A8_Metabarcoding_Assignment

Chenyang Wu

2022/3/16

Project Info

GitHub user name: Wuris

Date: 2022/3/16

GitHub Link: https://github.com/Wuris/Biol432_A8.git

Introduction

Garlic mustard was introduced to North America by European settlers in the 1800s for culinary and medicinal purposes. It is toxic or unpalatable to many native herbivores as well as some native Lepidoptera. Young first-year garlic mustard contains up to 100 ppm of cyanide, a level that is toxic to many vertebrates (*Cipollini and Gruner, 2007*). Since being brought to the United States by settlers, it has naturalized and expanded its range to include much of the Northeast and Midwest, as well as southeastern Canada. It is therefore necessary to conduct a study of this plant, focusing on its effects on the surrounding plant communities.

Biological questions:

1. What effect (if any) does garlic mustard have on the plant community?

2. What has a stronger effect on plant communities: the presence/absence of garlic mustard (in/out) or sampling population?

Method

```
# Load the packages we need
library(dplyr)
library(vegan)
library(reshape2)
library(ggplot2)
library(ape)
library(ggtree)
```

```
# Load the data set and check its structure
A8Data <- read.csv("./Data/FloristicSurvey.csv")
str(A8Data)
```

To test these questions, we will make a distance map for all quadrates with 34 different species data and make a "phylogeny" tree for all the quadrates to see whether the sites within same

population/location are close connect to each other or not. And we will also make a NMDS plot to check how far between each sites.

```
## 'data.frame':    30 obs. of  44 variables:
## $ Quadrate                 : chr  "7o3" "7o1" "7o2" "7i3" ...
## $ Population               : int  7 7 7 7 7 7 3 3 3 3 ...
## $ Location                 : chr  "o" "o" "o" "i" ...
## $ Rosettes                 : int  0 0 0 14 3 0 14 0 0 0 ...
## $ Bolting                  : int  0 0 0 8 18 3 3 12 26 0 ...
## $ Budding                  : int  0 0 0 157 184 122 11 23 19 0 ...
## $ Bud_Flw                  : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Flower                   : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Flw_Sil                  : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Sillique                 : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Claytonia_virginiana     : int  38 29 57 22 54 44 0 0 0 0 ...
## $ Anemone_hepatica         : int  8 7 3 0 0 0 0 0 0 0 ...
## $ Grass_tuft               : int  14 22 6 10 3 0 0 0 0 0 ...
## $ Trillium_grandifolium    : int  9 30 8 0 0 0 0 0 0 0 ...
## $ Erythronium.trout.lily.  : int  25 10 5 16 18 7 0 0 0 0 ...
## $ Acer_saccharum           : int  5 10 74 0 2 0 66 60 26 71 ...
## $ Dicentra_cucularia       : int  0 22 23 0 0 0 0 0 0 0 ...
## $ Bloodroot                : int  0 4 0 0 1 0 0 0 0 0 ...
## $ Gallium_aparine          : int  0 27 8 17 22 21 2 0 0 0 ...
## $ Ulmus_americana          : int  0 3 0 88 150 133 0 0 0 0 ...
## $ Unknown_1                : int  0 1 0 0 0 0 0 0 0 0 ...
## $ Unknown_2                : int  0 0 0 5 4 4 0 0 0 0 ...
## $ Unknown_3.rannunculaceae.: int  0 0 0 0 0 0 0 0 0 0 ...
## $ Unknown_4                : int  0 0 0 0 0 0 1 0 0 0 ...
## $ Dryopteris_marginalis    : int  0 0 0 0 0 0 0 1 0 0 ...
## $ Ostrich_fern             : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Plantago_lanceolata      : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Violet                   : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Rhamnus_frangula         : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Raspberry                : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Unknown_5                : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Unknown_6                : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Solidago_canadensis      : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Unknown_7                : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Dandelion                : int  0 0 0 0 0 0 0 0 0 0 ...
## $ grass                    : int  0 0 0 0 0 0 0 0 0 0 ...
## $ viccia_cracca            : int  0 0 0 0 0 0 0 0 0 0 ...
## $ herb_robert              : int  0 0 0 0 0 0 0 0 0 0 ...
## $ thorny_ash               : int  0 0 0 0 0 0 0 0 0 0 ...
## $ rhamnus_cathartica       : int  0 0 0 0 0 0 0 0 0 0 ...
## $ rhubarb                  : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Unknown_8                : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Unknown_9                : int  0 0 0 0 0 0 0 0 0 0 ...
## $ maianthenum_racemosum    : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
# We need Population column be treat as factor
A8Data$Population <- as.character(A8Data$Population)
```

