

БИ ДЗ-6 Микробиом

1. Kraken2 report

Data: /srv/common/microbiome/COLD_r09_F015_85.txt

```
cp ./common/microbiome/COLD_r09_F015_85.txt ./common/K
```

```
cd ./common/K
```

Теория (что в столбце №4):

U - Unclassified

R - Root: Обычно это обозначает основу или общее основание классификации.

D - Domain

P - Phylum

C - Class

O - Order

F - Family

G - Genus

S - Species

Какой это формат выходного файла?

… .kraken

Сколько микроорганизмов было определено с точностью до вида (и глубже)?

```
awk '{print $4}' COLD_r09_F015_85.txt | grep "S"
```

```
(base) jupyter-gorovenko-e@hse-students:~/common/K$ awk  
'  
{print $4}' COLD_r09_F015_85.txt | grep "S" | wc -l  
1032
```

… 1032

На какой определённый вид приходится больше всего фрагментов?

```
awk '$4 ~ /^S/ { if ($5 > max) { max = $5; name = $6 } } END { print name " - "  
" max}' COLD_r09_F015_85.txt
```

```
(base) jupyter-gorovenko-e@hse-students:~/common/K$ awk  
'  
$4 ~ /^S/ { if ($5 > max) { max = $5; name = $6 } } END {  
print name " - " max}' COLD_r09_F015_85.txt  
Mesorhizobium - 2725666
```

0.00	10	2	G	68287	Mesorhizobium
0.00	6	1	G1	325217	unclassified Mesorhizobium
0.00	1	1	S	2725666	Mesorhizobium sp. NIBRBAC000500504
0.00	1	1	S	2493668	Mesorhizobium sp. M9A.F.Ca.ET.002.03.1.2
0.00	1	1	S	2493671	Mesorhizobium sp. M2A.F.Ca.ET.043.05.1.1
0.00	1	1	S	2493676	Mesorhizobium sp. M3A.F.Ca.ET.080.04.2.1
0.00	1	1	S	2493681	Mesorhizobium sp. M7A.F.Ce.TU.012.03.2.1
0.00	2	0	S	593909	Mesorhizobium opportunistum
0.00	2	2	S1	536019	Mesorhizobium opportunistum WSM2075

… Mesorhizobium sp. NIBRBAC000500504 - 2725666 фрагментов

На какой вид бактерии приходится больше всего фрагментов?

… Mesorhizobium sp. NIBRBAC000500504 - 2725666 фрагментов (это определенно бактерия)

Какова доля фрагментов, не отнесённых с какому-либо таксону?

```
awk '$4 ~ /U/ {print $4}' COLD_r09_F015_85.txt | wc -l
```

1 - столько неклассифицированных

```
awk '$1' COLD_r09_F015_85.txt | wc -l
```

217 - столько всего строк

```
(base) jupyter-gorovenko-e@hse-students:~/common/K$ awk '$4 ~ /U/ {print $4}' COLD_r09_F015_85.txt | wc -l
1
(base) jupyter-gorovenko-e@hse-students:~/common/K$ awk '$1' COLD_r09_F015_85.txt | wc -l
217
```

… 1/217 = 0,005 (примерно такая доля неклассифицированных)

Сколько фрагментов было отнесено к роду Acetobacter, но ни к одному из конкретных видов Acetobacter?

39.75	173393	152065	G	434	Acetobacter
1.33	5782	5779	S	104102	Acetobacter tropicalis
0.00	3	3	S1	749388	Acetobacter tropicalis NBRC 101654
1.15	5029	5029	S	446692	Acetobacter senegalensis
0.73	3183	0	G1	151157	Acetobacter subgen. Acetobacter
0.73	3183	3176	S	435	Acetobacter aceti
0.00	7	7	S1	887700	Acetobacter aceti NBRC 14818
0.37	1631	530	S	438	Acetobacter pasteurianus
0.19	844	844	S1	1266844	Acetobacter pasteurianus 386B
0.05	200	125	S1	481145	Acetobacter pasteurianus subsp. pasteurianus
0.02	75	75	S2	634452	Acetobacter pasteurianus IF0 3283-01
0.01	57	57	S1	1006554	Acetobacter pasteurianus NBRC 101655
0.35	1530	1530	S	65959	Acetobacter pomorum
0.27	1176	0	G1	2628570	unclassified Acetobacter
0.22	939	939	S	2259883	Acetobacter sp. JWB
0.05	236	236	S	2592655	Acetobacter sp. KACC 21233
0.00	1	1	S	667111	Acetobacter sp. PG-2009
0.22	956	956	S	1633874	Acetobacter oryzifermentans
0.16	713	713	S	1076596	Acetobacter persici
0.13	547	547	S	481146	Acetobacter ascendens
0.09	377	377	S	431306	Acetobacter ghanensis
0.05	217	217	S	146474	Acetobacter orientalis
0.04	169	169	S	2500548	Acetobacter oryzoeni
0.00	7	7	S	85325	Acetobacter sicerae
0.00	4	0	G1	114713	environmental samples
0.00	4	4	S	114714	uncultured Acetobacter sp.
0.00	3	3	S	104100	Acetobacter lovaniensis
0.00	1	1	S	1076594	Acetobacter okinawensis
0.00	1	1	S	104101	Acetobacter indonesiensis
0.00	1	1	S	178901	Acetobacter malorum
0.00	1	1	S	104099	Acetobacter orleanensis

… 3 (тут проще глазами посмотреть, чем код придумывать)

2. Qiime2

We are using human microbiome samples from two individuals at four body sites at five timepoints, one of the individuals used antibiotics. This data came from the hypervariable region 4 (V4) of 16S rRNA and sequenced on an Illumina HiSeq.

