

Joint Conference

International Conference on

ENVIRONMENTAL MICROBIOLOGY AND MICROBIAL ECOLOGY

&

International Conference on

ECOLOGY AND ECOSYSTEMS

September 18-20, 2017 Toronto, Canada



Gregory B Gloor

University of Western Ontario, Canada

Cautionary tales from the microbiome: Finding what is real and reproducible

Statement of the Problem: Commonly-used methods of analyzing microbiome or RNA-seq datasets can be misleading and all the available information in a consistent manner are not in use. These results in many analyses being dominated by either the most abundant, or the rarest features: In fact, it is often the case that the most abundant taxa dominate multivariate outputs, and the rarest taxa dominate univariate outputs in the same dataset. Furthermore, these datasets have extraordinary properties that make the use of correlation and network analysis problematic.

Methodology and Theoretical Orientation: Data collected using high throughput sequencing (HTS) methods are sequence reads mapped to genomic intervals, and are commonly analyzed as either normalized count data or relative abundance data. One reason for these normalizations is to attempt to compensate for the problem that the sequencing instrument imposes an upper bound on the number of sequence reads. Positive data with an arbitrary bound are compositional data and are subject to the problem of spurious correlation. Thus, ordination, clustering and network analysis become unreliable. A second problem is that the data are sparse: i.e., contain many 0 values. A third problem is that the largest measurement error is at the low count margins in these datasets.

Conclusion & Significance: We use microbiome datasets to show how Bayesian estimation combined with compositional data approaches that examine the ratios between taxa give robust insights into the structure and function of microbial communities. I will present example datasets drawn from the human and ecological domains and show that ordination, differential abundance and correlation can be interpreted in an internally consistent manner that provides reproducible insights.

Biography

Greg Gloor is a professor of biochemistry with broad experience in molecular biology, genetics and genomics. Most recently, he has developed tools to investigate fundamentals of molecular evolution, microbial ecology and meta-transcriptomics. He is currently working on developing and adapting principled methods to characterize correlation and differential abundance in sparse, high throughput sequencing data as generated in 16S rRNA gene sequencing surveys, meta-genomics and meta-transcriptomics. He is the developer and maintainer of the ALDEx2 R package on Bioconductor.

ggloor@uwo.ca

Notes: