

Title:

Study the co-occurrence patterns of Metagenomics high-throughput sequencing data

Authors:

John Clay Heern, Jason Causey, Jake Qualls, Xiuzhen Huang

Abstract:

High-throughput sequencing technologies are providing unprecedented opportunities to characterize the microbial organisms of various environments. Here we study the microbial co-occurrence patterns of taxa or operational taxonomic units (OTUs) in a sample with Metagenomics high-throughput sequencing data. We have conducted the following analysis:

1. Study the Metagenomics compositional data;
2. Evaluate the accuracy and efficiency performance of our new mathematical models versus existing methods, such as REBACCA;
3. Implement the construction of covariance matrices;
4. Apply our new models to mouse skin data that includes three groups of individuals: groups of nonimmunized (Control), immunized-healthy (Healthy) and immunized-diseased (EBA) individuals. And visualize the identified pair-wise correlations of the data.

Result:

Compared with previous methods, our new models have consistent performance on various real and simulated Metagenomics datasets, with commensurate accuracy and improved efficiency. We anticipated that our new methods will be very useful for analyzing Metagenomics high-throughput sequencing data.

*Funding information*

*REU support:*

*III: EAGER: Novel algorithms for de novo transcriptome assembly using RNA-seq data and for metagenome assembly*

*Program Director:*

*Aidong Zhang*

*IIS Division of Information & Intelligent Systems*