

Balance Trees Reveal Microbial Niche Differentiation

James T. Morton, Jon Sanders, Robert A. Quinn, Daniel McDonald, Antonio Gonzalez, Yoshiki Vázquez-Baeza, Jose A. Navas-Molina, Se Jin Song, Jessica L. Metcalf, Embriette R. Hyde, Manuel Lladser, Pieter C. Dorrestein, Rob Knight



Presentation and slides by:

Yotam Nitzan

yotam.nitzan@gmail.com

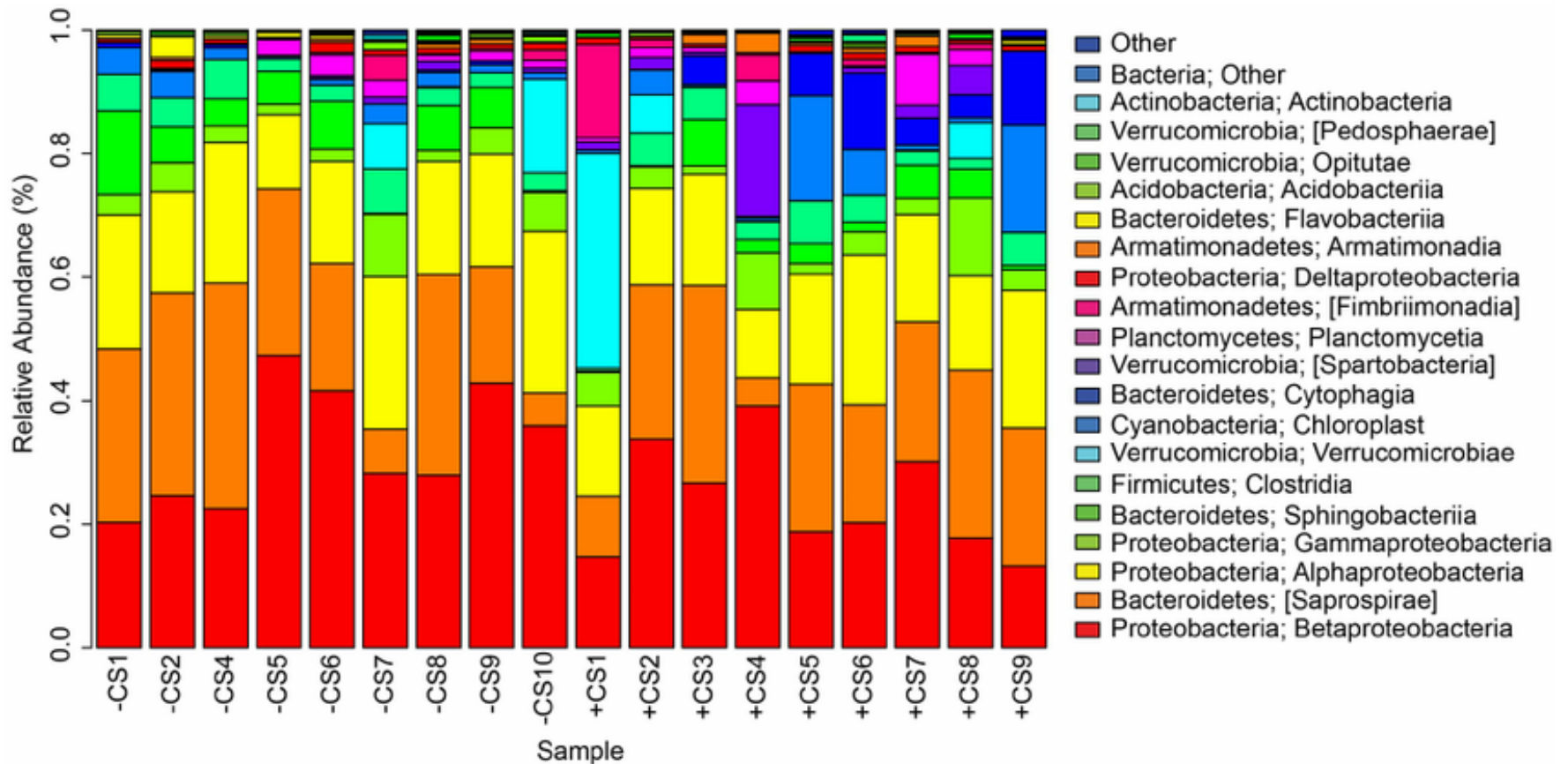
Outline

- Motivation
- Definitions
- Method
 - Balances
 - Balance Trees
 - ILR
- Results
 - Case study 1
 - Case study 2
- Implementation
- Summary

Outline

Compositional data

Represented by relative abundance and carry no meaning for absolute abundance of specific feature.



Outline

- **Motivation**
- Definitions
- Method
 - Balances
 - Balance Trees
 - ILR
- Results
 - Case study 1
 - Case study 2
- Implementation
- Summary

The problem with compositional data



But what happened ?

The problem with compositional data

All abundancies in a sample are dependent !

