BRIGHAM AND WOMEN'S HOSPITAL

HARVARD MEDICAL SCHOOL

181 Longwood Avenue
Boston, Massachusetts 02115-5804

Department of Medicine
Channing Division of Network Medicine

Channing Microbiome Seminar

March 28, 2014, 11am @ 4th-floor conference room

Jonathan Friedman, PhD
MIT

Inferring Correlation Networks from Genomic Survey Data

Abstract: High-throughput sequencing based techniques, such as 16S rRNA gene profiling, have the potential to elucidate the complex inner workings of natural microbial communities – be they from the world's oceans or the human gut. A key step in exploring such data is the identification of dependencies between members of these communities, which is commonly achieved by correlation analysis. However, it has been known since the days of Karl Pearson that the analysis of the type of data generated by such techniques (referred to as compositional data) can produce unreliable results since the observed data take the form of relative fractions of genes or species, rather than their absolute abundances. Using simulated and real data from the Human Microbiome Project, we show that such compositional effects can be widespread and severe: in some real data sets many of the correlations among taxa can be artifactual, and true correlations may even appear with opposite sign. Additionally, we show that community diversity is the key factor that modulates the acuteness of such compositional effects, and develop a new approach, called SparCC (available at https://bitbucket.org/yonatanf/sparcc), which is capable of estimating correlation values from compositional data. To illustrate a potential application of SparCC, we infer a rich ecological network connecting hundreds of interacting species across 18 sites on the human body. Using the SparCC network as a reference, we estimated that the standard approach yields 3 spurious species-species interactions for each true interaction and misses 60% of the true interactions in the human microbiome data, and, as predicted, most of the erroneous links are found in the samples with the lowest diversity.

Ref: http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi....

Hosted by Yang-Yu Liu