

A8_LI_ZHIJUN_Meta

Zhijun Li

2022-03-16

GitHub Link: https://github.com/zazauwu/A8_metabar

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'  
## The following objects are masked from 'package:stats':  
##  
##   filter, lag  
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)  
library(ape)  
library(ggtree)
```

```
## ggtree v3.2.1 For help: https://yulab-smu.top/treedata-book/  
##
```

```
## If you use ggtree in published research, please cite the most appropriate paper(s):  
##
```

```
## 1. Guangchuang Yu. Using ggtree to visualize data on tree-like structures. Current Protocols in Bioinformatics  
## 2. Guangchuang Yu, Tommy Tsan-Yuk Lam, Huachen Zhu, Yi Guan. Two methods for mapping and visualizing  
## 3. Guangchuang Yu, David Smith, Huachen Zhu, Yi Guan, Tommy Tsan-Yuk Lam. ggtree: an R package for visualizing
```

```
##  
## Attaching package: 'ggtree'
```

```
## The following object is masked from 'package:ape':  
##  
##   rotate
```

```
library(vegan)
```

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

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

```
## This is vegan 2.5-7
```

Import Data

```
Sample <- read.csv("./Data/FloristicSurvey.csv")  
Sample$Population <- as.character(Sample$Population)
```

Bray-Curtis dissimilarity

```
# investigate how the communities of plants differ across sample locations
## only keep Quadrates label and numbers of each species
SubSample <- subset(Sample[, 11:44])
row.names(SubSample) <- Sample$Quadrat
```

```
S_dist<- vegdist(SubSample, method = "bray", binary = F)
SDistMat <- as.matrix(S_dist)
```

```
library(reshape2)
SDat <- melt(SDistMat)
```

```
ggplot(SDat, aes(Var1, Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradientn(colours = rainbow(4), alpha(0.8)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))
```

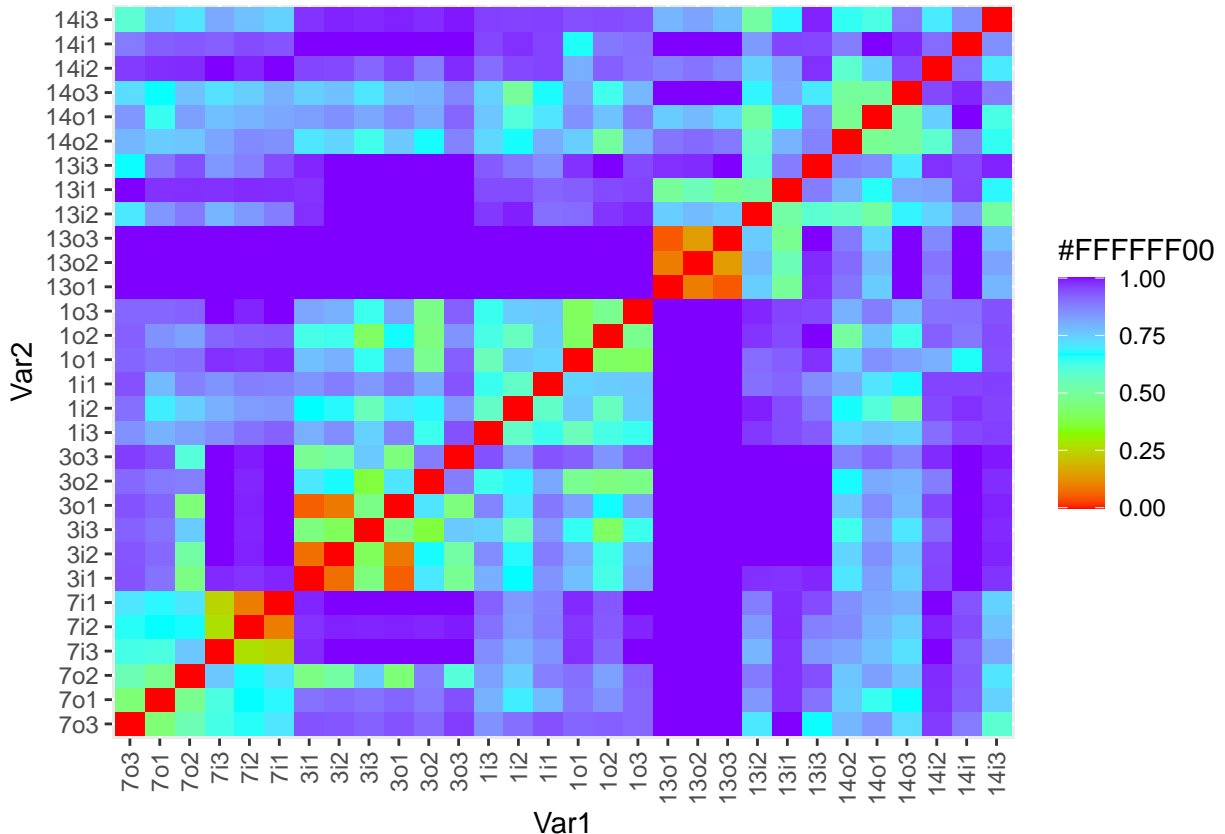


Figure 1. A tile plot that demonstrates similarity/dissimilarity of species abundance across sample locations at QUBS, color gradient from red to purple indicates value from 0 to 1. According to the distance matrix, communities of plants at location 13 with no garlic mustard present are mostly distinct from the rest.

NJ Tree grouped presence of garlic mustard