- It's impossible to certainly know if a growth or decline of a specie actually occurred.
- Analyses that assumes independence of variables, as many statistical approaches do, are prone to misinterpretation.
- Problematic to detect differently abundant microbes across environments.

- **Motivation**
- Definitions
- Method
 - Balances
 - Balance Trees
 - ILR
- Results
 - Case study 1
 - Case study 2
- Implementation
- Summary

Definitions

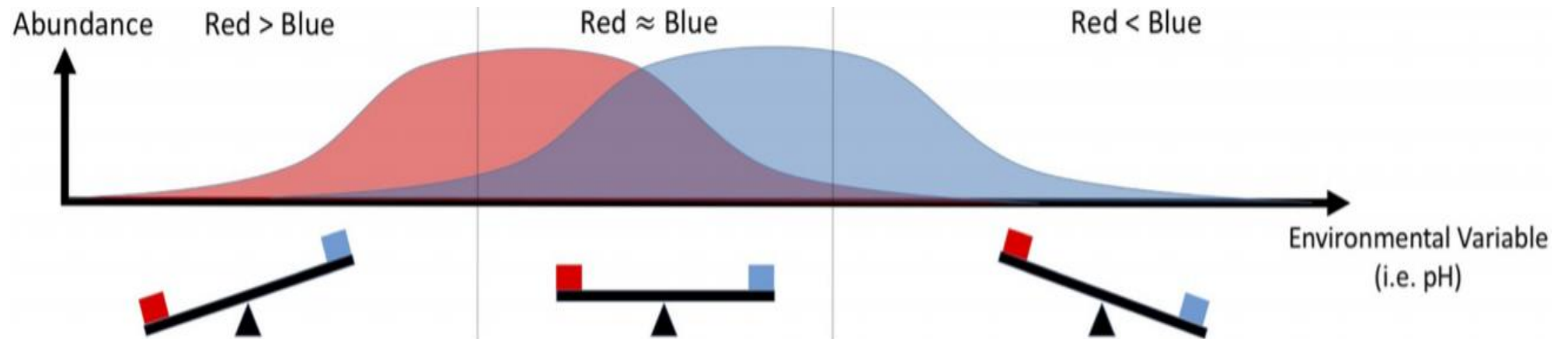
- Motivation
- **Definitions**
- Method
 - Balances
 - Balance Trees
 - ILR
- Results
 - Case study 1
 - Case study 2
- Implementation
- Summary

- Niche - the match of a specie to specific environmental condition, such as habitat, resources, competitors and more.
- Niche differentiation - the process by which competing species use the environment differently in a way that helps them to coexist.
- pH – scale from 0 to 14, used to specify how acidic or basic a water-based solution is.

