BIOL 386 Final Paper

Cecina Babich Morrow

May 12, 2017

Introduction

Biomes categorize global plant diversity into regions based on both climatic characteristics and dominant vegetation type (Woodward et al. 2004). Over the course of time, plant species can enter new biomes, either transitioning to a new habitat or expanding their range. These biome shifts can provide information about the assemblage processes that shape plant communities, as well as adaptations possibly enabling species to colonize these new environments (Donoghue and Edwards 2014). Since colonizing new biomes presents challenges in terms of adapting to novel climate constraints, biome transitions are thought to be relatively rare and most likely necessitate changes in functional traits (Kerkhoff et al. 2014). Thus by understanding these biome transition events, we can investigate the relative effects of ecological and evolutionary processes on assembling large-scale plant communities.

This semester, I took the first steps to investigate these transition events using the BIEN plant dataset. In order to study biome transition events, it is first necessary to identify the biomes inhabited by each extant species. I used the range maps of BIEN species in conjunction with maps of biomes to make these assignments. The following document outlines my methods and code to make these assignments, as well as the possibly simplifying assumptions I made along the way.

```
library(maptools)
library(maps)
library(mapproj)
library(rgeos)
library(rgdal)
library(sp)
library(BIEN)
library(mosaic)
library(raster)
```

Species Ranges

Brian Maitner sent me the following dataset, “bien_speciesrange_cells”. Tbis dataset includes the cell numbers occupied by each species in the BIEN dataset. These cells came from a 100x100 kilometer raster, here referred to as “emptyraster”, in the Lambert Azimuthal equal area projection. Each cell was counted as an occurrence for a species if the range map for that species covered at least 1% of the cell, making this occurrence list a fairly liberal estimate of which species inhabit which cells.

Below are the first 15 rows of this dataset, which gives the species name and the occupied cell numbers.

```
bien_8_25_2016_100km_1percent_occurrence_only <- read.csv("C:/Users/Cecina/Desk
top/plants/bien_8_25_2016_100km_1percent_occurrence_only.csv")
bien_speciesrange_cells<-bien_8_25_2016_100km_1percent_occurrence_only
head(bien_speciesrange_cells,15)
```

```
##    current_species occupied_cells
## 1   Aa_argyrolepis           7877
## 2    Aa_colombiana           7462
## 3    Aa_colombiana           9016
## 4    Aa_colombiana           9119
## 5    Aa_colombiana           9221
## 6    Aa_colombiana           9222
## 7    Aa_colombiana           9427
## 8    Aa_colombiana          10150
## 9    Aa_colombiana          11091
## 10  Aa_denticulata           9016
## 11  Aa_denticulata           9427
## 12    Aa_fiebrigii          11092
## 13    Aa_hieronymi          11502
## 14    Aa_hieronymi          11605
## 15    Aa_hieronymi          12532
```

```
emptyraster<-raster("blank_100km_raster.tif")
```

Biome Maps

The biome maps used for these analyses were imported from the World Wildlife Federation website. The biome maps needed to be converted into the same projection as the blank raster (Lambert Azimuthal equal area) so that the cell numbers for the biomes would match the cell numbers for the species ranges.

```
biome<-readOGR(dsn = "C:/Users/Cecina/Desktop/HypervolumeFiles/Terrestrial_Ecor
egions",layer ="wwf_terr_ecos" )
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "C:/Users/Cecina/Desktop/HypervolumeFiles/Terrestrial_Ecoregions", l
ayer: "wwf_terr_ecos"
## with 14458 features
## It has 21 fields
```

