established that this evolutionary information provides important insights into the structure and function of microbial communities. Here we use the phylogenetic tree as a sequential binary partition to define a phylogenetic isometric log-ratio transform which we term PhILR. We demonstrate how the PhILR transform overcomes challenges in the analysis of compositional microbiota data while enabling novel evolutionary analyses of these communities. To demonstrate the utility of this approach, we demonstrate how the PhILR transformed data can be used to find evolutionarily neighboring bacterial groups that may have recently differentiated to adapt to distinct body sites.