Data: /srv/common/microbiome/emp-single-end-sequences/sequences.fastq.gz

/srv/common/microbiome/emp-single-end-sequences/barcodes.fastq.gz

Metadata: /srv/common/microbiome/sample-metadata.tsv

Activate environment with Qiime2

```
cd
conda init
source ~/.bashrc
conda activate qiime2-amplicon-2024.10
```

```
(base) jupyter-gorovenko-e@hse-students:~$ conda init
no change      /opt/tljh/user/condabin/conda
no change      /opt/tljh/user/bin/conda
no change      /opt/tljh/user/bin/conda-env
no change      /opt/tljh/user/bin/activate
no change      /opt/tljh/user/bin/deactivate
no change      /opt/tljh/user/etc/profile.d/conda.sh
no change      /opt/tljh/user/etc/fish/conf.d/conda.fish
no change      /opt/tljh/user/shell/condabin/Conda.psml
no change      /opt/tljh/user/shell/condabin/conda-hook.ps1
no change      /opt/tljh/user/lib/python3.10/site-packages/xontrib/conda.xsh
no change      /opt/tljh/user/etc/profile.d/conda.csh
no change      /home/jupyter-gorovenko-e/.bashrc
No action taken.

(base) jupyter-gorovenko-e@hse-students:~$ source ~/.bashrc
(base) jupyter-gorovenko-e@hse-students:~$ conda activate qiime2-amplicon-2024.10
QIIME is caching your current deployment for improved performance. This may take a few moments and should only happen once per deployment.
Matplotlib is building the font cache; this may take a moment.
(qiime2-amplicon-2024.10) jupyter-gorovenko-e@hse-students:~$ █
```

Input data

1. How many reads in /srv/common/microbiome/emp-single-end-sequences/sequences.fastq.gz file?

```
cp ./common/microbiome/emp-single-end-sequences/sequences.fastq.gz ./common/K  
  
cd ./common/K  
  
gunzip sequences.fastq.gz  
  
head sequences.fastq  
  
grep -c '^@' sequences.fastq
```

... 304472 ридов

2. Is it single-end or paired-end fastq file?

Теория:

```
@HWI-EAS440_0386:1:23:14818:1533#0/1
```

Это уникальный идентификатор чтения. Перед #0/1 указывается номер чтения из пары (если это парное чтение, то /1 — первый конец, а /2 — второй).

Если ты видишь двойные записи с /1 и /2, значит, это **paired-end**.

Если записи идут по одиночке, это **single-end**.

```
(qiime2-amplicon-2024.10) jupyter-gorovenko-e@hse-students:~/common/K$ grep -c "#0/1" sequences.fastq
302581
(qiime2-amplicon-2024.10) jupyter-gorovenko-e@hse-students:~/common/K$ grep -c "#0/2" sequences.fastq
0
```

… это single-end, так как везде 0/1 (2 000 ридов какие-то левые)

3. QIIME uses QIIME artifacts as inputs to other commands which contain information about the type of data and the source of the data. So, the first thing we need to do is import these sequence data files into a QIIME 2 artifact

```
cd

cp -r ./common/microbiome/emp-single-end-sequences ./common/K

cd ./common/K

qiime tools import \
  --type 'EMPSingleEndSequences' \
  --input-path emp-single-end-sequences \
  --output-path emp-single-end-sequences/emp-single-end-sequences.qza
```

Demultiplexing sequences

4. To demultiplex sequences we need to know which barcode sequence is associated with each sample. This information is contained in the sample metadata file. Find the column name in sample metadata and put it instead of ?

```
cd emp-single-end-sequences

qiime demux emp-single --i-seqs emp-single-end-sequences.qza --m-barcodes-file ~/common/microbiome/sample-metadata.tsv --m-barcodes-column barcode-sequence --o-per-sample-sequences demux.qza --o-error-correction-details demux-details.qza
```

```
(qiime2-amplicon-2024.10) jupyter-gorovenko-e@hse-students:~/common/K/emp-single-end-sequences$ qiime demux emp-single --i-seqs emp-single-end-sequences.qza --m-barcodes-file ~/common/microbiome/sample-metadata.tsv --m-barcodes-column barcode-sequence --o-per-sample-sequences demux.qza --o-error-correction-details demux-details.qza
Saved SampleData[SequencesWithQuality] to: demux.qza
Saved ErrorCorrectionDetails to: demux-details.qza
```

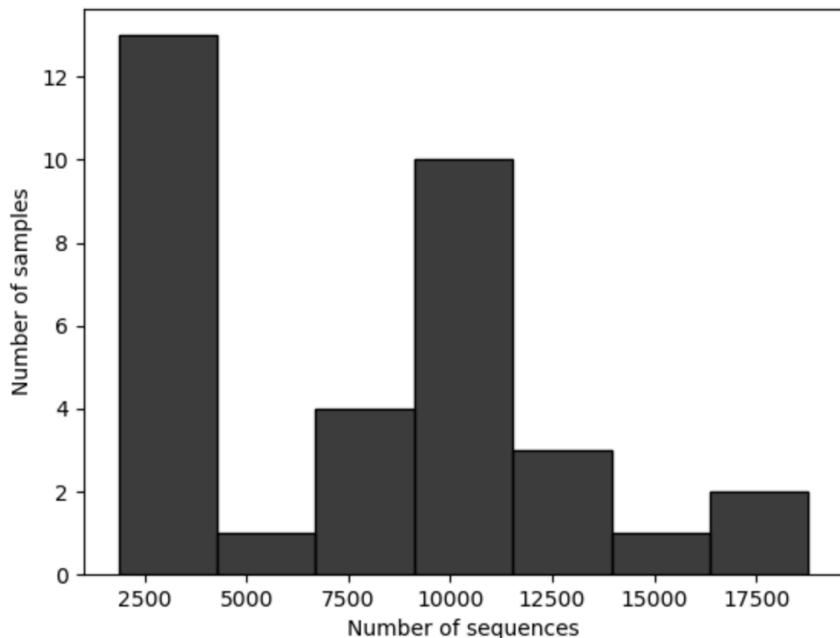
The demux.qza QIIME 2 artifact will contain the demultiplexed sequences. The second output (demux- details.qza) presents Golay error correction details, and will not be explored in this task.

5. After demultiplexing, it's useful to generate a summary of the demultiplexing results.

```
qiime demux summarize --i-data demux.qza --o-visualization demux.qzv
```

We cannot easily open qiime artifcat. Let's visualise demux.qzv using <https://view.qiime2.org/>

! Provide a screenshot of forward reads frequency histogram. What number of sequences occur **second most** among samples?



