Group Project BIOL432

R&B(ioinformatics)

06/04/2021

1 Project Info

Date: April 6th, 2021

Group members:

Amanda Zacharias (20043354)

Arjun Augustine (20060723)

Alexander Row (20067626)

Yudong Liu (20020036)

Cameron MacPhail (20012451)

GitHub Repository Link (https://github.com/alexrow15/BIOL432.GroupProject)

1.1 Dataset Description

The dataset selected for this project was originally published by Bahlai et al. (2015) (https://link-springer-com.proxy.queensu.ca/article/10.1007/s10530-014-0772-4#Sec2) and used data from a multi-year study conducted on a ladybeetle (Coccinellidae) community in southwestern Michigan. The study aimed to elucidate the mechanism behind the decline of native Coccinellidae species and generate potential conservation strategies. Data were collected in southern Michigan on an experimental farm and forest area. Forest sampling began in 1993, so data before this time does not include the forest community classification. Adult coccinellids were captured using un-baited yellow sticky pads suspended in the air. As this sampling technique may be more efficient in some habitats than other habitats, we will focus our analysis within habitat types to minimize bias. The time period covered by this dataset includes several

important ecological events, including the introduction of non-native competitors (alien coccinellid species) and prey (soybean aphids). The first, second and third variables describe collection dates in the form of year; day, month and year (DD-Month-YYYY); and ordinal day of the year, respectively. The fourth variable describes how the plot was managed (e.g. conventional tilling, organic, etc.), if at all. The fifth variable records the habitat type (i.e. the plant community growing in the plot). This differs from the management treatment due to crop rotation. The sixth variable is the plant community classification (annual, perennial, or forest). The seventh variable is the plot replicate number. The eighth variable is the sample station number; there are five stations per plot. The ninth variable is the species ID; there are 13 species in total. Finally, the tenth variable describes the total number of adults captured in a subsample during a week. Bahlai et al. (2015) concluded that the decreasing abundance of native Coccinellidae species was largely restricted to only two native species and the remaining native species were found in stable abundances. The study found that competitive exploitation driven by dietary overlap with alien Coccinellidae species was the main reason for the decline of the two native species. Our project will examine whether seasonal niche partitioning is occurring among and between native and non-native coccinellids.

There are 13 species of Ladybeetle documented in this study. 9 species are native, while the remaining 4 are non-native invase species.

Dataset reference:

Bahlai, C. A., M. Colunga-Garcia, S. H. Gage, and D. A. Landis. 2015. The role of exotic ladybeetles in the decline of native ladybeetle populations: evidence from long-term monitoring. Biological Invasions 17:1005–1024.

1.2 Native species

1. Two Spot ladybird - Adalia bipunctata (https://en.wikipedia.org/wiki/Adalia_bipunctata)

2. Spotted Lady Beetle - Coleomegilla maculata

(https://biocontrol.entomology.cornell.edu/predators/Coleomegilla.php)

3. Twice stabbed ladybug - Chilocorus stigma (https://www.discoverlife.org/mp/20q?

search=Chilocorus+stigma)

4. Three-banded ladybeetle - Coccinella trifasciata (https://www.discoverlife.org/mp/20q?

search=Coccinella+trifasciata)

5. Polished/spotless ladybeetle - Cycloneda munda (https://bugguide.net/node/view/472564)

6. Convergent ladybeetle - Hippodamia convergens

(https://en.wikipedia.org/wiki/Hippodamia_convergens)

7. Glacial ladybeetle - Hippodamia glacialis (https://www.discoverlife.org/mp/20q?

search=Hippodamia+glacialis)

8. Parenthesis ladybeetle -Hippodamia parenthesis (https://www.discoverlife.org/mp/20q?

search=Hippodamia+parenthesis&guide=Ladybug&cl=llp)

9. Thirteen-spot ladybeetle - Hippodamia tredecimpunctata
(https://www.jungledragon.com/specie/13740/thirteen-spot_ladybird.html)

1.3 Non-native Invasive Species

1. Seven-spot ladybeetle - Coccinella septempunctata

(https://kids.kiddle.co/Coccinella_septempunctata)

2. Harlequin/Asian ladybeetle - Harmonia axyridis (https://davesgarden.com/guides/bf/go/4139/#b)
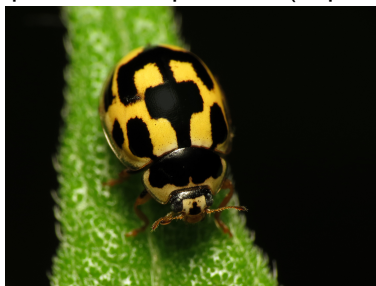
3. Variegated ladybeetle - Hippodamia variegata (https://bugguide.net/node/view/352609)

4. 14-spotted ladybeetle - Propylea quatuordecimpunctata (https://www.inaturalist.org/taxa/61532-

Propylea-quatuordecimpunctata)

We seek to address the following questions:

1. Within sampling years, does native and non-native species abundance change according to season?

2. Is there temporal niche partitioning between native and non-native species?

3. Is there a change in temporal partitioning before and after invasive species arrive?

Hypotheses for each question:

1. **H0**: Native and non-native species abundance does not change with time within sampling years.

H1: Native and non-native species abundance does change with time within sampling years. We suspect that native and non-native species abundance will change with time because abundance may be impacted by ecological conditions that vary across different stages of the growing season.

2. **H0**:There is no temporal niche partitioning between native and non-native species.

H1: There is temporal niche partitioning between native and non-native species. We suspect that there will be temporal niche partitioning because native and non-native species may compete with each other.

3. **H0**: There is no change in seasonal partitioning before and after invasive species arrive.

H1: There is a change in seasonal partitioning before and after invasive species arrive. We suspect that there will be changes in seasonal partitioning before and after the arrival of non-native ladybeetles and invasive soybean aphids because they represent major potential influences on lady beetle community dynamics. The arrival of soybean aphids may benefit certain species more than others as food source, and the arrival of non-native ladybeetles may impact different ladybeetle species to different degrees.

Predictions for each question:

1. If native and non-native species abundance changes with time within sampling years, then graphs should show seasonal changes in species abundance.

2. If there is temporal niche partitioning between native and non-native species, then graphs should show that when non-native species have high abundance, native species have low abundance and vice versa. Another way to see if there is temporal niche partitioning between native and non-native

species is by clustering the species by their temporal patterns in abundance within sampling years. If there is differences between native and non-native temporal patterns of abundance within sampling years, then they should cluster separately on a tree.

3. If there is a change in seasonal partitioning before and after non-native lady beetles and invasive soybean aphids arrive, then graphs will show changes in temporal niche partitioning. For example, non-native species may be at a high abundance during a particular season when they were previously low abundance, or more generally there might not be seasonal partitioning at all between native and non-native species in one time period but seasonal partitioning could be present in another time period. When clustering the species by their temporal patterns in abundance within sampling years, differences between time periods in the presence or absence of seasonal partitioning would be observed as the presence or absence of distinct native and non-native species clusters, respectively.

1.4 Load Required Packages

```
library(dplyr)
library(tidyr)
library(ggplot2)
library(vegan)
library(reshape2)
library(cowplot)
library(ape)
library(ggtree)
```

2 Preliminary Data Processing

2.1 Load Dataset

```
dat <- read.csv("BIOL432.Data.GroupProject.csv")
```

2.2 Visualise Raw Data

```
head(dat)
```

```
##    Year        Date DOY TREAT_DESC HABITAT        TYPE REPLICATE STATION SpeciesID
## 1 1989 24-May-89 144     Alfalfa alfalfa perennial         1       1     ABIPN
## 2 1989 24-May-89 144     Alfalfa alfalfa perennial         1       1     CSEPT
## 3 1989 24-May-89 144     Alfalfa alfalfa perennial         1       1      CMAC
## 4 1989 24-May-89 144     Alfalfa alfalfa perennial         1       1     CSTIG
## 5 1989 24-May-89 144     Alfalfa alfalfa perennial         1       1     CTRIF
## 6 1989 24-May-89 144     Alfalfa alfalfa perennial         1       1     CMUND
##    Adults
## 1      0
## 2      0
## 3      0
## 4      1
## 5      0
## 6      0
```

```
tail(dat)
```

```
##         Year        Date DOY  TREAT_DESC HABITAT    TYPE REPLICATE STATION
## 654785 2012 18-Sep-12 262 Succesional      SF forest         3       5
## 654786 2012 18-Sep-12 262 Succesional      SF forest         3       5
## 654787 2012 18-Sep-12 262 Succesional      SF forest         3       5
## 654788 2012 18-Sep-12 262 Succesional      SF forest         3       5
## 654789 2012 18-Sep-12 262 Succesional      SF forest         3       5
## 654790 2012 18-Sep-12 262 Succesional      SF forest         3       5
##         SpeciesID Adults
## 654785     CSTIG      0
## 654786     CMUND      0
## 654787      HAXY      0
## 654788     HPARN      0
## 654789      HVAR      0
## 654790      PQUA      0
```

```
summary(dat) #view summary for each variable in dataset
```

```
##       Year          Date              DOY          TREAT_DESC
##   Min.   :1989   Length:654790    Min.   :114.0   Length:654790
##   1st Qu.:1994   Class :character 1st Qu.:169.0   Class :character
##   Median :2001   Mode  :character Median :191.0   Mode  :character
##   Mean   :2000                    Mean   :192.3
##   3rd Qu.:2006                    3rd Qu.:215.0
##   Max.   :2012                    Max.   :289.0
##
##     HABITAT            TYPE           REPLICATE        STATION
##   Length:654790    Length:654790   Min.   :1.000   Min.   :1.000
##   Class :character Class :character 1st Qu.:2.000   1st Qu.:2.000
##   Mode  :character Mode  :character Median :3.000   Median :3.000
##                                    Mean   :3.268   Mean   :2.998
##                                    3rd Qu.:5.000   3rd Qu.:4.000
##                                    Max.   :6.000   Max.   :5.000
##
##    SpeciesID            Adults
##   Length:654790    Min.   : 0.00000
##   Class :character 1st Qu.: 0.00000
##   Mode  :character Median : 0.00000
##                    Mean   : 0.08829
##                    3rd Qu.: 0.00000
##                    Max.   :32.00000
##                    NA's   :11
```

2.3 Using Regex in Unix to change SpeciesID into modified scientific name

The code chunk below creates a new csv file with the updated scientific names converted using regex expressions

```bash
#!/bin/bash

#Date:04/01/2021
#Author: Yudong Liu
#Purpose: This bash script is written to test if we can substitute species ID with their n
ames

awk '{gsub(/ABIPN/, "A. bipunctata");print}' BashTest4.csv > New.csv &&
mv New.csv BashTest4.csv

awk '{gsub(/CMAC/, "C. maculata");print}' BashTest4.csv > New.csv &&
mv New.csv BashTest4.csv

awk '{gsub(/CSTIG/, "C. stigma");print}' BashTest4.csv > New.csv &&
mv New.csv BashTest4.csv

awk '{gsub(/CTRIF/, "C. trifasciata");print}' BashTest4.csv > New.csv &&
mv New.csv BashTest4.csv

awk '{gsub(/CMUND/, "C. munda");print}' BashTest4.csv > New.csv &&
mv New.csv BashTest4.csv

awk '{gsub(/HCONV/, "H. convergens");print}' BashTest4.csv > New.csv &&
mv New.csv BashTest4.csv

awk '{gsub(/HGLAC/, "H. glacialis");print}' BashTest4.csv > New.csv &&
mv New.csv BashTest4.csv

awk '{gsub(/HTRE/, "H. tredecimpunctata");print}' BashTest4.csv > New.csv &&
mv New.csv BashTest4.csv

awk '{gsub(/CSEPT/, "C. septempunctata");print}' BashTest4.csv > New.csv &&
mv New.csv BashTest4.csv

awk '{gsub(/HAXY/, "H. axyridis");print}' BashTest4.csv > New.csv &&
mv New.csv BashTest4.csv

awk '{gsub(/HVAR/, "H. variegata");print}' BashTest4.csv > New.csv &&
mv New.csv BashTest4.csv

awk '{gsub(/HPARN/, "H. parenthesis");print}' BashTest4.csv > New.csv &&
mv New.csv BashTest4.csv

awk '{gsub(/PQUA/, "P. quatuordecimpunctata");print}' BashTest4.csv > New.csv &&
mv New.csv BashTest4.csv


#End
```