```
STreeNJ <- nj(S_dist)
```

```
# use gsub to highlight the presence of garlic mustard as to see if it affects the species abundance
```

```

Pres <- gsub(".*([o|i]).*", "\\1", STreeNJ$tip.label)

# group 'tip.labels' by their corresponding treatment
TreatGroups <- split(STreeNJ$tip.label, Pres)

# use the groupOTU function to apply the grouping information for plotting
SDTcol <- groupOTU(STreeNJ, TreatGroups)

# plot
ggtree(SDTcol, layout = "rectangular", aes(color = group)) +
  geom_tiplab(size = 2) +
  theme(legend.position="right") +
  scale_colour_manual(name = "garlic mustard",
                      values = c("Black", "Blue"),
                      labels = c("present", "absent"))

```

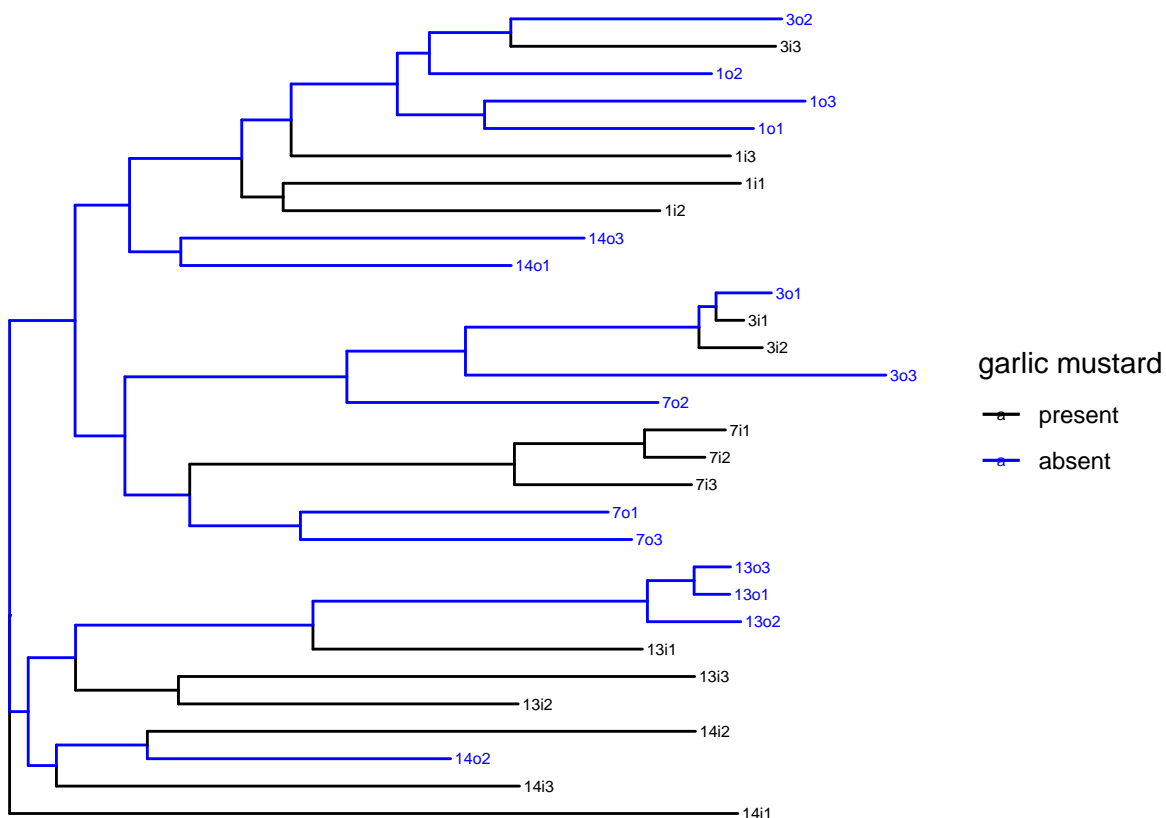


Figure 2. A neighbour-joining tree that clusters samples based on the similarity of their species abundance. The presence of garlic mustard is labeled as present (i.e. in black) and absent (i.e. in blue).

NJ Tree grouped by populations