… 10 000

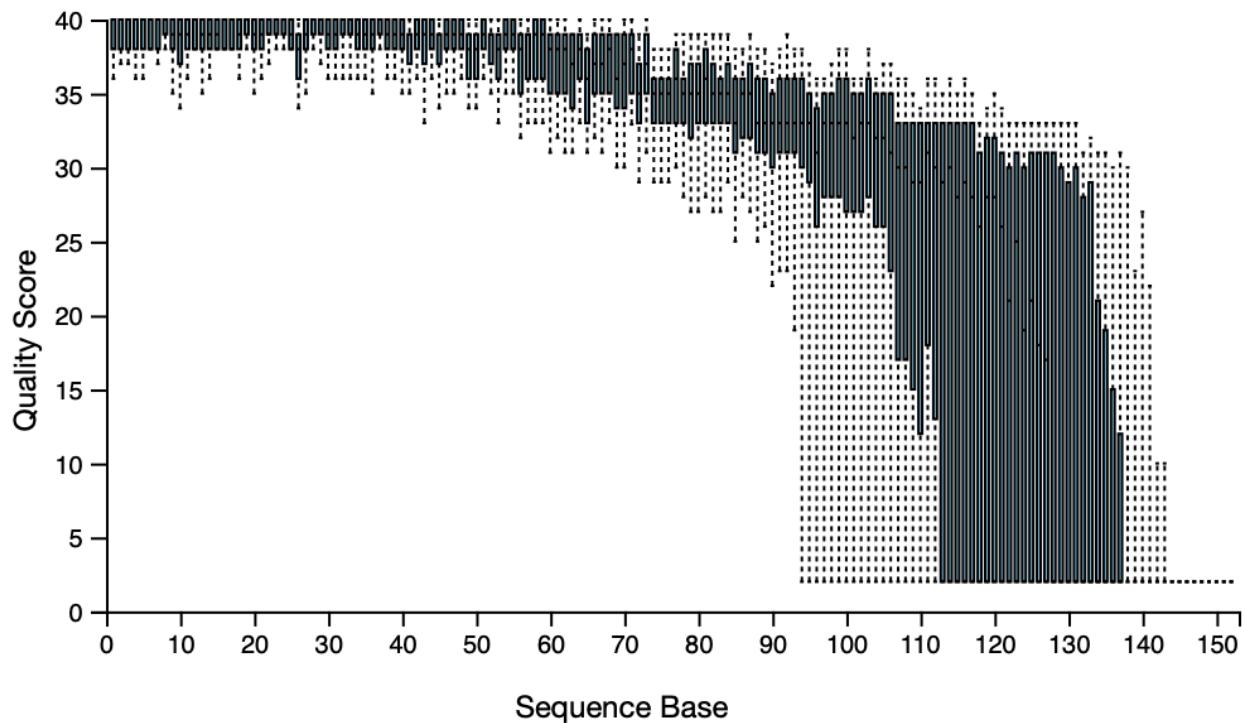
QC and feature table construction (relative abundance)

6. We are using tool DADA2 for this step. It will perform all the quality control for us and build table of relative abundance, which contains counts (frequencies) of each unique sequence in each sample in the dataset

For QC we will trim reads at the positions where quality of bases drops off (the flag --p-trunc-len n which truncates each sequence at position n)

Before running command, review the Interactive Quality Plot tab of the demux.qzv file in <https://view.qiime2.org/>. At what position the quality seems to drop off? Write the number instead of question mark

Forward Reads



… 105 позиция (примерно) - дроп оф

This next command may take up to 10 minutes to run, and is the slowest step in this task.

```
qiime dada2 denoise-single --i-demultiplexed-seqs demux.qza --p-trim-left 0  
--p-trunc-len 105 --o-representative-sequences rep-seqs-dada2.qza --o-table  
table.qza --o-denosing-stats stats.qza
```

Я сделяль:

```
ences$ qiime dada2 denoise-single --i-demultiplexed-seqs demux.qza --p-trim-left 0 --p-trunc-len 105 --o-representative-sequences rep-seqs-dada2.qza --o-table table.qza --o-denosing-stats stats.qza  
Saved FeatureTable[Frequency] to: table.qza  
Saved FeatureData[Sequence] to: rep-seqs-dada2.qza  
Saved SampleData[DADA2Stats] to: stats.qza
```

7. Command for generating file for visualization.

```
qiime metadata tabulate --m-input-file stats.qza --o-visualization stats.qzv
```

```
(qiime2-amplicon-2024.10) jupyter-gorovenko-e@hse-students:~/common/K/eqiime metadata tabulate --m-input-file stats.qza --o-visualization stats.qzv -visualization stats.qzv  
Saved Visualization to: stats.qzv
```

Чтобы ответить на последующие вопросы, делаю это:

Иду сюда <https://view.qiime2.org/> и вставляю файл stats.qzv (просто посмотреть, что там)

```
qiime tools export --input-path stats.qzv --output-path ./stats
```

```
# это типа разархивируем этот файл
```

```
nano ./stats/calculations.py
```

```
# в этот файл напишу код на питоне, который все посчитает
```

```
Этот код пишу в файле calculations.py
```

```
import pandas as pd

data = pd.read_csv('~/common/K/emp-single-end-sequences/stats/metadata.tsv',
sep='\t')

data.iloc[1:, 1] = pd.to_numeric(data.iloc[1:, 1], errors='coerce')
data.iloc[1:, 2] = pd.to_numeric(data.iloc[1:, 2], errors='coerce')
data.iloc[1:, 5] = pd.to_numeric(data.iloc[1:, 5], errors='coerce')

input_mean = round(data.iloc[1:, 1].mean(), 2)
filtered_mean = round(data.iloc[1:, 2].mean(), 2)
non_chim_mean = round(data.iloc[1:, 5].mean(), 2)

print(f'input_mean: {input_mean}')
print(f'filtered_mean: {filtered_mean}')
print(f'non_chim_mean: {non_chim_mean}')
```

```
python ./stats/calculations.py
```

```
ences$ python stats/calculations.py
input_mean: 7762.68
filtered_mean: 6337.41
non_chim_mean: 6173.74
```

What is mean number of input reads?

… 7762.68

What is the mean number of filtered reads?

… 6337.41

What is the mean percentage of input non-chimeric reads?

… 6173.74

Summaries of FeatureTable and FeatureData

8. After the quality filtering step completes, the feature-table summarize command will give you information on how many sequences are associated with each sample and with each feature, histograms of those distributions, and some related summary statistics. The feature-table tabulate-

seqs command will provide a mapping of feature IDs to sequences, and provide links to easily BLAST each sequence against the NCBI nt database.

