Project on systematic construction of 3-Dimensional microbial co-occurrence network

Idea:

Micro-organisms exist together as a community rather than a single entity for a better stability of the microbiota. Microbes in a community communicate with each other through the exchange of information (metabolites). This information exchange can be captured by the microbial co-occurrence network. The analysis suggests that the co-presence of microbial organisms across all the samples could explain the dependency of the one organism to another / interacts with one another. Most of the network analysis (like SparCC) provide insights on the relation or correlation between the two microbes in a community. But the microbial species assembly in a community depends on all the microbes present. So, to better understand the inter-relationship between the microbes, higher dimensional co-occurrence network could be helpful [For instance, if the microbe A and B are co-occurring in ~ 90% of the samples and A and C are co-occurring in 95% of the samples. Higher dimensional analysis could report that the microbial co-presence is influenced by the presence of another micro-organism).

Methodology:

Dataset:

Relative microbial abundance data obtained from 16s rRNA sequencing and Whole Genome sequencing (WGS - Higher resolution gives the information about the strain of the microbes) of healthy and diabetic children were collected from MicrobiomeDB - DIABIMMUNE study

Workflow:=

- Preprocess the taxonomic and OTU count data [Matrix with a) Species*samples and b) Genus * samples]
- Normalize / satndardize the abundance table [Common standardization is relative abundance, log-normalization]
- The relative abundance table (Species and Genus-based) can be used to generate a co-occurrence network (Based only on the presence / absence of the taxa) and correlation (Spearman or Pearson) network (Based on the abundance of the taxa)
- Retain only the significant microbial co-occurrence or co-presence in a healthy and disease network
- 2-D microbial relationship is obtained

Idea-2:

[Multiple healthy and disease based microbial abundance data can be analyzed to study the more susceptible microbial co-occurrence (Loss of connection in disease compared to the healthy) in a healthy network irrespective of the disease]

- Once the 2-D microbial co-occurrences was obtained, 3-D (Impact of co-occurrence of 2 microbes in presence of other microbe) co-occurrence network need to be built

- This can include information theory or the probability - based approach [Ref.Paper: Identification of multidimensional Boolean patterns in microbial communities]
- Have to read more on the other methods to identify the inter-relationship

Outcome:

- Better analysis of the microbial interplay in the healthy condition and how it impacts or impacted by the disease occurrence