Outline

Balances

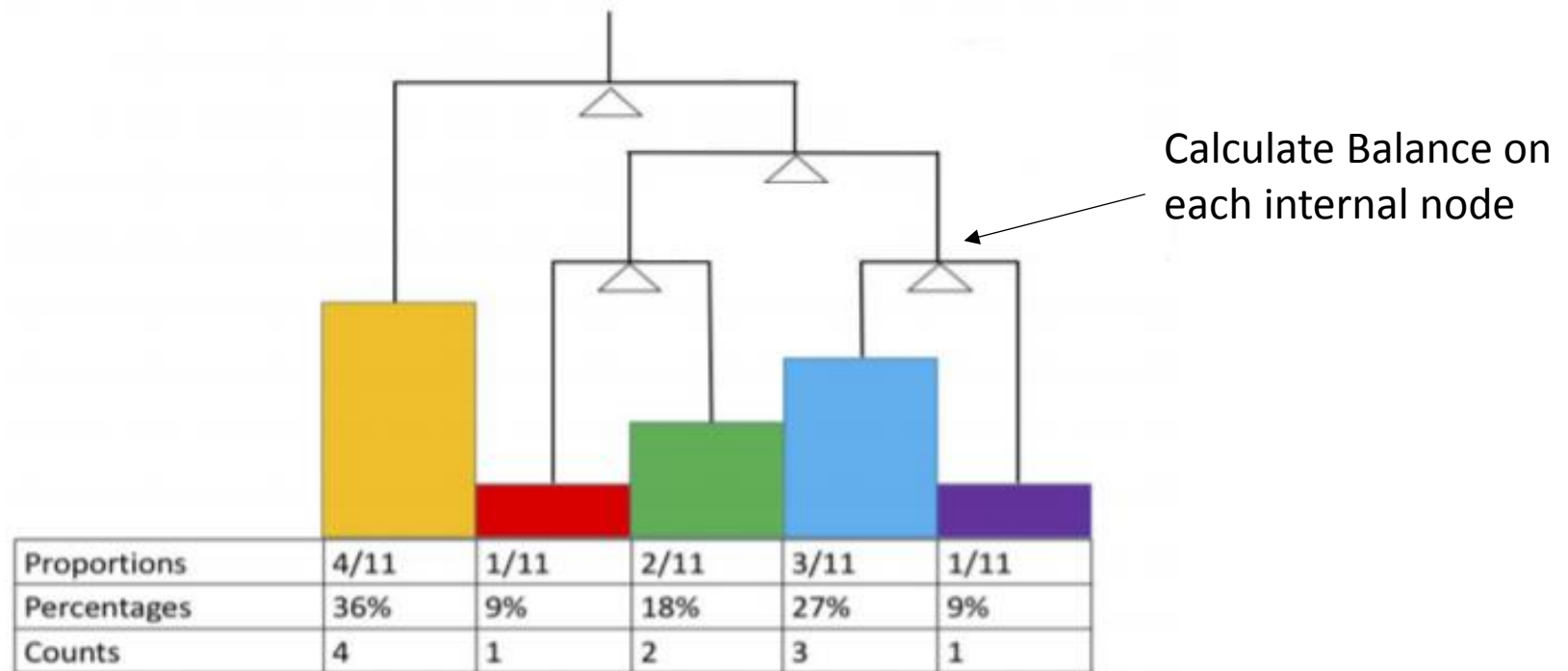
- Motivation
 - Definitions
 - **Method**
 - **Balances**
 - Balance Trees
 - ILR
 - Results
 - Case study 1
 - Case study 2
 - Implementation
 - Summary
- Avoid attempts to infer absolute changes in abundance.
 - Instead, Infer change in balances of abundances between two particular subsets of the community.



Outline

Balance Trees

- Motivation
- Definitions
- **Method**
 - Balances
 - **Balance Trees**
 - ILR
- Results
 - Case study 1
 - Case study 2
- Implementation
- Summary



Balance Trees

The tree could be built using any criterion, depending on the question in hand.