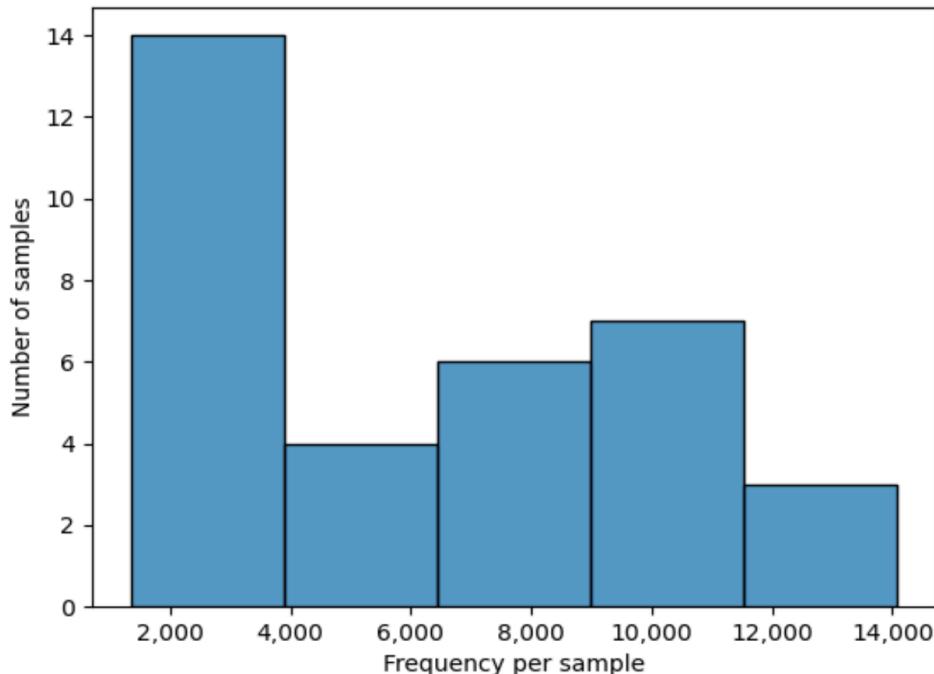
```
qiime feature-table summarize --i-table table.qza --o-visualization table.qzv  
--m-sample-metadata-file ~/common/microbiome/sample-metadata.tsv
```

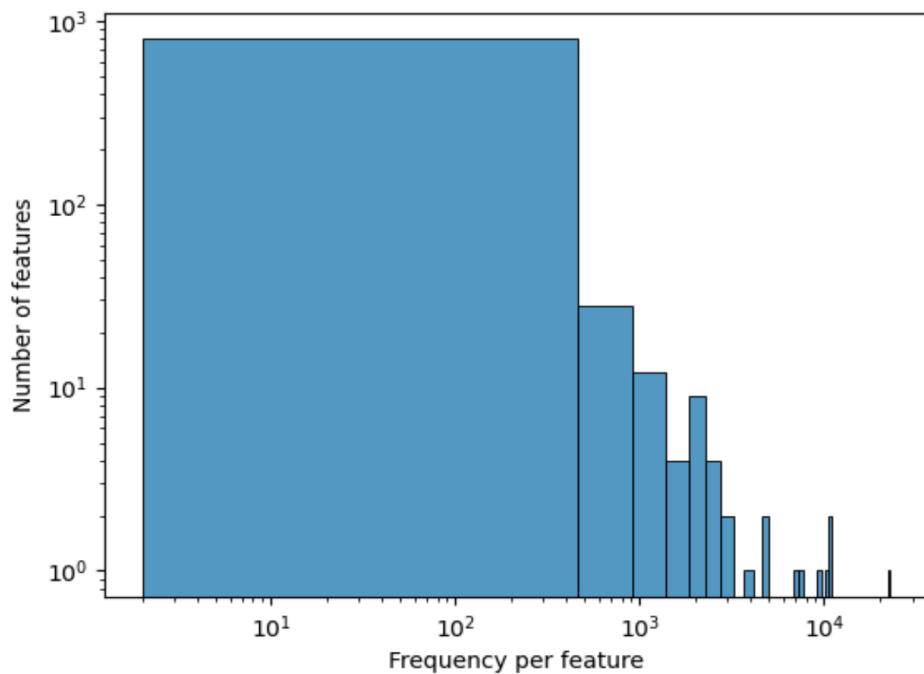
```
qiime feature-table tabulate-seqs --i-data rep-seqs-dada2.qza --o-  
visualization rep-seqs.qzv
```

```
ences$ qiime feature-table summarize --i-table table.qza --o-visualization table.qzv --m-  
sample-metadata-file ~/common/microbiome/sample-metadata.tsv  
Saved Visualization to: table.qzv  
(qiime2-amplicon-2024.10) jupyter-gorovenko-e@hse-students:~/common/K/emp-single-end-sequ-  
ences$ qiime feature-table tabulate-seqs --i-data rep-seqs-dada2.qza --o-visualization re-  
p-seqs.qzv  
Saved Visualization to: rep-seqs.qzv
```

Visualize table.qzv as usual.

💬 Иду сюда <https://view.qiime2.org/> и вставляю файл table.qzv





What is the minimum frequency per sample?

1 338

Frequency per sample

	Frequency
Minimum frequency	1,338
1st quartile	2,392.5
Median frequency	6,260
3rd quartile	9,277.5
Maximum frequency	14,088
Mean frequency	6,173.7

Generate tree for phylogenetic diversity analysis

```
qiime phylogeny align-to-tree-mafft-fasttree --i-sequences rep-seqs-dada2.qza
--o-alignment aligned-rep-seqs.qza --o-masked-alignment masked-aligned-rep-
seqs.qza --o-tree unrooted-tree.qza --o-rooted-tree rooted-tree.qza
```

```
ences$ qiime phylogeny align-to-tree-mafft-fasttree --i-sequences rep-seqs-dada2.qza --o-
alignment aligned-rep-seqs.qza --o-masked-alignment masked-aligned-rep-seqs.qza --o-tree
unrooted-tree.qza --o-rooted-tree rooted-tree.qza
Saved FeatureData[AlignedSequence] to: aligned-rep-seqs.qza
Saved FeatureData[AlignedSequence] to: masked-aligned-rep-seqs.qza
Saved Phylogeny[Unrooted] to: unrooted-tree.qza
Saved Phylogeny[Rooted] to: rooted-tree.qza
```

Alpha and beta diversity analysis

QIIME 2's diversity analyses are available through the `q2-diversity` plugin, which supports computing alpha and beta diversity metrics, applying related statistical tests, and generating interactive visualizations.

