

IllumFUN_Q2de: HMSC and follow-ups

Marissa Lee

12/16/2019

Q2. How are differences in fungal composition across the landscape explained by environmental variables?

Table of contents

0. Load data and pre-process ASV matrix

See IllumFUN_Q1.Rmd

A. Determine which environmental variables to include in path analysis

See IllumFUN_Q2a.Rmd

B. SEM using DPCoA1 score to represent community

See IllumFUN_Q2bc.Rmd

C. Investigate SEM results with bivariate plots

See IllumFUN_Q2bc.Rmd

D. Investigate SEM results with HMSC

1. Set up HMSC analyses on the HPC
2. Examine HMSC output
3. Prune to high confidence taxa-environment relationships
4. Make phylogenetic tree/heat map plots

E. Follow-ups

1. Varpart for leaf fungi
2. Variance explained by VST PCoA1

Load packages, functions, paths

Custom functions

D. Investigate SEM results with HMSC

1. Set up HMSC analyses on the HPC

Two general types of HMSC models were run for each plant compartment: (a) ASV presence/absence and (b) ASV relative abundance. Each type has its own folder that was uploaded to the HPC with relevant input data, code, and output. Folders are called “hmsc_fits_ab” and “hmsc_fits_pr”, respectively, and located at data_intermediates/Illum_analyses.

The general structure within each folder is as follows: 1. data folder with a phyloseq object for each plant compartment and taxonomy-based phylogenetic tree 2. R script for each plant compartment with the name “hmsc_compartment.R” 3. HPC scheduling script for each plant compartment with the name “runR_hmsc_.csh” 4. output folder: these items are used in the code below

2. Examine HMSC output

The code chunks below are mostly commented out because reading in some of these HMSC output files are very memory intensive and can take a while.

Examine mixing (commented out)

Examine model fit

```
## Warning in mean.default(x$R2, na.rm = T): argument is not numeric or logical:
## returning NA
```

```
## Warning in mean.default(x$R2, na.rm = T): argument is not numeric or logical:
## returning NA
```

```
## Warning in mean.default(x$R2, na.rm = T): argument is not numeric or logical:
## returning NA
```

```
## Warning in mean.default(x$AUC, na.rm = T): argument is not numeric or logical:
## returning NA
```

```
## Warning in mean.default(x$AUC, na.rm = T): argument is not numeric or logical:
## returning NA
```

```
## Warning in mean.default(x$AUC, na.rm = T): argument is not numeric or logical:
## returning NA
```

##	tissue	model	waic	rmse	r2	auc
## 1	Leaf	Abundance	70.03698	0.8728860	0.2536010	NA
## 2	Leaf	Presence	71.73838	0.2955561	NA	0.7922793
## 3	Root	Abundance	344.99796	0.8248864	0.2452250	NA
## 4	Root	Presence	118.31911	0.1907437	NA	0.8064638
## 5	Soil	Abundance	177.86975	0.8675905	0.1957414	NA
## 6	Soil	Presence	191.24815	0.2337007	NA	0.8165146

Examine “species niches” as structured by phylogeny (commented out)

Plot the distribution of the rho parameter for each model (commented out)

Print tables *leaf*

root

soil

3. Prune to high confidence taxa-environment relationships, i.e. 99% CI does not overlap zero

4. Make phylogenetic tree/heat map plots

Reformat the data matrix so that each row is an ASV. Also, add “ns” where ASV is present but not significant

Make the phylogenetic tree – define the data matrix, phylum nodes, and key classes

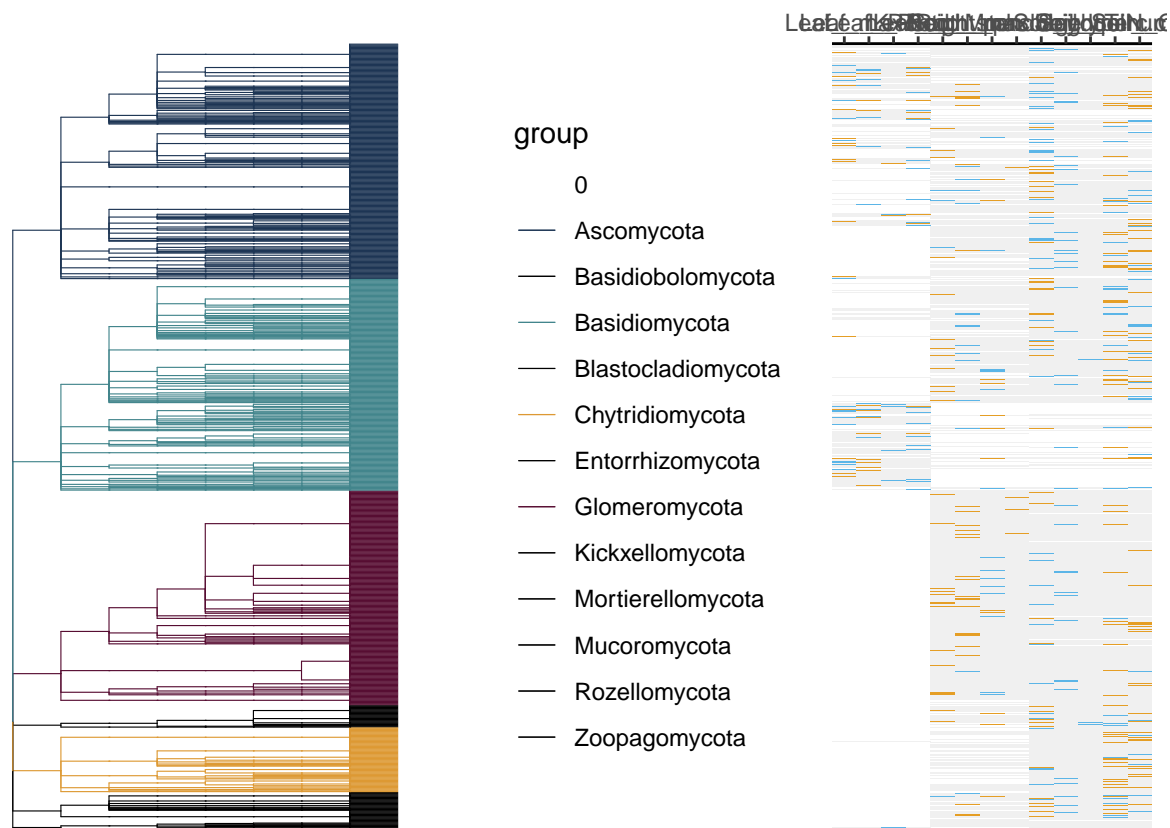
Make the tile plot and combine it with the tree

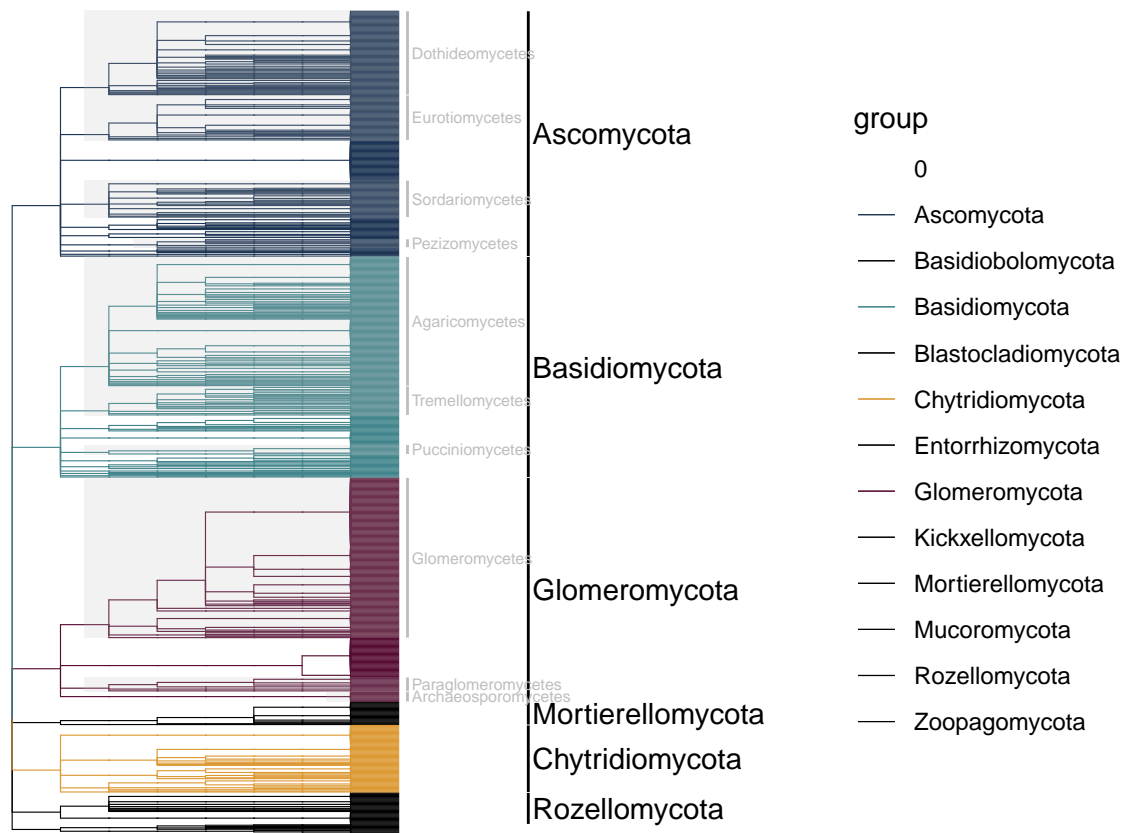
```
## Joining, by = "ASV"
```

```
## Scale for 'y' is already present. Adding another scale for 'y', which will
## replace the existing scale.
```

```
## Joining, by = "ASV"
```

```
## Scale for 'y' is already present. Adding another scale for 'y', which will
## replace the existing scale.
```





E. Followups

1. Varpart – Leaf fungi explained by lat lon versus environmental matrix (based on SEM)? (commented out)

2. Enough variance explained by VST PCoA1 to do SEM? (commented out)

Not very much variation in the community is explained by PC1 using VST. PC1 explains 9.4% (Leaf), 5.2% (Root), 6.0% (Soil).