##### Import and prepare sequencing data #####

rm(list=ls()); library (stringr); library (vegan);library (dplyr)

myBtaxa <- read.table("myBtaxa.FOL.txt", row.names=1, sep="\t", header=T, blank.lines.skip=F, check.names=F, stringsAsFactors=F)

myBotu <- read.table("myBotu.FOL.txt", row.names=1, sep="\t", header=T, blank.lines.skip=F, check.names=F)

mydesign <- read.table("mydesign.FOL.txt", header=T, row.names=1, stringsAsFactors=F, na.strings="NA")

dim(myBtaxa)

dim(myBotu)

dim(mydesign)

# create separate taxonomy label specifying classes of Proteobacteria

myBtaxa$labels <- myBtaxa$Phylum

myBtaxa[ rownames(myBtaxa)[myBtaxa$Class=="Alphaproteobacteria" ], ]$labels <- "Alphaproteobacteria"

myBtaxa[ rownames(myBtaxa)[myBtaxa $Class=="Betaproteobacteria" ], ]$labels <- "Betaproteobacteria"

myBtaxa[ rownames(myBtaxa)[myBtaxa $Class=="Gammaproteobacteria" ], ]$labels <- "Gammaproteobacteria"

myBtaxa[ rownames(myBtaxa)[myBtaxa $Class=="Deltaproteobacteria" ], ]$labels <- "Deltaproteobacteria"

#table(myBtaxa$labels)

##### Store Archaea, Cyanobacteria and mitrochrondia sequeces #####

#unique(myBtaxa$Kingdom)

#table(myBtaxa$Kingdom)

r1 <- rownames(myBtaxa[myBtaxa$Kingdom=="Archaea",]); r1

#table(myBtaxa$Phylum)

r2 <- rownames(myBtaxa[myBtaxa$Phylum=="Cyanobacteria",]); r2

##unique(myBtaxa$Family)

r3 <- rownames(myBtaxa[myBtaxa$Family=="Mitochondria",]); r3

#otus\_remove\_16s <- c(r1,r2)

otus\_remove\_16s <- c(r1, r2, r3)

## Remove these from otu table, tax table

otu\_filter\_16s <- myBotu[-which(rownames(myBotu) %in% otus\_remove\_16s),]; dim (myBotu); dim (otu\_filter\_16s)

tax\_filter\_16s <- myBtaxa[rownames(otu\_filter\_16s),]

design\_filter\_16s <- droplevels(mydesign[rownames(mydesign) %in% colnames(otu\_filter\_16s),])

design\_filter\_16s <- design\_filter\_16s[colnames(otu\_filter\_16s),]

dim(otu\_filter\_16s); dim(tax\_filter\_16s); dim(design\_filter\_16s)

###### 16S sequence and OTU counts ######

sum(colSums(otu\_filter\_16s))

sort(colSums(otu\_filter\_16s))

median(colSums(otu\_filter\_16s)); mean(colSums(otu\_filter\_16s))

nrow(tax\_filter\_16s)

table(tax\_filter\_16s$Kingdom)

## Order taxonmy file by OTU

otu\_order\_16s <- match(rownames(otu\_filter\_16s), rownames(tax\_filter\_16s))

tax\_filter\_16s <- tax\_filter\_16s[otu\_order\_16s,]

otu\_filter\_16s <- otu\_filter\_16s [rowSums(otu\_filter\_16s) > 1,]

dim(otu\_filter\_16s)

sum(otu\_filter\_16s); sort(colSums(otu\_filter\_16s))

tax\_filter\_16s <- tax\_filter\_16s[rownames(otu\_filter\_16s),] ; dim(tax\_filter\_16s)

write.table(otu\_filter\_16s, file="otu\_filter\_16s.txt", sep="\t",quote=F)

write.table(design\_filter\_16s, file="design\_filter\_16s.txt", sep="\t",quote=F)

write.table(tax\_filter\_16s, file="tax\_filter\_16s.txt", sep="\t",quote=F)

#####**rarefy###**

otu\_rarefy\_16s <-t(otu\_filter\_16s)

(raremax\_16s <- min(rowSums(otu\_rarefy\_16s)))

set.seed(315); otu\_rarefy\_16s <- **rrarefy**(otu\_rarefy\_16s, raremax\_16s)

otu\_rarefy\_16s <- otu\_rarefy\_16s[ ,colSums(otu\_rarefy\_16s)>0] ; dim(otu\_rarefy\_16s)

otu\_rarefy\_16s<-t(otu\_rarefy\_16s)

tax\_rarefy\_16s <- tax\_filter\_16s[rownames(otu\_rarefy\_16s),] ; dim(tax\_rarefy\_16s)

design\_rarefy\_16s <- design\_filter\_16s

write.table(otu\_rarefy\_16s, file="otu\_rarefy\_16s.txt", sep="\t",quote=F)

write.table(design\_rarefy\_16s, file="design\_rarefy\_16s.txt", sep="\t",quote=F)

write.table(tax\_rarefy\_16s, file="tax\_rarefy\_16s.txt", sep="\t",quote=F)

##### mean relative abundance of each OTU for each sample ####

otu\_16s\_RA <- t(t(otu\_rarefy\_16s)/colSums(otu\_rarefy\_16s))\*100

Sam.Rep=4

rep\_time <- length(colnames(otu\_16s\_RA))/ Sam.Rep ; t=1; otu\_16s\_RA\_AVE=c()

for (i in 1:rep\_time) { sub\_table= otu\_16s\_RA [,t:(t+ Sam.Rep -1)]

sub\_mean=apply(sub\_table, 1, mean)

otu\_16s\_RA\_AVE=cbind(otu\_16s\_RA\_AVE, sub\_mean)

t=t+ Sam.Rep }

colnames(otu\_16s\_RA\_AVE) <- unique (as.factor(design\_rarefy\_16s$Trt))

**write.table**(otu\_16s\_RA\_AVE, file="otu\_16s\_RA\_AVE.txt", sep="\t",quote=F)