Subset the dataset

```
# Create a new data set with just the 34 different species present.
A8SpecData <- A8Data %>%
  select(-c(Quadrate:Sillique))
row.names(A8SpecData) <- A8Data$Quadrate
```

Distance matrix

```
# Use package "vegan", "reshape2" and "ggplot2" here
# Bray-Curtis dissimilarity method
SpecDist <- vegdist(A8SpecData, method = "bray", binary = F)
SpecDistMat <- as.matrix(SpecDist)
SpecPDat <- melt(SpecDistMat)

# Visualize the distance matrix
ggplot(data = SpecPDat, aes(x = Var1, y = Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradientn(colours = c("white", "blue", "green", "red")) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))
```
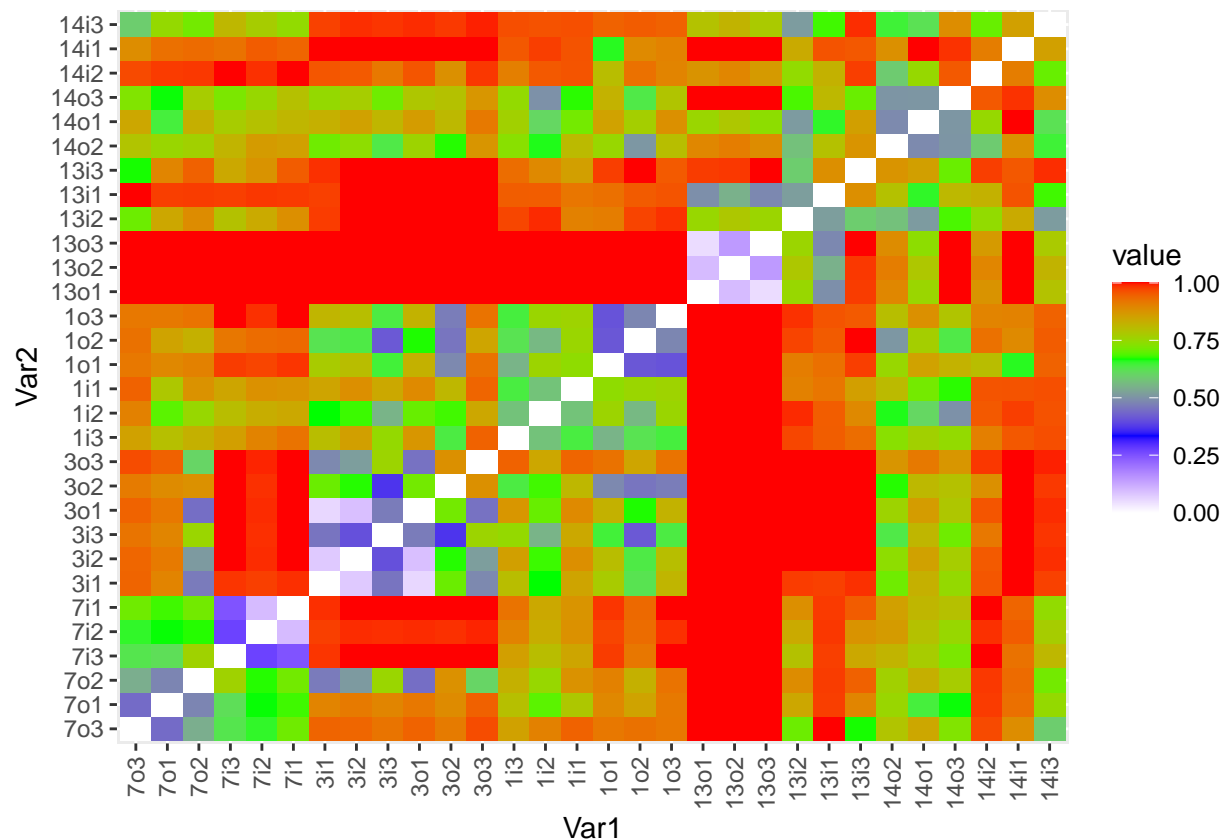

