```
library(vegan)

##Read in data file, samples in rows, taxa in columns
txal.all <- read.table (file = "TXvsAL_all.txt", header=TRUE,
row.names=1)

##Read in sample info file, with formation, lithology and sample
level for each sample
sample.info <- read.table (file = "Sample_info.txt", header=TRUE,
row.names=1)
attach(sample.info)

##RICHNESS ESTIMATION - ASSESS DEGREE OF UNDERSAMPLING
specpool(txal.all)

##SAMPLE DATA FOR TABLE 6.1
results <- dataframe(rowSums(txal.all), specnumber(txal.all),
t(rarefy(txal.all, sample = 150)))
colnames(results) <- c("N", "S", "R150")
write.table(results, "Table6.1.txt", sep = "\t", row.names = T,
col.names = T)

-----------------------------------------------------------------
-------------------------------------

##DATA ANALYSIS
##Remove uniques (taxa limited to one sample) - all further
analyses carried out with this dataset
txal.nosing <- txal.all[,colSums(txal.all > 0) != 1]

##ADDITIVE DIVERSITY PARTITIONING
##Read in sample level file for ADP, convert to factors, and carry
out ADP
fact.all <- read.table (file = "Sample_factors_all.txt",
header=TRUE, row.names=1)
fact.all.factors <- data.frame(
l1=as.factor(fact.all[,1]),
l2=as.factor(fact.all[,2]),
l3=as.factor(fact.all[,3]))

(txal.noSing.ADP <- adipart(txal.noSing ~ ., fact.all.factors,
index = "richness", weights = "prop", nsimul = 9999))

##Repeat for lignites only (limit dataset to lignite samples, and
remove uniques)
txal.noSing.lig <- txal.noSing[sample.info$Lithology ==
"Lignite",]
txal.noSing.lig <- txal.noSing.lig[,colSums(txal.noSing.lig > 0)
> 1]

fact.lig <- read.table (file = "Sample_factors_lignite.txt",
```

```
header=TRUE, row.names=1)
fact.lig.factors <- data.frame(
l1=as.factor(fact.lig[,1]),
l2=as.factor(fact.lig[,2]),
l3=as.factor(fact.lig[,3]))

(txal.noSing.lig.ADP <- adipart(txal.noSing.lig ~ .,
fact.lig.factors, index = "richness", weights = "prop", nsimul =
9999))

##Prepare results for plotting
txal.noSing.ADP.results <- cbind(txal.noSing.ADP[[1]],
apply(txal.noSing.ADP[[2]]$simulated, 1, median))
colnames(txal.noSing.ADP.results) = c("Observed", "Expected")

txal.noSing.lig.ADP.results <- cbind(txal.noSing.lig.ADP[[1]],
apply(txal.noSing.lig.ADP[[2]]$simulated, 1, median))
colnames(txal.noSing.lig.ADP.results) = c("Observed",
"Expected")

##Figure 6.4
par(mfrow = c(1,2))
barplot(txal.noSing.ADP.results[c(1,4,5),], ylim = c(0,120),
main = "All samples", ylab = "Number of taxa")
barplot(txal.noSing.lig.ADP.results[c(1,4,5),], legend.text = T,
ylim = c(0,120), main = "Lignite samples")

##NPMANOVA
##Whole dataset
(txal.adonis.form <- adonis(txal.noSing.dist ~ Formation, data =
sample.info, permutations = 9999))
(txal.adonis.form.lith <- adonis(txal.noSing.dist ~
Formation*Lithology, data = sample.info, permutations = 9999))

##Lignites only
sample.info.lig <- sample.info[sample.info$Lithology ==
"Lignite",]
txal.noSing.lig.dist <- vegdist(txal.noSing.lig, method =
"jaccard", binary = T)
(txal.lig.adonis <- adonis(txal.noSing.lig.dist ~ Formation, data
= sample.info.lig, permutations = 9999))


##NMDS ANALYSIS
##Whole dataset
meta.txal <- metaMDS(txal.nosing, distance = "jaccard", binary =
TRUE)

##Figure 6.5
plot(meta.txal, display = c("sites"), choices = c(1,2), type="n",
shrink=FALSE, xlab = "NMDS Axis 1", ylab = "NMDS Axis 2")
```

```
points(meta.txal, display = "sites", choices = c(1,2),
shrink=FALSE, pch = 1, Formation == "Tuscahoma" & Lithology ==
"Lignite")
points(meta.txal, display = "sites", choices = c(1,2),
shrink=FALSE, pch = 21, bg = "black", Formation == "Calvert.Bluff"
& Lithology == "Lignite")
points(meta.txal, display = "sites", choices = c(1,2),
shrink=FALSE, pch = 24, bg = "black", Formation == "Calvert.Bluff"
& Lithology == "Clastic")
points(meta.txal, display = "sites", choices = c(1,2),
shrink=FALSE, pch = 2, Formation == "Tuscahoma" & Lithology ==
"Clastic")

##NMDS for Calvert Bluff Formation