2.4 Upload updated file

Using the bash script above, a new csv file was created with the changed species names. In order to decrease computational time and difficulty, the resultant file is uploaded in the code chunk below.

```
LadyBugData <- read.csv("BashTest4.csv")
```

Inspect updated file

```
head(LadyBugData)
```

```
##   Year      Date DOY TREAT_DESC HABITAT      TYPE REPLICATE STATION
## 1 1989 24-May-89 144    Alfalfa alfalfa perennial         1       1
## 2 1989 24-May-89 144    Alfalfa alfalfa perennial         1       1
## 3 1989 24-May-89 144    Alfalfa alfalfa perennial         1       1
## 4 1989 24-May-89 144    Alfalfa alfalfa perennial         1       1
## 5 1989 24-May-89 144    Alfalfa alfalfa perennial         1       1
## 6 1989 24-May-89 144    Alfalfa alfalfa perennial         1       1
##           SpeciesID Adults
## 1     A. bipunctata      0
## 2 C. septempunctata      0
## 3      C. maculata      0
## 4        C. stigma      1
## 5    C. trifasciata      0
## 6         C. munda      0
```

```
tail(LadyBugData)
```

```
##          Year      Date DOY  TREAT_DESC HABITAT   TYPE REPLICATE STATION
## 654785 2012 18-Sep-12 262 Succesional      SF forest         3       5
## 654786 2012 18-Sep-12 262 Succesional      SF forest         3       5
## 654787 2012 18-Sep-12 262 Succesional      SF forest         3       5
## 654788 2012 18-Sep-12 262 Succesional      SF forest         3       5
## 654789 2012 18-Sep-12 262 Succesional      SF forest         3       5
## 654790 2012 18-Sep-12 262 Succesional      SF forest         3       5
##                      SpeciesID Adults
## 654785              C. stigma      0
## 654786               C. munda      0
## 654787            H. axyridis      0
## 654788         H. parenthesis      0
## 654789            H. variegata      0
## 654790 P. quatuordecimpunctata      0
```

```
summary(LadyBugData)
```

```
##       Year           Date                DOY          TREAT_DESC
##   Min.   :1989    Length:654790    Min.   :114.0    Length:654790
##   1st Qu.:1994    Class :character 1st Qu.:169.0    Class :character
##   Median :2001    Mode  :character Median :191.0    Mode  :character
##   Mean   :2000                     Mean   :192.3
##   3rd Qu.:2006                     3rd Qu.:215.0
##   Max.   :2012                     Max.   :289.0
##
##     HABITAT              TYPE              REPLICATE         STATION
##   Length:654790    Length:654790    Min.   :1.000    Min.   :1.000
##   Class :character Class :character 1st Qu.:2.000    1st Qu.:2.000
##   Mode  :character Mode  :character Median :3.000    Median :3.000
##                                     Mean   :3.268    Mean   :2.998
##                                     3rd Qu.:5.000    3rd Qu.:4.000
##                                     Max.   :6.000    Max.   :5.000
##
##     SpeciesID            Adults
##   Length:654790    Min.   : 0.00000
##   Class :character 1st Qu.: 0.00000
##   Mode  :character Median : 0.00000
##                    Mean   : 0.08829
##                    3rd Qu.: 0.00000
##                    Max.   :32.00000
##                    NA's   :11
```

2.5 Create native and non-native groups in the dataset

```
# Create another column in the dataset that assigns non-native species to the 'Non-native'
group and assign all other species to the 'Native' group
LadyBugData <- LadyBugData %>%
  mutate(Species_type = ifelse(SpeciesID == "C. septempunctata" |
                               SpeciesID == "H. axyridis" |
                               SpeciesID == "H. variegata" |
                               SpeciesID == "P. quatuordecimpunctata",
                         "Non-native", "Native"))
```

3 Use ggplot to Graph Abundance Data

We used the generalized additive model (GAM) method in order to discern general patterns in our data. While using GAM resulted in us being unable to plot certain habitats for certain periods of time due to limited data, other methods that would have allowed us to plot with limited data (e.g. LOESS) take large amounts of time and computer power.

3.1 Custom Function for ggplot Theme

This function will create a custom theme for our ggplots.

```
theme_custom <- function () {
  # Using theme_minimal as a base. Will replace elements that we want to change.
  theme_minimal() %+replace%
  theme(
    # Changing grid elements
    axis.ticks = element_line(colour = "black"),
    # Changing axis text
    axis.text = element_text(size = rel(0.8), colour="black"),
    # Changing facet title text
    strip.text = element_text(size = rel(1), colour = "black"),
    #add an axis line
    axis.line = element_line(colour = 'black', size = 1)
  )
}
```

3.2 General temporal abundance patterns

We plotted ladybeetle abundance against the day of the year to determine if abundance varies at all over time. We are displaying this data separately for each individual habitat type (annual, forest, perennial) because the sampling method used to capture the ladybeetles may have been more efficient in some habitats than other habitats.
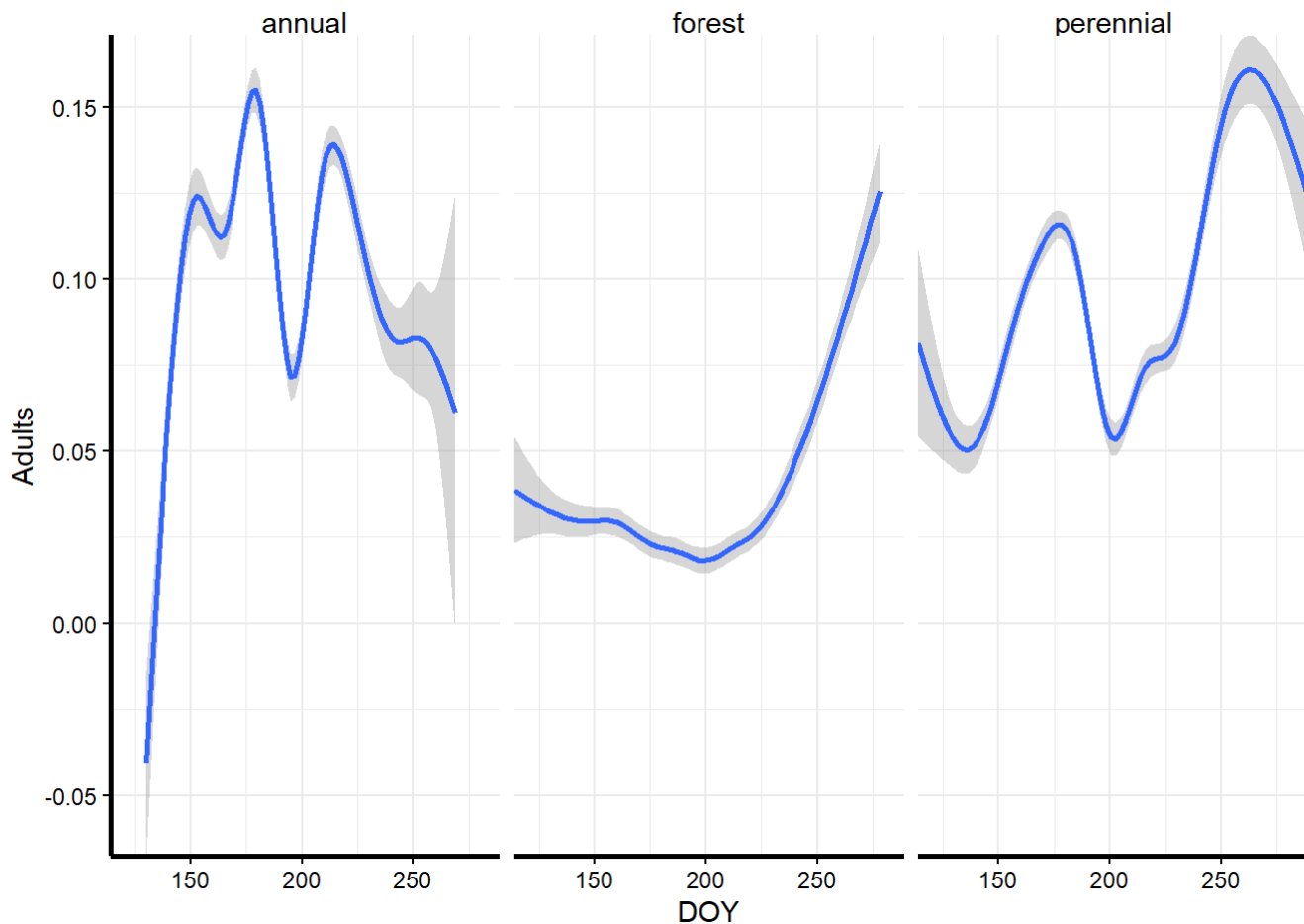
```
ggplot(LadyBugData, aes(x = DOY, y = Adults)) +
  scale_x_continuous(expand = c(0,0)) +
  scale_y_continuous(expand = c(0,0)) +
  geom_smooth(method = "gam") +
  facet_wrap(~ TYPE) +
  theme_custom()
```

```
## `geom_smooth()` using formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 11 rows containing non-finite values (stat_smooth).
```

Clearly abundance does vary over time. Let's examine how well this pattern holds for native vs non-native assemblages.

3.3 Temporal abundance of native vs non-native ladybeetles

Plotting abundance with native and non-native groupings

```
ggplot(LadyBugData, aes(x = DOY, y = Adults, colour = Species_type)) +
  scale_x_continuous(expand = c(0,0)) +
  scale_y_continuous(expand = c(0,0)) +
  geom_smooth(method = "gam") +
  facet_wrap(~ TYPE) +
  theme_bw()
```

```
## `geom_smooth()` using formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 11 rows containing non-finite values (stat_smooth).
```
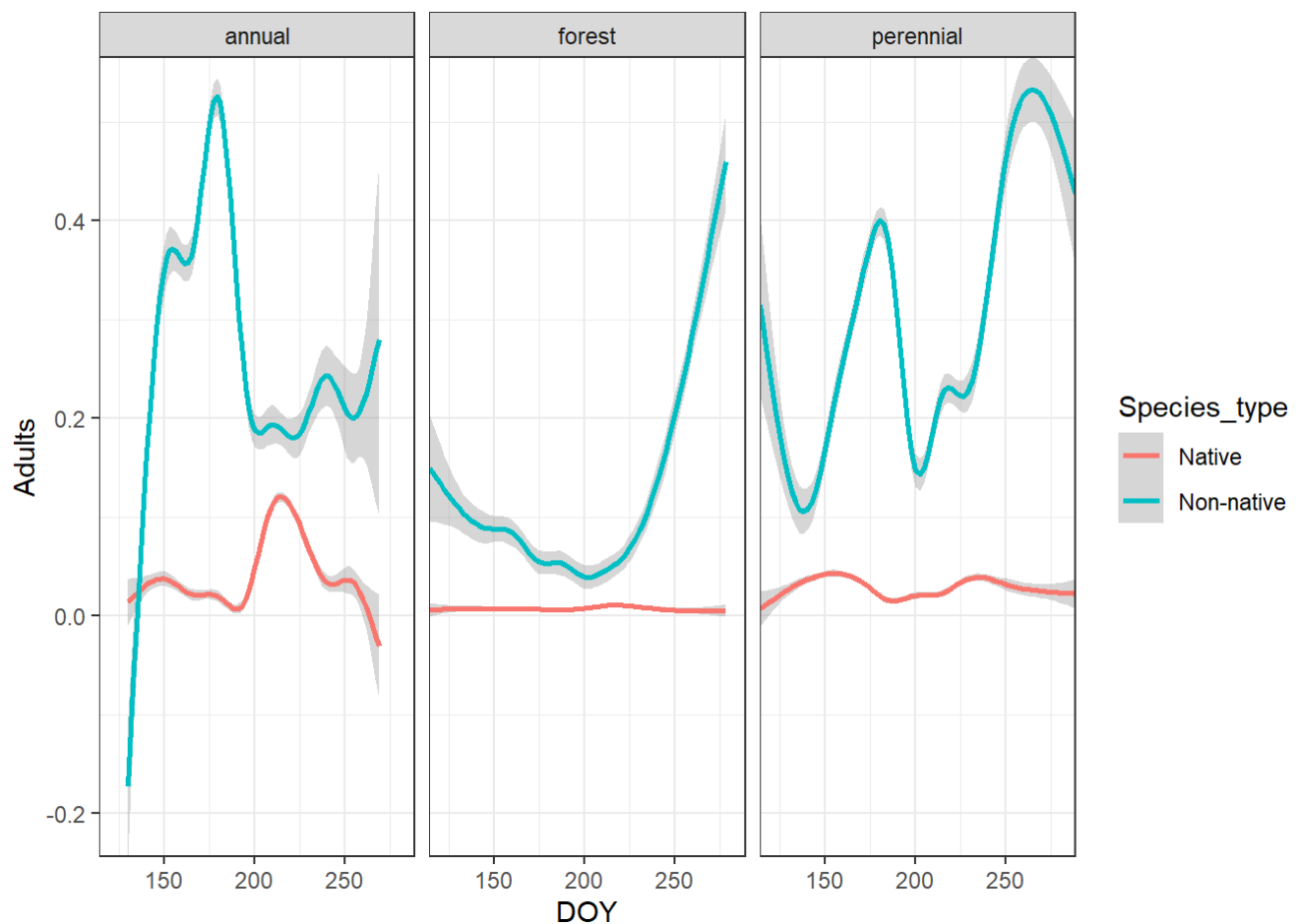
Non-native ladybeetles clearly have higher abundance throughout the entire year in most habitats. There doesn't seem to be strong patterns in forest habitat for native ladybeetles, but this might just be because they have such low abundance. Taking a look at abundance pre-invasion of the various exotic species should help show what's going on.

