library(plyr); library (reshape2); library (ggplot2)

design\_rarefy\_16s<- read.table("design\_rarefy\_16s.txt", row.names=1, sep="\t", header=T, blank.lines.skip=F, check.names=F)

otu\_rarefy\_16s <- read.table("otu\_rarefy\_16s.txt", row.names=1, sep="\t", header=T, blank.lines.skip=F, check.names=F)

tax\_rarefy\_16s<- read.table("tax\_rarefy\_16s.txt", row.names=1, sep="\t", header=T, blank.lines.skip=F, check.names=F)

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

PHYLAnames\_16s <- names(sort(table(tax\_rarefy\_16s[,"labels"]), decr=T))

length(PHYLAnames\_16s)

sort(table(tax\_rarefy\_16s[,"labels"]), decr=T)

y <- NULL

otunames <- rownames(otu\_16s\_RA)

for (i in PHYLAnames\_16s){

x <- array(colSums(otu\_16s\_RA [rownames(tax\_rarefy\_16s)[which(tax\_rarefy\_16s$labels == paste(i))],,drop=FALSE]))

y <- rbind(y,x) }

rownames(y) <- paste(PHYLAnames\_16s)

colnames(y) <- paste(colnames(otu\_16s\_RA))

PHYLUM\_mat\_16s <- y

colSums(PHYLUM\_mat\_16s)

PHYLUM\_mat\_16s\_mean <- sort(apply(PHYLUM\_mat\_16s,1,mean),decr=T)

PHYLUM\_mat\_16s <- PHYLUM\_mat\_16s[names(PHYLUM\_mat\_16s\_mean),]

tax\_rarefy\_16s$cols <- tax\_rarefy\_16s$labels

table(tax\_rarefy\_16s$cols)

tax\_rarefy\_16s$cols<-**as.character**(tax\_rarefy\_16s$cols)

table(apply(PHYLUM\_mat\_16s, 1, mean) < 1)

low\_count\_phyla\_16s <- rownames(PHYLUM\_mat\_16s)[sort(apply(PHYLUM\_mat\_16s, 1, mean), decr=T) < 1]

for(i in low\_count\_phyla\_16s){

tax\_rarefy\_16s[ rownames(tax\_rarefy\_16s)[tax\_rarefy\_16s$Phylum==paste(i) ], ]$cols <- "gray30"

}

table(tax\_rarefy\_16s$cols)

abundant\_phyla\_16s <- rownames(PHYLUM\_mat\_16s)[sort(apply(PHYLUM\_mat\_16s, 1, mean), decr=T) > 3]

Abundant.Phy = abundant\_phyla\_16s

Abundant.Phy = c("Alphaproteobacteria", "Betaproteobacteria", "Gammaproteobacteria", "Deltaproteobacteria", "Actinobacteria", "Acidobacteria", "Chloroflexi", "Gemmatimonadetes", "Bacteroidetes", "Firmicutes") #

PHYLUM\_mat\_16s.whole = as.data.frame(t(PHYLUM\_mat\_16s [ Abundant.Phy, ]))

PHYLUM\_mat\_16s.whole.bar <- PHYLUM\_mat\_16s.whole [, Abundant.Phy]

PHYLUM\_mat\_16s.whole.bar $ others <- 100-rowSums(PHYLUM\_mat\_16s.whole.bar)

PHYLUM\_mat\_16s.whole.bar$Trt <- design\_rarefy\_16s$ Trt

#PHYLUM\_mat\_16s.whole.bar$Cultivar <- design\_rarefy\_16s$Cultivar

Mydata.bar = melt(PHYLUM\_mat\_16s.whole.bar, id.vars=c("Trt"),

measure.vars=c(Abundant.Phy, "others"),

variable.name='Phylum', value.name='Abundance')

#################################################################################

mean <- aggregate(Mydata.bar$Abundance, by=list(Mydata.bar$Trt, Mydata.bar$Phylum), FUN=mean)

sd <- aggregate(Mydata.bar$Abundance, by=list(Mydata.bar$Trt, Mydata.bar$Phylum), FUN=sd)

len <- aggregate(Mydata.bar$Abundance, by=list(Mydata.bar$Trt, Mydata.bar$Phylum), FUN=length)

df\_res <- data.frame(mean, sd=sd$x, len=len$x)

colnames(df\_res) = c("Trt", "Phylum", "Mean", "Sd", "Count")

head(df\_res)

df\_res$Se <- df\_res$Sd/sqrt(df\_res$Count)

df\_res1 = ddply(df\_res, "Trt", transform, label\_y = cumsum(Mean ))

head(df\_res1)

###################################################################################

Barplot.whole = ggplot(df\_res1, aes(x=**factor**(Trt, levels=c("Control", "VOC")), y=Mean, fill=Phylum))+

geom\_bar(stat="identity", width=0.6, position = position\_stack(reverse = TRUE), color="transparent") +

scale\_fill\_manual(values=c("turquoise4", "springgreen3", "yellowgreen", "chartreuse", "brown2", "magenta", "darkorange", "dodgerblue", "tan4", "cyan1", "darkviolet", "grey50")) + ## "darkviolet",

geom\_errorbar(aes(ymin=label\_y-Se, ymax=label\_y+Se), width=0.4, linewidth=0.70, color="black") +

labs(x ="",y = "Relative abundance (%)")+

guides(fill = guide\_legend(reverse=TRUE))+

theme(axis.text=element\_text(size=10, color="black") )

Barplot.whole

Data = PHYLUM\_mat\_16s.whole

Data$Trt <- design\_rarefy\_16s$Trt

**stats::t.test**(Data $ Alphaproteobacteria ~ Data $Trt, p.adj="BH", var.equal=F) #Welch’s T test