

# Simple tutorial for alpha/beta diversity

Tom Fu

6/2/2022

## Intro

Here is a simple tutorial on how to do alpha diversity/beta diversity analyses, visualize the results, and conduct group-wise significance testing on the differences between means for the diversities in R. All statistical analyses are done with `vegan` and `betapart`, two known analytic packages for metagenomics analysis; and visualizations are done with `ggplot2`. While I can't believe that no one has put this together to streamline these analysis as there are so many papers out there containing such analysis, here it is!

```
library(betapart)
library(vegan)
```

```
## Loading required package: permute
```

```
## Loading required package: lattice
```

```
## This is vegan 2.6-2
```

```
library(ggplot2)
```

## 1. Load data

You can definitely load and use your own data, here we use a data given in the `vegan` package. As you can see below, the row names are sample names/numbers. and the column names are the metagenomic taxonomies.

```
data("varespec")
head(varespec)
```

```
##      Callvulg Empenigr Rhodtome Vaccmyrt Vaccviti Pinusylv Descflex Betupube
## 18      0.55      11.13      0.00      0.00      17.80      0.07      0.00      0
## 15      0.67      0.17      0.00      0.35      12.13      0.12      0.00      0
## 24      0.10      1.55      0.00      0.00      13.47      0.25      0.00      0
## 27      0.00      15.13      2.42      5.92      15.97      0.00      3.70      0
## 23      0.00      12.68      0.00      0.00      23.73      0.03      0.00      0
## 19      0.00      8.92      0.00      2.42      10.28      0.12      0.02      0
##      Vacculig Diphcomp Dicrsp Dicrfusc Dicrpoly Hylosple Pleuschr Polypili
## 18      1.60      2.07      0.00      1.62      0.00      0.0      4.67      0.02
## 15      0.00      0.00      0.33      10.92      0.02      0.0      37.75      0.02
## 24      0.00      0.00      23.43      0.00      1.68      0.0      32.92      0.00
## 27      1.12      0.00      0.00      3.63      0.00      6.7      58.07      0.00
## 23      0.00      0.00      0.00      3.42      0.02      0.0      19.42      0.02
## 19      0.00      0.00      0.00      0.32      0.02      0.0      21.03      0.02
##      Polyjuni Polycomm Pohlnuta Ptilcili Barbhatc Cladarbu Cladrang Cladstel
## 18      0.13      0.00      0.13      0.12      0.00      21.73      21.47      3.50
## 15      0.23      0.00      0.03      0.02      0.00      12.05      8.13      0.18
## 24      0.23      0.00      0.32      0.03      0.00      3.58      5.52      0.07
## 27      0.00      0.13      0.02      0.08      0.08      1.42      7.63      2.55
## 23      2.12      0.00      0.17      1.80      0.02      9.08      9.22      0.05
## 19      1.58      0.18      0.07      0.27      0.02      7.23      4.95      22.08
##      Cladunci Cladcocc Cladcorn Cladgrac Cladfimb Cladcris Cladchlo Cladbotr
## 18      0.30      0.18      0.23      0.25      0.25      0.23      0.00      0.00
## 15      2.65      0.13      0.18      0.23      0.25      1.23      0.00      0.00
## 24      8.93      0.00      0.20      0.48      0.00      0.07      0.10      0.02
## 27      0.15      0.00      0.38      0.12      0.10      0.03      0.00      0.02
## 23      0.73      0.08      1.42      0.50      0.17      1.78      0.05      0.05
## 19      0.25      0.10      0.25      0.18      0.10      0.12      0.05      0.02
##      Cladamau Cladsp Cetreric Cetrisla Flavnya Nepharct Sterisp Peltapht Icmaeric
## 18      0.08      0.02      0.02      0.00      0.12      0.02      0.62      0.02      0
## 15      0.00      0.00      0.15      0.03      0.00      0.00      0.85      0.00      0
## 24      0.00      0.00      0.78      0.12      0.00      0.00      0.03      0.00      0
## 27      0.00      0.02      0.00      0.00      0.00      0.00      0.00      0.07      0
## 23      0.00      0.00      0.00      0.00      0.02      0.00      1.58      0.33      0
## 19      0.00      0.00      0.00      0.00      0.02      0.00      0.28      0.00      0
##      Cladcerv Claddefo Cladphyl
## 18      0      0.25      0
## 15      0      1.00      0
## 24      0      0.33      0
## 27      0      0.15      0
## 23      0      1.97      0
## 19      0      0.37      0
```

Here we manually synthesize the metadata: two groups we are trying to compare: the “undisturbed” and “disturbed” treatment groups.

```
groups <- factor(c(rep(1,12), rep(2,12)), labels = c("undisturbed","disturbed"))
groups
```

```
## [1] undisturbed undisturbed undisturbed undisturbed undisturbed undisturbed
## [7] undisturbed undisturbed undisturbed undisturbed undisturbed undisturbed
## [13] disturbed disturbed disturbed disturbed disturbed disturbed
## [19] disturbed disturbed disturbed disturbed disturbed disturbed
## Levels: undisturbed disturbed
```

## 2. alpha diversity

Alpha diversity measures the within-sample microbiome diversity. Comparing the alpha diversity of two groups of samples thus look at whether two groups differ significantly in terms of each member's (sample's) species composition. We first generate all the alpha diversity using `vegan`, then compare across the two groups using one-way anova. Note that here we use the Shannon index, you can easily use other statistics such as the Fisher index if you prefer.