Also, in annual crop habitats there seems to be a decline in non-native ladybeetle abundance and a concurrent increase in native ladybeetle abundance around 2/3 of the way through the year, lending support to our hypothesis of temporal niche partitioning.

3.4 Invasion time series temporal abundance patterns

We are summarizing our findings here in a time series graph created using the cowplot package for concision. Code for the individual graphs and joining the graphs can be found in the "time_series_graphing.Rmd" file in the project folder.
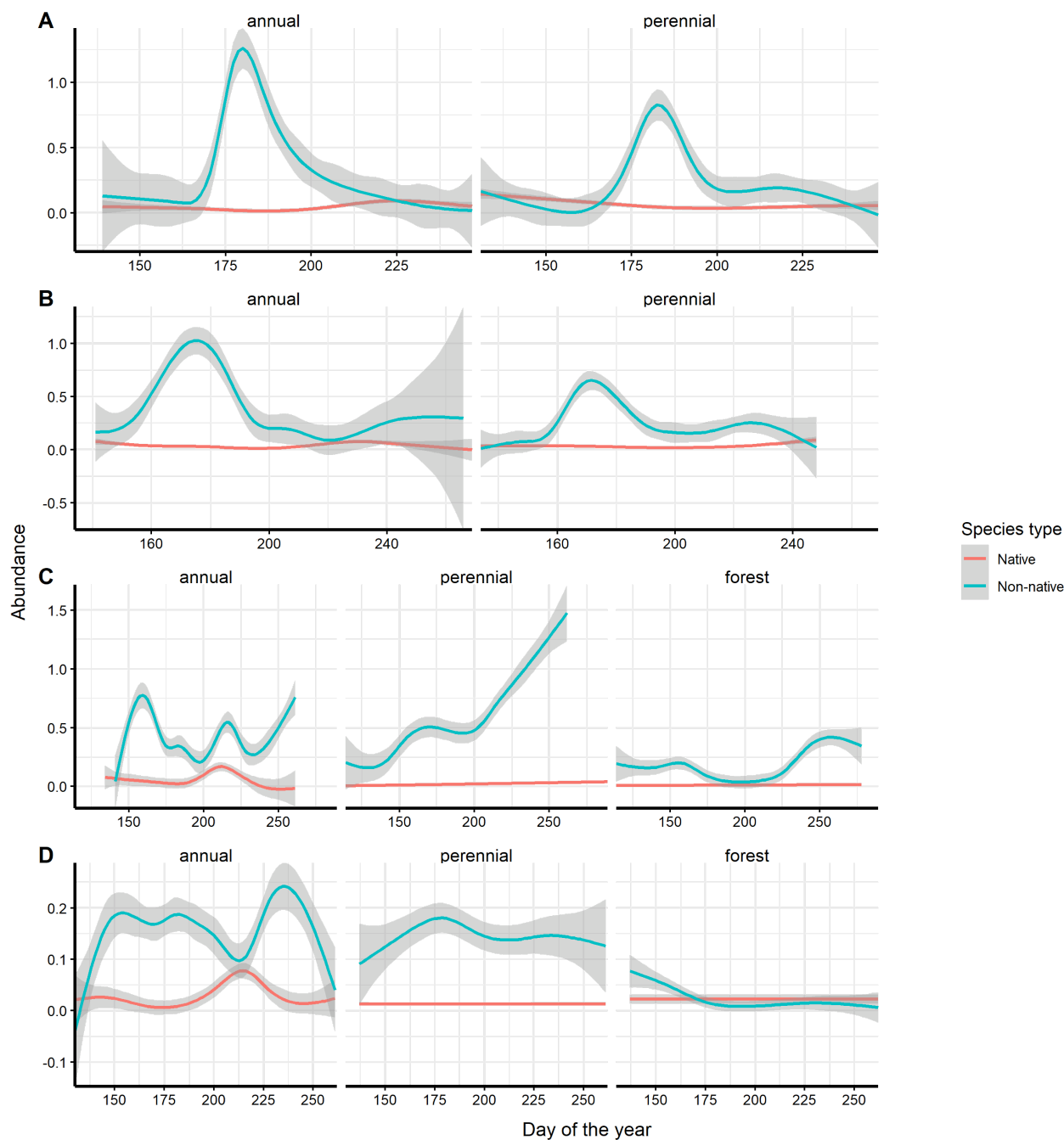
A = 1989 - 1993: one non-native ladybeetle, 1994 - 1998: two non-native ladybeetles, C = 1999 - 2006: three non-native ladybeetles + soybean aphid outbreaks, D = 2007 - 2012: four non-native ladybeetles + soybean aphids present

Overall, native ladybeetle abundance is low compared to non-native species. In the early stages of invasion especially, the shoulder seasons seem to benefit native species. In addition, native species often reach peak abundance in annual habitat ~215 - 225 days into the year. In many cases, this increase in native species abundance corresponds with a decrease in non-native species abundance. This does not occur during the 1999-2006 period, however, possibly due to soybean aphid outbreaks during that time. For the most part in most habitats, however, native ladybeetle abundance do not seem to respond strongly to changes in non-native ladybeetle abundance.

4 Cluster the species by DOY abundances

As an alternative way to test hypotheses 2 and 3 we sought to cluster the species based off patterns in their abundances over days of the year.

To determine if there is temporal niche partitioning between species, we clustered the species by differences in abundance for each day of the year (DOY) (pooled across all years within a given time frame). The clustering method we chose was the neighbour-joining (NJ) tree method. We predicted that this would reveal any differences in the temporal patterns of abundance between native and non-native species. This does not tell us if species partition resources 'seasonally', i.e. the clusters will not reveal specific time periods when certain species become more abundant and other species become less abundant. However, the trees should indicate which species are similar to each other and which are dissimilar in terms of their temporal patterns of abundance.

4.1 Plot histograms of the abundance data

```
# Plot counts for the raw abundance data (for each observation)
ggplot(LadyBugData, aes(x=(Adults), y=..count..))+
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 11 rows containing non-finite values (stat_bin).
```
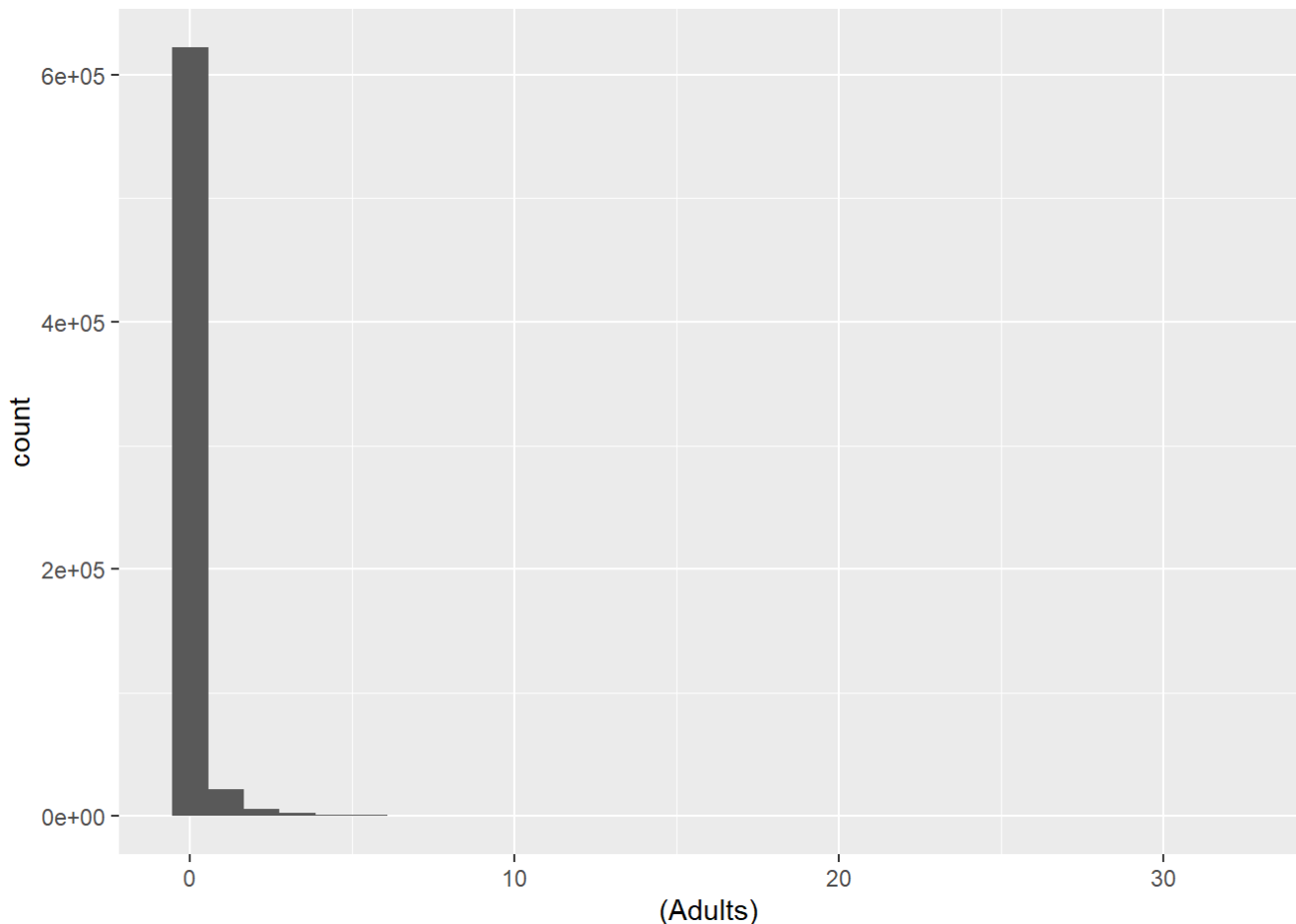
There is a wide range in abundances. It is possible that certain species might be less abundant in the dataset compared to others. To get an idea of how often each species is represented in the dataset:

```
# Sum up the abundances for each species in the entire dataset
LadyBugData %>% group_by(SpeciesID) %>% summarise(totalAbun = sum(Adults, na.rm = TRUE))
```

```
## # A tibble: 13 x 2
##    SpeciesID               totalAbun
##  * <chr>                       <int>
##  1 A. bipunctata                 486
##  2 C. maculata                 10764
##  3 C. munda                     2457
##  4 C. septempunctata           23816
##  5 C. stigma                     485
##  6 C. trifasciata                 55
##  7 H. axyridis                 16124
##  8 H. convergens                 238
##  9 H. glacialis                  400
## 10 H. parenthesis               1412
## 11 H. tredecimpunctata            21
## 12 H. variegata                  814
## 13 P. quatuordecimpunctata       741
```

Certain species are indeed far more common in the dataset than others. Log-transforming the dataset might allow us to de-emphasize the importance of highly abundant species so that rare species can have more influence in the analysis. This is especially important for comparing native and non-native species because two of the most abundant species in the dataset are non-native (*C. septempuncta* and *H. axyridis*). If raw abundance data is analyzed, the clusters in the trees might just reflect differences in abundance rather than differences in temporal patterns of abundance between species.

