

# Analyzing microbial communities: from sequences to patterns and processes

*A workshop on the analysis of microbial communities assessed using 16S rRNA tag-sequencing*

22-26 October 2012

Yale University, Bass Center 405

8:30 -10:30 a.m.

Contact:

Ashley Shade, Jo Handelsman Lab

[ashley.shade@yale.edu](mailto:ashley.shade@yale.edu)

## **Day 5**

**Analyzing community patterns: The contributions of individual OTUs to community dynamics**

# Review from Day 4

- Community patterns fall into two broad categories: those that **cluster** by qualitative assignment , and those that change along a quantitative **gradient**.
- We use **hypothesis tests** to statistically validate community patterns visualized by heatmaps, ordinations, etc.
- We validate community patterning across environmental gradients using correlation of matrices (**Mantel** test) or **vector fitting** to an ordination

# Questions from yesterday?



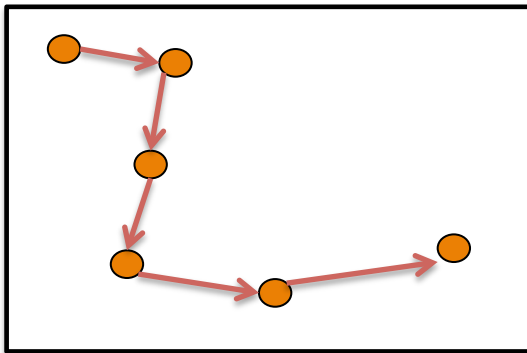
# Overview Day 5

- Asking whether communities have synchronous dynamics, and identifying specific temporal dynamics
- Contributions of individual OTUs to community patterns: taxonomic **Venn** and **hierarchical clustering**
- Rare and prevalent taxa: Species abundance distributions (**SADs**)
- Do rare taxa matter for patterns in beta diversity?  
**MultiCOLA**
- Persistent and transient taxa: Species occurrence distributions (**SODs**)
- General recommendations

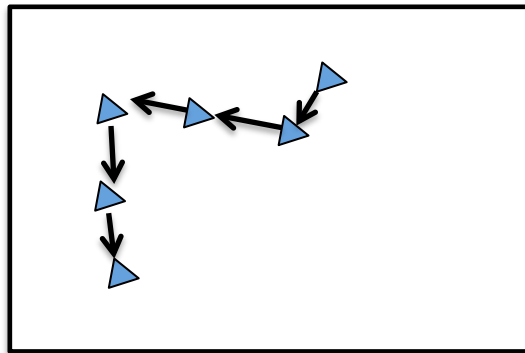
# Do different communities have synchronous (coherent) dynamics?

- PROTEST (Procrustean superimposition)