calvert <- txal.noSing[Formation == "Calvert.Bluff",]
calvert <- calvert[,colSums(calvert > 0) > 1]
calvert.info <- sample.info[Formation == "Calvert.Bluff",]
meta.calvert <- metaMDS(calvert, distance = "jaccard", binary =
T)

##Figure 6.6A
plot(meta.calvert, display = "sites", choices = c(1,2), type="n",
shrink=FALSE, xlab = "NMDS Axis 1", ylab = "NMDS Axis 2")
points(meta.calvert, display = "sites", choices = c(1,2),
type="p", shrink=FALSE, pch = 24, bg = "black",
(calvert.info$Locality == "NAC" | calvert.info$Locality == "UCB"
| calvert.info$Locality == "BGC") & calvert.info$Lithology ==
"Clastic")
points(meta.calvert, display = "sites", choices = c(1,2),
type="p", shrink=FALSE, pch = 21, bg = "black",
(calvert.info$Locality == "NAC" | calvert.info$Locality == "UCB"
| calvert.info$Locality == "BGC") & calvert.info$Lithology ==
"Lignite")
points(meta.calvert, display = "sites", choices = c(1,2),
type="p", shrink=FALSE, pch = 2, bg = "black",
(calvert.info$Locality == "WCM" | calvert.info$Locality == "SM")
& calvert.info$Lithology == "Clastic")
points(meta.calvert, display = "sites", choices = c(1,2),
type="p", shrink=FALSE, pch = 1, bg = "black",
(calvert.info$Locality == "WCM" | calvert.info$Locality == "SM")
& calvert.info$Lithology == "Lignite")

##NMDS for Tuscahoma Formation
tuscahoma <- txal.noSing[Formation == "Tuscahoma",]
tuscahoma <- tuscahoma[,colSums(tuscahoma > 0) > 1]
tusc.info <- sample.info[Formation == "Tuscahoma",]
meta.tusc <- metaMDS(tuscahoma, distance = "jaccard", binary = T)

##Figure 6.6B
plot(meta.tusc, display = "sites", choices = c(1,2), type="n",
shrink=FALSE, main = "NMDS plot showing Tuscahoma samples", xlab
```

```
= "NMDS Axis 1", ylab = "NMDS Axis 2")
text(meta.tusc, display = "sites", choices = c(1,2), labels =
c(26:1, "GH00/AC/012", "GH00/AC/009"), shrink=FALSE)
points(meta.tusc, display = "sites", choices = c(1,2), pch=21, cex
= 4, tusc.info$Lithology == "Lignite", shrink=FALSE)


##CLUSTER ANALYSIS
##Distance matrix and cluster analysis for samples
txal.sites.dist <- vegdist(txal.noSing, method = "jaccard",
binary = TRUE)
txal.sites.clust <- hclust(txal.sites.dist, method = "average")

##Distance matrix and cluster analysis for taxa
txal.species.dist <- vegdist(t(txal.noSing), method = "jaccard",
binary = TRUE)
txal.species.clust <- hclust(txal.species.dist, method =
"average")

##Figure 6.7 - dendrograms exported to Illustrator for figure
preparation
plot(txal.sites.clust)
plot(txal.sites.clust, labels = Formation)
plot(txal.sites.clust, labels = Lithology)

plot(txal.species.clust)

##COMPARISON OF COMMON TAXA
calvert <- txal.noSing[Formation == "Calvert.Bluff",]
calv.inc <- colSums(calvert > 0)
calv.inc <- sort(calv.inc, decreasing = T)

tuscahoma <- txal.noSing[Formation == "Tuscahoma",]
tusc.inc <- colSums(tuscahoma > 0)
tusc.inc <- sort(tusc.inc, decreasing = T)

##Figure 6.8
par(mfrow = c(2,1))
barplot(calv.inc[1:15], las = 2, cex.names = 0.7, font.axis = 3,
ylim = c(0,30), ylab = "Number of samples", main = "Calvert Bluff
Fm., TX")
barplot(tusc.inc[1:15], las = 2, cex.names = 0.7, font.axis = 3,
ylim = c(0,30), ylab = "Number of samples", xlab = "Taxa", main
= "Tuscahoma Fm., MS/AL")

------------------------------------------------------------
-----------------------------------

##COMPARISON WITH HOLOCENE DATA
##Construct geographical inter-sample distance matrices
library(spdep)
```

```
##Temperate data, coordinates for each lake core in rows
temp <- read.table (file = "Temperate_coords.txt", header=TRUE,
row.names=1)
temp <- as.matrix(temp)
temp.knn <- knearneigh(temp, k=10, longlat=TRUE)
temp.nb <- knn2nb(temp.knn)
temp.dist <- nbdists(temp.nb, temp, longlat = TRUE)
temp.dist.matrix <- matrix(unlist(temp.dist), byrow = T, nrow =
11)

##Save matrix for formatting in Excel (e.g. adding in zero
distances on the diagonal)
write.table(temp.dist.matrix, "Temp_dist.txt", sep = "\t",
row.names = T, col.names = T)