See how the distribution of abundances looks with different log transformations:

```
# Add 1 to each abundance value prior to log transforming to prevent 0 values from becomin
g undefined. Since logx(1), where x is any base, equals 0, the 0 abundance values will sim
ply return to being 0.
# Higher log bases give less weight to more abundant species and let rare species have mor
e influence.

# Try log base 2
ggplot(LadyBugData, aes(x=log((Adults + 1), base = 2), y=..count..)) +
  geom_histogram() +
  scale_x_continuous(name = "log2(Adults)")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 11 rows containing non-finite values (stat_bin).
```
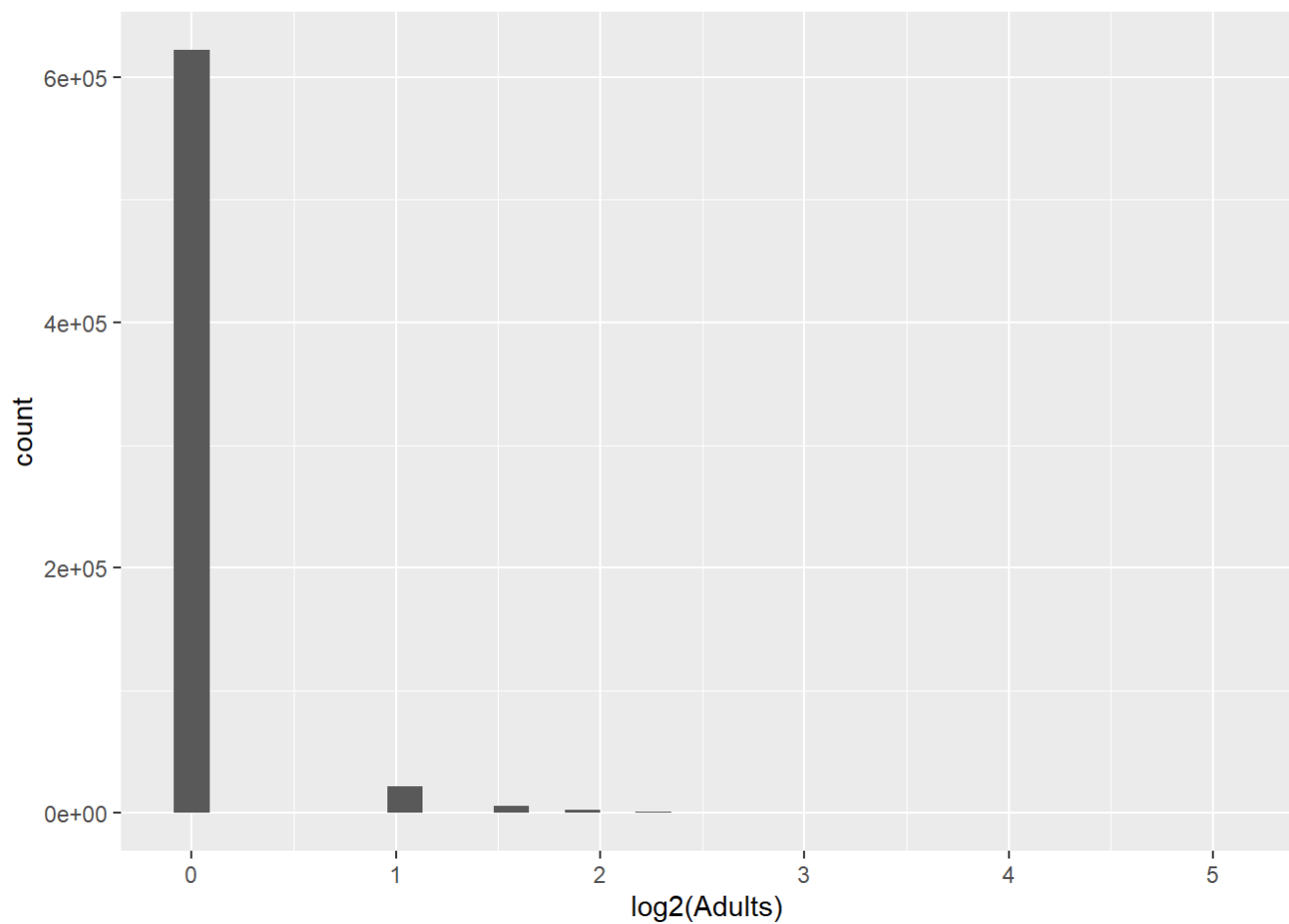
```
# Try log base 5
ggplot(LadyBugData, aes(x=log((Adults + 1), base = 5), y=..count..)) +
  geom_histogram() +
  scale_x_continuous(name = "log5(Adults)")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
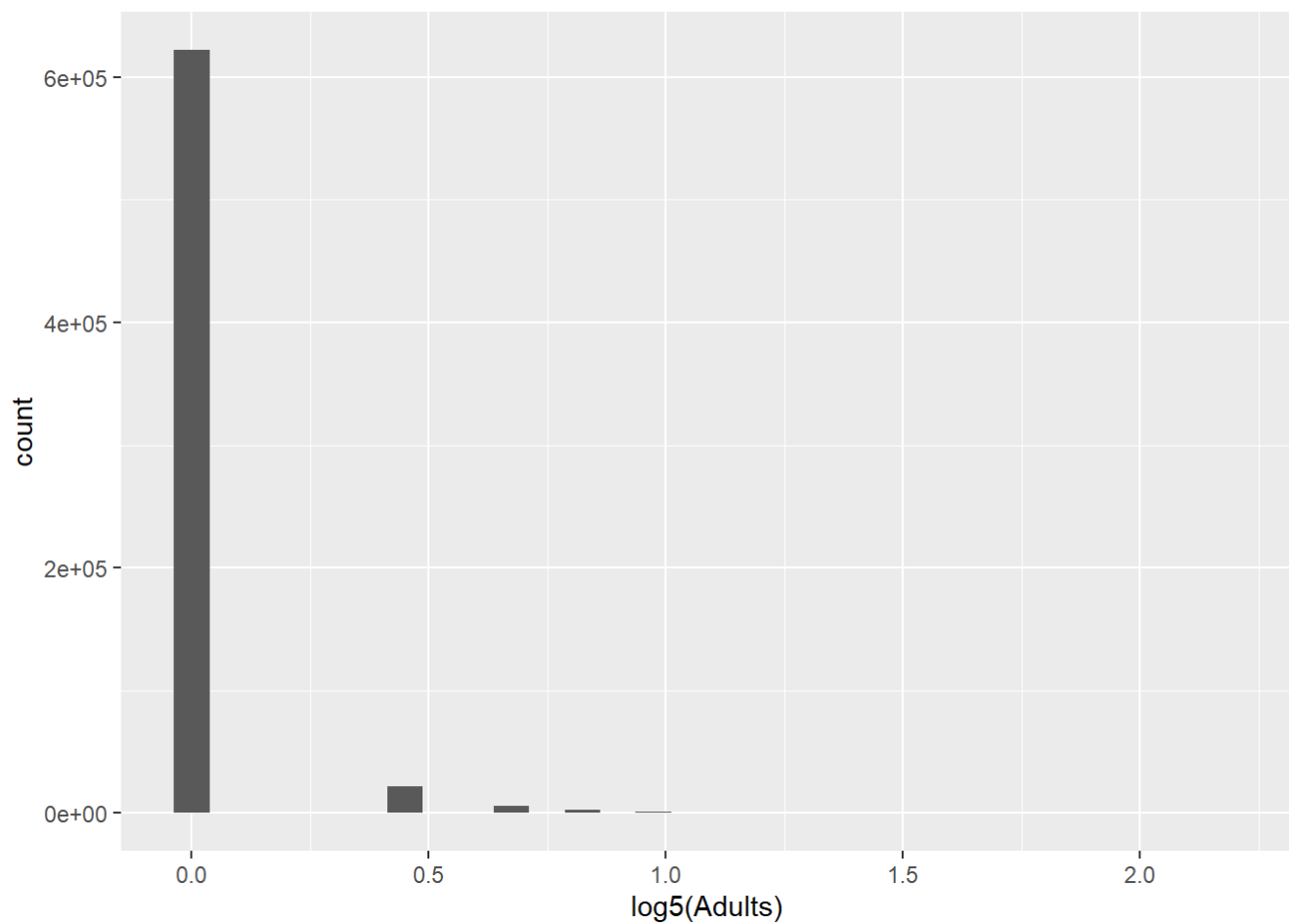
```
## Warning: Removed 11 rows containing non-finite values (stat_bin).
```
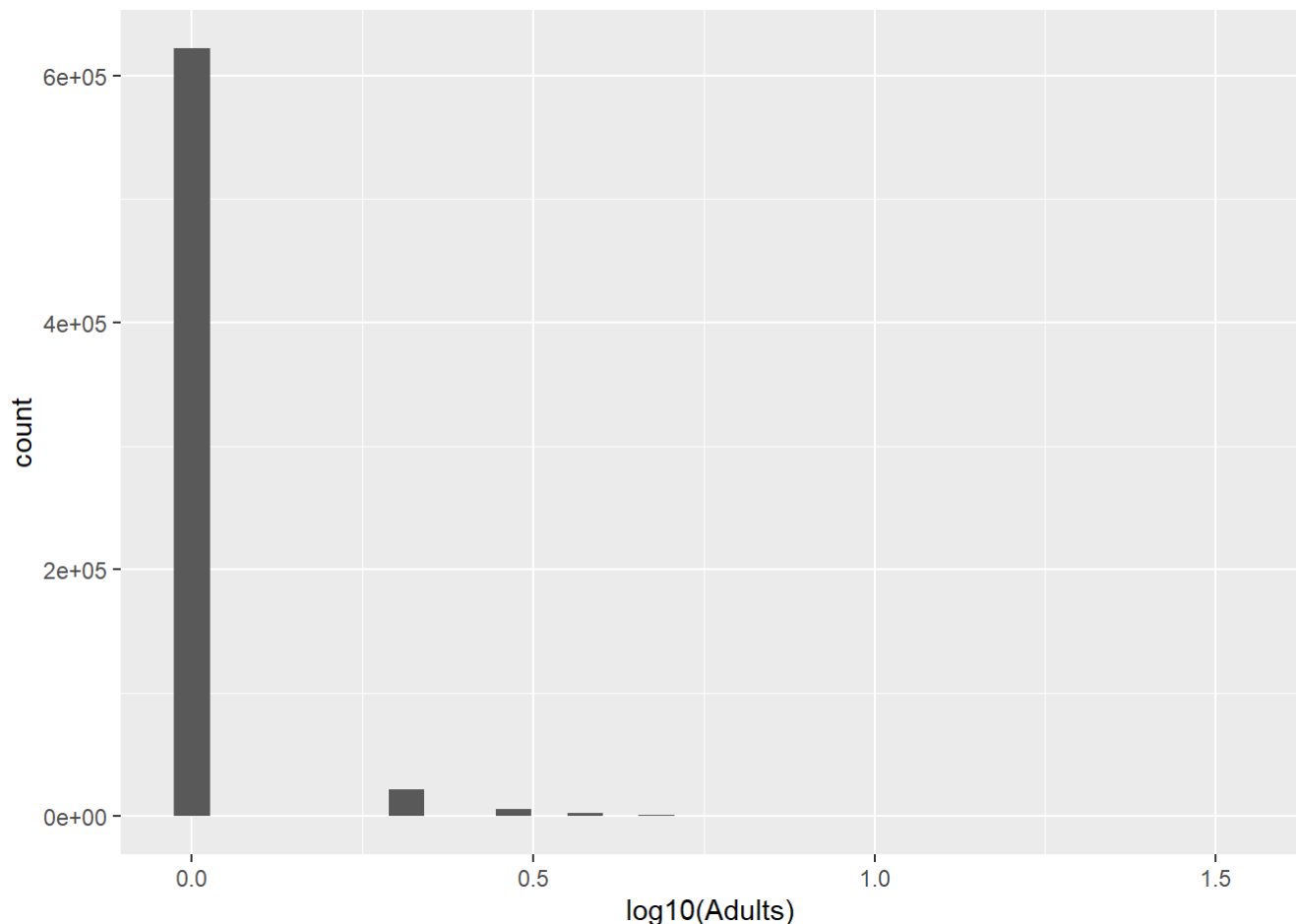
```r
# Try log base 10
ggplot(LadyBugData, aes(x=log((Adults + 1), base = 10), y=..count..)) +
  geom_histogram() +
  scale_x_continuous(name = "log10(Adults)")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 11 rows containing non-finite values (stat_bin).
```

For most trees, the only effect of switching to the raw abundance data is that 2 non-native species (*C. septempunctata* and *H. axyridis*) became closer to each other, probably reflecting their high abundance relative to most of the other species (data not shown).For at least one tree (forest data, including all years), switching from log transformed data to the raw abundance data had little effect on the topology.

Below, we present the analyses using the log5(x+1) transformation of the abundance data, as it appears to have limited the impact of the two most abundant species and had little other impact.

