Monte_carlo_pip

Cecilia Wang

5/4/2017

```r
library(ggplot2)
library(phyloseq)
library(plyr)
library(vegan)
```

```
## Loading required package: permute
```

```
## Loading required package: lattice
```

```
## This is vegan 2.5-2
```

```r
library("biom")
library(ggpubr)
```

```
## Loading required package: magrittr
```

```r
NZGL_taxonomy<-import_qiime_sample_data("~/PIP2018/primary_data/taxonomy.tsv")
  # the imported taxonomy data should have each sample as a row and each variable or taxonomy as a colu
Taxonomy_filter_file<-NZGL_taxonomy # make a copy
  #First make a plot of unfiltered taxonomy data, showing E coli abundance for each age group.
NZGL_taxonomy$time<-as.factor(NZGL_taxonomy$time) # to separate boxplot by different age category, type
### SEE PLOT 1: Supplemental 1: Abundance of E. coli x stratified by age in unfiltered data (chunk3)

# Then filter those samples out of all data, and use these data for every downstream analysis.

# select the useful part and find the interquartile range for E. coli, filter out samples that E. coli
Taxonomy_filter_file$E_coli<-NZGL_taxonomy$k__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Ente
# Split the dataset by time/age
E_coli_abundance_AtBirth<-subset(Taxonomy_filter_file, time==0)
E_coli_abundance_3_month<-subset(Taxonomy_filter_file, time==3)
E_coli_abundance_12_month<-subset(Taxonomy_filter_file, time==12)
E_coli_abundance_24_month<-subset(Taxonomy_filter_file, time==24)
# Calculate IQR by each time
E_coli_abundance_IOR_AtBirth<-IQR(E_coli_abundance_AtBirth$E_coli)
E_coli_abundance_IQR_3_month<-IQR(E_coli_abundance_3_month$E_coli)
E_coli_abundance_IQR_12_month<-IQR(E_coli_abundance_12_month$E_coli)
E_coli_abundance_IQR_24_month<-IQR(E_coli_abundance_24_month$E_coli)
# Filter the whole dataset at each time on E.coli > 1.5IQR
Taxonomy_filtered_AtBirth<-subset(E_coli_abundance_AtBirth, E_coli<=(1.5*E_coli_abundance_IOR_AtBirth))
Taxonomy_filtered_3_month<-subset(E_coli_abundance_3_month, E_coli<=(1.5*E_coli_abundance_IQR_3_month))
Taxonomy_filtered_12_month<-subset(E_coli_abundance_12_month, E_coli<=(1.5*E_coli_abundance_IQR_12_mont
Taxonomy_filtered_24_month<-subset(E_coli_abundance_24_month, E_coli<=(1.5*E_coli_abundance_IQR_24_mont
Taxonomy_filtered<-rbind(Taxonomy_filtered_AtBirth,Taxonomy_filtered_3_month,Taxonomy_filtered_12_month
```

Plot4: Our power to detect effects

```r
#Monte Carlo for BUG and MODULE beta diversity: C-section, time, antibiotics, treatment #bf
# source the functions for power test use Monte-carlo simulations
source("~/PIP2018/src/Monte_carlo_power_test.R")
```

```
## Loading required package: tcltk
```

```r
Meta_pip<-as.data.frame(Taxonomy_filtered[,c(1:28)])
# NOTE: any category that tested here need to be factor. Therefore, transform every category as factors
Meta_pip[]<- lapply(Meta_pip, function(x){as.factor(x)})
# load metadata (pcl file does not have the info)
Metadata_pip<-read.csv("~/PIP2018/primary_data/metadata.csv")
# assign BF info to Meta_pip and reset the subsets
Meta_pip$ageanyformula<-Metadata_pip$ageanyformula[match(Meta_pip$Studyid,Metadata_pip$Studyid)]
####================Taxonomy test===================####