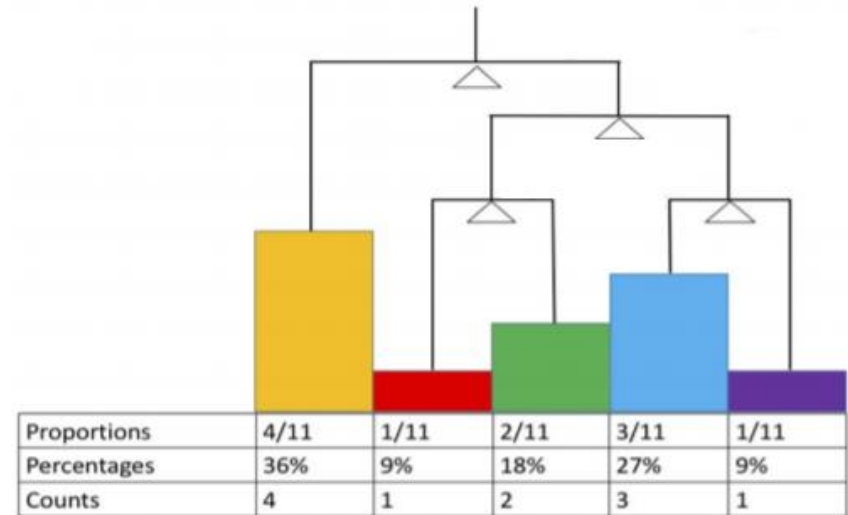
- A phylogenetic tree could be used to investigate evolutionary relationships.
- A hierarchal clustering of environmental variables could be used to explore environmental niches of microbes.
- You name it !

- Motivation
- Definitions
- **Method**
 - Balances
 - **Balance Trees**
 - ILR
- Results
 - Case study 1
 - Case study 2
- Implementation
- Summary

Balance function

- We'll use isometric log ratio (ILR) as the balance function.

$$b_i = \sqrt{\frac{|i_L||i_R|}{|i_L| + |i_R|}} \log \left[\frac{g(i_L)}{g(i_R)} \right]$$



* Add pseudocount of 1 to all counts to mitigate the problem of zero-counts.

- Motivation
- Definitions
- **Method**
 - Balances
 - Balance Trees
 - **ILR**
- Results
 - Case study 1
 - Case study 2
- Implementation
- Summary

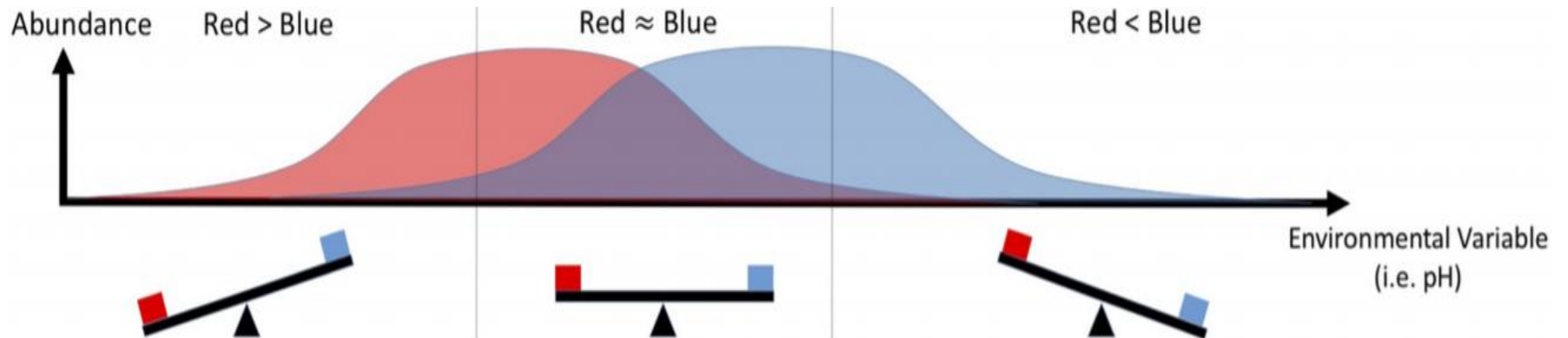
Outline

Balance function

- We'll use isometric log ratio (ILR) as the balance function.