Fig.1 Distance map for each Quadrate used Bray-Curtis dissimilarity method.

From this figure we could found that there are higher distance gap between different populations and with different location (o and i) it also have different distance value in between. And

we could also easily found that there are three sites are far away from others (13o1, 13o2, 13o3). According to all of these, we could say that garlic mustard do have effect on the plant community.

Tree Building

```
# Use "ape" and "ggtree" package here
SpecTree <- nj(SpecDist)
ggtree(SpecTree, layout = "rectangular") %<+% A8Data +
  geom_tiplab(aes(colour = Population)) +
  geom_label(aes(x = branch, label = Location), fill = 'yellow') +
  theme(legend.position = "right")
```
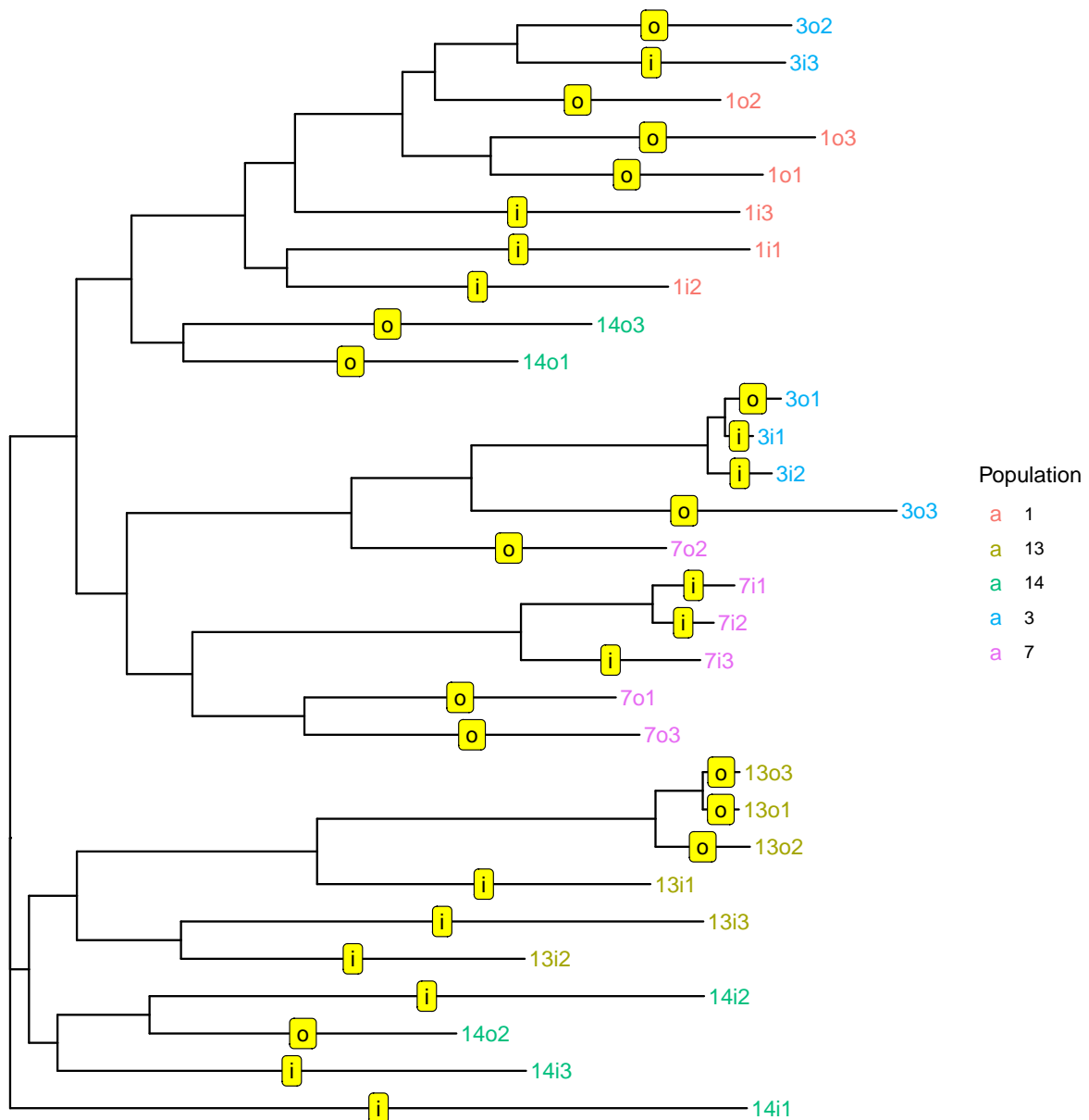
Fig.2 The tree of each Quadrates labeled with different locations and names indicated by different colors showing the populations. (o means there are no garlic mustard within the site, i means there are garlic mustard within the site.)

