

Mini project report

Mike Nsubuga

6/4/2021

MSB7102 Mini-project, Semester 1, 2021

Background

This is an assessed exercise that will contribute to coursework for this module. It is aimed at providing an experience with all that has been covered throughout the semester. The tasks are based on two Bioconductor packages; phyloseq and DEseq2. Endeavor to look at the documentation and the links indicated below may be useful. The data files are located in the shared google drive folder.

Data sources and description

The data used in this exercise is derived and was generated by Kolistic et al in their study “The dynamics of the human infant gut microbiome in development and progression toward Type 1 Diabetes”. See full publication at <https://doi.org/10.1016/j.chom.2015.01.001>. Briefly, this was a prospective analysis of developing gut microbiome in infants en route to type 1 diabetes. Infants from Finland and Estonia were recruited at birth based on predisposition to autoimmunity determined by human leukocyte antigen (HLA) genotyping. The cohort consists of 33 infants, 11 of which seroconverted to serum autoantibody positivity and of those, four developed T1D within the three-year time-frame of this study.

Tasks

1. Import the data described above into R, provide descriptive summaries of the subject data (using appropriate graphics and statistical summary measures) given in the `diabimmune_16s_t1d_metadata.csv` file. In addition, use appropriate test(s) to check for association/independency between disease status and other variables (delivery mode, gender and age). Note that age is given in days.
2. Using phyloseq, create a phyloseq object. This will comprise the OTU abundance, taxonomy (provided in the .txt file) and sample data (provided in the .csv file).
3. Generate Alpha diversity plots and ordination plots. Examine any observed patterns by delivery mode, gender and disease status.
4. Perform differential abundance using DEseq2 ### Useful links
 - Importing data: <https://joey711.github.io/phyloseq/import-data.html>
 - Ordination: https://joey711.github.io/phyloseq/plot_ordination-examples.html
 - Alpha diversity: https://joey711.github.io/phyloseq/plot_richness-examples.html
 - Differential abundance: <http://joey711.github.io/phyloseq-extensions/DESeq2.html>

Descriptive summaries of the subject data using appropriate graphics and statistical summary measures

Using the summary function to get a summary of descriptive statistics which include mean, median, 25th and 75th quartiles, min, max

```
data <- read.csv("diabimmune_16s_t1d_metadata.csv");
summary(data)
```

```
## Sample_ID      Subject_ID      Case_Control      Gender
## Length:777     Length:777     Length:777     Length:777
## Class :character Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character Mode  :character
##
##
##
## Delivery_Route  Age_at_Collection
## Length:777     Min.      : 6.0
## Class :character 1st Qu.: 229.0
## Mode  :character Median : 452.0
##                      Mean   : 482.9
##                      3rd Qu.: 702.0
##                      Max.   :1233.0
```

Appropriate tests to check for association/independency between disease status and other variables

Phyloseq

Using phyloseq, create a phyloseq object. this will comprise the OTU abundance, taxonomy(provided in the .txt file) and sample data(provided in the .csv file)

```
summary(cars)
```

```
##      speed      dist
## Min.   : 4.0    Min.   : 2.00
## 1st Qu.:12.0    1st Qu.: 26.00
## Median :15.0    Median : 36.00
## Mean   :15.4    Mean   : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
## Max.   :25.0    Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.