## separate relative abundance data based on their associated metadata by time
t<-as.data.frame(t(Taxonomy_filtered[,c(-1:-28)]))
# subset metadata for further use of time stratefication
Meta_at_birth<-subset(Meta_pip,time==0)
Meta_3_month<-subset(Meta_pip, time ==3)
Meta_12_month<-subset(Meta_pip, time ==12)
Meta_24_month<-subset(Meta_pip, time ==24)
Monte_carlo_time<-Power_CI_figure(t,seq(5,50,5), Meta_pip, "time", 30, 10 )

pd <- position_dodge(0.1)
time_tax<-ggplot(Monte_carlo_time,aes(Sampling_depth,mean)) + geom_errorbar(aes(ymin=CI_low, ymax=CI_hig
#ggsave("~/PIP2018/results/time_tax.pdf", plot = last_plot(), device = NULL, path = NULL,
 # scale = 1, width = 9, height = 6, units = c("in", "cm", "mm"),
#  dpi = 300, limitsize = FALSE)

## caesar
 # figure out maximum sampling depth
summary(Meta_at_birth$caesar)
```

```
##  0  1
## 89 49
```

```r
summary(Meta_3_month$caesar)
```

```
##  0  1
## 90 33
```

```r
summary(Meta_12_month$caesar)
```

```
##   0   1
## 114  45
```

```r
summary(Meta_24_month$caesar)
```

```
##  0  1
## 75 23
```

```r
caeser_AB<-Power_CI_figure(t, seq(5,45,5), Meta_at_birth, "caesar", 30, 10 )
caeser_AB$age<-"0"
caeser_3month<-Power_CI_figure(t, seq(5,30,5), Meta_3_month, "caesar", 30, 10 )
caeser_3month$age<-"3"
```

```
caeser_12month<-Power_CI_figure(t, seq(5,45,5), Meta_12_month, "caesar", 30, 10 )
caeser_12month$age<-"12"
caeser_24month<-Power_CI_figure(t, seq(5,20,5), Meta_24_month, "caesar", 30, 10 )
caeser_24month$age<-"24"

Caesar_all_time<-rbind(caeser_AB,caeser_3month,caeser_12month,caeser_24month)
Caesar_all_time$age<-as.factor(Caesar_all_time$age)

caesar_tax<-ggplot(Caesar_all_time,aes(Sampling_depth,mean, colour=age)) + geom_errorbar(aes(ymin=CI_low
#ggsave("~/PIP2018/results/caesartax.pdf", plot = last_plot(), device = NULL, path = NULL,
#  scale = 1, width = 9, height = 6, units = c("in", "cm", "mm"),
#  dpi = 300, limitsize = FALSE)

## Antibiotics before 3 months
  # figure out maximum sampling depth
summary(Meta_at_birth$Antibiotics_before_6_months)
```

```
##   0   1
## 117  21
```

```
summary(Meta_3_month$Antibiotics_before_6_months)
```

```
##   0   1
## 107  16
```

```
summary(Meta_12_month$Antibiotics_before_6_months)
```

```
##   0   1
## 138  21
```

```
summary(Meta_24_month$Antibiotics_before_6_months)
```

```
##  0  1
## 86 12
```

```
Anti6_AB<-Power_CI_figure(t,seq(5,20,5),Meta_at_birth, "Antibiotics_before_6_months", 30, 10 )
Anti6_AB$age<-"0"
Anti6_3month<-Power_CI_figure(t,seq(5,15,5),Meta_3_month, "Antibiotics_before_6_months", 30, 10 )
Anti6_3month$age<-"3"
Anti6_12month<-Power_CI_figure(t,seq(5,20,5),Meta_12_month, "Antibiotics_before_6_months", 30, 10 )
Anti6_12month$age<-"12"
Anti6_24month<-Power_CI_figure(t,seq(5,10,5),Meta_24_month, "Antibiotics_before_6_months", 30, 10 )
Anti6_24month$age<-"24"

Anti6_all_time<-rbind(Anti6_AB,Anti6_3month,Anti6_12month,Anti6_24month)
Anti6_all_time$age<-as.factor(Anti6_all_time$age)
abxtax<-ggplot(Anti6_all_time,aes(Sampling_depth,mean, colour=age)) + geom_errorbar(aes(ymin=CI_low, yma

#ggsave("~/PIP2018/results/abx-tax.pdf", plot = last_plot(), device = NULL, path = NULL,
