8. Worksheet: Among Site (Beta) Diversity – Part 1

Brooke Peckenpaugh; Z620: Quantitative Biodiversity, Indiana University

06 February, 2019

OVERVIEW

In this worksheet, we move beyond the investigation of within-site α-diversity. We will explore β-diversity, which is defined as the diversity that occurs among sites. This requires that we examine the compositional similarity of assemblages that vary in space or time.

After completing this exercise you will know how to:

1. formally quantify β-diversity
2. visualize β-diversity with heatmaps, cluster analysis, and ordination
3. test hypotheses about β-diversity using multivariate statistics

Directions:

1. In the Markdown version of this document in your cloned repo, change "Student Name" on line 3 (above) with your name.
2. Complete as much of the worksheet as possible during class.
3. Use the handout as a guide; it contains a more complete description of data sets along with examples of proper scripting needed to carry out the exercises.
4. Answer questions in the worksheet. Space for your answers is provided in this document and is indicated by the ">" character. If you need a second paragraph be sure to start the first line with ">". You should notice that the answer is highlighted in green by RStudio (color may vary if you changed the editor theme).
5. Before you leave the classroom today, it is *imperative* that you **push** this file to your GitHub repo, at whatever stage you are. Ths will enable you to pull your work onto your own computer.
6. When you have completed the worksheet, **Knit** the text and code into a single PDF file by pressing the `Knit` button in the RStudio scripting panel. This will save the PDF output in your '8.BetaDiversity' folder.
7. After Knitting, please submit the worksheet by making a **push** to your GitHub repo and then create a **pull request** via GitHub. Your pull request should include this file (**8.BetaDiversity_1_Worksheet.Rmd**) with all code blocks filled out and questions answered) and the PDF output of `Knitr` (**8.BetaDiversity_1_Worksheet.pdf**).

The completed exercise is due on **Wednesday, February 6th, 2019 before 12:00 PM (noon)**.

1) R SETUP

Typically, the first thing you will do in either an R script or an RMarkdown file is setup your environment. This includes things such as setting the working directory and loading any packages that you will need.

In the R code chunk below, provide the code to:

1. clear your R environment,
2. print your current working directory,
3. set your working directory to your "*/8.BetaDiversity*" folder, and
4. load the `vegan` R package (be sure to install if needed).

```
rm(list = ls())
setwd("~/GitHub/QB2019_Peckenpaugh/2.Worksheets/8.BetaDiversity")

package.list <- c('vegan', 'ade4', 'viridis', 'gplots', 'BiodiversityR', 'indicspecies')
for (package in package.list) {
  if (!require(package, character.only=TRUE, quietly=TRUE)) {
    install.packages(package)
    library(package, character.only=TRUE)
  }
}
```

```
## This is vegan 2.5-3
```

```
##
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:stats':
##
##     lowess
```

```
## BiodiversityR 2.11-1: Use command BiodiversityRGUI() to launch the Graphical User Interface;
## to see changes use BiodiversityRGUI(changeLog=TRUE, backward.compatibility.messages=TRUE)
```

2) LOADING DATA

Load dataset

In the R code chunk below, do the following:

1. load the `doubs` dataset from the `ade4` package, and
2. explore the structure of the dataset.

```
# note, pleae do not print the dataset when submitting
data(doubs)
str(doubs, max.level = 1)
```

```
## List of 4
##  $ env    :'data.frame': 30 obs. of  11 variables:
##  $ fish   :'data.frame': 30 obs. of  27 variables:
##  $ xy     :'data.frame': 30 obs. of  2 variables:
##  $ species:'data.frame': 27 obs. of  4 variables:
```

```
head(doubs$env)
```

```
##   dfs alt   slo flo pH har pho nit amm oxy bdo
## 1   3 934 6.176  84 79  45   1  20   0 122  27
## 2  22 932 3.434 100 80  40   2  20  10 103  19
## 3 102 914 3.638 180 83  52   5  22   5 105  35
## 4 185 854 3.497 253 80  72  10  21   0 110  13
## 5 215 849 3.178 264 81  84  38  52  20  80  62
## 6 324 846 3.497 286 79  60  20  15   0 102  53
```

Question 1: Describe some of the attributes of the `doubs` dataset.

a. How many objects are in `doubs`?
b. How many fish species are there in the `doubs` dataset?
c. How many sites are in the `doubs` dataset?

Answer 1a: 4 objects. ***Answer 1b***: 27 species. ***Answer 1c***: 30 sites.

Visualizing the Doubs River Dataset

Question 2: Answer the following questions based on the spatial patterns of richness (i.e., α-diversity) and Brown Trout (*Salmo trutta*) abundance in the Doubs River.

a. How does fish richness vary along the sampled reach of the Doubs River?
b. How does Brown Trout (*Salmo trutta*) abundance vary along the sampled reach of the Doubs River?
c. What do these patterns say about the limitations of using richness when examining patterns of biodiversity?