From this tree graph, we could see that all sites in same population group are close to the sites from same population, except population 3 and popultion 14. And all sites within different locations are partly convergent to each other. This shows that both population and location have effect on the plant community.

Non-Metric Multidimensional Scaling (NMDS)

```
set.seed(1117)
NMDSdat <- metaMDS(SpecDist, k = 2)
```

```
## Run 0 stress 0.1461135
## Run 1 stress 0.1534441
## Run 2 stress 0.1461135
## ... Procrustes: rmse 0.0001081138  max resid 0.0002747381
## ... Similar to previous best
## Run 3 stress 0.1461135
## ... New best solution
## ... Procrustes: rmse 3.63743e-05  max resid 0.0001179687
## ... Similar to previous best
## Run 4 stress 0.1461136
## ... Procrustes: rmse 0.0001411065  max resid 0.0004774209
## ... Similar to previous best
## Run 5 stress 0.1560521
## Run 6 stress 0.161178
## Run 7 stress 0.1461135
## ... Procrustes: rmse 7.547414e-05  max resid 0.000236258
## ... Similar to previous best
## Run 8 stress 0.141439
## ... New best solution
## ... Procrustes: rmse 0.06681613  max resid 0.3092865
## Run 9 stress 0.1461135
## Run 10 stress 0.1414392
## ... Procrustes: rmse 0.0006100653  max resid 0.002264194
## ... Similar to previous best
## Run 11 stress 0.1560521
## Run 