R Packages: Phyloseq, ade4, ape, biom, Biostrings, cluster, data.table, DESeq2, foreach, ggplot2, igraph, methods, multtest, plyr, reshape2, scales, shiny, vegan, doParallel, Knitr, genefilter, testthat

```
library("phyloseq")
```

```
library("biom")
```

Loading required libraries (packages)

```
mydata<-import_biom("C:/Users/SaiDheeraj/Desktop/Phyloseq/Biome file/QIIME.organism2.data.biom")
```

Naming the dataset as mydata

```
new=read_biom("C:/Users/SaiDheeraj/Desktop/Phyloseq/Biome file/QIIME.organism2.data.biom")
```

reading the data and stroing it under name “new”

```
otumat = as(biom_data(new), "matrix")
```

```
OTU = otu_table(otumat, taxa_are_rows=TRUE)
```

Defining the out_table data class

```
taxmat = as.matrix(observation_metadata(new), rownames.force=TRUE, byrows=FALSE)
```

```
TAX = tax_table(taxmat)
```

Defining the tax_table data class

```
physeq = phyloseq(OTU, TAX)
```

Defining the phyloseq class

```
Physeq
```

Displays the data of the new phyloseq class

```
most_abundant_taxa <- sort(taxa_sums(physeq), TRUE)[1:10]
```

```
ex2 <- prune_taxa(names(most_abundant_taxa), physeq)
```

```
ex2
```

```
topFamilies <- tax_table(ex2)[, "taxonomy3"]
```

```
as(topFamilies, "vector")
```

Displays the top 10 abundant taxa, with their families

Tree diagram

```
library("ape")

physeq = phyloseq(newOTU, newTAX)

newrandom_tree = rtree(ntaxa(physeq), rooted = TRUE, tip.label = taxa_names(physeq))

plot(random_tree)

physeq1 = phyloseq(newOTU, newTAX, newrandom_tree)
```

Bar plot

```
plot_bar(physeq1, fill="taxonomy3")
```

Ordination

```
GP.chl = subset_taxa(physeq1)

GP.chl = prune_samples(names(which(sample_sums(GP.chl) >= 20)), GP.chl)

GP.chl.r = rarefy_even_depth(GP.chl)

GP.chl

gpsfb.wUF = distance(GP.chl.r, "unifrac", weighted = TRUE)

gpsfb.NMDS = ordinate(GP.chl.r, "CCA", gpsfb.wUF)

library("ggplot2")

p1 = plot_ordination(GP.chl.r, gpsfb.NMDS, "taxa", color = "taxonomy5", title = "plot_ordination, CCA,
wUF")

'?'(plot_ordination)

p1

p1 + facet_wrap(~taxonomy5, 3)

(p <- plot_heatmap(GP.chl.r, "NMDS", "bray", "taxa", "taxonomy5"))

p2 <- plot_net(GP.chl.r)

p2
```

Network diagram

```
g <- make_network(physeq1, type = "taxa",color = "taxonomy3", max.dist = 0.4)

g

plot_network(g, physeq1, type="taxa")
```

Code for Group1 (There are in total 4 groups, code is similar for rest 3 groups)

```
library("phyloseq")

library("biom")
```

Above code calls the installed packages ‘phyloseq’ and ‘biom’. That means, we can use all the functionalities of those packages.

```
Grp1<-import_biom("C:/Users/SaiDheeraj/Desktop/Phyloseq/Project files/Mouse/QIIME.grp1.biom")
```

Import_biom is a function that can be used to import the data from a biom file. After ‘import_biom’ function, give the path of the biom file which has extension as ‘.biom’. ‘Grp1’ is the name assigned to the result of import_biom functionality for the given biom file.

```
a=read_biom("C:/Users/SaiDheeraj/Desktop/Phyloseq/Project files/Mouse/QIIME.grp1.biom")
```

read_biom is a function that can be used to read the data of biom file. Similar to the above, path of biom file must be given. ‘a’ is assigned to the result of read_biom functionality for the given biom file. This will be used further to refer the data of biom file.

```
otumat = as(biom_data(a), "matrix")
```