Answer 2a: Fish richness is highest downstream and in the middle of the stream, at high Y-coordinates. ***Answer 2b***: Brown trout abundance is highest upstream, at lower Y-coordinates. ***Answer 2c***: These patterns suggest that while richness gives you an overall idea about species diversity, it does not convey information about individual species (which may not follow that pattern).

3) QUANTIFYING BETA-DIVERSITY

In the R code chunk below, do the following:

1. write a function (`beta.w()`) to calculate Whittaker's β-diversity (i.e., β_w) that accepts a site-by-species matrix with optional arguments to specify pairwise turnover between two sites, and
2. use this function to analyze various aspects of β-diversity in the Doubs River.

```
beta.w <- function(site.by.species = ""){
  SbyS.pa <- decostand(site.by.species, method = "pa")
  S <- ncol(SbyS.pa[,which(colSums(SbyS.pa) > 0)])
  a.bar <- mean(specnumber(SbyS.pa))
  b.w <- round(S/a.bar, 3)
  return(b.w)
}

beta.w <- function(site.by.species = "", sitenum1 = "", sitenum2 = "", pairwise = FALSE){
  if (pairwise == TRUE){
    if (sitenum1 == "" | sitenum2 == "") {
      print("Error: please specify sites to compare")
      return(NA)}
    site1 = site.by.species[sitenum1,]
    site2 = site.by.species[sitenum2,]
    site1 = subset(site1, select = site1 > 0)
    site2 = subset(site2, select = site2 > 0)
    gamma = union(colnames(site1), colnames(site2))
    s = length(gamma)
    a.bar = mean(c(specnumber(site1), specnumber(site2)))
    b.w = round(s/a.bar - 1, 3)
    return(b.w)
  }
  else{
    SbyS.pa <- decostand(site.by.species, method = "pa")
    S <- ncol(SbyS.pa[,which(colSums(SbyS.pa) > 0)])
    a.bar <- mean(specnumber(SbyS.pa))
    b.w <- round(S/a.bar, 3)
```

```
    return(b.w)
  }
}

beta.w(doubs$fish, 1, 2, pairwise = TRUE)
```

```
## [1] 0.5
```

```
beta.w(doubs$fish, 1, 10, pairwise = TRUE)
```

```
## [1] 0.714
```

Question 3: Using your `beta.w()` function above, answer the following questions:

a. Describe how local richness (α) and turnover (β) contribute to regional (γ) fish diversity in the Doubs.
b. Is the fish assemblage at site 1 more similar to the one at site 2 or site 10?
c. Using your understanding of the equation $\beta_w = \gamma/\alpha$, how would your interpretation of β change if we instead defined beta additively (i.e., $\beta = \gamma - \alpha$)?

> ***Answer 3a***: Local richness and turnover both increase regional fish diversity. ***Answer 3b***: The fish assemblage at site 1 is more similar to the one as site 2 than site 10. ***Answer 3c***: My interpretation would be similar (lower values = more homogeneous sites), but it would no longer be on a scale from 0 to 1.

The Resemblance Matrix

In order to quantify β-diversity for more than two samples, we need to introduce a new primary ecological data structure: the **Resemblance Matrix**.

Question 4: How do incidence- and abundance-based metrics differ in their treatment of rare species?

> ***Answer 4***: I think incidence-based metrics define rarity by how many times a species is identified among sites, whereas abundance-based metrics define rarity by how many times a species is identified within each site.

In the R code chunk below, do the following:

1. make a new object, `fish`, containing the fish abundance data for the Doubs River,
2. remove any sites where no fish were observed (i.e., rows with sum of zero),
3. construct a resemblance matrix based on Sørensen's Similarity ("fish.ds"), and
4. construct a resemblance matrix based on Bray-Curtis Distance ("fish.db").

Question 5: Using the distance matrices from above, answer the following questions:

a. Does the resemblance matrix (`fish.db`) represent similarity or dissimilarity? What information in the resemblance matrix led you to arrive at your answer?
b. Compare the resemblance matrices (`fish.db` or `fish.ds`) you just created. How does the choice of the Sørensen or Bray-Curtis distance influence your interpretation of site (dis)similarity?

> ***Answer 5a***:
> ***Answer 5b***:

4) VISUALIZING BETA-DIVERSITY

A. Heatmaps

In the R code chunk below, do the following:

1. define a color palette,
2. define the order of sites in the Doubs River, and
3. use the `levelplot()` function to create a heatmap of fish abundances in the Doubs River.