We'll first apply the `core-metrics-phylogenetic` method, which rarefies a `FeatureTable[Frequency]` to a user-specified depth, computes several alpha and beta diversity metrics, and generates principle coordinates analysis (PCoA) plots using Emperor for each of the beta diversity metrics. The metrics computed by default are:

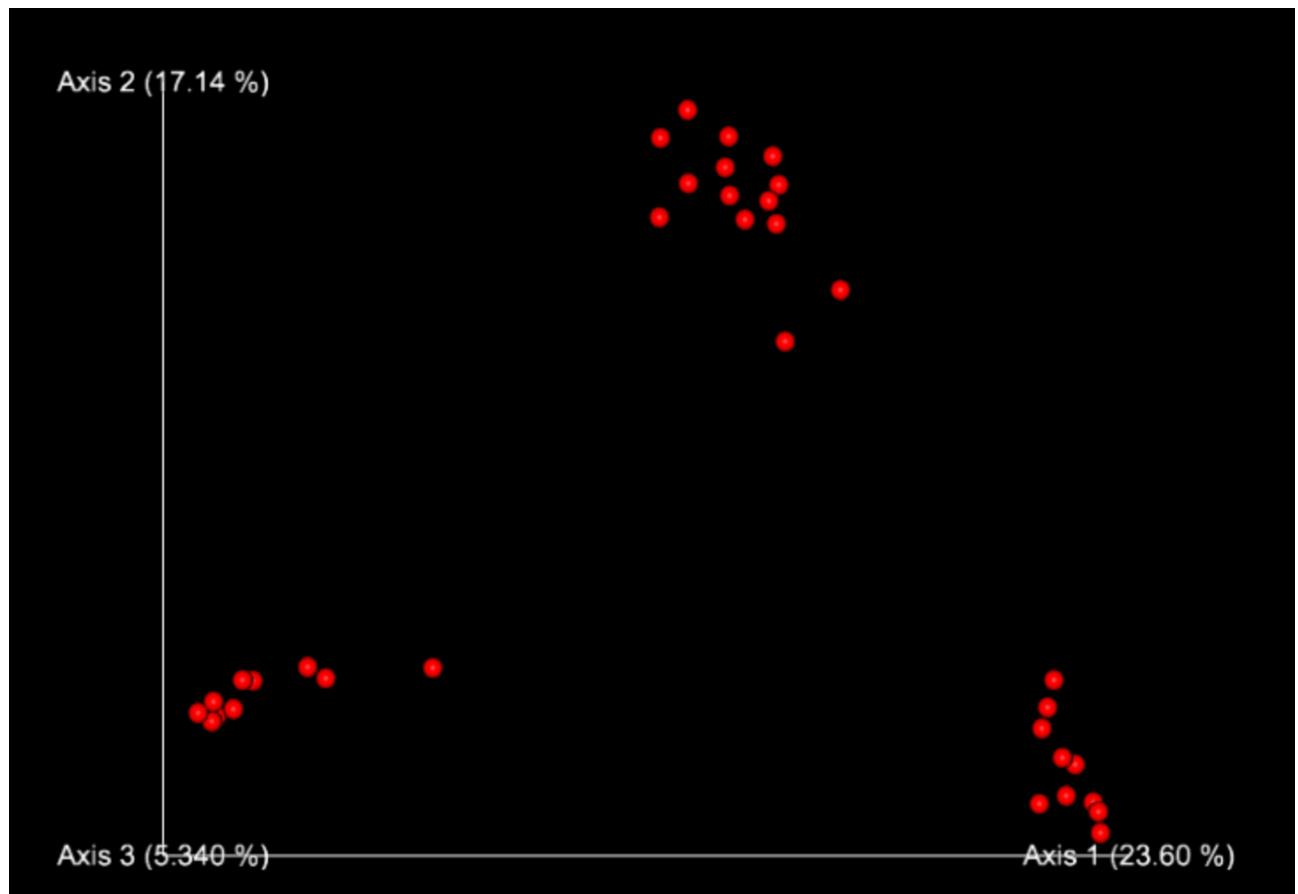
- Alpha diversity
 - Shannon's diversity index (a quantitative measure of community richness)
 - Observed Features (a qualitative measure of community richness)
 - Faith's Phylogenetic Diversity (a qualitative measure of community richness that incorporates phylogenetic relationships between the features)
 - Evenness (or Pielou's Evenness; a measure of community evenness)
- Beta diversity
 - Jaccard distance (a qualitative measure of community dissimilarity)
 - Bray-Curtis distance (a quantitative measure of community dissimilarity)
 - unweighted UniFrac distance (a qualitative measure of community dissimilarity that incorporates phylogenetic relationships between the features)
 - weighted UniFrac distance (a quantitative measure of community dissimilarity that incorporates phylogenetic relationships between the features)

```
qiime diversity core-metrics-phylogenetic --i-phylogeny rooted-tree.qza --i-table table.qza --p-sampling-depth 1103 --m-metadata-file ~/common/microbiome/sample-metadata.tsv --output-dir core-metrics-results
```