12 stress 0.1676489
## Run 13 stress 0.1534441
## Run 14 stress 0.161178
## Run 15 stress 0.159307
## Run 16 stress 0.1461135
## Run 17 stress 0.1461135
## Run 18 stress 0.1560521
## Run 19 stress 0.1534441
## Run 20 stress 0.1666365
## *** Solution reached
```

```
A8PDat <- data.frame(NMDS1 = NMDSdat$points[,1],
                     NMDS2 = NMDSdat$points[,2],
                     Quadrate = row.names(A8SpecData))
```

```
A8PDat <- merge(A8PDat, A8Data, by = "Quadrate", all.x = T, all.y = F)

qplot(x = NMDS1, y = NMDS2, colour = Population, shape = Location, alpha = I(0.6), data = A8PDat) +
    theme_bw()
```
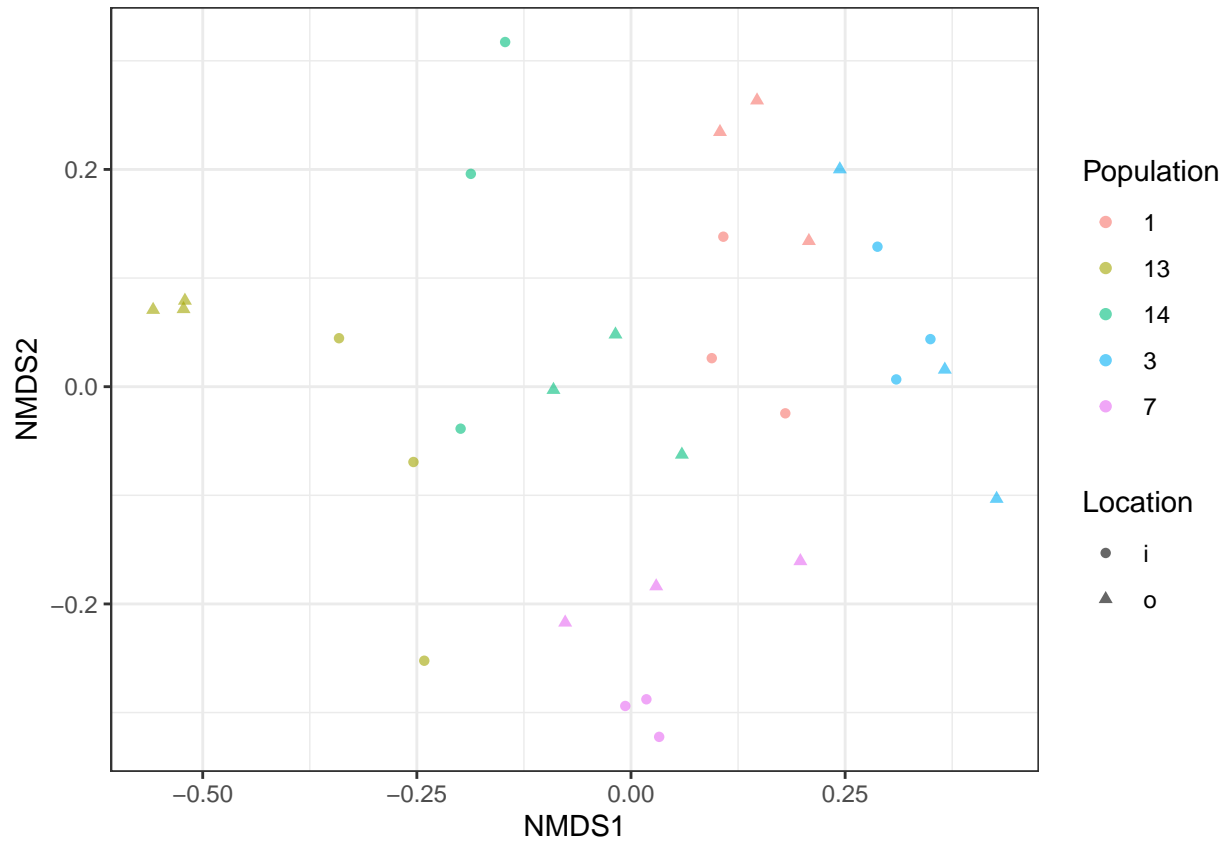

Fig.3 Two dementional NMDS plot with different colors showing the population groups and different shapes showing the location groups.

From this figure we can see that the points with different colors are clustered in different areas without much overlap. And the points with different shapes, on the other hand, do not show a clear group aggregation, they overlaped to each other. Thus, we could say that sampling population has a stronger effect on plant communities.

Reference

1. Cipollini, D., Gruner, B. Cyanide in the Chemical Arsenal of Garlic Mustard, Alliaria petiolata . J Chem Ecol 33, 85–94 (2007). https://doi.org/10.1007/s10886-006-9205-x