Lab_10_Metagenomics_MBI3100A_2022_Assignment_Questions

2022-11-12

R Markdown

```
# clear the R environment
rm(list = ls())
```

Install required libraries

Installing these libraries may take some time. Try to update all other dependencies when prompted (type "a" and enter).

```
if (!require("BiocManager")) install.packages("BiocManager")
```

```
## Loading required package: BiocManager
```

```
if (!require("phyloseq")) BiocManager::install("phyloseq")
```

```
## Loading required package: phyloseq
```

```
if (!require("microbiomeMarker")) BiocManager::install("microbiomeMarkera")
```

```
## Loading required package: microbiomeMarker
```

```
## Registered S3 method overwritten by 'gplots':
##   method         from
##   reorder.factor DescTools
```

```
##
## Attaching package: 'microbiomeMarker'
```

```
## The following object is masked from 'package:phyloseq':
##
##     plot_heatmap
```

```
if (!require("tidyverse")) install.packages("tidyverse")
```

```
## Loading required package: tidyverse
```

```
## Found more than one class "atomicVector" in cache; using the first, from namespace 'Matrix'
## Also defined by 'Rmpfr'
## Found more than one class "atomicVector" in cache; using the first, from namespace 'Matrix'
## Also defined by 'Rmpfr'
## Found more than one class "atomicVector" in cache; using the first, from namespace 'Matrix'
## Also defined by 'Rmpfr'
## Found more than one class "atomicVector" in cache; using the first, from namespace 'Matrix'
## Also defined by 'Rmpfr'
## Found more than one class "atomicVector" in cache; using the first, from namespace 'Matrix'
## Also defined by 'Rmpfr'
## Found more than one class "atomicVector" in cache; using the first, from namespace 'Matrix'
## Also defined by 'Rmpfr'
## Found more than one class "atomicVector" in cache; using the first, from namespace 'Matrix'
## Also defined by 'Rmpfr'
## Found more than one class "atomicVector" in cache; using the first, from namespace 'Matrix'
## Also defined by 'Rmpfr'
## Found more than one class "atomicVector" in cache; using the first, from namespace 'Matrix'
## Also defined by 'Rmpfr'
## Found more than one class "atomicVector" in cache; using the first, from namespace 'Matrix'
## Also defined by 'Rmpfr'
## Found more than one class "atomicVector" in cache; using the first, from namespace 'Matrix'
## Also defined by 'Rmpfr'
## Found more than one class "atomicVector" in cache; using the first, from namespace 'Matrix'
## Also defined by 'Rmpfr'
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
if (!require("dendextend")) install.packages("dendextend")
```

```
## Loading required package: dendextend
## Registered S3 method overwritten by 'dendextend':
##   method     from
##   rev.hclust vegan
##
## ---------------------
## Welcome to dendextend version 1.16.0
## Type citation('dendextend') for how to cite the package.
##
## Type browseVignettes(package = 'dendextend') for the package vignette.
## The github page is: https://github.com/talgalili/dendextend/
##
## Suggestions and bug-reports can be submitted at: https://github.com/talgalili/dendextend/issues
## You may ask questions at stackoverflow, use the r and dendextend tags:
##   https://stackoverflow.com/questions/tagged/dendextend
##
##  To suppress this message use:  suppressPackageStartupMessages(library(dendextend))
## ---------------------
##
##
## Attaching package: 'dendextend'
##
## The following object is masked from 'package:stats':
##
##     cutree
```

Load Libraries

```
library(phyloseq)
library(ggplot2)
library(dplyr)
library(dendextend)
library(microbiomeMarker)
```

Data import

Question 1: Import the three files named as 'GP_sp_assignment_otu_table_df.csv', 'GP_sp_assignment_sample_data_df.csv', and 'GP_sp_assignment_tax_table_df.csv' and make a phyloseq object named 'asgmt_physeq'? (2 points)

Please provide the correct file/folder path

Print asgmt_physeq

Question 2: How many taxa, samples, and sample variables are there in asgmt_physeq? (1 point)

Question 3: List the catagories present in the sample variable "SampleOrigin". (1 point)

Question 4: Generate a bar plot at "Phylum" level for sample vs abundance and facet it for the categories in SampleOrigin (1 point)

Question 5: Based on the plot generated in question 4, name all the phylum which have a big difference in abundance between "Feces" and "Freshwater" samples? (1 point)

Transform the absolute abundance into relative abundance and filter the taxa which have mean relative abundance less than 0.0001

```
# To convert to relative abundance

# Keep the taxa which have a mean values at least 0.0001
```

Question 6: How many taxa are left after the above filtering? (1 point)

For question 7 to 12, use dataset 'asgmt_physeq'.

Question 7: Generate a Hierarchical clustering plot using the distance "ward.D2". (2 points)

It will be a four step process

Step1: Extract OTU table as data frame

```
# transpose the table (required by vegdist)
```

Step2: Transpose the table (required by vegdist package)

```
#compute Bray-Curtis dissimilarity
```

Step3: Compute Bray-Curtis dissimilarity

```
#Save as dendrogram

#Plot
```

Step4: Save as dendrogram

Question 8: Plot for alpha diversity using two measures, "Observed" and "Shannon". (1 points)

Question 9: Apply wilcox.test to see if the Observed diversity is significantly different for SampleOrigin. (2 points)

```
# Make a dataframe to combine the ouputs of Observed, Shannon and SampleOrigin
```

Step 1: Make a dataframe to combine the outputs of Observed and SampleOrigin.

```
#Wilcoxon test for Shannon diversity for categories in SampleOrigin
```

Step 2: Check the significance level for wilcox.test

```

```

Note down the p-value? is the difference significant i.e is p-value less than 0.05?

Question 10: Make a PCoA plot using the "bray" method as distance for beta diversity. (1 point)

Question 11: Apply DESeq2 method to identify the differentially abundant taxa based on SampleOrigin column. (1 point)

```
set.seed(2345)
# run_deseq2 command run the program deseq2 to identify DA taxa
# Running this command takes a few seconds
```

Question 12: Plot the differentially abundant taxa identified by deseq2 method . (1 point)

Bonus Question: plot a heatmap for the differentially abundant taxa identified by deseq2 method