```
# Log transform the abundance data with the transformation log5(x+1)
logTLadyBug <- LadyBugData %>%
  mutate(logTAdults = log((Adults + 1), base = 5)) %>%  # transform with log5
  select(-Adults) # Replace Adults with logTAdults
```

4.2 Make trees to cluster the species by differences in abundance by DOY

4.2.1 Make a data.frame indicating the native/non-native status of each species

Will need this later to colour-code the species trees by native and non-native species to see if there are any clusters in abundance by DOY

```
# Select the species names and their corresponding status as native or non-native from the
original dataset, and remove the duplicate rows using the distinct() function from dplyr
SpeciesOrigin <- LadyBugData %>% select(SpeciesID, Species_type) %>% distinct()
```

4.2.2 Custom Functions for Creating the Trees

This function creates distance matrices for pairwise comparisons in abundance by DOY (day of year) between species:

```
# This function is applied after the abundance data is log transformed, pooled across the
 years within groups of DOY + SpeciesID + TYPE, and subsetted by habitat type and/or time
 period

mk_distmat <- function(sum_abun_dat){

  # Make a matrix of species abundances where the rows are the individual species and the
 columns are days of the year (DOY) for each habitat type:

  sum_abun_dat <- sum_abun_dat %>% pivot_wider(names_from = DOY, values_from = SumAbun) #
 Rearrange the dataset so abundance data is arranged as separate DOY columns
  sum_abun_dat <- data.frame(sum_abun_dat) # Convert from grouped data to a regular data.f
rame object
  row.names(sum_abun_dat) <- sum_abun_dat$SpeciesID # Set the SpeciesIDs as the row names
 in the data.frame
  sum_abun_dat <- sum_abun_dat %>% select(-SpeciesID, -TYPE) # Remove speciesID and habita
t type from dataset so vegdist() can be run on it

  # Make the distance matrix:

  distmat <- vegdist(sum_abun_dat, method = "bray", na.rm = TRUE) # Use Bray-Curtis dissim
ilarity to calculate distance matrix. Na values were removed to allow vegdist() to work

  return(distmat)
}
```

This function plots the distance matrices as heat maps:

```
# This function makes heat maps that display the distance matrices with white representing
no difference in temporal abundance between species and red representing the highest pairw
ise distance between species

p_dist <- function(distmat){
  Pdat <- melt(as.matrix(distmat)) # Linearize the distance matrix so that it lists the pa
irwise combinations of species with their distances
  hMap <- ggplot(data = Pdat, aes(x=Var1, y=Var2, fill = value)) +
    geom_tile() + # Makes heat map
    scale_fill_gradientn(colours=c("white","blue","green","red")) + # Assigns colours to d
istance values
    theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5)) # Makes axis lab
els readable
  return(hMap)
}
```

This function creates neighbour-joining (NJ) trees from a distance matrix:

```
# This function makes the NJ trees and plots them as well

mk_tree <- function(distmat, dat = SpeciesOrigin){
  tree <- njs(distmat) # njs allows for missing values in the distance matrix
  pTree <- ggtree(tree, layout = "rectangular") %<+% dat + # Use SpeciesOrigin data.frame
 to colour-code the species by Species_type (Native or Non-native)
    geom_tiplab(aes(colour = Species_type)) +
    scale_colour_discrete(direction = -1) + # Make non-natives red and natives blue, which
 is the opposite of the default colour scheme
    xlim_tree(1) # Set x-axis limits for the tree so that the tip-labels don't get cut off
  return(pTree)
}
```

4.2.3 Subset by habitat type

The sampling technique used to capture adult coccinellids may be more efficient in some habitats than other habitats (see methods of Bahlai et al. 2015), so we chose to focus our analysis within habitat types.

There are 3 habitat types:

- forest

- annual cropland

- perennial cropland

```
# Pool the abundance data across the years within groups of DOY + SpeciesID + TYPE
datAbunSum <- logTLadyBug %>% group_by(DOY, SpeciesID, TYPE) %>% summarize(SumAbun = sum(l
ogTAdults, na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'DOY', 'SpeciesID'. You can override using the `.gr
oups` argument.
```

```
# Subset the data by habitat type
forestAbunSum <- datAbunSum %>% filter(TYPE == "forest")
annualAbunSum <- datAbunSum %>% filter(TYPE == "annual")
perennialAbunSum <- datAbunSum %>% filter(TYPE == "perennial")
```

Create distance matrices and neighbor-joining trees:

```
# Make distance matrices using the custom mk_distmat() function
forest.dist <- mk_distmat(forestAbunSum)
annual.dist <- mk_distmat(annualAbunSum)
perennial.dist <- mk_distmat(perennialAbunSum)

# Make heat maps that display the distance matrices with white representing no difference
 in temporal abundance between species and red representing the highest pairwise distance
 between species
# Use the custom p_dist() function to do this
forest.hMap <- p_dist(forest.dist)
annual.hMap <- p_dist(annual.dist)
perennial.hMap <- p_dist(perennial.dist)

# Plot the heat maps for each habitat type subset together as a single figure. A = forest;
B = annual; C = perennial.
plot_grid(forest.hMap, annual.hMap, perennial.hMap, labels = c("A", "B", "C"))
```
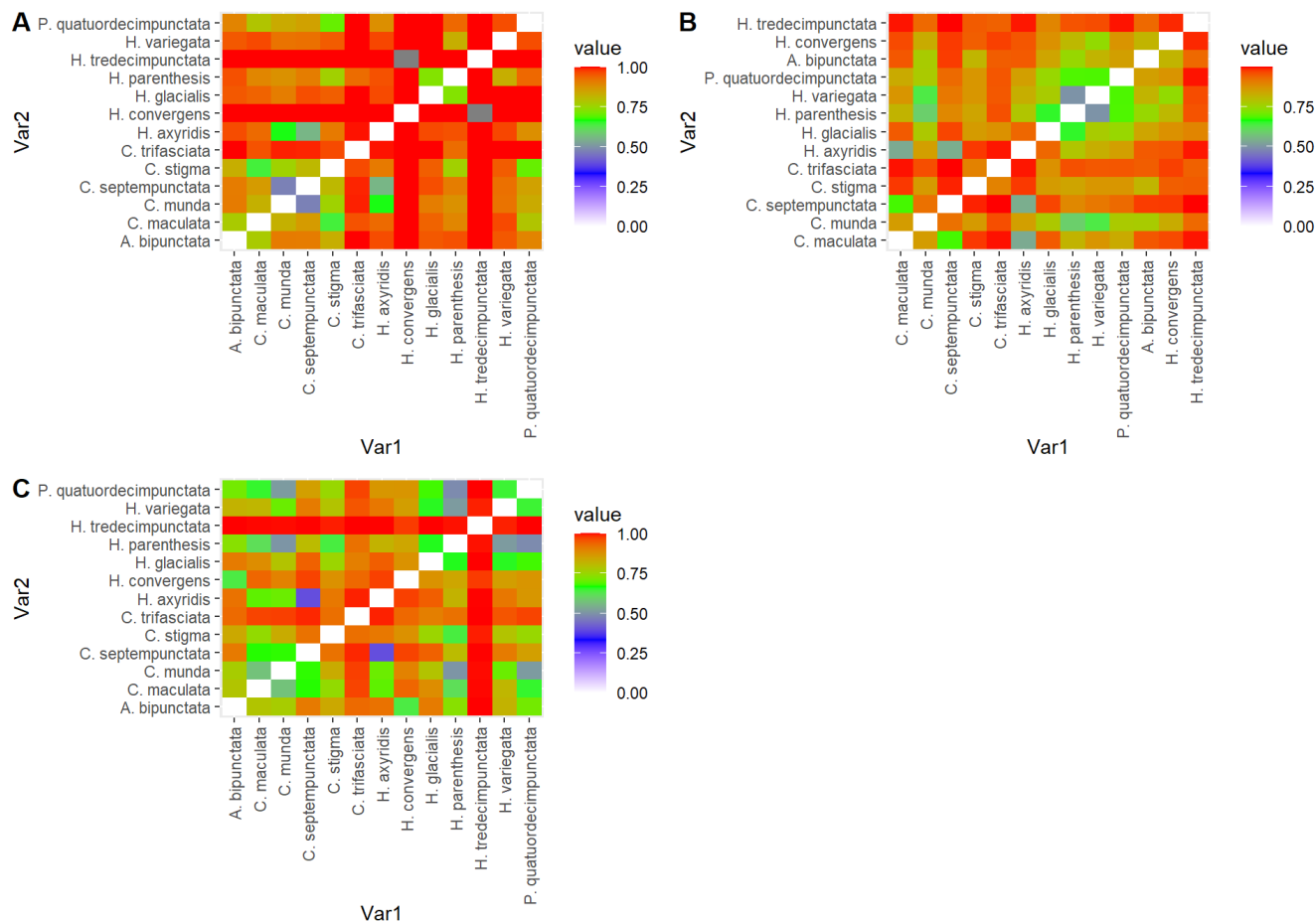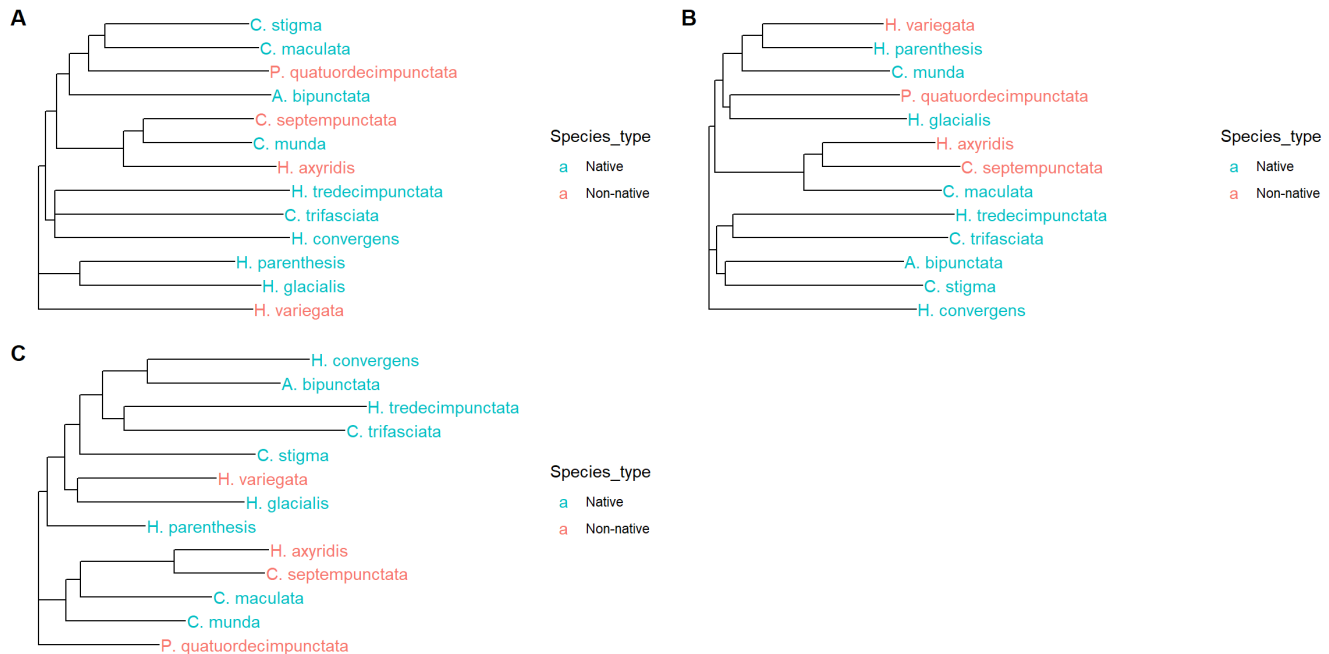
```
# Make the trees
forest.pT <- mk_tree(forest.dist)
annual.pT <- mk_tree(annual.dist)
perennial.pT <- mk_tree(perennial.dist)