```

# use gsub to highlight the presentence of garlic mustard as to see if it affects the species abundance
Pop <- gsub("(.*)[o|i].*", "\\1", STreeNJ$tip.label)

# group 'tip.labels' by their corresponding treatment
PopGroups <- split(STreeNJ$tip.label, Pop)

# use the groupOTU function to apply the grouping information for plotting

```

```
SDTcol2 <- groupOTU(STreeNJ, PopGroups)
```

```
# plot
ggtree(SDTcol2, layout = "rectangular", aes(color = group)) +
  geom_tiplab(size = 2) +
  theme(legend.position="right") +
  scale_colour_manual(name = "population",
                     values = rainbow(5),
                     labels = c("1", "3", "7", "13", "14"))
```

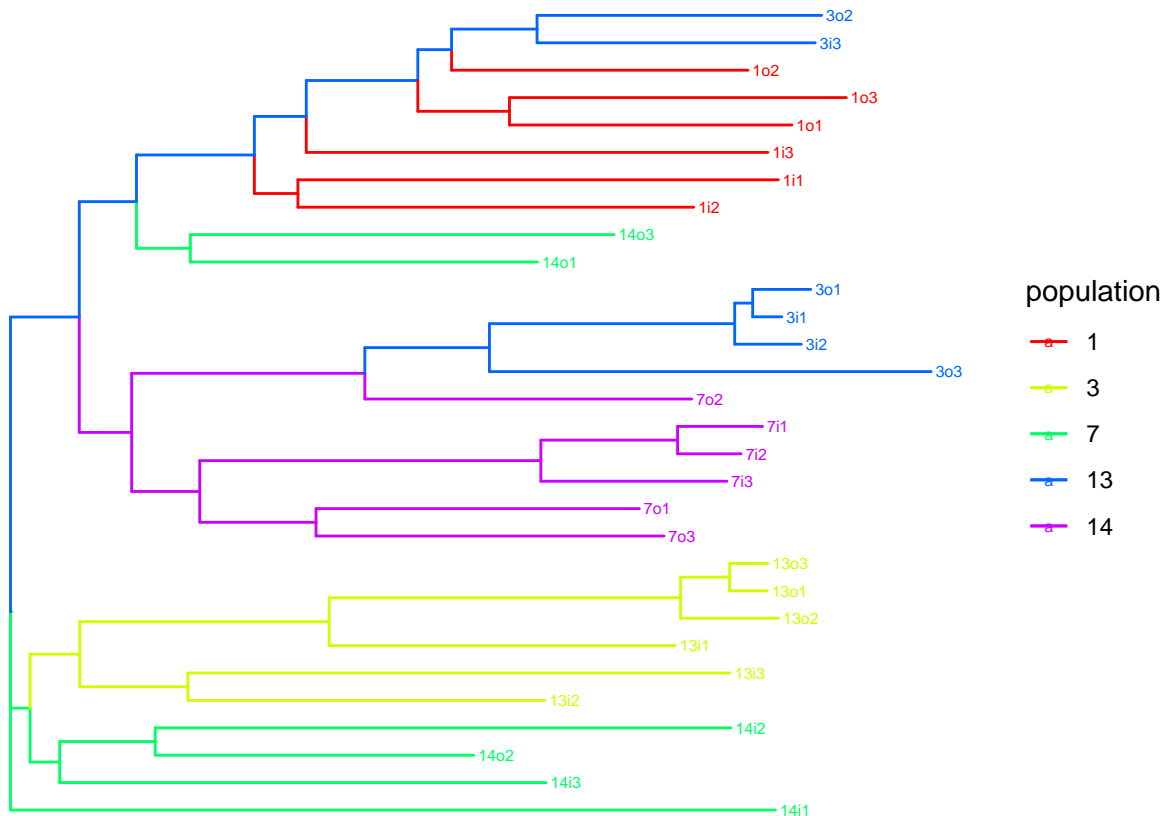


Figure 3. A neighbour-joining tree that clusters samples based on the similarity of their species abundance. The 5 sampling populations are colored in red (1), light green (3), green (7), blue (13), and purple (14).

NMDS