$$b_i = \sqrt{\frac{|i_L||i_R|}{|i_L| + |i_R|}} \log \left[\frac{g(i_L)}{g(i_R)} \right]$$

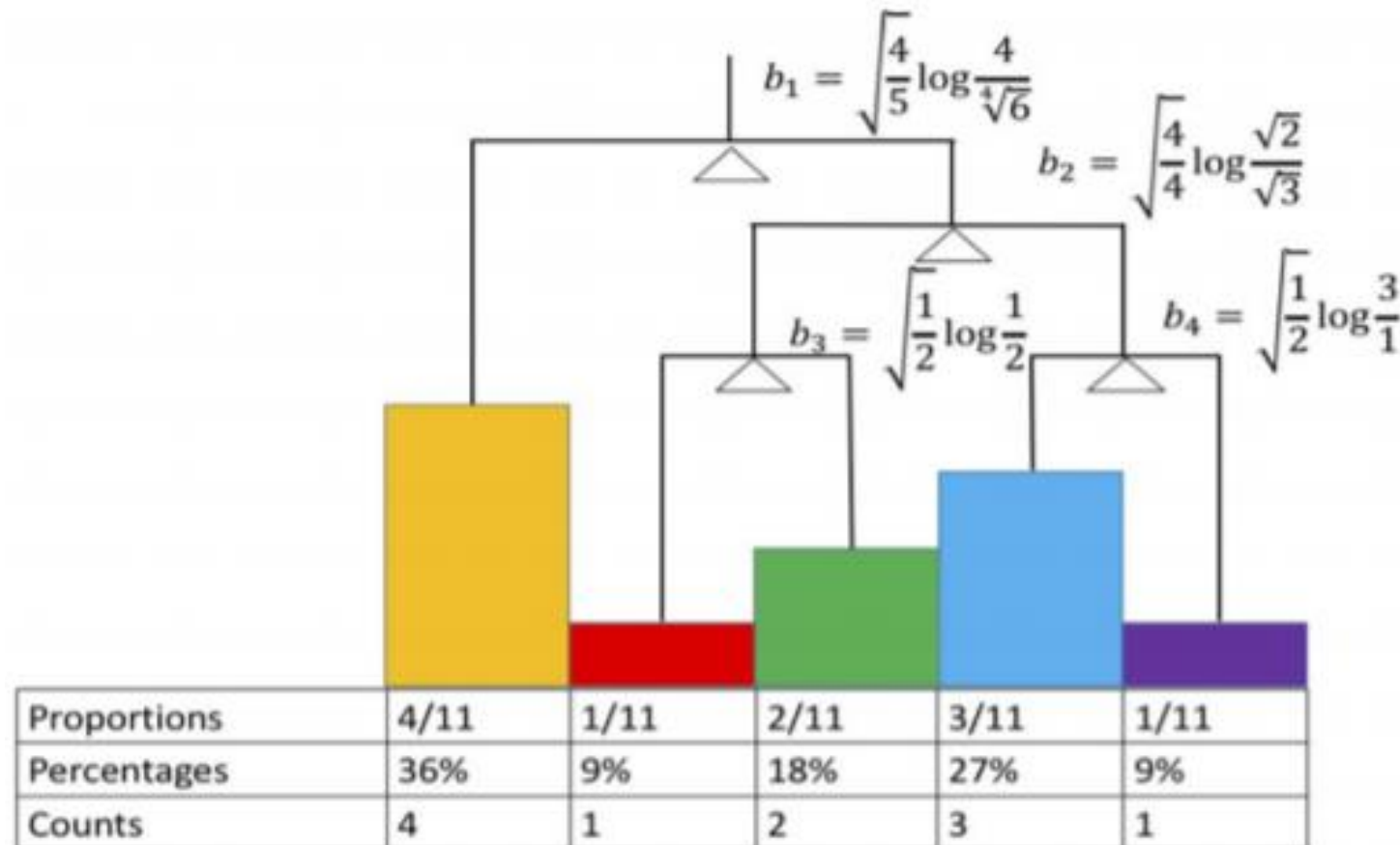
- Back to our 2D toy example:



ILR Example

$$b_i = \sqrt{\frac{|i_L||i_R|}{|i_L| + |i_R|}} \log \left[\frac{g(i_L)}{g(i_R)} \right]$$

- Motivation
- Definitions
- **Method**
 - Balances
 - Balance Trees
 - **ILR**
- Results
 - Case study 1
 - Case study 2
- Implementation
- Summary

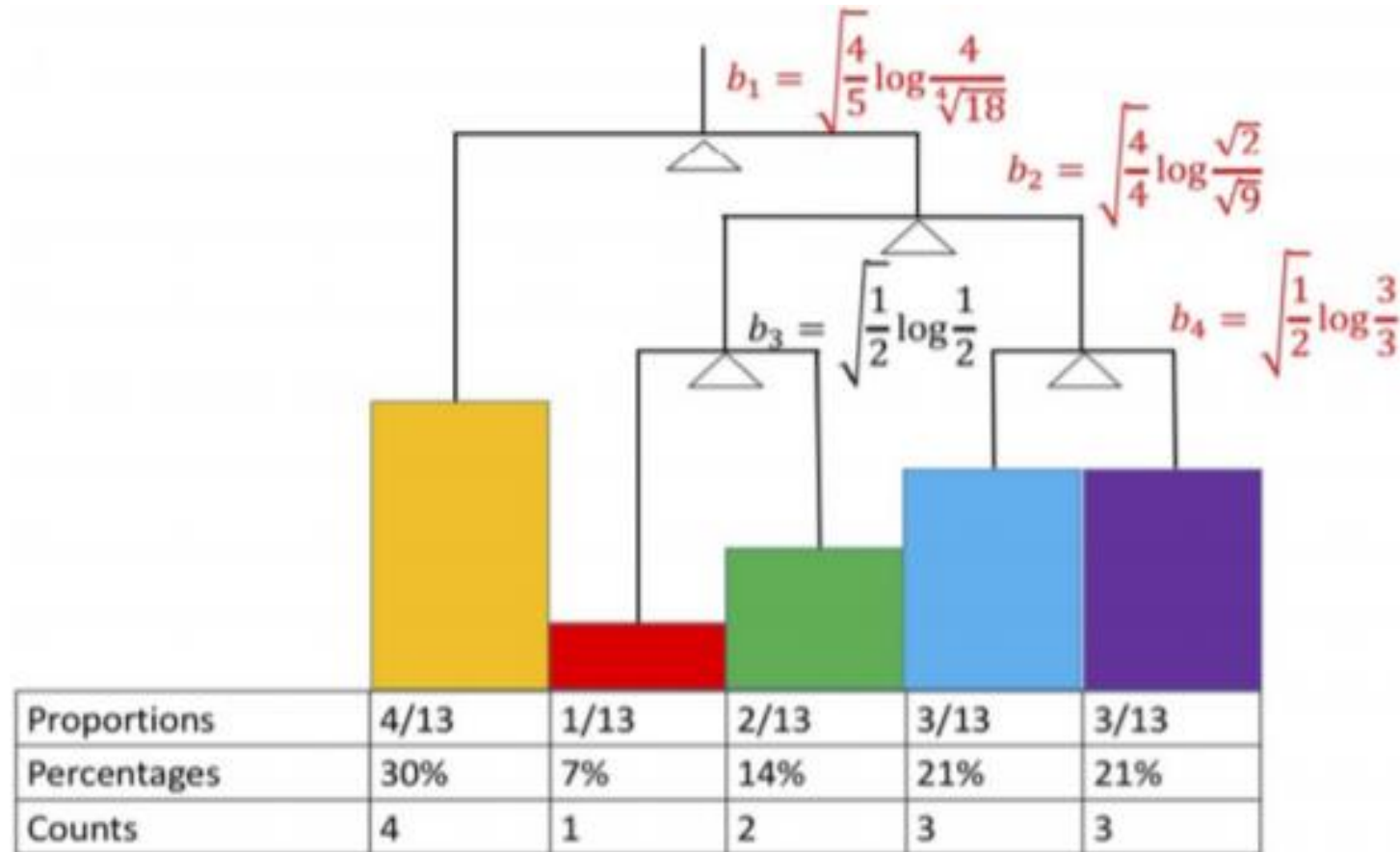


Outline

ILR Example

$$b_i = \sqrt{\frac{|i_L||i_R|}{|i_L| + |i_R|}} \log \left[\frac{g(i_L)}{g(i_R)} \right]$$

- Motivation
- Definitions
- **Method**
 - Balances
 - Balance Trees
 - **ILR**
- Results
 - Case study 1
 - Case study 2
- Implementation
- Summary



Outline

- Motivation
- Definitions
- **Method**
 - Balances
 - Balance Trees
 - ILR
- Results
 - Case study 1
 - Case study 2
- Implementation
- Summary

Some benefits to note

- Scale invariant.
- Subcompositionally coherent, which means that changes in nonoverlapping subcommunities do not impact each other.