 # scale = 1, width = 9, height = 6, units = c("in", "cm", "mm"),
#  dpi = 300, limitsize = FALSE)

## Treatment/Studygroup
  # figure out maximum sampling depth
summary(Meta_at_birth$studygroup)
```

```
## bifido DR10 lactob DR20      placeb
##          14          58          66
```

```
summary(Meta_3_month$studygroup)
```

```
## bifido DR10 lactob DR20      placeb
##           9          51          63
```

```
summary(Meta_12_month$studygroup)
```

```
## bifido DR10 lactob DR20      placeb
##          16          67          76
```

```
summary(Meta_24_month$studygroup)
```

```
## bifido DR10 lactob DR20      placeb
##           7          43          48
```

```
  # bifido group has very limited number of samples
  # test with lactob and placeb groups
Studygroup_meta<-subset(Meta_pip, studygroup != "bifido DR10")
# reset the factor category studygroup
Studygroup_meta$studygroup<-as.factor(as.character(Studygroup_meta$studygroup))
SGmeta_AB<-subset(Studygroup_meta, time =="0")
SGmeta_3month<-subset(Studygroup_meta, time == "3")
SGmeta_12month<-subset(Studygroup_meta, time == "12")
SGmeta_24month<-subset(Studygroup_meta, time =="24")

SG_AB<-Power_CI_figure(t, seq(5,55,5), SGmeta_AB, "studygroup", 30, 10 )
SG_AB$age<-"0"
SG_3month<-Power_CI_figure(t, seq(5,50,5), SGmeta_3month, "studygroup", 30, 10 )
SG_3month$age<-"3"
SG_12month<-Power_CI_figure(t, seq(5,65,5), SGmeta_12month, "studygroup", 30, 10 )
SG_12month$age<-"12"
SG_24month<-Power_CI_figure(t, seq(5,40,5), SGmeta_24month, "studygroup", 30, 10 )
SG_24month$age<-"24"

SG_all_time<-rbind(SG_AB,SG_3month,SG_12month,SG_24month)
SG_all_time$age<-as.factor(as.character(SG_all_time$age))
tax_studygroup<-ggplot(SG_all_time,aes(Sampling_depth,mean, colour=age)) + geom_errorbar(aes(ymin=CI_lo
#ggsave("~/PIP2018/results/tax-studygroup.pdf", plot = last_plot(), device = NULL, path = NULL,
 # scale = 1, width = 9, height = 6, units = c("in", "cm", "mm"),
#  dpi = 300, limitsize = FALSE)
```

```
####================Module test===================####
# read the filtered modules file and modify it to fit the requirement for power test
Module<-import_qiime_sample_data("~/PIP2018/primary_data/modules.pcl")
Module<-as.data.frame(t(Module))
```

```
## Warning in class(X) <- NULL: Setting class(x) to NULL; result will no
## longer be an S4 object
```

```
Module_filtered<-Module[rownames(Module)%in%Taxonomy_filtered$Sample,]
t_modules<-Module_filtered[,c(-1:-27)] # remove metadata on top of the datset
t_modules<-as.data.frame(t(t_modules))
t_modules[]<- lapply(t_modules, function(x){as.numeric(as.character(x))})
t_modules[is.na(t_modules)]<-0 # replaced all the NA element in the dataset to 0

# Time test for modules
```

```
Monte_carlo_time_module<-Power_CI_figure(t_modules,seq(5,50,5), Meta_pip, "time", 30, 10 )
pd <- position_dodge(0.1)
modtime<-ggplot(Monte_carlo_time_module,aes(Sampling_depth,mean)) + geom_errorbar(aes(ymin=CI_low, ymax=

#ggsave("~/PIP2018/results/mod-time.pdf", plot = last_plot(), device = NULL, path = NULL,
 # scale = 1, width = 9, height = 6, units = c("in", "cm", "mm"),
#  dpi = 300, limitsize = FALSE)

## caesar
caeser_AB_module<-Power_CI_figure(t_modules, seq(5,45,5), Meta_at_birth, "caesar", 30, 10 )