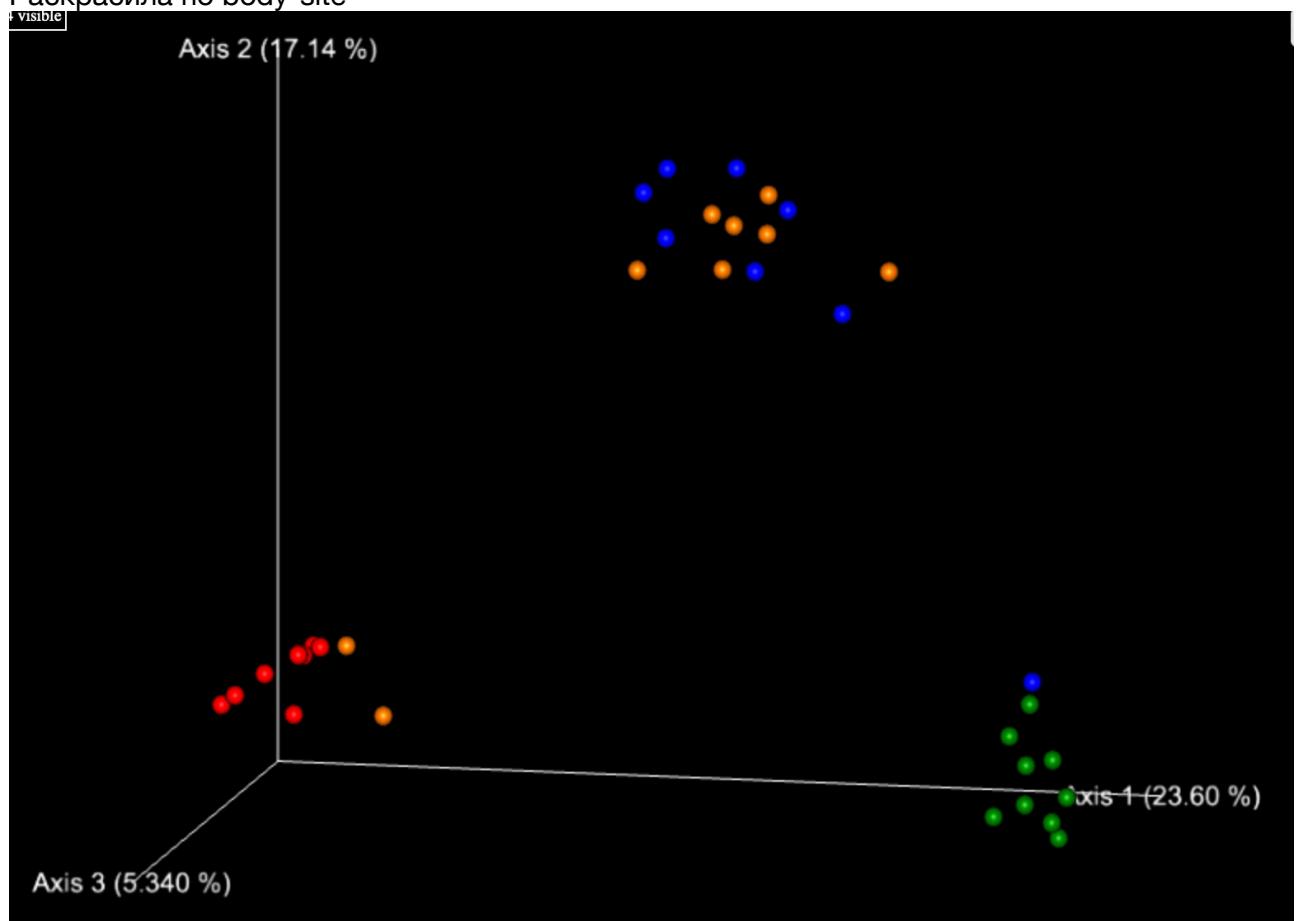
```
ences$ qiime diversity core-metrics-phylogenetic --i-phylogeny rooted-tree.qza --i-table table.qza --p-sampling-depth 1103 --m-metadata-file ~/common/microbiome/sample-metadata.tsv --output-dir core-metrics-results
Saved FeatureTable[Frequency] to: core-metrics-results/rarefied_table.qza
Saved SampleData[AlphaDiversity] to: core-metrics-results/faith_pd_vector.qza
Saved SampleData[AlphaDiversity] to: core-metrics-results/observed_features_vector.qza
Saved SampleData[AlphaDiversity] to: core-metrics-results/shannon_vector.qza
Saved SampleData[AlphaDiversity] to: core-metrics-results/evenness_vector.qza
Saved DistanceMatrix to: core-metrics-results/unweighted_unifrac_distance_matrix.qza
Saved DistanceMatrix to: core-metrics-results/weighted_unifrac_distance_matrix.qza
Saved DistanceMatrix to: core-metrics-results/jaccard_distance_matrix.qza
Saved DistanceMatrix to: core-metrics-results/bray_curtis_distance_matrix.qza
Saved PCoAResults to: core-metrics-results/unweighted_unifrac_pcoa_results.qza
Saved PCoAResults to: core-metrics-results/weighted_unifrac_pcoa_results.qza
Saved PCoAResults to: core-metrics-results/jaccard_pcoa_results.qza
Saved PCoAResults to: core-metrics-results/bray_curtis_pcoa_results.qza
Saved Visualization to: core-metrics-results/unweighted_unifrac_emperor.qzv
Saved Visualization to: core-metrics-results/weighted_unifrac_emperor.qzv
Saved Visualization to: core-metrics-results/jaccard_emperor.qzv
Saved Visualization to: core-metrics-results/bray_curtis_emperor.qzv
```

Visualise `jaccard_emperor.qzv` in `core-metrics-results`. Provide a screenshot. What do you see?

… Иду сюда <https://view.qiime2.org/>, скачиваю и загружаю на сайт файл common/K/emp-single-end-sequences/core-metrics-results/jaccard_emperor.qzv



Раскрасила по body-site



Вижу три группы

3. Alpha and beta diversity

```
library(microbiome)
library(knitr)
library(tidyverse)
library(MicrobeDS)
library(vegan)
library(ggplot2)
library(reshape)
# alpha diversity
data(dietswap)
# beta (intra-individual and inter-individual)
data(peerj32)
# beta (intra-overtime)
data(MovingPictures)
```