# Plot the trees for each habitat type subset together as a single figure. A = forest; B =
annual; C = perennial.
plot_grid(forest.pT, annual.pT, perennial.pT, labels = c("A", "B", "C"))
```

There appears to be negligible differentiation between native and non-native species in forest habitats (A) and perennial habitats (C) in terms of their DOY abundance patterns.

Annual habitats (B) show some association of non-native species with each other in their DOY abundance patterns, however there are also native species within the same cluster. Also, the distinction between the cluster that holds the non-native species and the cluster that contains most of other native species appears to be poorly resolved, as indicated by the short branch lengths near the root of the tree.

Overall there seems to be little to no temporal partitioning between native and non-native ladybeetle species in any habitat, except for perhaps some weak temporal partitioning in annual habitat.

4.2.4 Subset the habitat types by time period

We chose to subset the data for each habitat type by different time periods to see if there was any ephemeral temporal partitioning between native and non-native species that may not have been captured by the analysis above, which looked at the entire time-span of the dataset. We predicted that there may be differences in temporal partitioning during different stages of the non-native ladybeetle invasion, which were defined by the number of non-native ladybeetles present in the study area. We also predicted that temporal partitioning might vary before and after the outbreak of invasive soybean aphids, as this could favour certain species more than others and alter the population dynamics of ladybeetle species in the study area.

There are 4 time periods defined by the presence of non-native ladybeetles and soybean aphid outbreaks:

- 1989 - 1993 (T1): one non-native ladybeetle present

- 1994 - 1998 (T2): two non-native ladybeetles present

- 1999 - 2006 (T3): three non-native ladybeetles present + soybean aphid outbreaks

- 2007 - 2012 (T4): four non-native ladybeetles present + soybean aphids present but not at outbreak levels

See Bahlai et al. (2015) for more details on which non-native ladybeetle species arrived when.

Subset the data:

```
# Split the dataset into subsets for relevant time periods. Then repeat the above analyses
for each habitat type in each time period

T1 <- logTLadyBug %>% filter(Year == c(1989:1993)) %>% filter(SpeciesID != "H. axyridis")
# Also removing the non-native species "H. axyridis" from this subset. Bahlai et al. (201
5) mentions that this species was first detected in 1994, and all entries for this species
from 1989 - 1993 have 0 abundance
T2 <- logTLadyBug %>% filter(Year == c(1994:1998))
T3 <- logTLadyBug %>% filter(Year == c(1999:2006))
```

```
## Warning in Year == c(1999:2006): longer object length is not a multiple of
## shorter object length
```

```
T4 <- logTLadyBug %>% filter(Year == c(2007:2012))
```

```
## Warning in Year == c(2007:2012): longer object length is not a multiple of
## shorter object length
```

```
# Pool the abundance data for station, plot, treatment, across the years within groups of
 DOY + SpeciesID + TYPE
T1AbunSum <- T1 %>% group_by(DOY, SpeciesID, TYPE) %>% summarize(SumAbun = sum(logTAdults,
na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'DOY', 'SpeciesID'. You can override using the `.gr
oups` argument.
```

```
T2AbunSum <- T2 %>% group_by(DOY, SpeciesID, TYPE) %>% summarize(SumAbun = sum(logTAdults,
na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'DOY', 'SpeciesID'. You can override using the `.gr
oups` argument.
```

```
T3AbunSum <- T3 %>% group_by(DOY, SpeciesID, TYPE) %>% summarize(SumAbun = sum(logTAdults,
na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'DOY', 'SpeciesID'. You can override using the `.gr
oups` argument.
```

```
T4AbunSum <- T4 %>% group_by(DOY, SpeciesID, TYPE) %>% summarize(SumAbun = sum(logTAdults,
na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'DOY', 'SpeciesID'. You can override using the `.gr
oups` argument.
```

```
# Subset the pooled abundance data for each time period by habitat type
T1.F.AbunSum <- T1AbunSum %>% filter(TYPE == "forest") # This subset will not be analyzed.
The forest site only started getting sampled in 1993. There are only 16 observations, and
 the pooled abundances are all 0 values (see the T1.F.AbunSum object), so it is not possib
le to see how the species cluster in DOY abundance.
T1.A.AbunSum <- T1AbunSum %>% filter(TYPE == "annual")
T1.P.AbunSum <- T1AbunSum %>% filter(TYPE == "perennial")

T2.F.AbunSum <- T2AbunSum %>% filter(TYPE == "forest") # This subset could not be analyzed
with an NJ tree. See below (where the trees are made) for details.
T2.A.AbunSum <- T2AbunSum %>% filter(TYPE == "annual")
T2.P.AbunSum <- T2AbunSum %>% filter(TYPE == "perennial")

T3.F.AbunSum <- T3AbunSum %>% filter(TYPE == "forest")
T3.A.AbunSum <- T3AbunSum %>% filter(TYPE == "annual")
T3.P.AbunSum <- T3AbunSum %>% filter(TYPE == "perennial")

T4.F.AbunSum <- T4AbunSum %>% filter(TYPE == "forest")
T4.A.AbunSum <- T4AbunSum %>% filter(TYPE == "annual")
T4.P.AbunSum <- T4AbunSum %>% filter(TYPE == "perennial")
```

Make distance matrices from DOY species abundances for each time period in each habitat type:

```
# Make distance matrices using the custom mk_distmat() function
T1.A.dist <- mk_distmat(T1.A.AbunSum)
T1.P.dist <- mk_distmat(T1.P.AbunSum)

T2.F.dist <- mk_distmat(T2.F.AbunSum)
T2.A.dist <- mk_distmat(T2.A.AbunSum)
T2.P.dist <- mk_distmat(T2.P.AbunSum)

T3.F.dist <- mk_distmat(T3.F.AbunSum)
T3.A.dist <- mk_distmat(T3.A.AbunSum)
T3.P.dist <- mk_distmat(T3.P.AbunSum)

T4.F.dist <- mk_distmat(T4.F.AbunSum)
T4.A.dist <- mk_distmat(T4.A.AbunSum)
T4.P.dist <- mk_distmat(T4.P.AbunSum)

# Make heat maps that display the distance matrices with white representing no difference
 in temporal abundance between species and red representing the highest pairwise distance
 between species
# Use the custom p_dist() function to do this
T1.A.hMap <- p_dist(T1.A.dist)
T1.P.hMap <- p_dist(T1.P.dist)

T2.F.hMap <- p_dist(T2.F.dist)
T2.A.hMap <- p_dist(T2.A.dist)
T2.P.hMap <- p_dist(T2.P.dist)

T3.F.hMap <- p_dist(T3.F.dist)
T3.A.hMap <- p_dist(T3.A.dist)
T3.P.hMap <- p_dist(T3.P.dist)

T4.F.hMap <- p_dist(T4.F.dist)
T4.A.hMap <- p_dist(T4.A.dist)
T4.P.hMap <- p_dist(T4.P.dist)
```

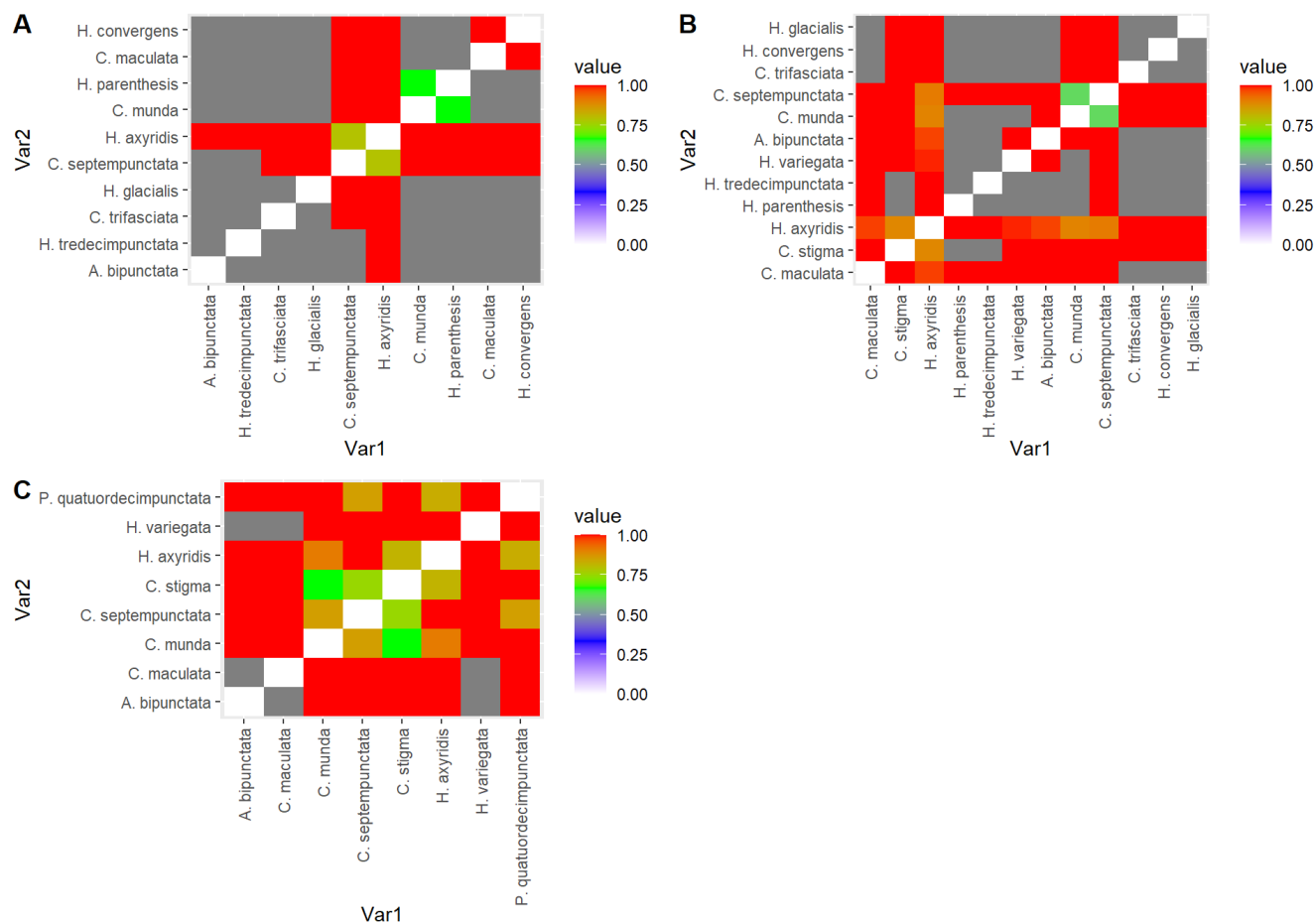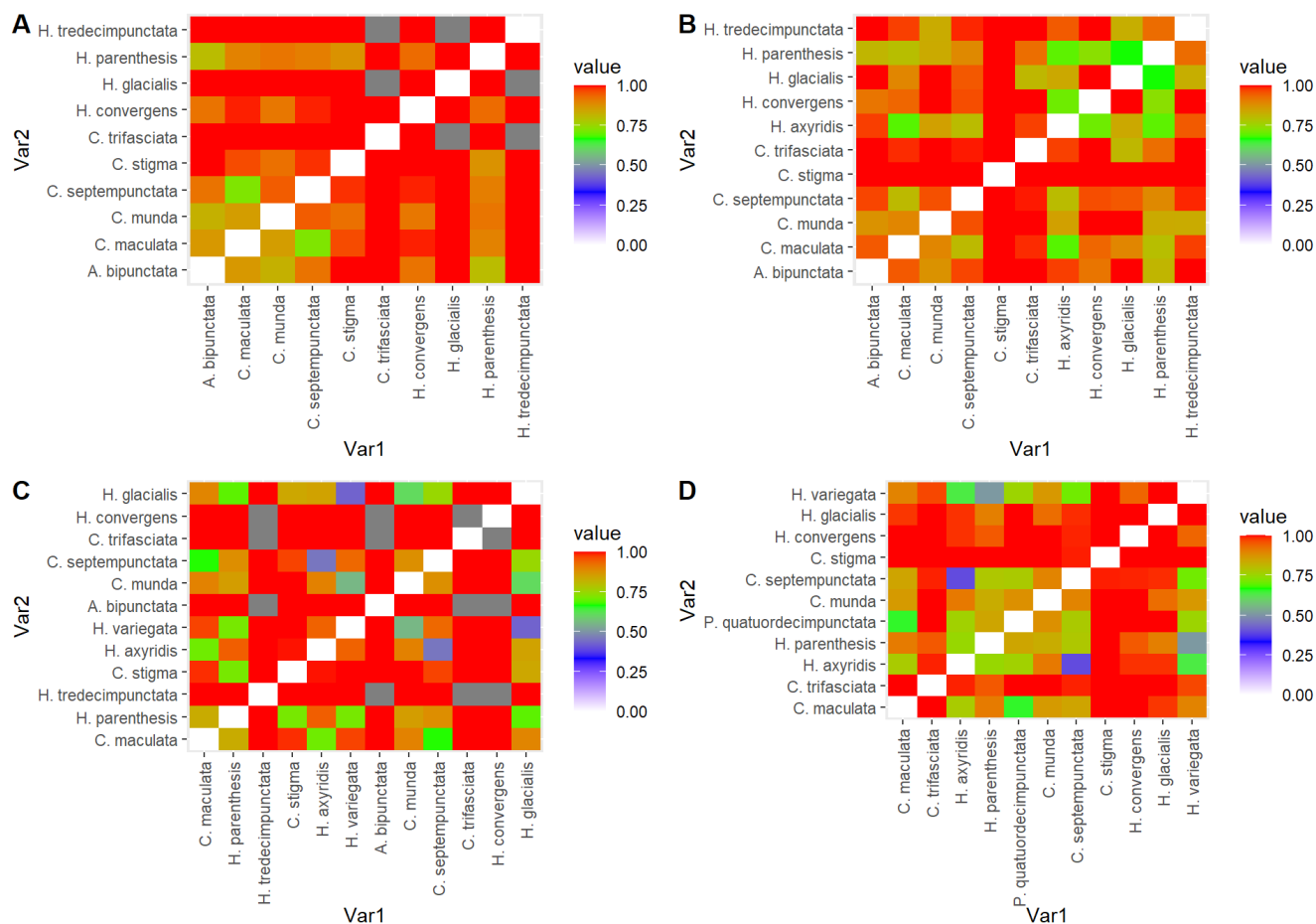Plot the heat maps by time period for each habitat type together as a single figure