We are referring to the data present in matrix format of our file. If you notice we are using ‘a’ which we have named for our file. Finally we are naming that matrix data as ‘otumat’.

```
taxmat = as.matrix(observation_metadata(a), rownames.force=TRUE)
```

Similar to the above, we are using the observation_metadata of our file, and making them to incline with our ‘otumat’ and as these are present in form of rows, we are using the function ‘rownames.force’ as ‘TRUE’. Finally we are naming this particular data as ‘taxmat’.

```
OTU = otu_table(otumat, taxa_are_rows=TRUE).
```

Now after the preliminary work done above, we are defining the classes used in phyloseq. First we are using the ‘otu_table’ class for the otu’s present in our file and the corresponding ‘taxa’ are present in rows of the file, hence used function ‘taxa_are_rows’ and status is set to ‘TRUE’. The result of the class out_table is named as ‘OTU’.

```
TAX = tax_table(taxmat)
```

Similar to the above, now we are using the other class ‘tax_table’ and the result is named to ‘TAX’.

```
no1 = phyloseq(OTU, TAX)
```

Once defining the otu and tax classes, we are creating a new phyloseq class using the classes defined. The new phyloseq object is named as ‘no1’.

```
library("ape")

tree = rtree(ntaxa(no1), rooted = TRUE, tip.label = taxa_names(no1))
```

Calling the package ‘ape’ and using the ‘rtree’ function we are creating a tree for our data. This is other required class. We are naming the result as ‘tree’.

```
new1 = phyloseq(OTU,TAX,tree)
```

Using the new ‘tree’ created, we are using the phyloseq and creating a new object ‘new1’.

```
sampleDF <- read.csv("C:/Users/SaiDheeraj/Desktop/Phyloseq/Project files/Mouse/Grp1.csv", row.names = "Metagenome_ID")
```

Now to include the metadata for the file, we are using the data in csv file and naming it as ‘sampleDF’. Path of the file must be provided, and the connecting key i.e. common to the biom file and csv file, must be included. Here it is ‘Metagenome_ID’.

```
SD<-sample_data(sampleDF)
```

Using the sample_data class, we are using the already defined ‘sampleDF’ and the result is named as ‘SD’.

```
new1 = merge_phyloseq(new1, SD)
```

As the sample data and the phyloseq object created before, are two different data types. One in ‘csv file’ and other in ‘biom file’, we are using the ‘merge_phyloseq’ to merge both. The result is named as ‘new1’ again.

```
bacteria1<-subset_taxa(new1, taxonomy2=="Bacteria")
```

Using the ‘subset_taxa’ function in phyloseq, we are separating all the rows whose kingdom is ‘Bacteria’. Here taxonomy2 is kingdom. So all the rows separated are named as ‘bacteria1’.

```
b1 = prune_taxa(taxa_sums(bacteria1) > 0, bacteria1)
```

Above code eliminates the rows whose abundance is zero. Though there are very minimal chances to have this, it is a good practice to clear out. We are naming this as ‘b1’.

```
OTU10 = names(sort(taxa_sums(b1), TRUE)[1:10])
```

Above code do the sorting and extracts the top 10 OTU based on their abundance values.

```
b10=prune_taxa(OTU10, b1)
```

We are combining both the objects using the function prune_taxa which should be used when there are more than one object for OTU’s. Naming it as ‘b10’.

```
trial1 <- transform_sample_counts(b10, function(OTU) OTU/sum(OTU))
```

Function ‘transform_sample_counts’ helps us to convert the abundance values to percentages. This can be done using the function for the specified data ‘b10’. Naming it as ‘trial1’.

```
p<-plot_bar(trial1, "taxonomy3", fill = "taxonomy3", facet_grid=~Mouse.Type)
```

plot_bar function is used to get the bar plot for the specified taxonomy rank. Here taxonomy3 is used, which is ‘phylum’. Facet_grid can be used to get the plot for each ‘Mouse.Type’.

```
p + geom_bar(aes(color=taxonomy3, fill= taxonomy3), stat="identity", position="stack")
```

Using the above function, we can remove the unwanted space (black color bar).

The above steps are done for first file for the group1 mice. Similarly all the steps can be repeated for all the other groups, whose data is saved in other files.