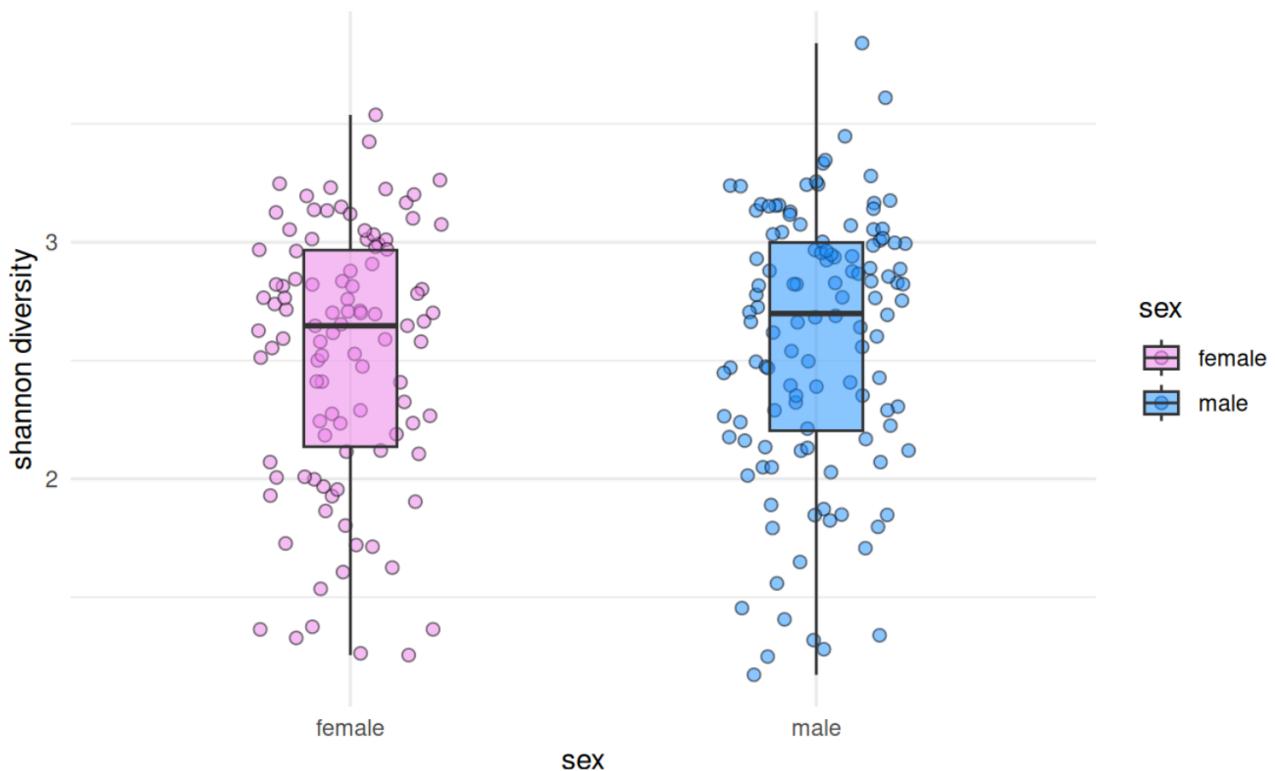
1. calculate the alpha diversity, provide a Shannon plot and perform a Kolmogorov-Smirnov test... provide the p-value

Теория:
Shannon diversity - мера разнообразия

```
data(dietswap)
```

```
# считаем альфа дайверсити
ad = microbiome::alpha(dietswap, index = "all") # calculate the alpha
diversity
```

```
# график
boxplot_alpha(dietswap, index = "shannon", x_var = "sex",
               fill.colors = c(female="violet", male="dodgerblue")) +
theme_minimal() +
labs(x="sex", y="shannon diversity")
```



```
# тест К-С
d = meta(dietswap)
d$diversity = microbiome::diversity(dietswap, "shannon")$shannon
spl = split(d$diversity, d$sex)

pv = ks.test(spl$female, spl$male)$p.value

pv
```

💬 p.value = 0.2463578 (типа разницы у мужчин и женщин нет)

2. provide a plot between cases and controls, showing divergence within each group samples

👉 Спасибо пайплану за все последующие графики: <https://microbiome.github.io/tutorials/Betadiversity.html>

```
data(peerj32)
```

```
pseq <- peerj32$phyloseq

betas <- list()
groups <- as.character(unique(meta(pseq)$group))
for (g in groups) {
```

```

df <- subset(meta(pseq), group == g)
beta <- c()

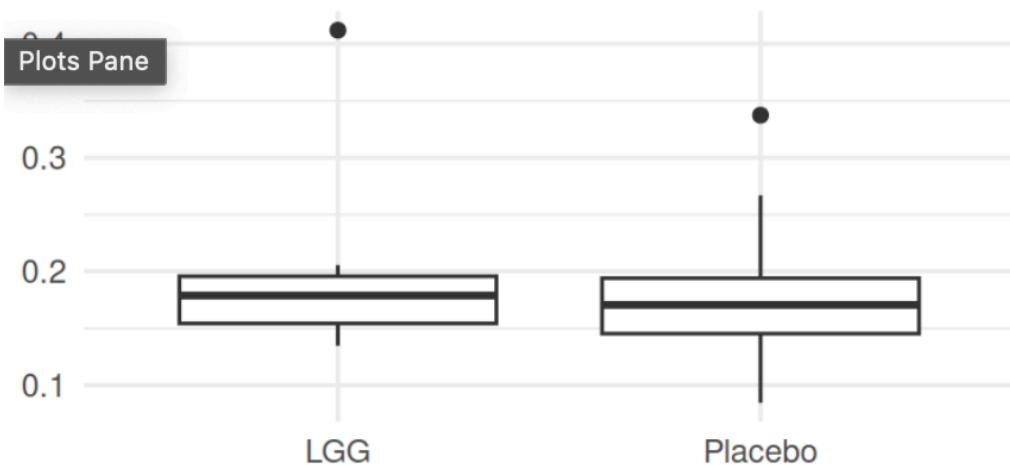
for (subj in df$subject) {
  # Pick the samples for this subject
  dfs <- subset(df, subject == subj)
  # Check that the subject has two time points
  if (nrow(dfs) == 2) {
    s <- as.character(dfs$sample)
    # Here with just two samples we can calculate the
    beta[[subj]] <- divergence(abundances(pseq)[, s[[1]]], abundances(pseq)
    [, s[[2]]], method = "bray")
  }
}
betas[[g]] <- beta
}

# boxplot
df <- as.data.frame(unlist(betas))
s<- rownames(df)
si<- as.data.frame(s)
si<- separate(si, s, into = c('names','s'))
df1<- bind_cols(df, si)
rownames(df1)<- df1$s ; df1$s<- NULL

p<- ggplot(df1, aes(x = names, y = `unlist(betas)`))+ geom_boxplot() +
ylab('') + xlab('') +
theme_minimal()

plot(p)