```
# Plot the heat maps for each time period in the forest habitat together as a single figur
e
# Note that there was too little data for the forest habitat in the first time period so i
t is not plotted.
# Forest habitat. A = 1994 - 1998; B = 1999 - 2006; C = 2007 - 2012.
plot_grid(T2.F.hMap, T3.F.hMap, T4.F.hMap, labels = c("A", "B", "C"))
```

```
# Plot the heat maps for each time period in the annual habitat together as a single figure
# Annual habitat. A = 1989 - 1993; B = 1994 - 1998; C = 1999 - 2006; D = 2007 - 2012.
plot_grid(T1.A.hMap, T2.A.hMap, T3.A.hMap, T4.A.hMap, labels = c("A", "B", "C", "D"))
```

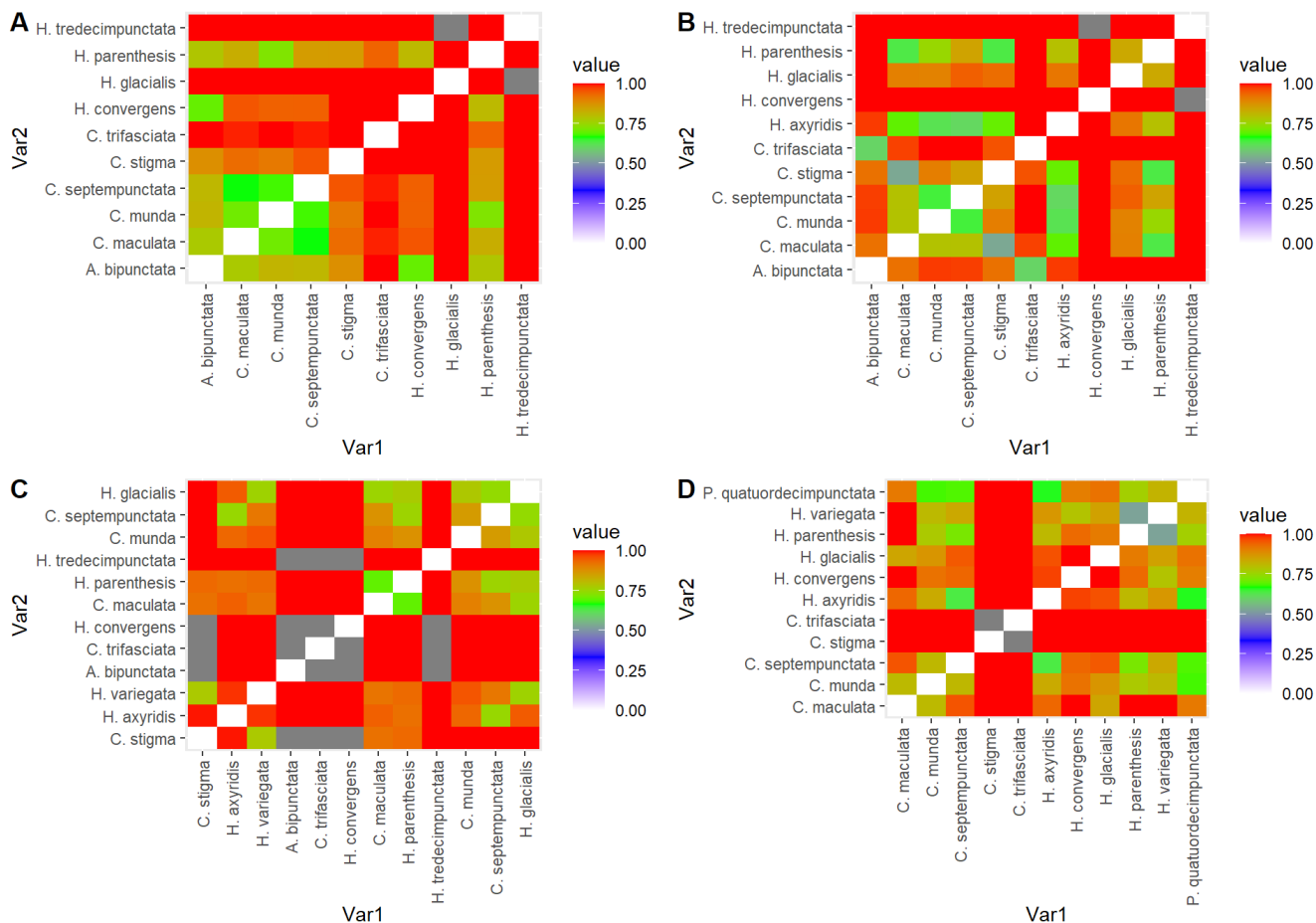# Plot the heat maps for each time period in the perennial habitat together as a single fi
gure
# Perennial habitat. A = 1989 - 1993; B = 1994 - 1998; C = 1999 - 2006; D = 2007 - 2012.
plot_grid(T1.P.hMap, T2.P.hMap, T3.P.hMap, T4.P.hMap, labels = c("A", "B", "C", "D"))

Make trees from the distance matrices for each time period in each habitat type:

```
# Make the trees. Use the custom mk_tree() function to do this
forest.pT <- mk_tree(forest.dist)
annual.pT <- mk_tree(annual.dist)
perennial.pT <- mk_tree(perennial.dist)

T1.A.tree <- mk_tree(T1.A.dist)
T1.P.tree <- mk_tree(T1.P.dist)

# T2.F.tree <- mk_tree(T2.F.dist) An njs tree could not be constructed for the forest data
in time period 2. This is because the distance information was insufficient to construct a
tree (see the high number of pairwise distances that could not be calculated for this subs
et — gray boxes in the corresponding heat map)
T2.A.tree <- mk_tree(T2.A.dist)
T2.P.tree <- mk_tree(T2.P.dist)

T3.F.tree <- mk_tree(T3.F.dist)
T3.A.tree <- mk_tree(T3.A.dist)
T3.P.tree <- mk_tree(T3.P.dist)

T4.F.tree <- mk_tree(T4.F.dist)
T4.A.tree <- mk_tree(T4.A.dist)
T4.P.tree <- mk_tree(T4.P.dist)
```
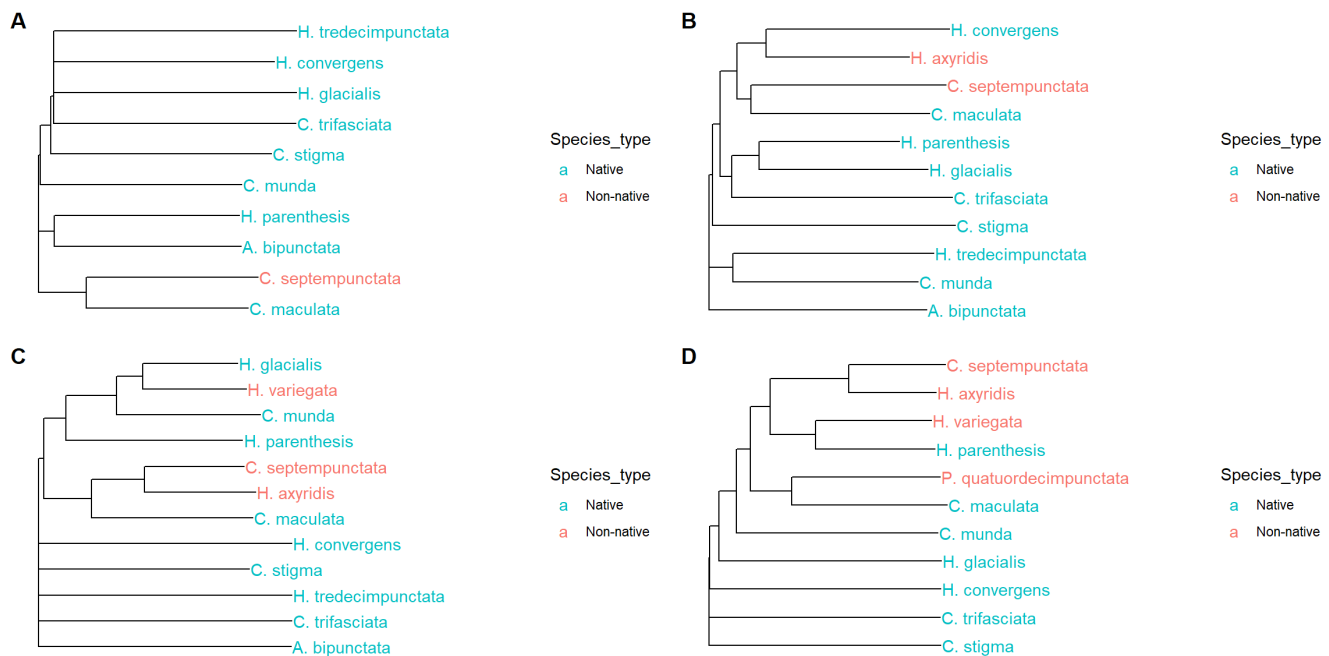
```
# Plot the trees for each time period in the forest habitat together as a single figure
# Forest habitat. A = 1999 - 2006; B = 2007 - 2012.
plot_grid(T3.F.tree, T4.F.tree, labels = c("A", "B"))
```

A

B
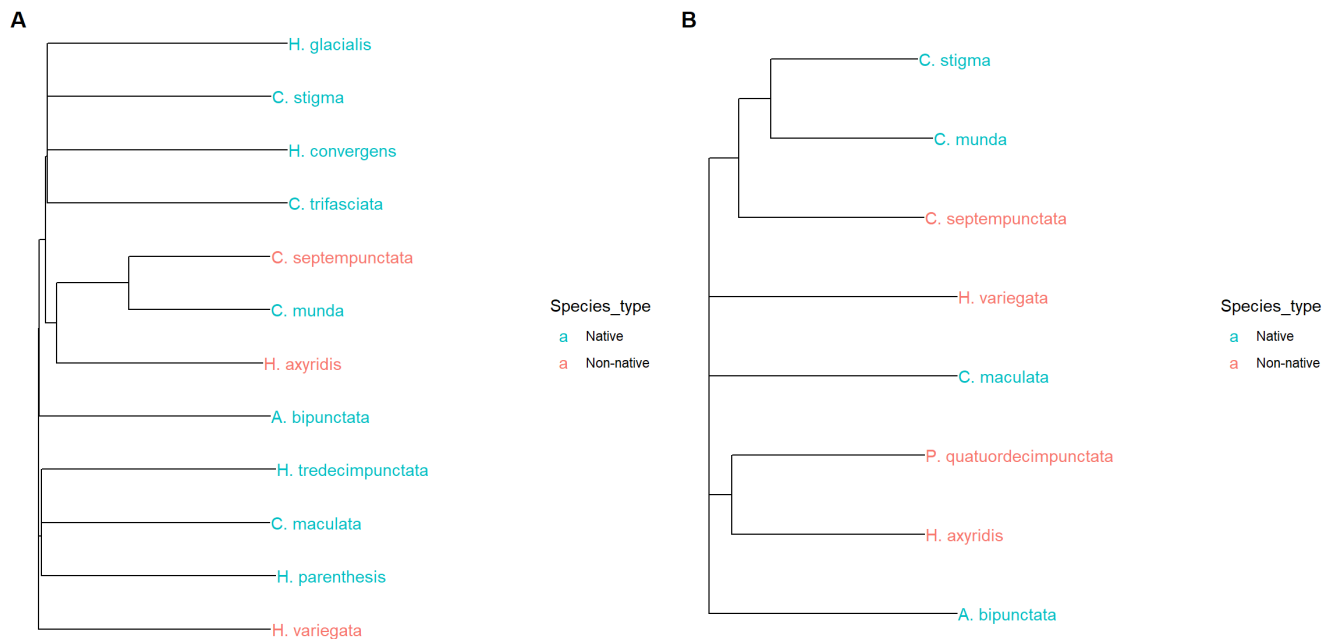
As noted in the comments in the code above, the 1989 - 1993 forest data could not be analyzed because no ladybeetles were observed in this habitat during that time period (see the T1.F.AbunSum object), and the 1994 - 1998 forest data could not be analyzed with a NJ tree because there was not enough distance information to construct the tree.

There is no distinct clustering of native and non-native ladybeetle species in either the 1999 - 2006 (A) or the 2007 - 2012 (B) time periods for forest data. Both trees have a lot of polytomies, which suggests that the differences between species in their DOY abundance patterns are poorly resolved.