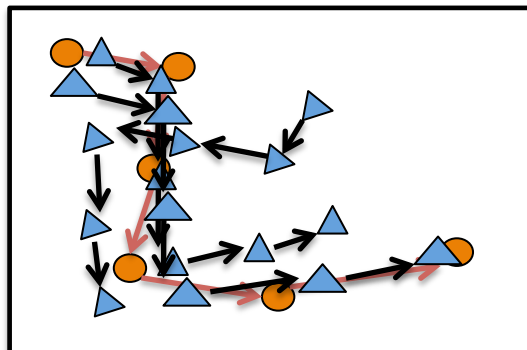
Community A



Community B



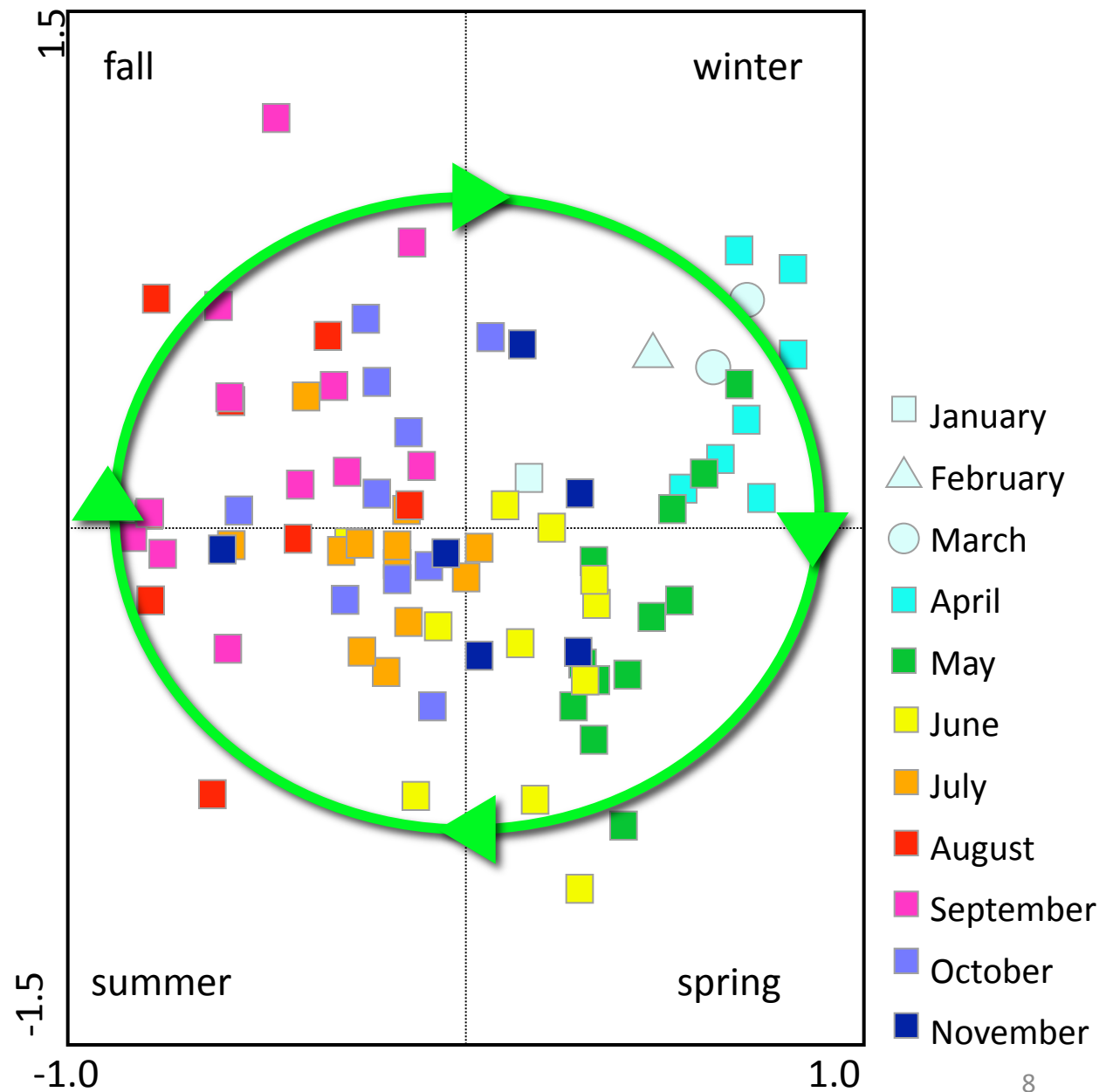
Procrustes rotation



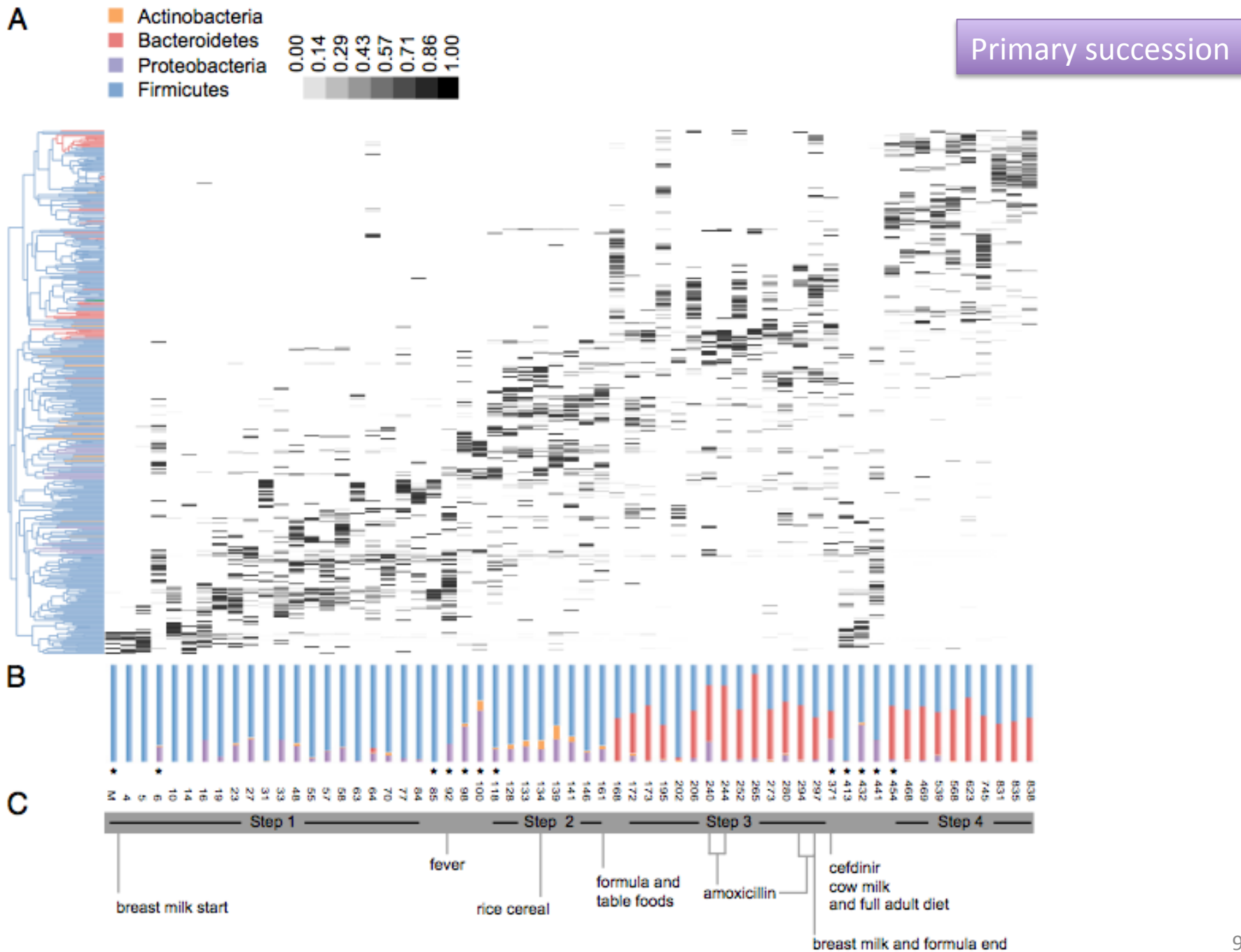
# Do communities exhibit temporal patterns?