B. Cluster Analysis

In the R code chunk below, do the following:

1. perform a cluster analysis using Ward's Clustering, and
2. plot your cluster analysis (use either `hclust` or `heatmap.2`).

Question 6: Based on cluster analyses and the introductory plots that we generated after loading the data, develop an ecological hypothesis for fish diversity the `doubs` data set?

Answer 6:

C. Ordination

Principal Coordinates Analysis (PCoA)

In the R code chunk below, do the following:

1. perform a Principal Coordinates Analysis to visualize beta-diversity
2. calculate the variation explained by the first three axes in your ordination
3. plot the PCoA ordination,
4. label the sites as points using the Doubs River site number, and
5. identify influential species and add species coordinates to PCoA plot.

In the R code chunk below, do the following:

1. identify influential species based on correlations along each PCoA axis (use a cutoff of 0.70), and
2. use a permutation test (999 permutations) to test the correlations of each species along each axis.

Question 7: Address the following questions about the ordination results of the `doubs` data set:

a. Describe the grouping of sites in the Doubs River based on fish community composition.
b. Generate a hypothesis about which fish species are potential indicators of river quality.

Answer 7a:
Answer 7b:

SYNTHESIS

Using the jelly bean data from class (i.e., JellyBeans.Source.txt and JellyBeans.txt):

1) Compare the average pairwise similarity among subsamples in group A to the average pairswise similarity among subsamples in group B. Use a t-test to determine whether compositional similarity was affected by the "vicariance" event. Finally, compare the compositional similarity of jelly beans in group A and group B to the source community?

2) Create a cluster diagram or ordination using the jelly bean data. Are there any visual trends that would suggest a difference in composition between group A and group B?

```
setwd("~/GitHub/QB2019_Peckenpaugh/2.Worksheets/6.DiversitySampling")
jb <- read.table("JellyBeans.txt", sep="\t", header=T)
jbsource <- read.table("JellyBeans.Source.txt", sep="\t", header=T)

#Adjusting jelly bean data to conform to source data labels
```

```r
jb$GreenTrans <- jb$GreenTrans + jb$GreenTrans2
jb <- jb[-15]
jb$Rainbow <- jb$Rainbow + jb$WhiteSolid
jb <- jb[-29]

#Dividing data into groups, getting rid of factors
jb.a <- subset(jb, Group == "A")
jb.b <- subset(jb, Group == "B")

jb.a <- jb.a[3:28]
jb.b <- jb.b[3:28]

#Creating resemblance matrices for groups A and B
jb.rm.a <- vegdist(jb.a, method = "bray")
jb.rm.b <- vegdist(jb.b, method = "bray")

#T test for groups A and B
t.test(jb.rm.a, jb.rm.b)
```

```
##
##  Welch Two Sample t-test
##
## data:  jb.rm.a and jb.rm.b
## t = -2.5912, df = 7.5291, p-value = 0.03372
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.124214114 -0.006556611
## sample estimates:
## mean of x mean of y
## 0.2649123 0.3302977
```

```r
#Reformatting jelly bean source data for Bray-Curtis dissimilarity. Also reordering the columns to matc
jbsource <- t(jbsource)
jbsource <- jbsource[3:4, ]
colnames(jbsource) = jbsource[1, ]
jbsource = jbsource[-1, ]
jbsource <- jbsource[c("Red", "RedShinny", "OrangeSpe", "OrangeDrk", "OrangeBrgt", "OrangeShn", "Orange
jbsource <- as.numeric(jbsource)
jb.rm.source <- vegdist(jbsource, method = "bray")

#T tests between groups A, B, and source data
t.test(jb.rm.a, jb.rm.source)
```

```
##
##  Welch Two Sample t-test
##
## data:  jb.rm.a and jb.rm.source
## t = -4.1655, df = 48.102, p-value = 0.0001283
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.10685861 -0.03728623
## sample estimates:
## mean of x mean of y
## 0.2649123 0.3369848
```