```
set.seed(20)
NMDSdat <- metaMDS(S_dist, k = 2, trymax = 100)

## Run 0 stress 0.1461135
## Run 1 stress 0.1414388
## ... New best solution
## ... Procrustes: rmse 0.06689963 max resid 0.309254
## Run 2 stress 0.1666366
## Run 3 stress 0.1602287
## Run 4 stress 0.1414388
## ... Procrustes: rmse 8.157245e-05 max resid 0.0002927585
## ... Similar to previous best
## Run 5 stress 0.1561092
```

```

## Run 6 stress 0.1666369
## Run 7 stress 0.1461135
## Run 8 stress 0.1414389
## ... Procrustes: rmse 0.000167074  max resid 0.0006280928
## ... Similar to previous best
## Run 9 stress 0.1534441
## Run 10 stress 0.153444
## Run 11 stress 0.1461136
## Run 12 stress 0.1489965
## Run 13 stress 0.1560522
## Run 14 stress 0.1489971
## Run 15 stress 0.1665458
## Run 16 stress 0.1461135
## Run 17 stress 0.1589506
## Run 18 stress 0.1461135
## Run 19 stress 0.1489954
## Run 20 stress 0.1414389
## ... Procrustes: rmse 0.0001979677  max resid 0.0007305364
## ... Similar to previous best
## *** Solution reached

SDat2 <- data.frame(NMDS1 = NMDSdat$points[,1],
                    NMDS2 = NMDSdat$points[,2],
                    Quadrate = row.names(SubSample))

SDat2 <- merge(SDat2, Sample, by = "Quadrate", all.x = T, all.y = F)

