SUPPLEMENTARY ONLINE MATERIAL 4

```
#####################################################################
#  SCRIPT: Multivariate analyses Azzarone et al., 2020                #
#############################m#########################################

getwd()
install.packages("vegan")
install.packages("ggplot2")
install.packages("ggrepel")
#install.packages("ggdendro")

library(vegan)
library(ggplot2)
library(ggrepel)
#library(ggdendro)

gps <- 2                              # number of grouping variables
if (!exists("n")) n <- 20   # minimum acceptable sample size
if (!exists("s")) s <- 2      # minimum acceptable number of occurrences

data <- read.csv2("molldata.csv", header = T, row.names = 1)[,-(1:gps)]

####remove low abundance samples####
lowab.s <- which(apply(data,1,sum)<n)
paste(length(lowab.s), 'low abundance samples found')
data1 <- data[-lowab.s,]

####remove low occurences species####
rare.sp <- which(apply(data1>0,2,sum)<s)         # find rare species (singleton occurrences)
paste(length(rare.sp), 'rare species found')
data2 <- data1[,-rare.sp]
```

```
####check if removal of low occurrence species affect n of samples####
lowab.s1 <- which(apply(data2,1,sum)<n)
paste(length(lowab.s1), 'low abundance samples found')
data3 <- data2[-lowab.s1,]


####check if the removal of samples affect n of low occurences species####
rare.sp1 <- which(apply(data3>0,2,sum)<s)       # find rare species (singleton occurrences)
paste(length(rare.sp1), 'rare species found')
data4 <- data3[,-rare.sp1]

####to log transform the abundance matrix####
data5<-log(data4+1)

#Use log transformed matrix to run DCA
DCA_b=decorana(data5)
summary(DCA_b)

####to save the species and sample scores from the DCA####
species.scores=as.data.frame(scores(DCA_b, "species"))
species.scores$species=rownames(species.scores)
data.scores=as.data.frame(scores(DCA_b,"sites"))
data.scores$sites=rownames(data.scores)

####to plot core labels####
label<- read.csv2("molldata.csv", header = T)[,1:2]
data.scores<-merge(data.scores, label, by=c("sites"))

####to plot DCA using ggplot####
l=ggplot(data=data.scores, aes(x=DCA1,y=DCA2)) +
  geom_point(size = 3, aes(shape = core_ID, colour = core_ID)) +
```

```
 geom_point(data=species.scores,aes(x=DCA1,y=DCA2),shape=16,colour="grey40",size=2) + # species symbol
 scale_shape_manual(values=c(18,16,15,18,18)) +
 coord_equal() +
 theme_bw()+
 theme(panel.grid.major = element_blank(),
       panel.grid.minor = element_blank())
k=l+geom_text_repel(data=data.scores,aes(x=DCA1,y=DCA2,label=sites),alpha=0.7,size=2.2)
o=k+geom_text_repel(data=species.scores,aes(x=DCA1,y=DCA2,label=species),alpha=0.7,fontface="bold",size=3)
o

####R-cluster####

install.packages("factoextra")
library(factoextra)

data4a<-decostand(data4, "total")
data4b<-decostand(data4a, "max")
data4c<-t(data4b)
dist.cor <- get_dist(data4c, method = "pearson") #to create the matrix using a correlation distance
dist.clust <- hclust(dist.cor, method="average") #to run the R-cluster analysis with UPGMA method
hclu1 <- plot(dist.clust, main="UPGMA - R cluster", hang=-1, cex=0.6) # plot R-cluster

####END####
```