Comparing two groups:

```
Dist2 = distance(bacteria2, "unifrac", weighted = TRUE)

Ordinate2 = ordinate(bacteria2, "CCA", Dist2)

library("ggplot2")

o1 = plot_ordination(bacteria2, Ordinate2, "samples", color = "Mouse.Type", title = "plot_ordination, CCA, wUF")

o1

o1+facet_wrap(~taxonmy3, 3)

o1+geom_polygon(aes(fill=Grp))+geom_point(size=2)+ggtitle("samples")

o3=plot_ordination(bacteria2, Ordinate2, type="split", color = "Mouse.Type")

o3

plot_bar(bacteria2, "Grp", fill="Mouse.Type", facet_grid=~taxonomy3)

'?'(merge_phyloseq_pair)

merge12 <- merge_phyloseq(bacteria1, bacteria2)

m1<-merge_samples(b10new, "Grp")

sample_names(b10new)

m1

OTU12 <-merge_phyloseq_pair(OTU, OTU1)

c1<-phyloseq(OTU,TAX)

c11<- merge_phyloseq(c1, SD)

sample_names()

TAX12 <- merge_phyloseq_pair(TAX, TAX1)

TAX12

SD12 <- merge_phyloseq_pair(SD, SD1)

SD12
```

```
tree12<-merge_phyloseq_pair(tree, tree1)
bacteria12 <- phyloseq(OTU12, TAX12, SD12)
bacteria12
bacteria12 <- subset_taxa(bacteria12, taxonomy2=="Bacteria")
myTaxa12 = names(sort(taxa_sums(bacteria12), decreasing = TRUE)[1:10])
myTaxa
b12 = prune_taxa(taxa_sums(bacteria12) > 0, bacteria12)
mergedb12=merge_samples(b12, "Grp")
mergedb12
SD12 = merge_samples(sample_data(b12), "Mouse.Type")
SD12
identical(SD12, sample_data(mergedb12))
b12OTU10 = names(sort(taxa_sums(b12), TRUE)[1:30])
any(taxa_sums(b12OTU10)== 0)
b10_12=prune_taxa(b12OTU10, b12)
b10new = prune_taxa(taxa_sums(b10_12)>0, b10_12)
any(sample_sums(b10new) == 0)
b10new = prune_samples(sample_sums(b10_12)>0, b10_12)
trial12 <- transform_sample_counts(b10new, function(x) 100* x/sum(x))

trial12
'?'(transform_sample_counts)
// b12mb10 = prune_taxa(b12OTU10, mergedb12)
rowSums(otu_table(b10_12))
'?'(rowSums)
//p1111<-plot_bar(b10_12, "taxonomy3", fill = "taxonomy3", facet_grid=~Grp)
p1111
p12<-plot_bar(trial12, "Grp", fill = "Grp", facet_grid=~taxonomy3)
p12
```

```
library("ggplot2")
p12 + geom_bar(aes(color=Grp, fill= Grp), stat="identity", position="stack")


ss <- subset_taxa(trial12, taxonomy3 %in%
c('Actinobacteria','Bacteroidetes','Chloroflexi','Firmicutes','Proteobacteria','Tenericutes'))
glom <- tax_glom(ss, taxrank = 'taxonomy3')


dat <- psmelt(glom)
library("ggplot2")
ggplot(dat, aes(x=Grp, y=Abundance, fill=taxonomy3)) + geom_boxplot()
plot_richness(b10new, Mouse.Type)
'?'(geom_boxplot)
// p2 <- plot_richness(b10new, "Grp1", "Grp")
// p2 <- p2+geom_boxplot(data=p2$data, aes(x=Grp1, y=value, color=NULL))
// p2
//plot_net(b10new, maxdist=0.3, point_label="Mouse.Type")
//plot_heatmap(b10new)
dist = distance(m1, "unifrac", weighted = TRUE)
ord = ordinate(bacteria2, "CCA",dist)
library("ggplot2")
b.ord <- ordinate(m1, "CCA", "unifrac")
p0 = plot_ordination(m1, b.ord, "samples", color = "Mouse.Type")
p0+geom_jitter(aes(fill=Mouse.Type)) + geom_point(size=10)
```