MSSQ

Multi-Sample Poisson Model for Species Abundance Quantification

Introduction

The human microbiome, which includes the collective microbes residing in or on the human body, has a profound influence on the human health. DNA sequencing technology has made the large-scale human microbiome studies possible by using shotgun metagenomic sequencing. One important aspect of data analysis of such metagenomic data is to quantify the bacterial abundances based on the metagenomic sequencing data. Existing methods almost always quantify such abundances one sample at a time, which ignore certain systematic differences in read coverage along the genomes due to GC contents, copy number variation and the bacterial origin of replication.

In order to account for such differences in read counts, we propose a multi-sample Poisson model to quantify microbial abundances based on read counts that are assigned to species-specific taxonomic markers. Our model takes into account the marker-specific effects when normalizing the sequencing count data in order to obtain more accurate quantification of the species abundances. Compared to currently available methods on simulated data and real data sets, our method has demonstrated an improved accuracy in bacterial abun- dance quantification, which leads to more biologically interesting results from downstream data analysis.

Installation

You can install our MSSQ package from Github

```r
install.packages("devtools")
devtools::install_github("chvlyl/MSSQ")
library(MSSQ)
```

Basic Usage

```r
est <- poisson_estimate_parameter_LS(X=X,tc=tc,l=l,N=N,S=S,
                                     species.names=species.names,
                                     estimate.phi=TRUE)
```