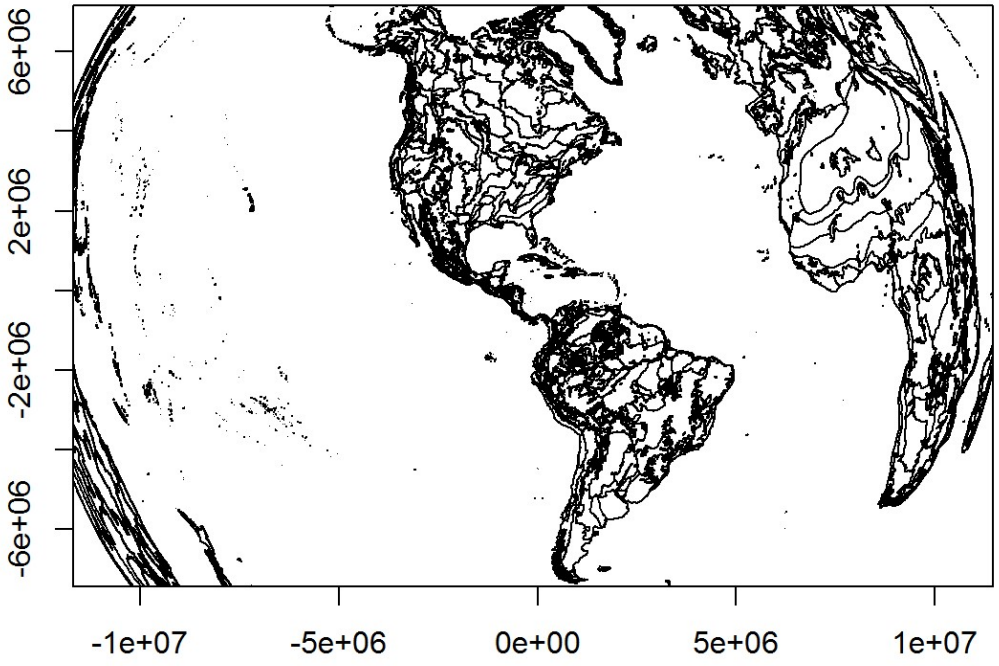
```
#Get projection of the raster
crs(emptyraster)
```

```
## CRS arguments:
##  +proj=laea +lat_0=15 +lon_0=-80 +x_0=0 +y_0=0 +datum=WGS84
## +units=m +no_defs +ellps=WGS84 +towgs84=0,0,0
```

```
#transform biomes to same projection as the raster
biome_transform<-spTransform(x = biome,CRSobj = emptyraster@crs)
```

The resulting transformed biome maps are now in the same projection as the raster and the species range maps. This map shows the WWF biome maps in this projection, which is specialized to reduce distortion in the New World specifically.

```
plot(emptyraster)
plot(biome_transform,add=T)
```