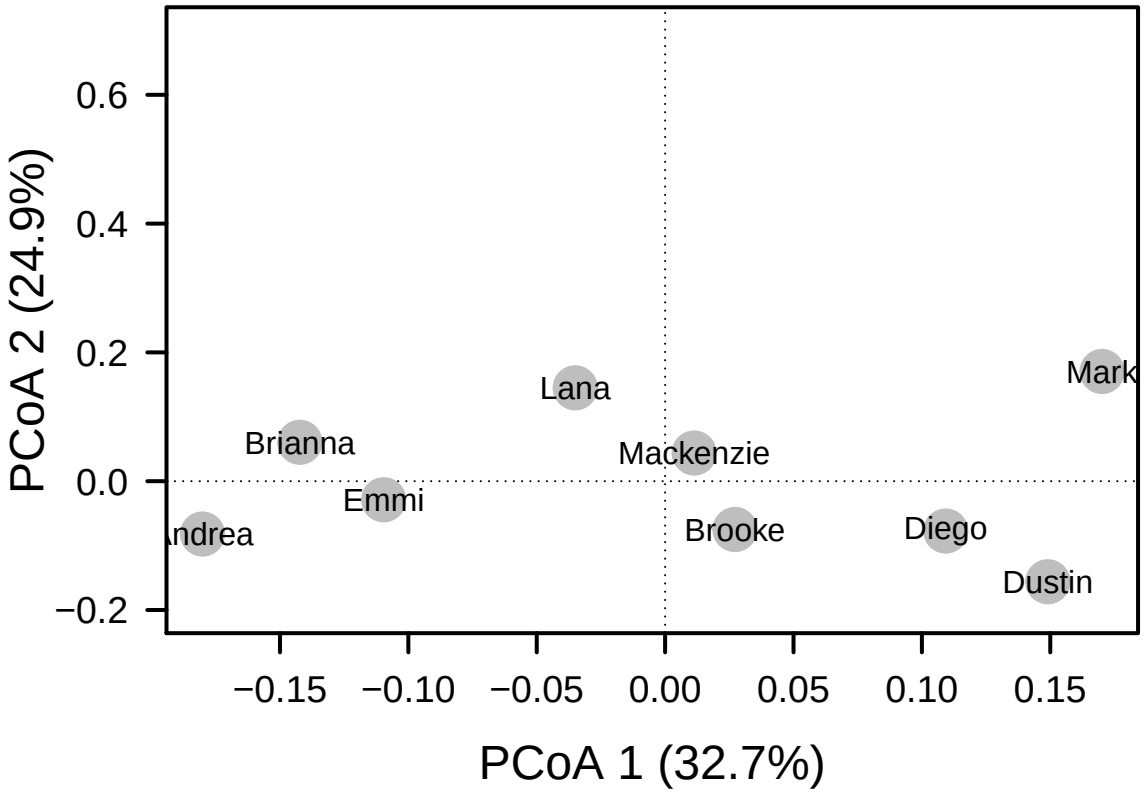
```r
t.test(jb.rm.b, jb.rm.source)
```

```
##
##  Welch Two Sample t-test
##
## data:  jb.rm.b and jb.rm.source
## t = -0.25585, df = 8.9472, p-value = 0.8039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.06586566  0.05249155
## sample estimates:
## mean of x mean of y
## 0.3302977 0.3369848
```

```r
#Reformatting jelly bean data for Bray-Curtis dissimilarity & PCoA
rownames(jb) = jb[ ,2]
jb <- jb[3:28]
jb.rm <- vegdist(jb, method = "bray")
jb.pcoa <- cmdscale(jb.rm, eig = TRUE, k = 3)

explainvar1 <- round(jb.pcoa$eig[1] / sum(jb.pcoa$eig), 3) * 100
explainvar2 <- round(jb.pcoa$eig[2] / sum(jb.pcoa$eig), 3) * 100
explainvar3 <- round(jb.pcoa$eig[3] / sum(jb.pcoa$eig), 3) * 100
sum.eig <- sum(explainvar1, explainvar2, explainvar3)

#Plotting PCoA
par(mar = c(5, 5, 1, 2) + 0.1)
plot(jb.pcoa$points[ ,1], jb.pcoa$points[ ,2], ylim = c(-0.2, 0.7),
     xlab = paste("PCoA 1 (", explainvar1, "%)", sep = ""),
     ylab = paste("PCoA 2 (", explainvar2, "%)", sep = ""),
     pch = 16, cex = 2.0, type = "n", cex.lab = 1.5, cex.axis = 1.2, axes = FALSE)
axis(side = 1, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
axis(side = 2, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
abline(h = 0, v = 0, lty = 3)
box(lwd = 2)
points(jb.pcoa$points[ ,1], jb.pcoa$points[ ,2],
       pch = 19, cex = 3, bg = "gray", col = "gray")
text(jb.pcoa$points[ ,1], jb.pcoa$points[ ,2],
     labels = row.names(jb.pcoa$points))
```



Answer 1): My first t-test suggests that compositional similarity between group A and B was affected by the vicariance event ($p = 0.0337$). I also used t-tests to compare similarity between groups A and B and the source, and found that while group A was significantly different from the source ($p < 0.001$), group B was not ($p = 0.804$).

Answer 2): I created a PCoA for our jelly bean data. Visually, it appears that group A (Andrea, Brianna, Brooke, Emmi, and Mackenzie) are compositionally different from group B (Diego, Dustin, Lana, and Mark).