##Repeat for each of other datasets - warm mixed, tropical,
temperate with combined sites, warm mixed with combined sites

##Regression of inter-sample compositional similarity onto
inter-sample geographic distance
##Temperate compositional data matrix
temperate.comm <- read.table (file = "Temperate.txt",
header=TRUE, row.names=1)
temperate.comm <- temperate.comm[,colSums(temperate.comm > 0) !=
1]
temperate.comm.dist <- vegdist(temperate.comm, method =
"jaccard", binary = TRUE)
##Take natural log and convert from distance to similarity
temperate.comm.dist <- log(1-temperate.comm.dist)
temperate.comm.dist.vec <- as.vector(temperate.comm.dist)

##Temperate geographic distance matrix (now correctly formatted)
temperate.geog <- read.table (file = "Temperate_dist.txt",
header=TRUE, row.names=1)
temperate.geog.dist <- as.dist(temperate.geog)
temperate.geog.dist.vec <- as.vector(temperate.geog.dist)

##Regression
temperate.obs.ind <- lm(temperate.comm.dist.vec ~
temperate.geog.dist.vec)
summary(temperate.obs.ind)

##Randomization test of significance
temperate.r.squared <- numeric(9999)
temperate.F <- numeric(9999)
for(i in 1:9999) {
    model <- lm(sample(temperate.comm.dist.vec,
length(temperate.comm.dist.vec)) ~ temperate.geog.dist.vec)
    temperate.r.squared[i] <- summary(model)[[8]]
    temperate.F[i] <- summary(model)[[10]][1]
```

```
}

##Significance of R^2 value
temperate.r.squared <- c(temperate.r.squared,
summary(temperate.obs.ind)[[8]])
1-sum(temperate.r.squared<summary(temperate.obs.ind)[[8]])/len
gth(temperate.r.squared)

##Significance of F value
temperate.F <- c(temperate.F,
summary(temperate.obs.ind)[[10]][1])
1-(sum(temperate.F<summary(temperate.obs.ind)[[10]][1])/length
(temperate.F))

##Should be the same!
##Again, repeated for each of the other Holocene datasets, so that
each one has a regression model with significance

##Calculate GCP Paleocene inter-formation, inter-sample
compositional similarity
##Limit to clastic samples, and remove unique taxa (i.e. now
limited to one sample)
txal.nosing.clast <- txal.nosing[Lithology == "Clastic",]
txal.nosing.clast <-
txal.nosing.clast[,colSums(txal.nosing.clast > 0) > 1]
txal.nosing.clast.dist <- vegdist(txal.nosing.clast, method =
"jaccard", binary = TRUE)
txal.nosing.clast.dist.matrix <-
as.matrix(txal.nosing.clast.dist)
##Only inter-formation distances wanted:
txal.nosing.clast.dist.vec <-
as.vector(txal.nosing.clast.dist.matrix[23:45,1:22])
txal.nosing.clast.dist.vec <- log(1-txal.nosing.clast.dist.vec)

##Mean similarity
mean(txal.nosing.clast.dist.vec)

##Test for normality
shapiro.test(txal.nosing.clast.dist.vec)

##Non-parametric (bootstrapped) 95% CIs
boot.mean.clast <- numeric(10000)
for(i in 1:10000) boot.mean.clast[i] <-
mean(sample(txal.nosing.clast.dist.vec,
length(txal.nosing.clast.dist.vec), replace = T))
quantile(boot.mean.clast, c(.025, .975))

##Also repeated with whole dataset (e.g. with clastics and
lignites, using txal.nosing rather than txal.nosing.clast)

##Figure 6.10A
```

```
plot(tropics.geog.dist.vec, tropics.comm.dist.vec, col =
"grey40", ylim = c(-1.4, 0), xlim  = c(0, 1300), xlab = "Distance
(km)", ylab = "ln Jaccard similarity")
points(warm.mixed.geog.dist.vec, warm.mixed.comm.dist.vec,
pch=0, col = "grey40")
points(temperate.geog.dist.vec, temperate.comm.dist.vec, pch=2,
col = "grey40")
abline(tropics.obs.ind)
abline(warm.mixed.obs.ind, lty = 2)
abline(temperate.obs.ind, lty = 3)
points(819, -0.8317677, pch = 4)
lines(c(819, 819), c(-0.8442460, -0.8194573))

##Figure 6.10B
plot(temperate.comb.geog.dist.vec,
temperate.comb.comm.dist.vec, pch=2, col = "grey40", ylim= c(-1,
0), xlab = "Distance (km)", ylab = "ln Jaccard similarity")
points(warm.mixed.comb.geog.dist.vec,
warm.mixed.comb.comm.dist.vec, pch=0, col = "grey40")
abline(temperate.comb.obs.ind, lty = 3)
abline(warm.mixed.comb.obs.ind, lty = 2)
points(819, -0.8317677, pch = 4)
lines(c(819, 819), c(-0.8442460, -0.8194573))
```