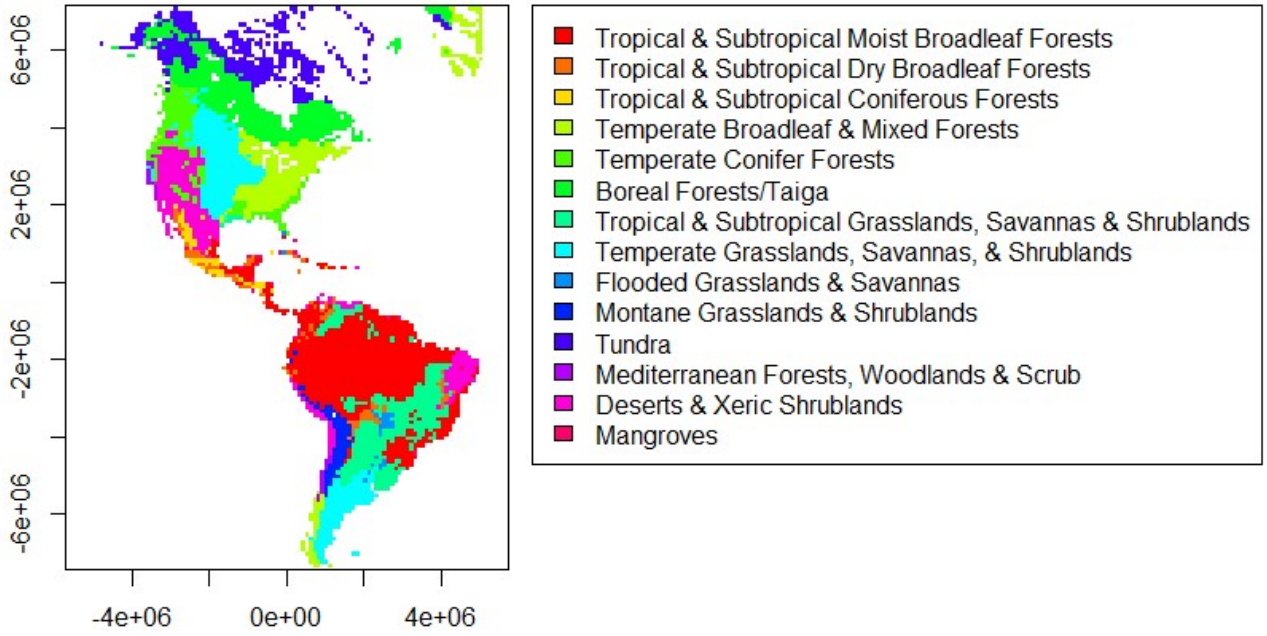
After putting the biome maps into the correct projection, I developed a dataset with a list of the cells occupied by each biome in a similar format to the species list. Below are the first 15 rows of this dataframe.

```
biome_cells<-data.frame(Biome=character(),occupied_cells=character(),stringsAsF
actors = FALSE)
for(i in 1:length(unique(biome_transform@data$BIOME))) {
  biome_i<-biome_transform[biome_transform@data$BIOME==unique(biome_transform@d
ata$BIOME)[i],]
  biome_id<-unique(biome_transform@data$BIOME)[i]
  biomeraster_i<-rasterize(biome_i,emptyraster)
  biome_cells_i<-Which(biomeraster_i>0,cells=TRUE)
  biomedata_i<-data.frame(Biome=rep(biome_id,length(biome_cells_i)),occupied_ce
lls=biome_cells_i)
  biome_cells<-rbind(biomedata_i,biome_cells)
}

colnames(biome_cells)<-c("Biome","occupied_cells")

head(biome_cells,15)
```

```
##    Biome occupied_cells
## 1     98           1268
## 2     98           1269
## 3     98           1372
## 4     98           1373
## 5     98           1374
## 6     98           1785
## 7     98           1888
## 8     98           1889
## 9     98           2511
## 10    98           2613
## 11    98           2924
## 12    98           3026
## 13    98           3027
## 14    98           3234
## 15    98           3652
```

By setting the value of each raster cell to the number of the biome covering that cell, we can plot the following biome map, where the colors correspond to the biome numbers.

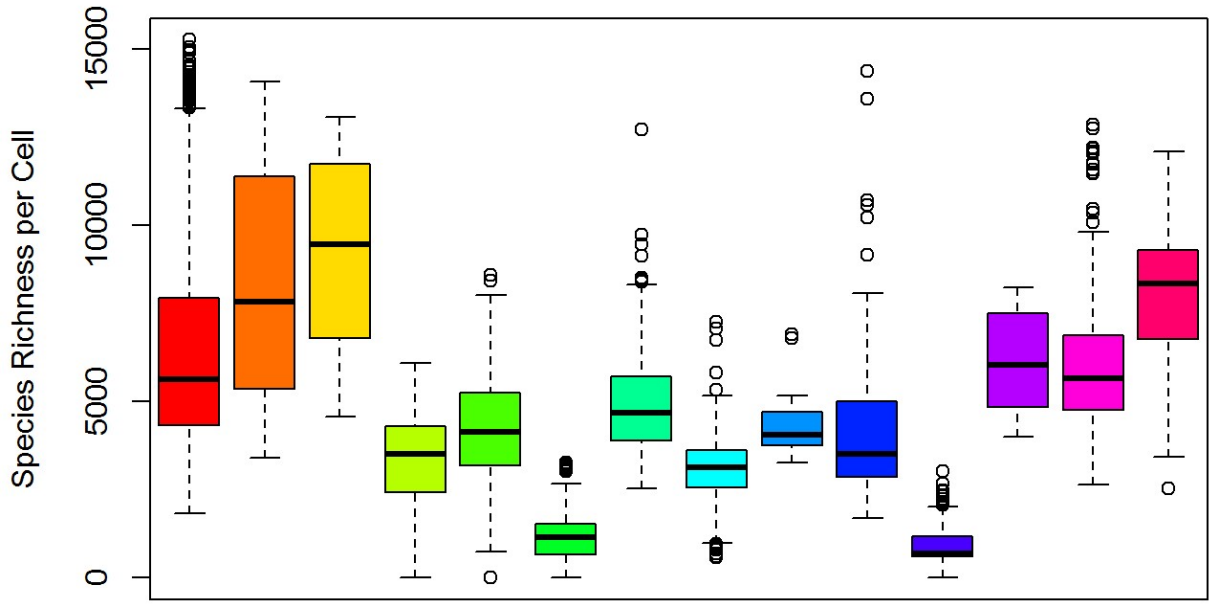
```
biomeraster_test<-emptyraster
for(i in 1:14){
  biomeraster_test[biome_cells$occupied_cells[which(biome_cells$Biome==i)]]<-i
}
biomenames<-c("Tropical & Subtropical Moist Broadleaf Forests", "Tropical & Sub
tropical Dry Broadleaf Forests", "Tropical & Subtropical Coniferous Forests",
"Temperate Broadleaf & Mixed Forests", "Temperate Conifer Forests", "Boreal For
ests/Taiga", "Tropical & Subtropical Grasslands, Savannas & Shrublands", "Tempe
rate Grasslands, Savannas, & Shrublands", "Flooded Grasslands & Savannas", "Mon
tane Grasslands & Shrublands", "Tundra", "Mediterranean Forests, Woodlands & Sc
rub", "Deserts & Xeric Shrublands", "Mangroves","","")
par(mar=c(2,2.5,2,25))
plot(biomeraster_test,col=rainbow(14),legend = FALSE,xlim=c(-5000200,5000200))
legend("topright",inset=c(-1.7,0),legend=biomenames[1:14],fill=rainbow(14),xpd=
TRUE)
```



Species Density in Each Biome

From these cell numbers, I was able to determine the species richness of each cell. Thus each biome had a range of species richnesses in each of the cells occupied by the biome. The following boxplot gives the range of species richness per cell for each biome, with colors corresponding to the map above.

```
biome_cells$speciesrichness<-NA
for(i in 1:length(biome_cells$Biome)){
  biome_cells$speciesrichness[i]<-length(which(bien_speciesrange_cells$occupied
_cells==biome_cells$occupied_cells[i]))
}
boxplot(biome_cells$speciesrichness~biome_cells$Biome, xlim=c(1,14),col=rainbow
(14),ylab="Species Richness per Cell",xaxt="n")
```



Species Occurrence by Biome

Given the list of cell numbers occupied by both species and occupied by each biome, I was then able to determine the biome for each cell inhabited by the species.

```
for(i in 1:14){
  biome_cells_i<-biome_cells$occupied_cells[which(biome_cells$Biome==i)]
  bien_speciesrange_cells$biome[which(is.element(bien_speciesrange_cells$occupi
ed_cells,biome_cells_i))]<-i
}
```

Now bien_species_range_cells has a column for the biome number:

```
head(bien_speciesrange_cells,15)
```

```
##    current_species occupied_cells biome
## 1   Aa_argyrolepis           7877     1
## 2    Aa_colombiana           7462     3
## 3    Aa_colombiana           9016     1
## 4    Aa_colombiana           9119     1
## 5    Aa_colombiana           9221     1
## 6    Aa_colombiana           9222    10
## 7    Aa_colombiana           9427    10
## 8    Aa_colombiana          10150    10
## 9    Aa_colombiana          11091     2
## 10  Aa_denticulata           9016     1
## 11  Aa_denticulata           9427    10
## 12    Aa_fiebrigii          11092     2
## 13    Aa_hieronymi          11502     2
## 14    Aa_hieronymi          11605    10
## 15    Aa_hieronymi          12532     7
```

From this list of biomes in each cell, I determined the total number of cells occupied by each species that lay in each biome.

```
speciesbybiome<-tally(current_species~biome, data=bien_speciesrange_cells)
speciesbybiome<-as.data.frame(speciesbybiome)
head(speciesbybiome,15)
```

```
##     current_species biome Freq
## 1    Aa_argyrolepis     1    1
## 2     Aa_colombiana     1    3
## 3    Aa_denticulata     1    1
## 4      Aa_fiebrigii     1    0
## 5      Aa_hieronymi     1    0
## 6      Aa_leucantha     1    3
## 7          Aa_macra     1    1
## 8        Aa_maderoi     1   71
## 9      Aa_mathewsii     1   54
## 10    Aa_microtidis     1    0
## 11      Aa_paleacea     1  131
## 12     Aa_riobambae     1    2
## 13 Aa_sphaeroglossa     1    0
## 14   Aa_trilobulata     1    1
## 15   Aa_weddelliana     1    0
```

From this dataframe, I determined the biome that covered the greatest number of cells occupied by each species. Some cells were not covered by a biome, resulting in NA. For some species with limited ranges, NA was thus the biome with the highest coverage, so I removed these rows from the dataset in

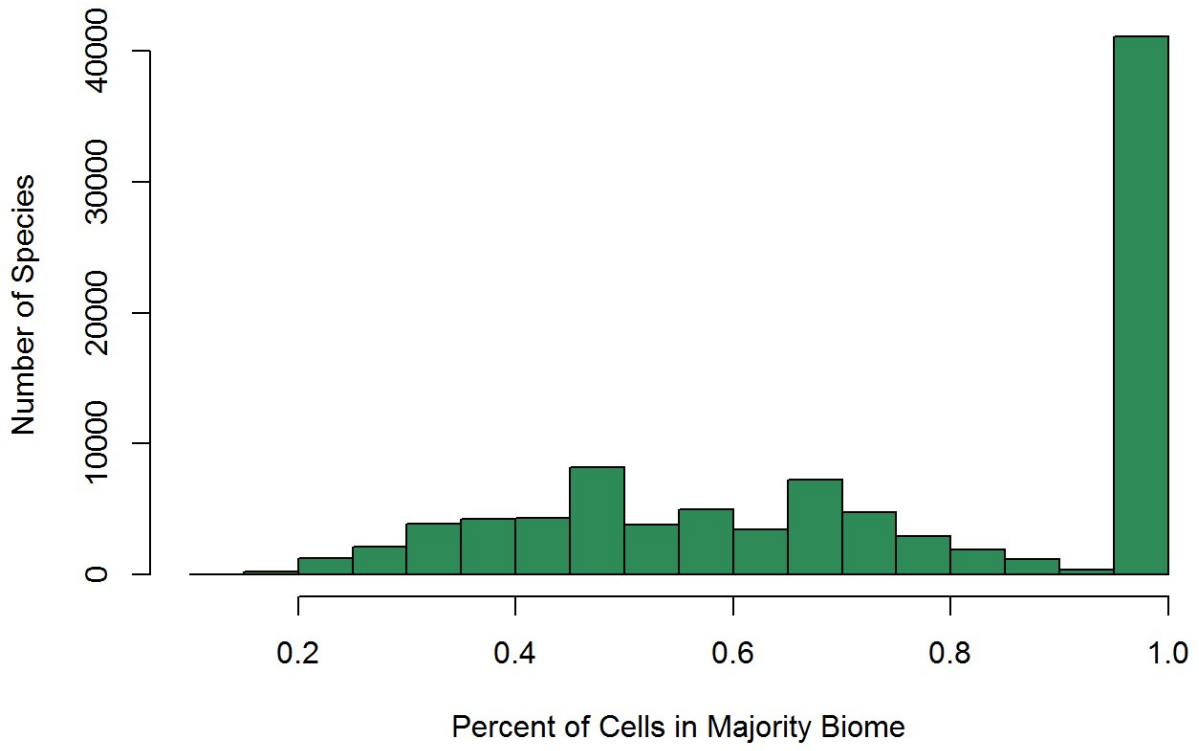
order to determine the actual biome that a species belonged to most frequently. In some cases, multiple biomes tied for the greatest coverage, so both biomes were included in the biome column, separated by a comma.

Finally, I added a column with the percentage of the cells occupied by the species that were within the biome that contained the highest number of cells.

```
species_biome_frequencies<-speciesbybiome[which(!is.na(speciesbybiome$biom
e)),] %>% group_by(current_species) %>%
  summarize(freqbiome = paste(biome[which(Freq == max(Freq))], collapse = ",
"),maxfreq=max(Freq),totalcells=sum(Freq))
#get rid of the species that occur in 0 cells
species_biome_frequencies<-species_biome_frequencies[which(species_biome_freque
ncies$totalcells!=0)]
#Add column for the percentage of total cells that fall within the biome of hig
hest frequency
species_biome_frequencies$percentage<-species_biome_frequencies$maxfreq/species
_biome_frequencies$totalcells
```

The following histogram shows the distribution of percentages of cells in the biome of maximum cell coverage for each species.

```
hist(na.omit(species_biome_frequencies$percentage),xlab="Percent of Cells in Ma
jority Biome", prob=FALSE,col="seagreen",main="",ylab="Number of Species")
```



For most species, all of the cells of occurrence fall within the majority biome. This allowed us to feel fairly confident in assigning each species to its biome of maximum coverage.

Multiple Biomes of Maximum Coverage

7,278 species in the dataset have two or more biomes that cover an equal and maximum number of cells in their range. I extracted these species into their own dataset.

```
species_biome_frequencies_tied<-species_biome_frequencies[grep(",", species_bio
me_frequencies$freqbiome),]
#split up the column freqbiome by the commas
split<-strsplit(species_biome_frequencies_tied$freqbiome, split=",")
species_biome_frequencies_tied<-data.frame(current_species = rep(species_biome_
frequencies_tied$current_species, sapply(split, length)), freqbiome = unlist(sp
lit), maxfreq = rep(species_biome_frequencies_tied$maxfreq,sapply(split,lengt
h)),totalcells = rep(species_biome_frequencies_tied$totalcells,sapply(split,len
gth)))
head(species_biome_frequencies_tied,15)
```

```
##      current_species freqbiome maxfreq totalcells
## 1      Aa_colombiana         1       3          8
## 2      Aa_colombiana        10       3          8
## 3     Aa_denticulata         1       1          2
## 4     Aa_denticulata        10       1          2
## 5       Aa_hieronymi         2       1          3
## 6       Aa_hieronymi         7       1          3
## 7       Aa_hieronymi        10       1          3
## 8  Abies_cephalonica         4       1          2
## 9  Abies_cephalonica         5       1          2
## 10       Abies_firma         1       0          0
## 11       Abies_firma         2       0          0
## 12       Abies_firma         3       0          0
## 13       Abies_firma         4       0          0
## 14       Abies_firma         5       0          0
## 15       Abies_firma         6       0          0
```

I then extracted the species without “ties” (i.e. those with one and only one biome of maximum coverage).

```
species_biome_frequencies_noties<-species_biome_frequencies[grep(",",species_bi
ome_frequencies$freqbiome,invert = TRUE),]
head(species_biome_frequencies_noties[,1:4],15)
```

```
## Source: local data frame [15 x 4]
##
##        current_species freqbiome maxfreq totalcells
##                  (fctr)     (chr)   (int)      (int)
## 1       Aa_argyrolepis         1       1          1
## 2         Aa_fiebrigii         2       1          1
## 3         Aa_leucantha         1       3          4
## 4             Aa_macra         1       1          1
## 5           Aa_maderoi         1      71        101
## 6         Aa_mathewsii         1      54        111
## 7        Aa_microtidis         2       1          1
## 8          Aa_paleacea         1     131        216
## 9         Aa_riobambae         1       2          2
## 10    Aa_sphaeroglossa         2       1          1
## 11      Aa_trilobulata         1       1          1
## 12      Aa_weddelliana        10       1          1
## 13     Abarema_acreana         1       6          7
## 14  Abarema_adenophora         1     729        993
## 15 Abarema_adenophorum         1       1          1
```

Then I put these two dataframes together into one frequency list with both tied and non-tied species and their biome(s) of maximum coverage.

```
total_species_biome_frequencies<-rbind(species_biome_frequencies_tied,species_biome_frequencies_noties)
write.csv(total_species_biome_frequencies,"total_species_biome_frequencies.csv")
```

```
total_species_biome_frequencies <- read.csv("C:/Users/Cecina/OneDrive/Documents/Kenyon College/Kerkhoff Lab/kerkhofflab/Biome-Transitions/total_species_biome_frequencies.csv")
```

Species List By Biome

Finally, I was able to create a species list for each biome. For each biome, the dataset has a list of all species whose ranges are covered more by that biome than by any other. In the case of ties, species are assigned to all of the biomes covering the maximum number of cells in their ranges.

```
species_biome_list<-data.frame(Biome=character(),species=character(),stringsAsFactors = FALSE)
#adding all species (tied and not tied)
for(i in 1:14) {
  rows_i<-total_species_biome_frequencies[which(total_species_biome_frequencies$freqbiome==i),]
  species_list_i<-total_species_biome_frequencies$current_species[which(total_species_biome_frequencies$freqbiome==i)]
  species_biome_data_i<-data.frame(Biome=rep(i, length(length(species_list_i))),species=species_list_i)
  species_biome_list<-rbind(species_biome_data_i,species_biome_list)
}
```

Conclusion

I was able to complete my goal of assigning a biome or biomes to each species in the BIEN dataset, which will enable me to reconstruct biome inhabitation over evolutionary time. As part of this next step in the process, I will examine what model of ancestral reconstruction to use–whether one that permits multiple biome assignments or not–and apply it to the BIEN phylogeny. Alternatively, I may select a clade or multiple clades containing species in multiple biomes and having a well-resolved phylogeny on which to perform the reconstruction.

The resulting ancestral reconstruction will permit me to identify nodes at which species have transitioned into novel biomes. These evolutionary events may be correlated with the evolution of new functional traits or tolerances enabling the species to invade a new habitat. Alternatively, certain clades may be better able to radiate into new environments than others. I hope to use my results to investigate the answers to these and other questions over the summer.

Works Cited

Donogue, M.J. & Edwards, E.J. (2014) Biome shifts and niche evolution in plants. Annual Review of Ecology, Evolution, and Systematics, 45, 547-572.

Kerkhoff, A.J., Moriarty, P.E. & Weiser, M.D. (2014). The latitudinal species richness gradient in New World wood angiosperms is consistent with the tropical conservatism hypothesis. Proceedings of the National Academy of Sciences of the United States of America, 111, 8125-8130.

Woodward, F.I., Lomas, M.R., & Kelly, C.K. (2004) Global climate and the distribution of plant biomes. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 359, 1465-1476.