- Seasonal
- Primary succession

# Seasonality





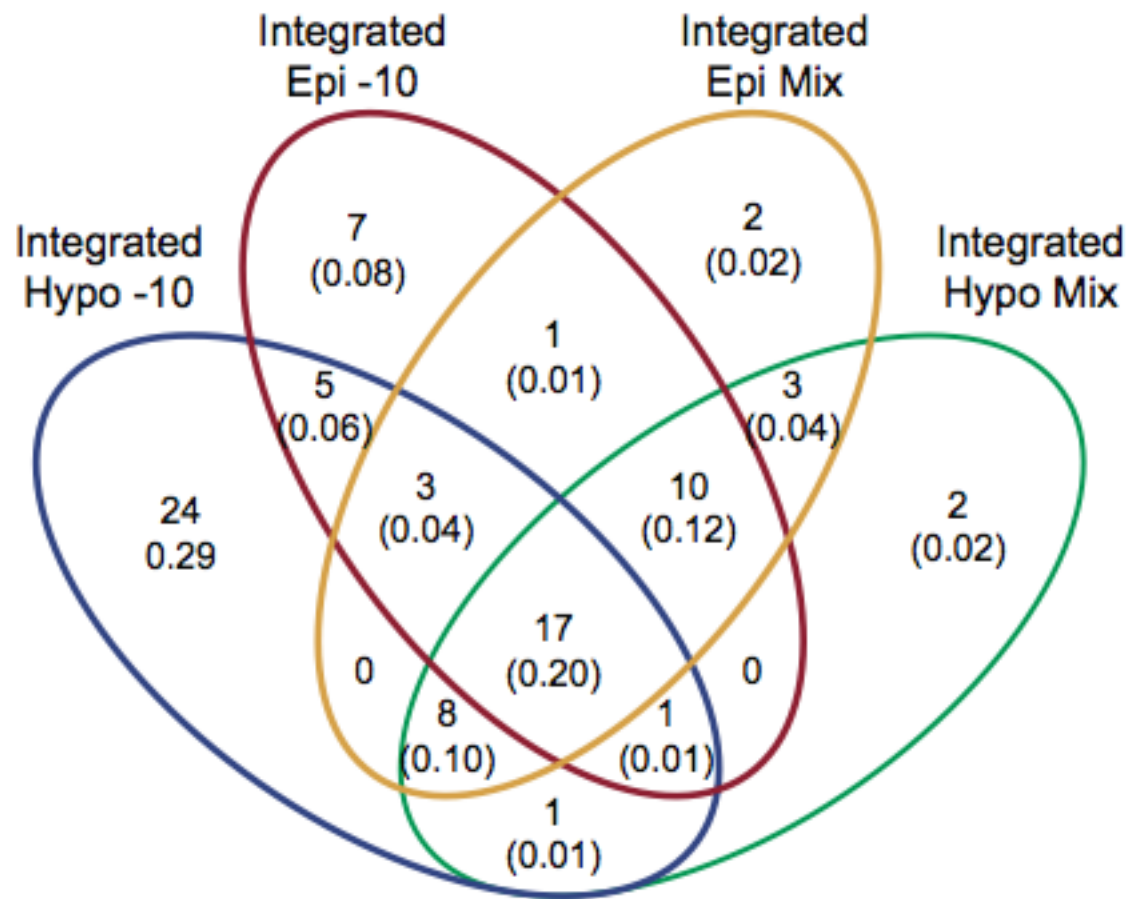


# Contribution of individual OTUs to community patterns

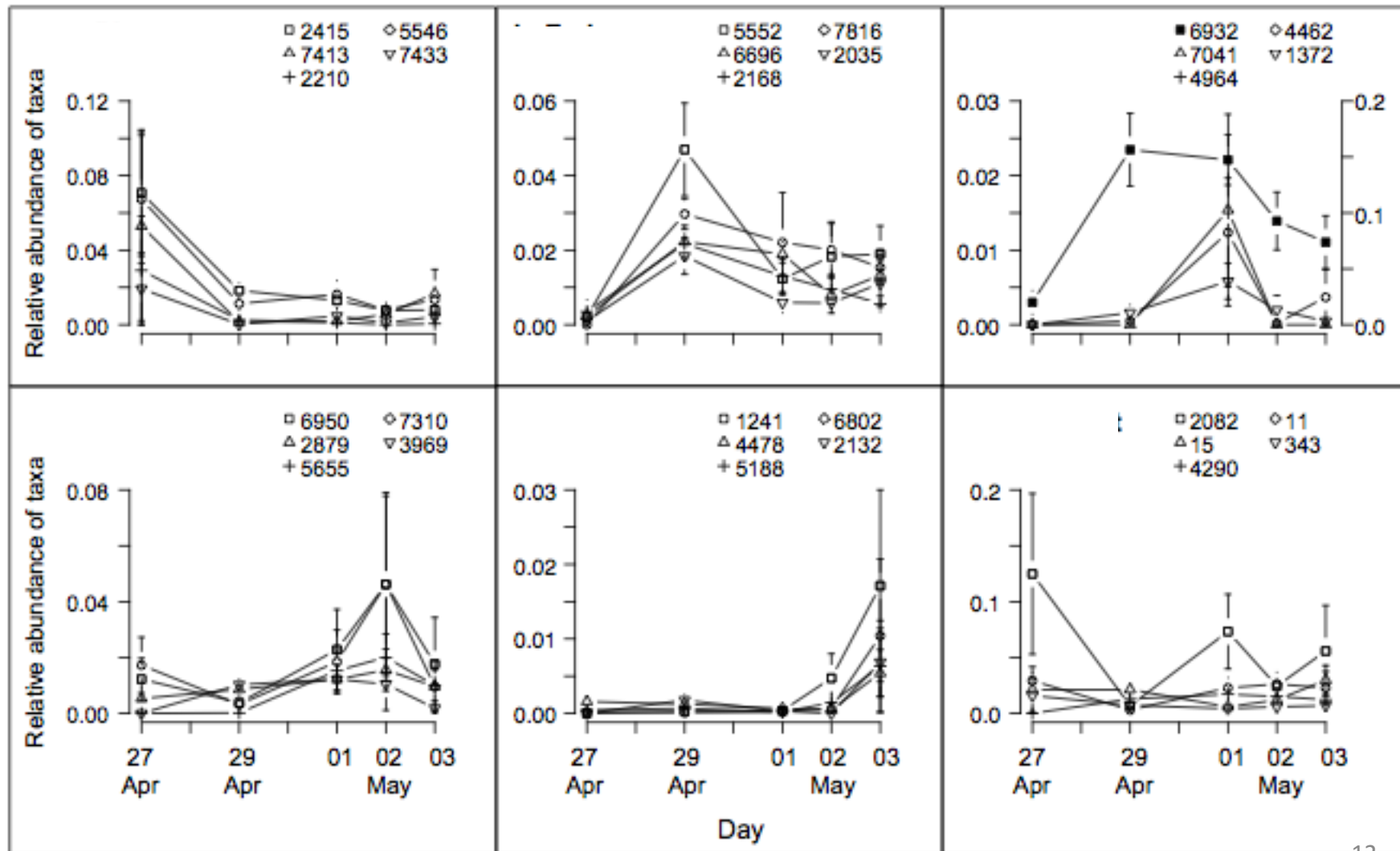
- Find OTUs that are unique to and shared among groups: taxonomic **Venn**
- Determine OTUs that have similar dynamics: hierarchical **clustering**

Find OTUs that are unique to and shared among groups: taxonomic Venn analysis

**a**



Determine OTUs that have similar dynamics:  
(without staring at thousands of line graphs)



# Finding OTUs with similar occurrence patterns : hierarchical clustering

## 1. OTU table

|       | T1   | T2   | T3   | Row sum |
|-------|------|------|------|---------|
| OTU A | 50   | 25   | 10   | 85      |
| OTU B | 500  | 250  | 100  | 850     |
| OTU C | 1    | 2    | 3    | 6       |
| OTU D | 1000 | 2000 | 3000 | 6000    |
| OTU E | 0    | 50   | 0    | 50      |
| OTU F | 25   | 500  | 30   | 555     |