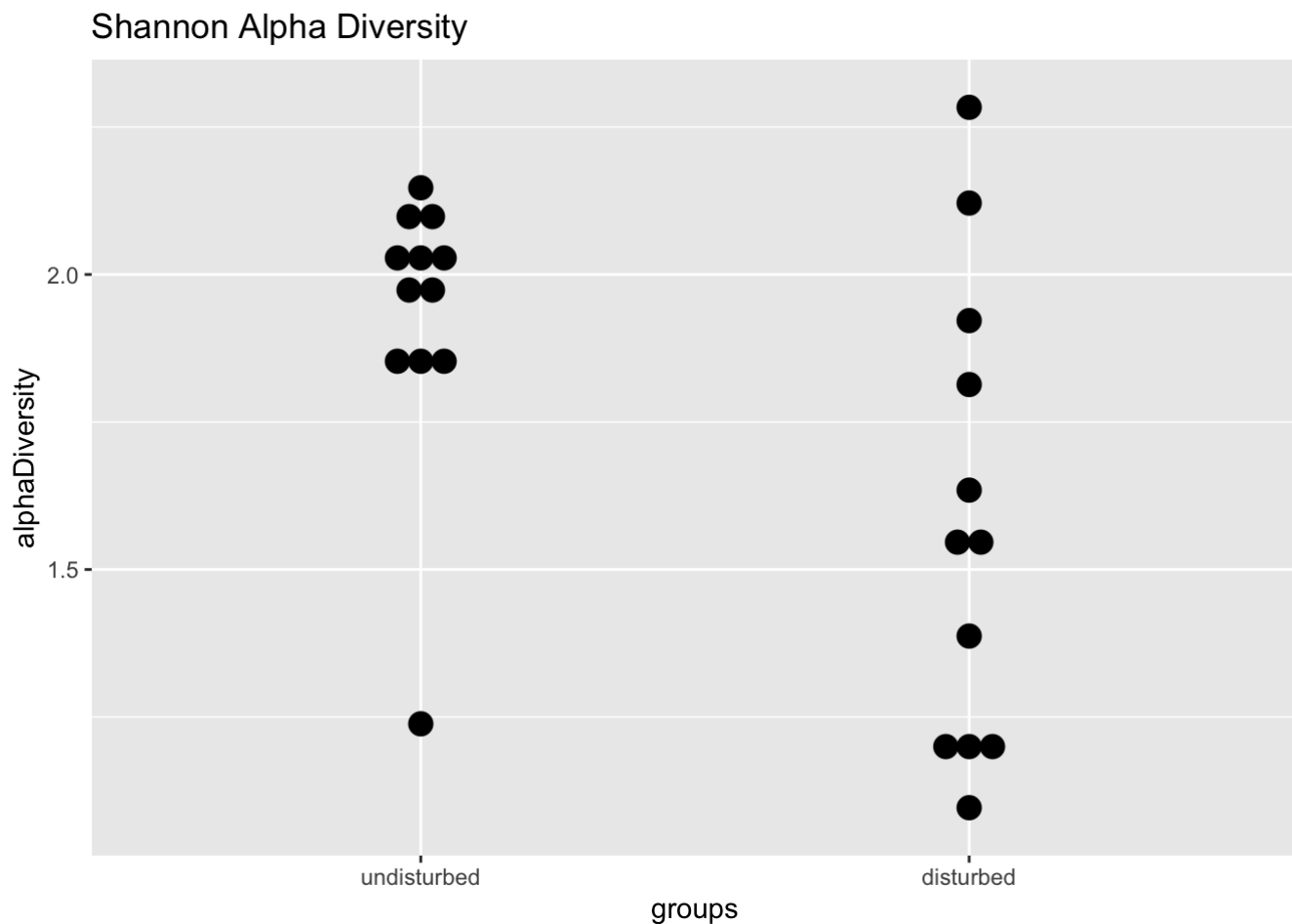
```
alphaDiversity <- diversity(varespec, index="shannon")
summary(aov(alphaDiversity ~ groups))
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## groups      1  0.7362   0.7362    7.011 0.0147 *
## Residuals   22  2.3101   0.1050
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We then visualize the two groups using `ggplot2`.