```
# Plot the trees for each time period in the annual habitat together as a single figure
# Annual habitat. A = 1989 - 1993; B = 1994 - 1998; C = 1999 - 2006; D = 2007 - 2012.
plot_grid(T1.A.tree, T2.A.tree, T3.A.tree, T4.A.tree, labels = c("A", "B", "C", "D"))
```

A
- H. tredecimpunctata
- H. convergens
- H. glacialis
- C. trifasciata
- C. stigma
- C. munda
- H. parenthesis
- A. bipunctata
- C. septempunctata
- C. maculata

Species_type
- a   Native
- a   Non-native

B
- H. convergens
- H. axyridis
- C. septempunctata
- C. maculata
- H. parenthesis
- H. glacialis
- C. trifasciata
- C. stigma
- H. tredecimpunctata
- C. munda
- A. bipunctata

Species_type
- a   Native
- a   Non-native

C
- H. glacialis
- H. variegata
- C. munda
- H. parenthesis
- C. septempunctata
- H. axyridis
- C. maculata
- H. convergens
- C. stigma
- H. tredecimpunctata
- C. trifasciata
- A. bipunctata

Species_type
- a   Native
- a   Non-native

D
- C. septempunctata
- H. axyridis
- H. variegata
- H. parenthesis
- P. quatuordecimpunctata
- C. maculata
- C. munda
- H. glacialis
- H. convergens
- C. trifasciata
- C. stigma

Species_type
- a   Native
- a   Non-native

In the 1989 - 1993 (A) and 1999 - 2006 (C) annual habitat data there is no clustering of native and non-native species. The 1999 - 2006 annual data appears at first glance to have a lot of native species that segregate separately from the non-native species, however these native species do not belong together in a common cluster and each of these native species has little to no records in this subset of the dataset (see code chunk below). This suggests that most of the native ladybeetles in this subset simply segregate separately from the non-native species because they are absent or nearly absent from the dataset.

The other 2 time periods, 1994 - 1998 (B) and 2007 - 2012 (D), seem to show some segregation between native and non-native species on the tree. In both time periods however, there are one or two native species within the non-native species cluster.

In annual habitats, these trees suggest that there was no temporal partitioning between native and non-native species in the 1989 - 1993 and 1999 - 2006 time periods, but they do suggest that there might be some partitioning between native and non-native species in the 1994 - 1998 and 2007 - 2012 time periods.

```
# For annual habitat, 1999 - 2006 (C) subset, count the total amount of log-transformed ab
undances for each species
# Notice that many of the native species that appear to segregate separately from the non-
native species actually have little to no presence in this subset
T3.A.AbunSum %>% group_by(SpeciesID) %>% summarise(tot = sum(SumAbun))
```

```
## # A tibble: 12 x 2
##    SpeciesID              tot
##  * <chr>                <dbl>
##  1 A. bipunctata            0
##  2 C. maculata           101.
##  3 C. munda              3.88
##  4 C. septempunctata     46.5
##  5 C. stigma            0.431
##  6 C. trifasciata           0
##  7 H. axyridis           389.
##  8 H. convergens            0
##  9 H. glacialis          7.14
## 10 H. parenthesis        10.6
## 11 H. tredecimpunctata      0
## 12 H. variegata          8.18
```

```
# Plot the trees for each time period in the perennial habitat together as a single figure
# Perennial habitat. A = 1989 - 1993; B = 1994 - 1998; C = 1999 - 2006; D = 2007 - 2012.
plot_grid(T1.P.tree, T2.P.tree, T3.P.tree, T4.P.tree, labels = c("A", "B", "C", "D"))
```
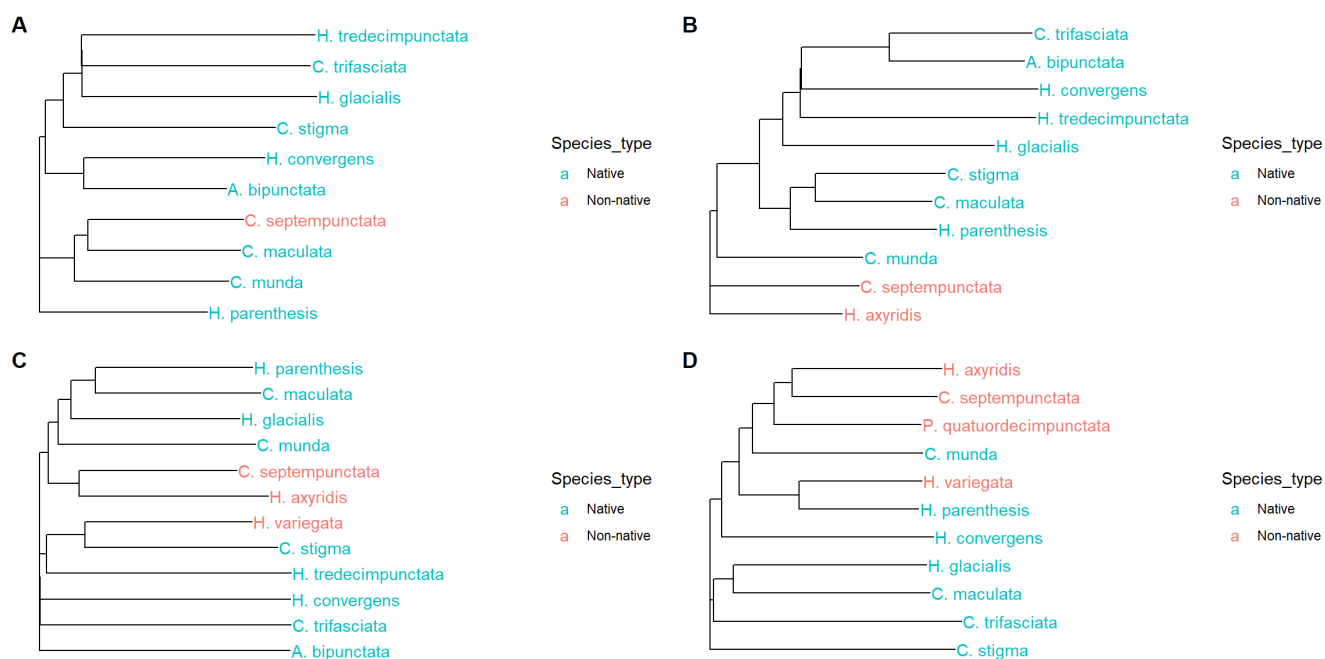

There is no distinct clustering of native and non-native ladybeetle species in the first two time periods for perennial habitat data. However, in the 1999 - 2006 (C) and 2007 - 2012 (D) time periods there does appear to be some segregation between native and non-native species. In 1999 - 2006 only *C. septempunctata* and *H. axyridis* segregate separately from the other species, while the third invasive species is in a separate cluster. In 2007 - 2012 all the non-native species are in the same cluster along with a couple native species.

Overall, this suggests there may be some partitioning between native and non-native species in perennial habitat during the soybean aphid outbreaks.

4.2.5 Potential issues with clustering the species by DOY abundance

Even though we log transformed the abundance data to reduce the importance of extremely abundant species on the clustering analysis, the clusters are still probably biased towards finding differences between abundant and non-abundant species because more abundant species are more likely to be caught by the traps and rare species are more likely to be missed.

5 Conclusion

We used graphical and tree clustering analysis to analyze how ladybeetle abundance changes seasonally and whether niche partitioning is occurring between native and non-native ladybeetles.

Using graphical analysis on adult abundance, we found that abundance changes with seasonality for both native and non-native species, except for native species found in the forest habitat. Seasonal variation in abundance may reflect environmental changes. For example, activity levels of ladybeetles may vary with temperature, increasing or decreasing each species risk of being trapped at different times of the year. This may explain seasonal differences in abundance of all ladybeetles, but does not account for differences between species.

Across the perennial and forest habitat, graphical analysis found little to no evidence for temporal niche partitioning between native and non-native ladybeetles. In contrast, cluster analysis indicates that niche partitioning may occur in the perennial habitat during the final 2 time periods (1999 - 2006 and 2007 – 2012). These differing results may be explained by the overall low abundance of native species throughout the year leading to difficulty in graphically assessing proportional changes in abundance. For example, native ladybeetle abundance might increase by a 1/3 in response to a decrease in non-native ladybeetle abundance, but, due to the scale of the graph and the high abundance of non-native ladybeetles generally, we would have had difficulty visually detecting this change because it would represent such a small change in height of the line.

Within the annual habitat, which is a rotating monoculture, evidence for temporal niche partitioning was found using both techniques. Varying native and non-native species abundance indicates temporal niche partitioning ~2/3 of the way through the year, where there is a concurrent decline of non-native abundance and increase of native abundance. Using cluster analysis, the weak segregation between native and non-native species on the tree for annual habitats in the time periods of 1994 - 1998 and 2007 - 2012 corresponds with the weak opposing increases and decreases between native an non-natives in the same subsets in the abundance time-series graph. It is interesting to note that we only see strong evidence for the abundances of native and non-native ladybeetles impacting each other during a limited time of the year. These results suggest that species-species interactions and competitive exclusion only influence abundance at a certain time of year, potentially driven by a seasonally-shared resource. Food preference and availability may overlap during this period of time and increase species-species interactions. This would suggest that native ladybeetles are competitively dominant, as they are able to exclude non-native ladybeetles from this habitat during this time, driving down non-native abundance.

However, the similarity in niches of native and non-native species (they are all aphidophagous) as well as the overall high abundance of non-native ladybeetles throughout the year suggests otherwise: it is likely the non-native species are competitively dominant and exclude native species from these habitats for most of the year. In this scenario, the rise in native ladybeetle abundance for a limited time can be explained by a decline in non-native abundance driven by other abiotic or biotic factors, such as changes in temperature, allowing native species to occupy annual habitat they would otherwise be excluded from.

A third possibility combines the two previous hypotheses. It is possible that, while non-native ladybeetles are competitively dominant throughout most of the year, a seasonal, shared resource (such as a prey species) over which native species are competitively dominant emerges around ~2/3 of the way through the sampled

season. Thus, native species would be excluded from annual habitat at all times except this single period, when they can better monopolize a specific shared resource.

Interestingly, during soybean aphid outbreak years (1999 - 2006), both native and non-native abundance increases ~2/3 of the way through the sampling season, when they normally change inversely. This result lends support to the possibility of a shared resource; it is possible that soybean aphids, a prey source many ladybeetle species analyzed here, emerge during this time. During outbreak years, aphids may have become so abundant that they stopped being a limiting resource, allowing both native and non-native ladybeetles to thrive.

This study was able to build off of the results of previous researchers and provide a better understanding of how native and non-native ladybeetles coexist. Understanding species-species interactions and abundance variation is important to understand how invasive ladybeetles impact native ladybeetles, which is crucial to properly managing their impact. Further research should examine interactions at a species level, which may deliver more concrete explanations of the patterns we observed. In addition, experimental work to determine dominance relationships among this system are required to determine whether changes in abundance are a result of competitive exclusion.