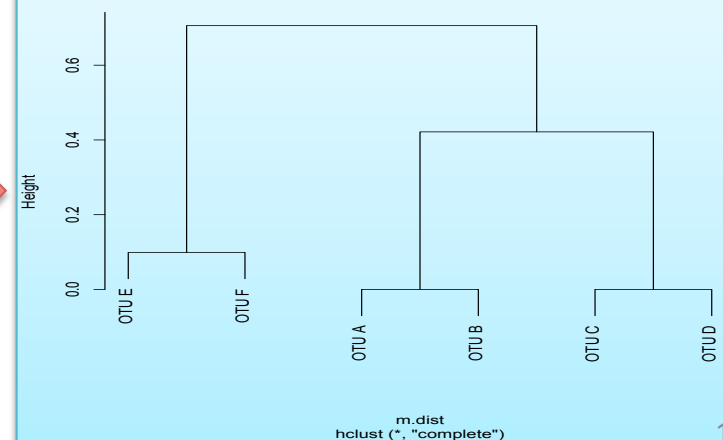
## 2. Standardized OTU table

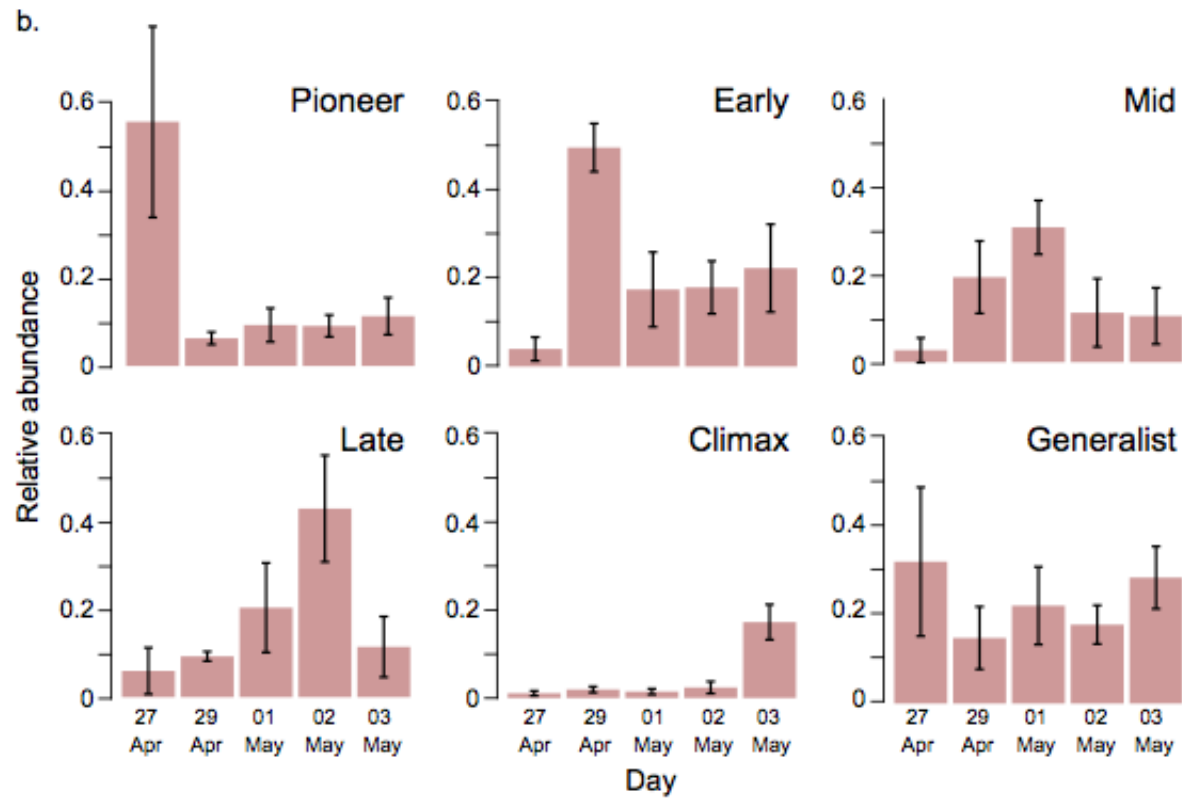
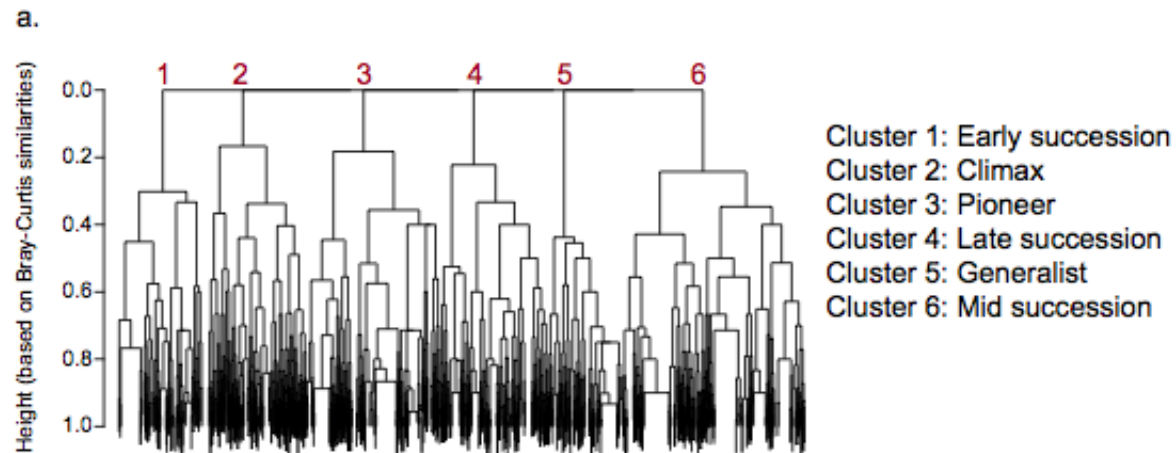
|       | T1   | T2   | T3   | Row sum |
|-------|------|------|------|---------|
| OTU A | 0.59 | 0.29 | 0.12 | 1       |
| OTU B | 0.59 | 0.29 | 0.12 | 1       |
| OTU C | 0.17 | 0.33 | 0.50 | 1       |
| OTU D | 0.17 | 0.33 | 0.50 | 1       |
| OTU E | 0.00 | 1.00 | 0.00 | 1       |
| OTU F | 0.05 | 0.90 | 0.05 | 1       |

## 3. Distance matrix (Bray-Curtis)

|       | OTU A | OTU B | OTU C | OTU D | OTU E |
|-------|-------|-------|-------|-------|-------|
| OTU B | 1.00  |       |       |       |       |
| OTU C | 0.58  | 0.58  |       |       |       |
| OTU D | 0.58  | 0.58  | 1.00  |       |       |
| OTU E | 0.29  | 0.29  | 0.33  | 0.33  |       |
| OTU F | 0.39  | 0.39  | 0.43  | 0.43  | 0.90  |

