

Aging - human gut 16S - microbial correlation Networks - microbial co-occurrences - time series (relative to Aging) - functional profiles from correlates

Abstract

Bla bla bla

Introduction

Our aim is to identify a correlation between species withing ecological communities and their dynamics over time.

Materials and Methods

Counts and taxonomical assignment of OTU

Stool samples from 244 unrelated individuals were processed using... Lab methods. using "Mothur - Lotus" software (ref).

16S rRNA data was processed using the software Mothur v 1.34. (Schloss et al., 2009), to minimize effects of sequencing errors, we trimmed sequences that contained more than one undetermined nucleotide (N), reads that had more than eight homopolymers and those with the Q-score average below 25 in a window of 50 bp. Trimmed reads were aligned using the Needleman-Wunch algorithm and SILVA SEED reference alignment database (Schloss, 2009). To maximize the number of sequences that overlap over the longest span, an algorithm that allows keeping sequence that started after the position that 95% of the sequences was used. A pre-clustering step (Huse et al., 2010) and the Uchime algorithm in de novo mode (<http://drive5.com/uchime>) was included to reduce chimeras and sequencing noise. was used to detect (Gomez-Alvarez et al., 2009) and subsequently remove them. Afterward, the sequences were classified using Wang method (Wang et al., 1999) against SILVA reference database with 1000 iterations. The high-quality sequences were clustered into OTUs using the average algorithm (Schloss, 2011) based on 3% of dissimilarity cutoff.

-Output: OTU counts

Data treatment

The OTU counts revealed a high diversity environment where most OTU appears in a few individuals with low abundance and zero values dominate most count values. It is thought that this phenomenon is explained by nonsufficient sampling effort which zero values refer to unobserved values caused by the limited size of the sample. Additionally, this data structure comprises the relative abundance of the community components falling into special data class named as compositional data (Aitchison J 2003). Under the previous premises, the data was treated according to compositional techniques. These methods are based on the log-ratio methodology (Pawlowsky-Glahn et. al. 2011), but log-ratio transformations require data with positive values. Therefore, for replacing the zero counts by a proper value, we managed the data using the Bayesian-multiplicative approach (Pawlowsky-Glahn et. al. 2011-chap4).

A Bayesian estimation with Dirichlet distribution as a prior, $\theta = Dir(st)$ was used to replace the composition vector $x_i = (c_{i1}/N_i, \dots, c_{ik}/N_i)$ by the vector $r_i = (r_{i1}, \dots, r_{ik})$ (Pawlowsky-Glahn et. al. 2011-chap4), where N_i is the number of trials of the sample i (relative to sample read counts), and the outcomes from each trial falls in any of k mutually exclusive categories (OTU), with c_{ij} number of observations for category j , so then the transformed count r_{ij} is denoted by

$$r_{ij} = \begin{cases} \frac{t_{ij} s_i}{N_i + s_i} & \text{if } x_{ij} = 0, \\ x_{ij} \left(1 - \frac{s_i}{N_i + s_i} \sum_{k|x_{ik}=0} t_{ik} \right) & \text{if } x_{ij} > 0. \end{cases} \quad (1)$$

s_i and t_i stands for prior total strength and expectation of the probability distribution, $t_i = (t_{i1}, \dots, t_{ik})$, with $t_{ij} > 0$, $\sum_j t_{ij} = 1$ and t_i belongs to the simplex S^k (Aitchison J. 2003). In this study was used a uniform prior with strength based on square root Dirichlet model due to $\sqrt{N_i} > k$ (Martin-Fernandez 2015), so then $s_i = \sqrt{N_i}$ and $t_{ij} = 1/k$.

Microbial community analysis

The Inverse Simpson Index was used to estimate the OTU richness (other analyzes are coming). A Covariance matrix was calculated using SPIEC-EASI algorithm.

1. react

2. diffuse free particles
3. increment time by dt and go to 1

Results

Discussion

Supporting Information

References