```
alphaDataframe <- data.frame(alphaDiversity, groups)
p<-ggplot(alphaDataframe, aes(x=groups, y=alphaDiversity)) +
  geom_dotplot(binaxis='y', stackdir='center')+
  ggtitle("Shannon Alpha Diversity")
p
```

```
## Bin width defaults to 1/30 of the range of the data. Pick better value with `binwidth`.
```

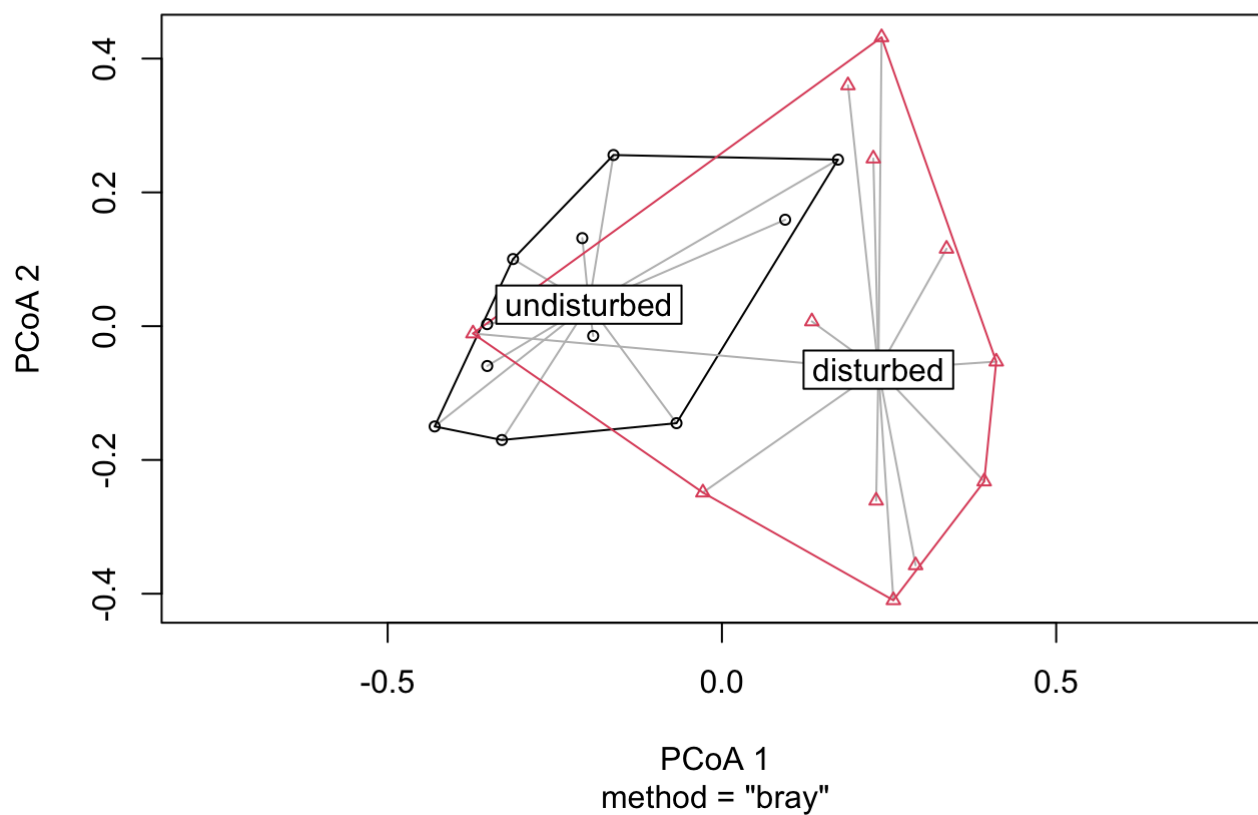


### 3. beta diversity

Beta diversity measures the between-sample similarity/dissimilarity in terms of microbiome diversity. Comparing the beta diversity of two groups of samples thus look at whether two groups differ significantly in terms of how much the microbial diversity samples within each group differ from each other. We now calculate beta diversity, visualize it, and conduct statistical test to compare between two groups.

```
# first compute pairwise bray-curtis similarity matrix
dist <- vegdist(varespec, method = "bray")
# decompose the matrix into 2d dispersion values that are plottable and analyzable
betaDiversityDispersion<-betadisper(dist,groups)
# plot beta diversity and analyze
plot(betaDiversityDispersion)
```

## betaDiversityDispersion



```
anova(betaDiversityDispersion)
```

```
## Analysis of Variance Table
##
## Response: Distances
##          Df Sum Sq Mean Sq F value Pr(>F)
## Groups    1 0.00354  0.0035387   0.1946 0.6634
## Residuals 22 0.39998  0.0181810
```