## 4. Hierarchical clustering

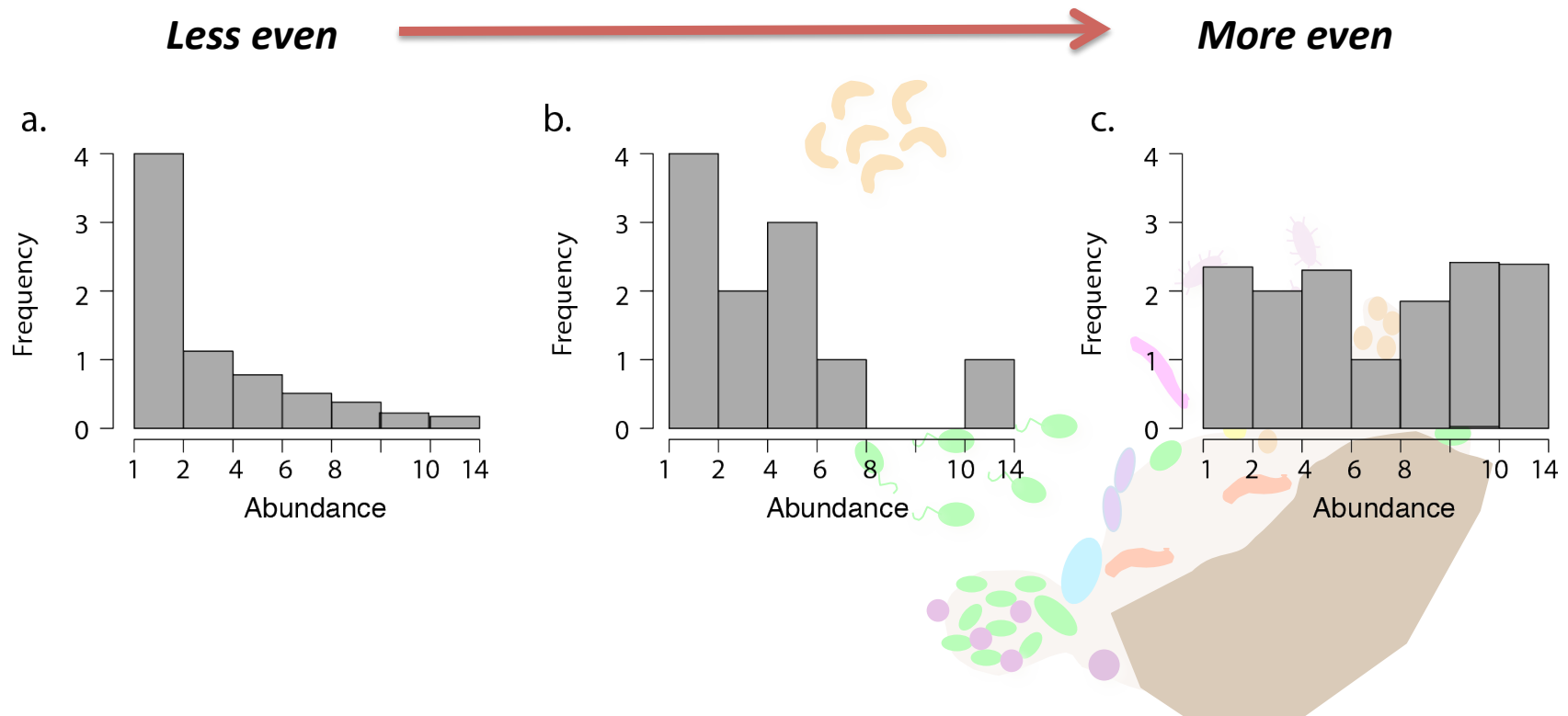




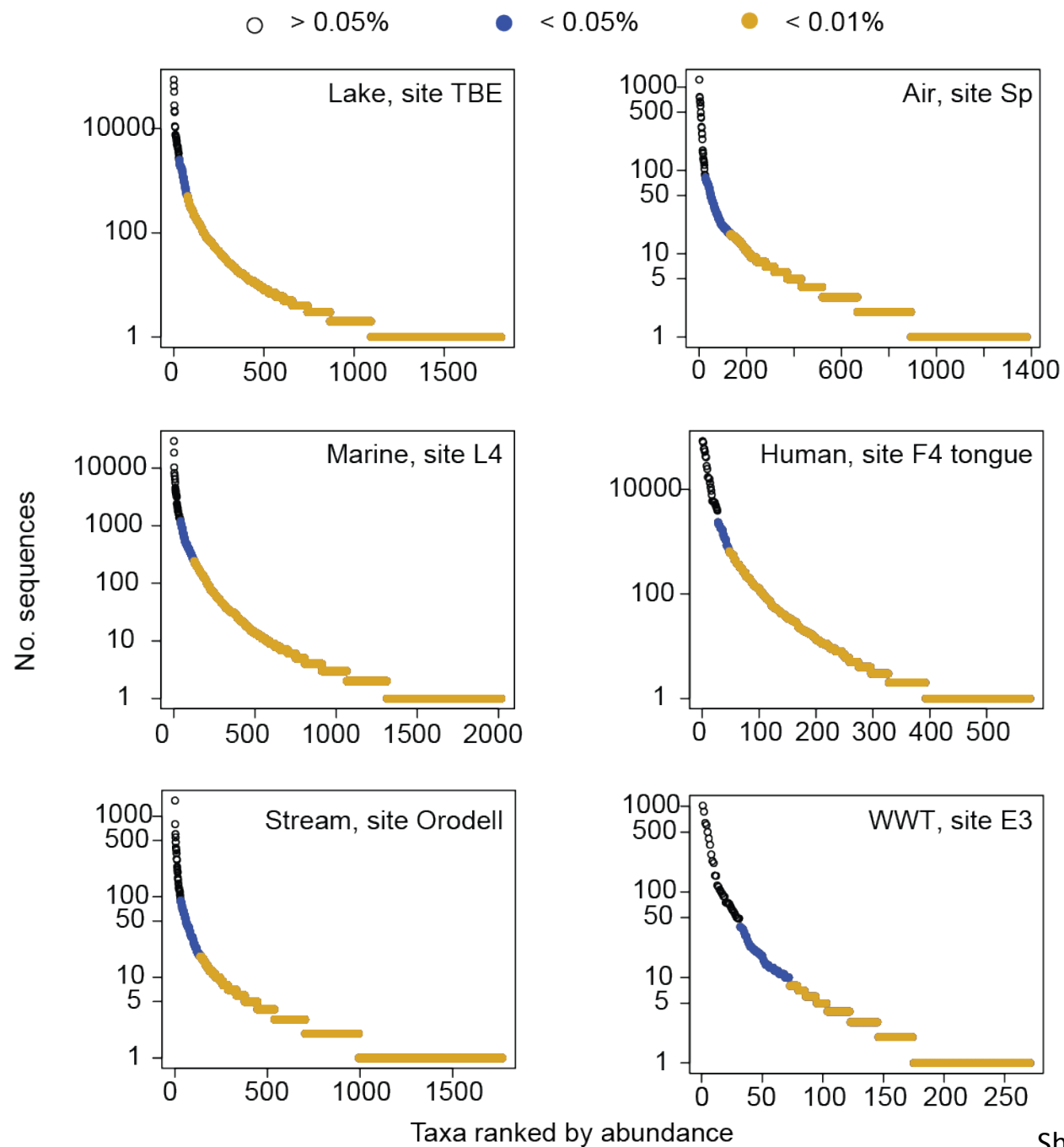
Primary succession on apple flowers

# Alpha diversity

**Evenness:** What is the distribution of abundances in the community?



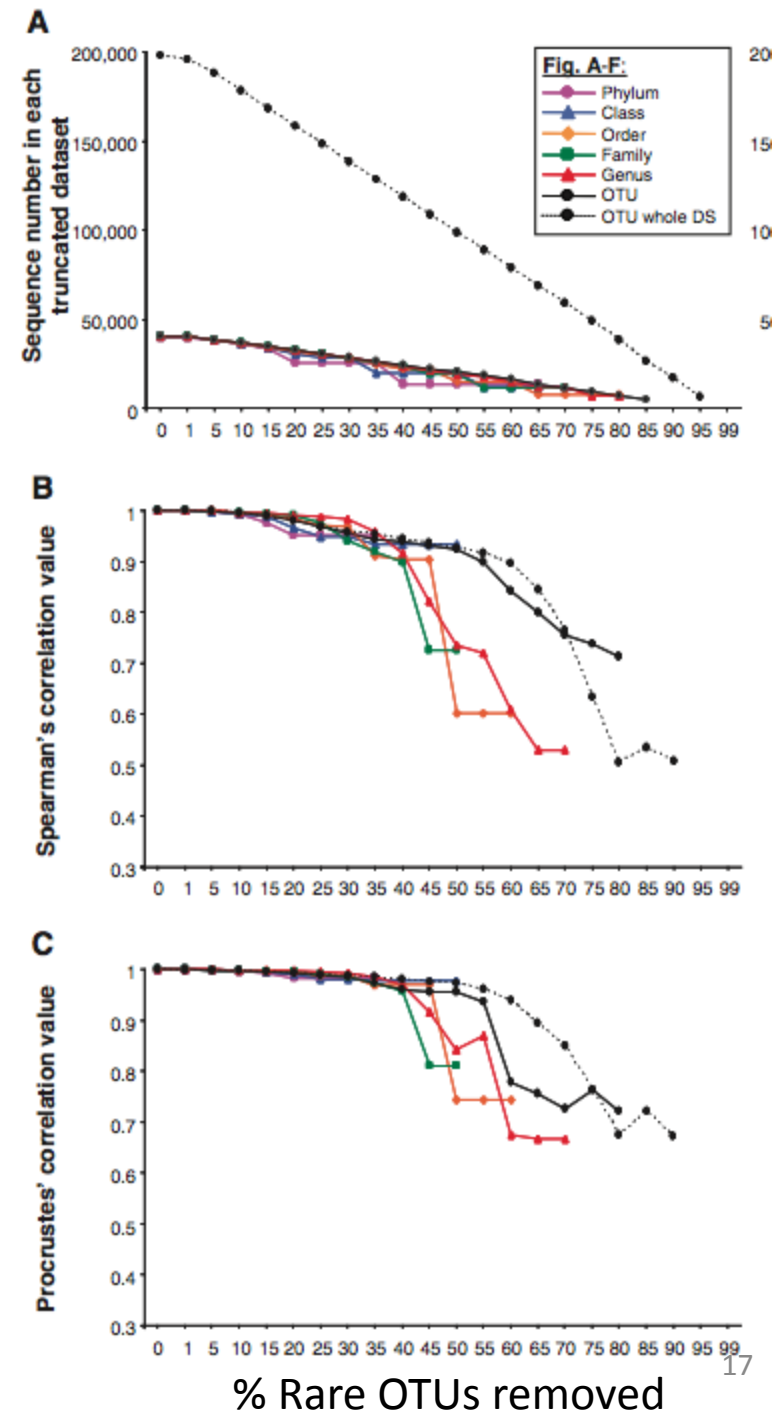
Common metric = Pielou's evenness





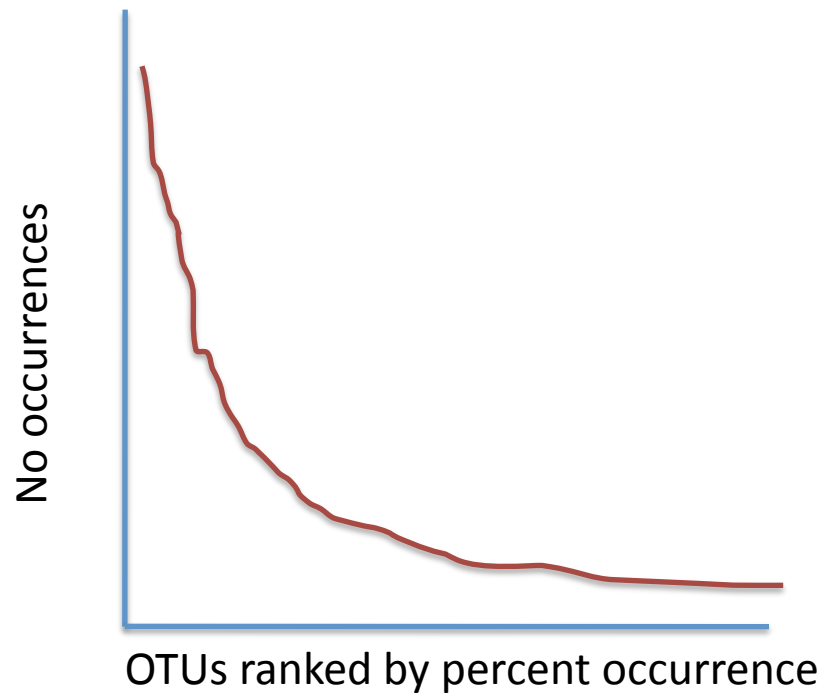
# Do rare taxa matter for beta diversity?

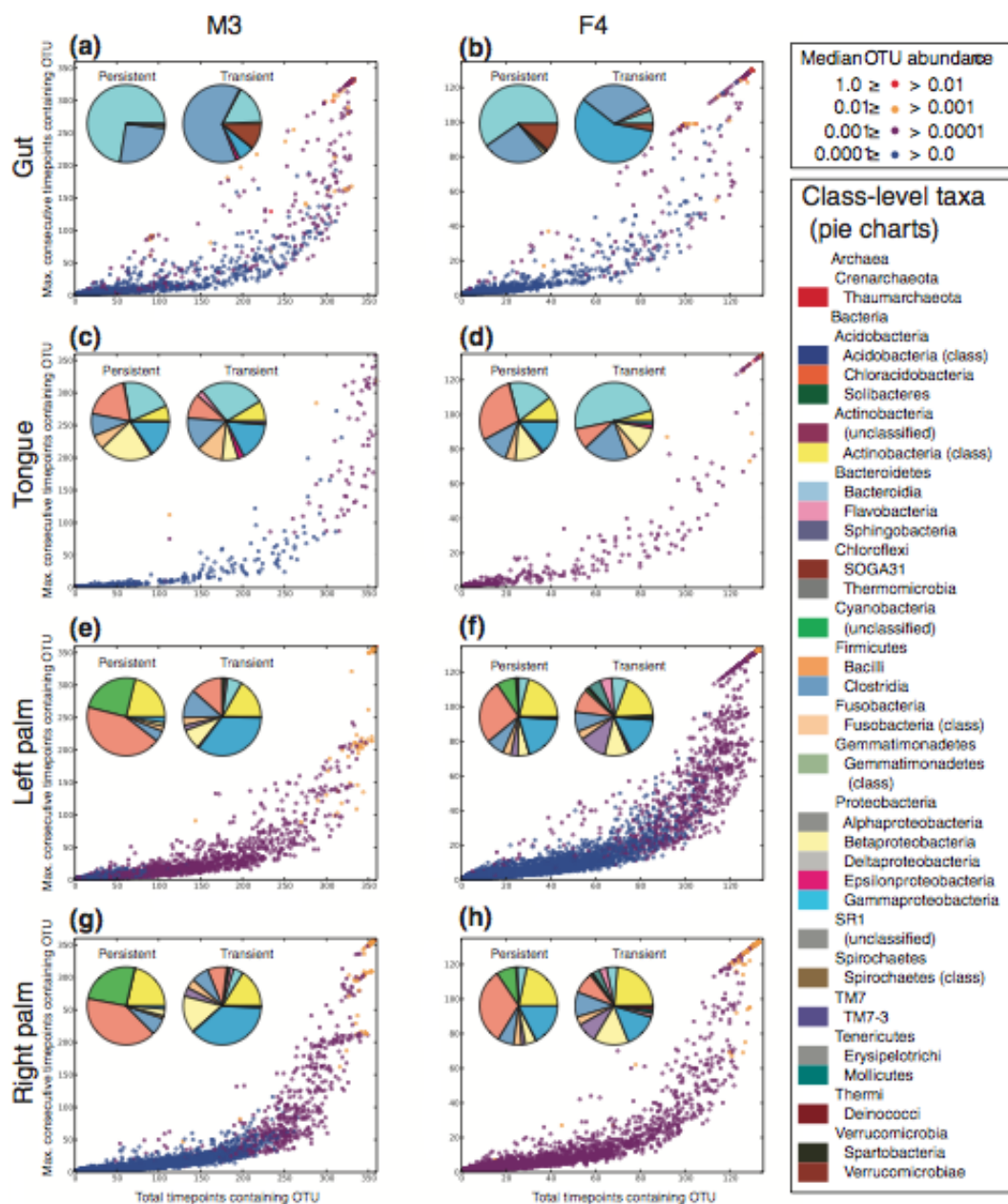
Multivariate cut-off level analysis (MultiCOLA)



# Persistent and transient taxa

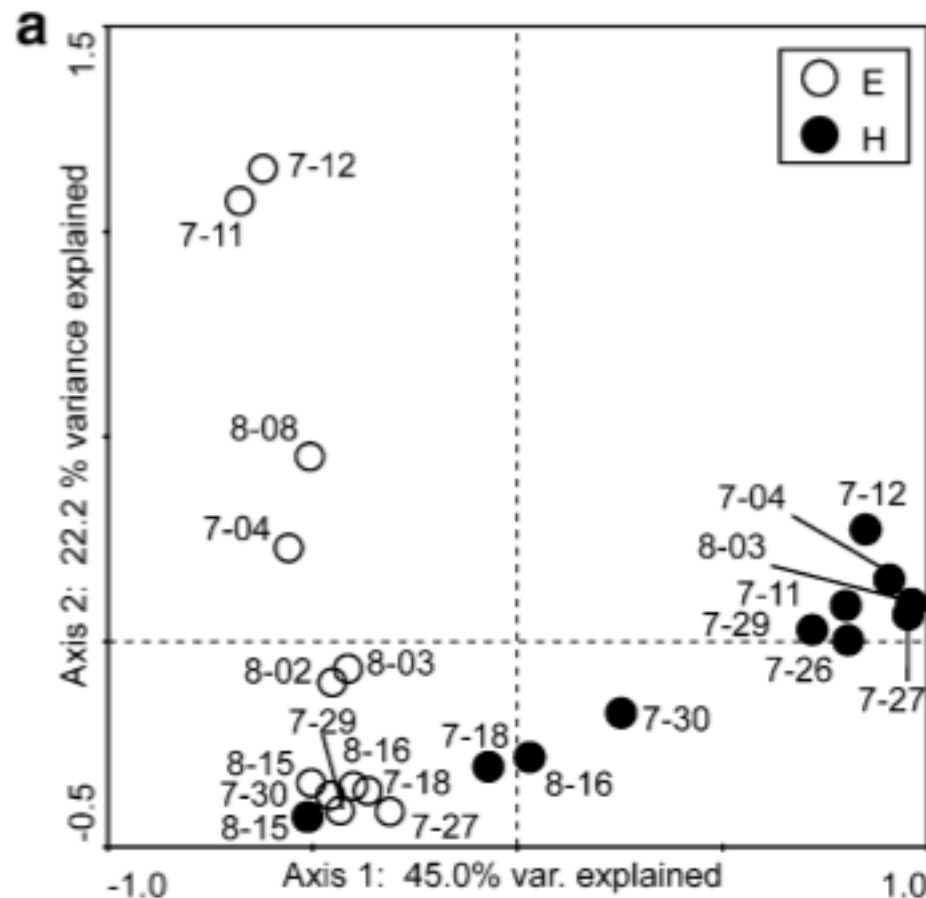
- Species occurrence distributions

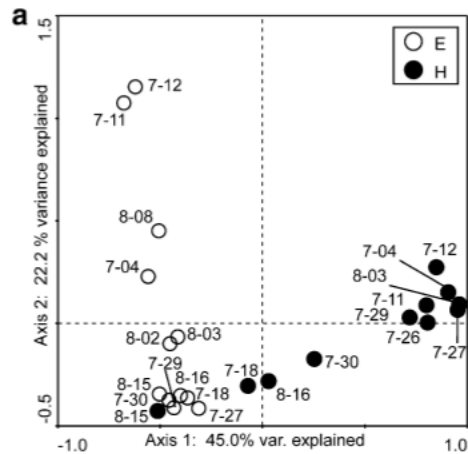




**Figure 3 Community membership.** Community membership summary for all OTUs in (a) M3 gut, (b) F4 gut, (c) M3 tongue, (d) F4 tongue, (e) M3 left palm, (f) F4 left palm, (g) M3 right palm, and (h) F4 right palm. Points are OTUs colored by their median relative abundance computed over all samples where they occur, and pie charts summarize the class-level taxa observed as persistent and transient OTUs.

# Rare & prevalent, persistent & transient





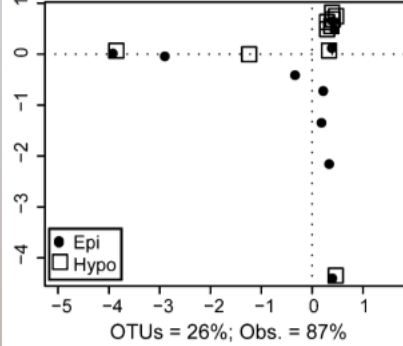
**Transient**

**Rare**

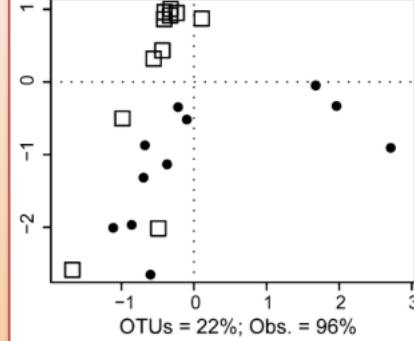
**Common**

**Prevalent**

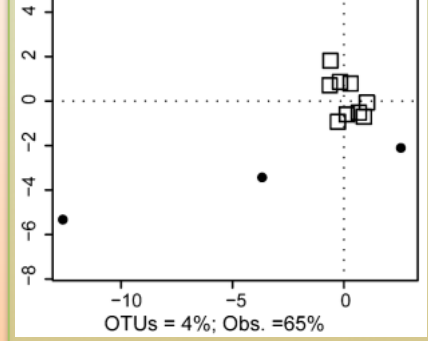
**A. RT06**



**B. CT06**



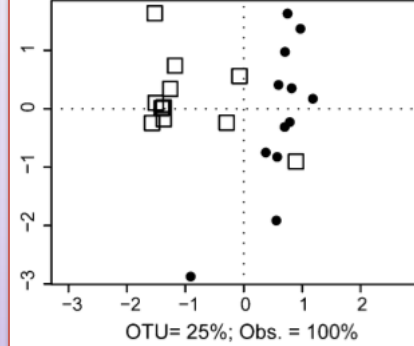
**D. DT06**



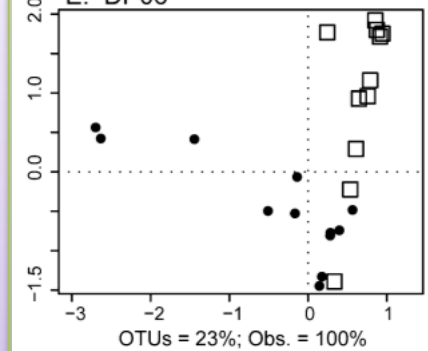
**Persistent**

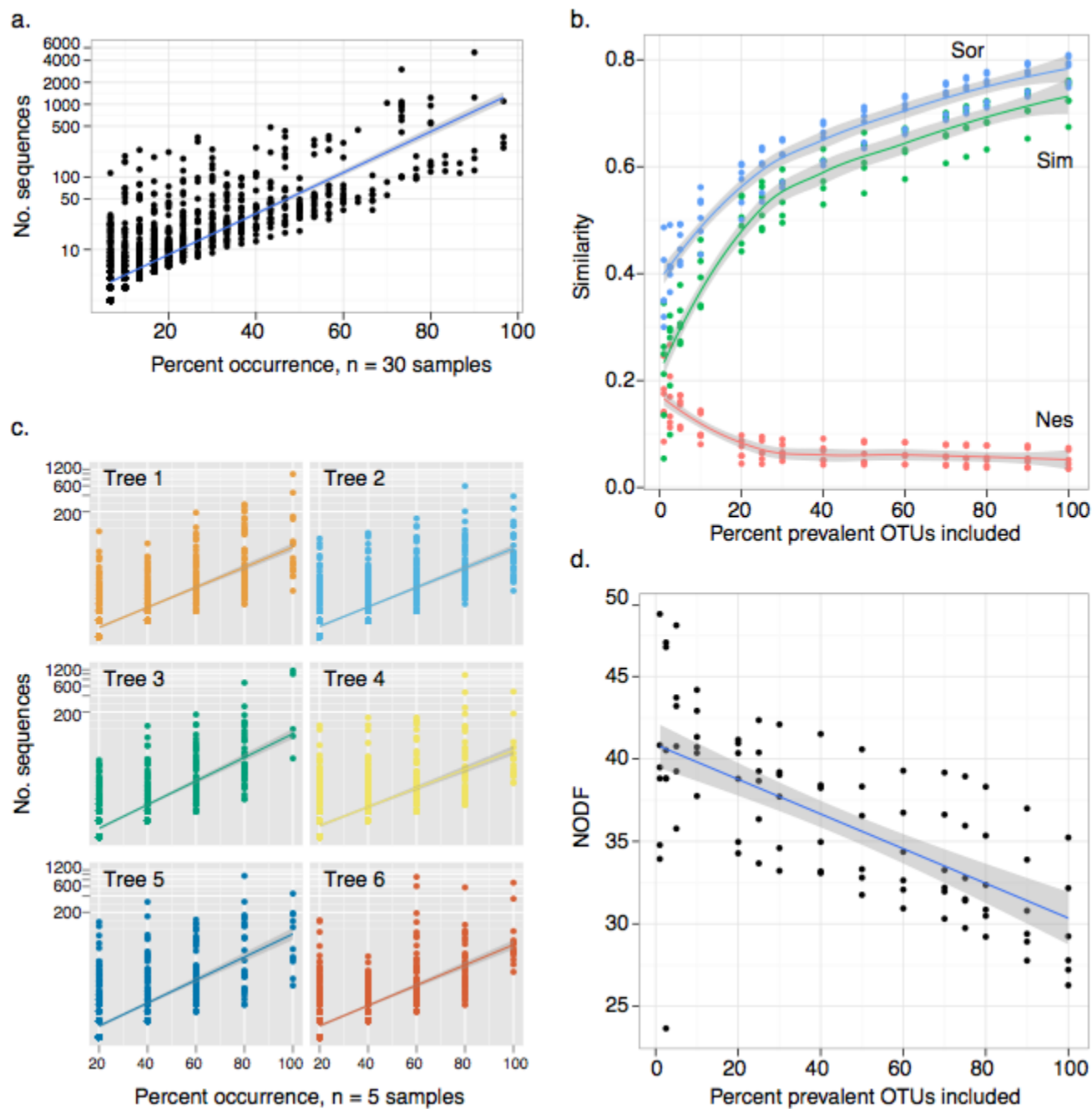
**NA**

**C. CP06**



**E. DP06**





# General recommendations

- Think first, generate gobs of data second.
  - Think the whole way through the experiment, starting with the hypothesis, the design, the data processing, and the statistical analyses. Don't stop at the experimental design.
  - Make sure you'll be able to answer your research question.
  - Make sure you have the expertise in the lab or with collaborators to handle the data properly, and if not, seek out the expertise or get training
  - Otherwise, you may end up with gobs of expensive data stored on hard drive for an indefinite period, waiting for someone to come along and analyze it. (By that time, the sequencing technology may be obsolete and the work may not be as impactful).
  - **Analyzing high-throughput data is not something that can be done well in an afternoon. It takes time, strategy, perseverance, and commitment.**

# More recommendations

- Measure/observe as many quantitative variables as possible.
  - It is hard to publish data that do not support a hypothesis – measuring a lot of variables gives provides “back-up” for exploration in the event that a hypothesis is not supported.
- Get familiar with finding help on the interwebs!
  - Lots of blogs, discussion groups, help forums for R (vegan) and QIIME and many other analysis tools
  - Subscribe, get tips, and start to navigate self-help and troubleshooting. It is tricky at first because it is not intuitive what the search terms should be, but this get easier



# Tutorial

Get R started

Open the [Day 5 tutorial + the R scripts in R studio](#)

Open a directory browser and navigate to the files we'll be analyzing

Get started if you'd like to work independently, follow along if you'd like to work with the group