qplot(x = NMDS1, y = NMDS2, colour = Population, shape = Location, alpha = I(0.6), data = SDat2) +
  theme_bw()

```

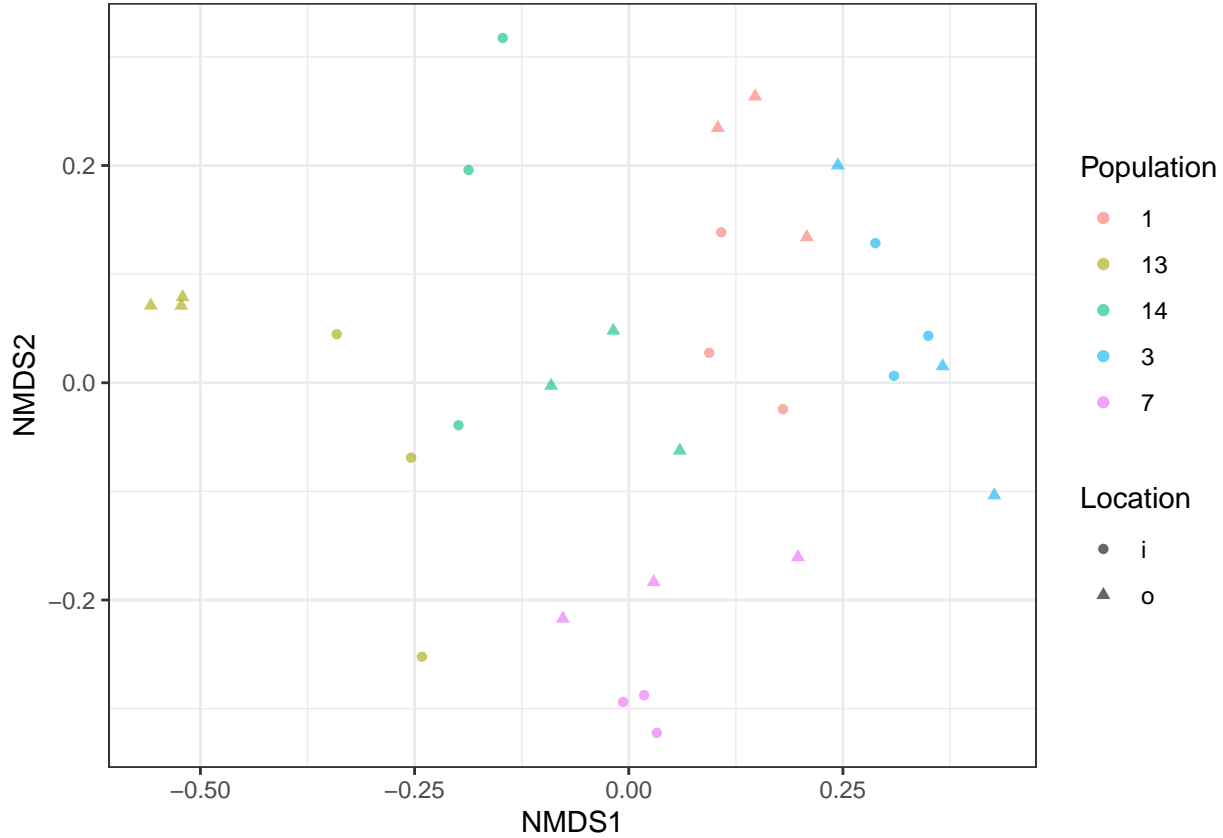


Figure 4. A Non-Metric Multidimensional Scaling (NMDS) plot to visualize similarity/differences among species abundance across sampling population and presence of garlic mustard. Different populations are indicated in colors and the presence (i.e. location) is distinguished in shapes.

Report

Garlic mustard, once introduced, is an invasive plant that could negatively influence both resident communities and ecological interactions. Such biological invasion is widely assumed to be one dominant causation of biodiversity decline. Hence, it is critical to investigate the relationship between garlic invasion and native community richness. In this study, we sampled 6 plant communities from 5 locations at QUBS, respectively, and subgrouped them into 3 by presence of garlic mustard. Each community is identified by distinct Quadrant code. Based on distance matrix visualization, plant species abundance of the communities differ across sample locations, especially the populations at location 13 with no garlic mustard present are mostly distinct from the rest (Figure 1). Given by the neighbour-joining tree, presence of garlic mustard has the potential to drive the differentiation of species abundance among local communities (Figure 2). Similarly, another tree shows that locations can also lead to various biodiversity, regardless of whether garlic mustard is present (Figure 3). However, in accordance with the NMDS analysis, because points from the same populations stay much closer to each other, sampling location has a stronger effect on the plant communities than the invasion of garlic mustard (Figure 4).