```



3. provide a timeseries of divergence over time

```
data(MovingPictures)
```

```
# Pick the metadata for this subject and sort the
# samples by time

# Pick the data and modify variable names
pseq <- MovingPictures
s <- "F4" # Selected subject
b <- "UBERON:feces" # Selected body site

# Let us pick a subset
pseq <- subset_samples(MovingPictures, host_subject_id == s & body_site == b)

# Rename variables
sample_data(pseq)$subject <- sample_data(pseq)$host_subject_id
sample_data(pseq)$sample <- sample_data(pseq)$X.SampleID

# Tidy up the time point information (convert from dates to days)
sample_data(pseq)$time <- as.numeric(as.Date(gsub(" 0:00", "", 
as.character(sample_data(pseq)$collection_timestamp)), "%m/%d/%Y") - 
as.Date("10/21/08", "%m/%d/%Y"))

# Order the entries by time
df <- meta(pseq) %>% arrange(time)

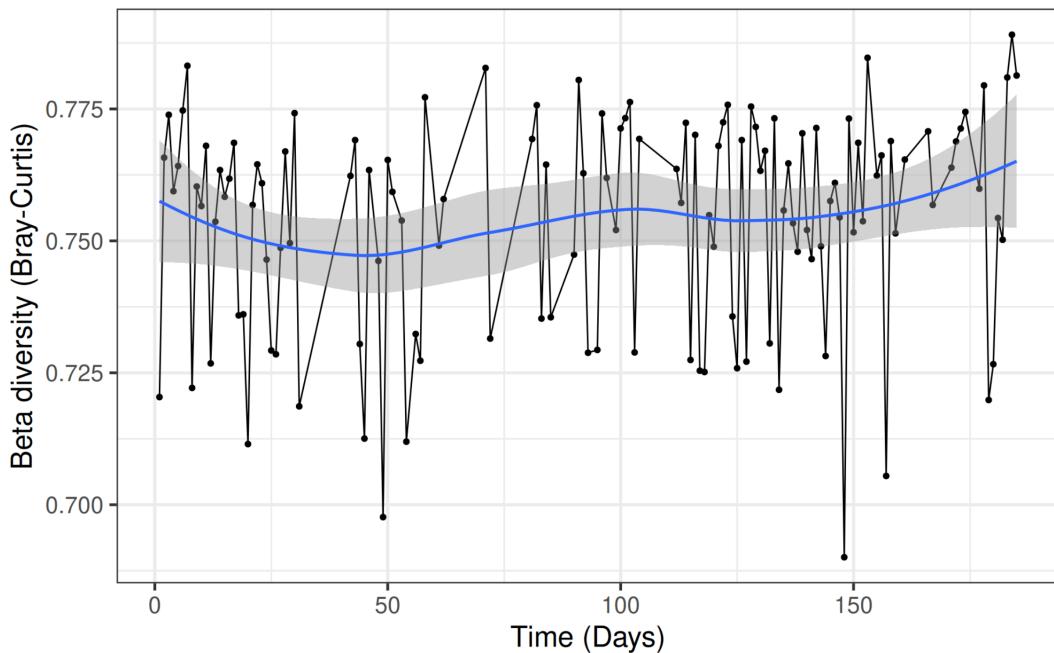
# Calculate the beta diversity between each time point and
# the baseline (first) time point
beta <- c() # Baseline similarity
s0 <- subset(df, time == 0)$sample
# Let us transform to relative abundance for Bray-Curtis calculations
a <- microbiome::abundances(microbiome::transform(pseq, "compositional"))
for (tp in df$time[-1]) {
  # Pick the samples for this subject
  # If the same time point has more than one sample,
  # pick one at random
  st <- sample(subset(df, time == tp)$sample, 1)
  # Beta diversity between the current time point and baseline
  b <- vegdist(rbind(a[, s0], a[, st]), method = "bray")
  # Add to the list
  beta <- rbind(beta, c(tp, b))
}
```

```

colnames(beta) <- c("time", "beta")
beta <- as.data.frame(beta)

theme_set(theme_bw(20))
library(ggplot2)
p <- ggplot(beta, aes(x = time, y = beta)) +
  geom_point() +
  geom_line() +
  geom_smooth() +
  labs(x = "Time (Days)", y = "Beta diversity (Bray-Curtis)")
print(p)

```



4. provide a plot for inter-individual divergance for each groups as a comparison

```

pseq <- peerj32$phyloseq

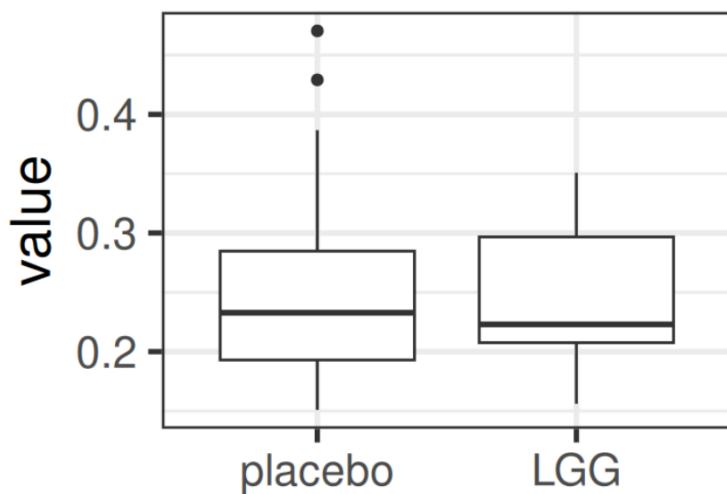
b.pla <- divergence(subset_samples(pseq, group == "Placebo"),
apply(abundances(subset_samples(pseq, group == "Placebo")), 1, median))

b.lgg <- divergence(subset_samples(pseq, group == "LGG"),
apply(abundances(subset_samples(pseq, group == "LGG")), 1, median))

l<- list(b.pla, b.lgg)
df<- melt(l)
df$L1[df$L1 == '1']<- 'placebo'
df$L1[df$L1 == '2']<- 'LGG'
df$L1<- factor(df$L1, levels = c('placebo','LGG'))

```

```
p<- ggplot(df, aes(x = L1, y = value)) + geom_boxplot() + xlab('')  
plot(p)
```



Bcë...