caeser_AB_module$age<-"0"
caeser_3month_module<-Power_CI_figure(t_modules, seq(5,30,5), Meta_3_month, "caesar", 30, 10 )
caeser_3month_module$age<-"3"
caeser_12month_module<-Power_CI_figure(t_modules, seq(5,45,5), Meta_12_month, "caesar", 30, 10 )
caeser_12month_module$age<-"12"
caeser_24month_module<-Power_CI_figure(t_modules, seq(5,20,5), Meta_24_month, "caesar", 30, 10 )
caeser_24month_module$age<-"24"

Caesar_all_time_module<-rbind(caeser_AB_module,caeser_3month_module,caeser_12month_module,caeser_24montl
Caesar_all_time_module$age<-as.factor(Caesar_all_time_module$age)

mod_caesar<-ggplot(Caesar_all_time_module,aes(Sampling_depth,mean, colour=age)) + geom_errorbar(aes(ymin
#ggsave("~/PIP2018/results/mod-caesar.pdf", plot = last_plot(), device = NULL, path = #NULL, scale = 1,

## Antibiotics before 6 months
Anti6_AB_module<-Power_CI_figure(t_modules,seq(5,20,5),Meta_at_birth, "Antibiotics_before_6_months", 30
Anti6_AB_module$age<-"0"
Anti6_3month_module<-Power_CI_figure(t_modules,seq(5,15,5),Meta_3_month, "Antibiotics_before_6_months",
Anti6_3month_module$age<-"3"
Anti6_12month_module<-Power_CI_figure(t_modules,seq(5,20,5),Meta_12_month, "Antibiotics_before_6_months
Anti6_12month_module$age<-"12"
Anti6_24month_module<-Power_CI_figure(t_modules,seq(5,10,5),Meta_24_month, "Antibiotics_before_6_months
Anti6_24month_module$age<-"24"

Anti6_all_time_module<-rbind(Anti6_AB_module,Anti6_3month_module,Anti6_12month_module,Anti6_24month_modu
Anti6_all_time_module$age<-as.factor(Anti6_all_time_module$age)
mod_anti<-ggplot(Anti6_all_time_module,aes(Sampling_depth,mean, colour=age)) + geom_errorbar(aes(ymin=C

#ggsave("~/PIP2018/results/mod-anti.pdf", plot = last_plot(), device = NULL, path = NULL, #scale = 1, w
```

```
## Studygroup
SG_AB_module<-Power_CI_figure(t_modules, seq(5,55,5), SGmeta_AB, "studygroup", 30, 10 )
SG_AB_module$age<-"0"
SG_3month_module<-Power_CI_figure(t_modules, seq(5,50,5), SGmeta_3month, "studygroup", 30, 10 )
SG_3month_module$age<-"3"
SG_12month_module<-Power_CI_figure(t_modules, seq(5,65,5), SGmeta_12month, "studygroup", 30, 10 )
SG_12month_module$age<-"12"
SG_24month_module<-Power_CI_figure(t_modules, seq(5,40,5), SGmeta_24month, "studygroup", 30, 10 )
SG_24month_module$age<-"24"

SG_all_time_module<-rbind(SG_AB_module,SG_3month_module,SG_12month_module,SG_24month_module)
SG_all_time_module$age<-as.factor(as.character(SG_all_time_module$age))
mod_studygroup<-ggplot(SG_all_time_module,aes(Sampling_depth,mean, colour=age)) + geom_errorbar(aes(ymin
#ggsave("~/PIP2018/results/mod-studygroup.pdf", plot = last_plot(), device = NULL, path = #NULL, scale
cfig<-ggarrange(caesar_tax, mod_caesar, labels = c("A", "B"), common.legend = TRUE, legend="bottom" )
ggsave("~/PIP2018/results/fig4.pdf", plot = last_plot(), device = NULL, path = NULL,
  scale = 1, width = 9, height = 6, units = c("in", "cm", "mm"),
  dpi = 300, limitsize = FALSE)

supfig5<-ggarrange(time_tax, modtime, tax_studygroup, mod_studygroup, abxtax, mod_anti, labels=c("A", "
ggsave("~/PIP2018/results/supfig5.pdf", plot = last_plot(), device = NULL, path = NULL,
  scale = 1, width = 9, height = 6, units = c("in", "cm", "mm"),
  dpi = 300, limitsize = FALSE)
```