NIASRA
NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

UNIVERSITY OF WOLLONGONG
AUSTRALIA

Microbiome Data Analysis

A short course presented by
Olivier Thas (Hasselt University, Belgium, Ghent University, Belgium and University of Wollongong, Australia) and Stijn Hawinkel (Ghent University Belgium)

Monday and Tuesday, 4 - 5 February 2019
Building 21 Room G08
University of Wollongong Campus

Overview

High-throughput sequencing technologies allow easy characterisation of the microbiome, but the data analysis faces many particular issues and difficulties. The data analysis starts with the processing of the raw read counts to turn them into an OTU table. In this process, quality control, filtering and clustering into OTUs are essential steps. Once the OTU count table is ready, the choice of data analysis method depends on the research objectives, but very often a first visual data exploration is performed. Ordination methods, which often originate from ecology, are well suited for this purpose, but new methods tailored to microbiome data behave better for the over dispersed, zero inflated sequencing data. Taxon correlation networks can be helpful for the identification of co-occurring taxa. Formal statistical data analysis methods are required for identifying species that are differentially abundant between several conditions; again there is a need for special methods that can deal with over dispersion, zero-inflation, library size variability and potentially with the compositional nature of microbiome data. The data analysis becomes even more elaborated for longitudinal data when studying the evolution of the microbiome over time. These analyses may focus on either individual taxa or on diversity of the microbial community (richness, alpha and beta diversity ...).

The Course

This 2-day course provides an overview of the data analysis pipeline for microbiome data, particularly 16S rRNA amplicon sequencing data. The focus is on data visualisation, and testing for differential abundance and diversity. Lectures will be alternated with hands-on R sessions.

The course starts with a brief overview of the processing of raw reads data into an OTU table (including filtering, trimming and clustering into OTUs). We continue with summarising, exploring and plotting the high dimensional data with ordination and clustering methods. Next we focus on the estimation of diversity (including evenness, richness, and beta diversity) and relative abundances, while spending attention on normalisation issues. We discuss several methods for estimating taxon correlation networks and testing for differential abundance and diversity, including methods for longitudinal data analysis. Finally, we briefly illustrate methods for sample size calculation.

Methods included in course:

- Dada2 for clustering into OTUs
- PCoA, Coda and RCM ordination methods for visualisation
- Normalisation: CCS, TSS, TMM, RLE, rarefying,...
- diversity metrics: alpha diversity (richness, Shannon and Simpson index, ...), beta diversity (Bray-Curtis, Jaccard, UniFrac, Chao1, ...)
- taxon correlation networks: SparCC, SpiecEasi
- Testing for differential abundance: DESeq2, EdgeR, Limma-voom, ALDex2, SAMSeq,...

Participants are assumed to have a good basic knowledge of R.

NIASRA
NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

UNIVERSITY OF WOLLONGONG
AUSTRALIA

Participants will receive a printed copy of the notes and slides used in the presentations and of the example computer programs. Participants are required to bring a laptop with a recent version of R loaded. Loaner laptops may be available with sufficient prior notice. Additional material for R will be made available during the course.

Target Audience

This course is appropriate for any statistician, data scientist or biologist with a good basic statistics background and with experience with the R software.

The Instructor

Olivier Thas is a Professor of Biostatistics in the Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-Biostat) of Hasselt University (Belgium), and in the Department of Data Analysis and Mathematical Modelling of Ghent University (Belgium). He is also an Honorary Professorial Fellow in the National Institute for Applied Statistics Research Australia (NIASRA) at the University of Wollongong.
Olivier's research covers many aspects of nonparametric and semiparametric statistics, with a particular focus on hypothesis testing and statistical methods for high-throughput and high-dimensional genomics data, including microbiome and (sc)RNA-Seq. He is (co)author of more than 100 journal papers and 2 monographs.

Stijn Hawinkel is a PhD student at Ghent University (Belgium). Under the supervision of Olivier Thas and Prof. Luc Bijnens (Janssen Pharmaceutica, Johnson and Johnson, and Hasselt University) he works on the development and application of new data analysis methods for microbiome data.

Fees and Information

Course	Date	Fee	SSAI Members	Students
Microbiome Data Analysis	4 - 5 Feb	$880	$800	$440

Location: Early Start Building (Building 21) Room G08
Duration: Monday 4 Feb: *Registration at 9:00 am; course from 9.30 am to 4.30 pm*
Tuesday 5 Feb: *Course from 9:30 am to 4:30 pm*

Morning and afternoon coffee/tea and a sandwich lunch are included in the course fee.

Places are strictly limited and registrations will be processed as they are received.

To register and for further information please contact Michele Boatswain on micheleb@uow.edu.au or visit the NIASRA website at niasra.uow.edu.au