- Permutation invariant.

Outline

- Motivation
- Definitions
- Method
 - Balances
 - Balance Trees
 - ILR
- **Results**
 - Case study 1
 - Case study 2
- Implementation
- Summary

Results

- The authors did not benchmark their method against other “compositional aware tools”.
- As for results, the authors demonstrate novel insight realized using balance trees on two existing papers.

Introduction – Case study 1

- Motivation
- Definitions
- Method
 - Balances
 - Balance Trees
 - ILR
- **Results**
 - **Case study 1**
 - Case study 2
- Implementation
- Summary

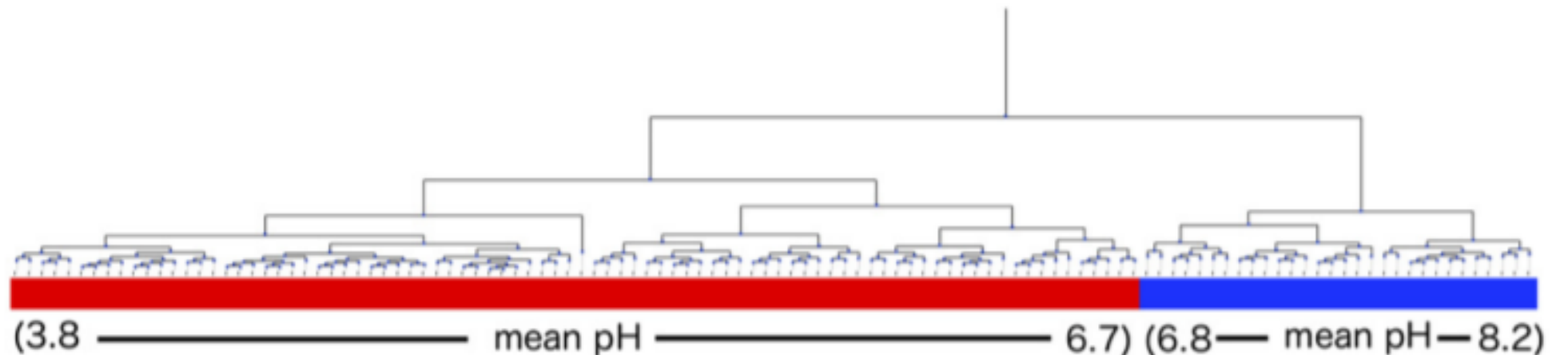
- 88 soil samples collected along with many measurements.
- The study reported the existence or inexistence of a few bacterial phyla to pH levels.
 - Acidobacteria negatively correlated with pH
 - Actinobacteria, Bacteroidetes correlated with pH
 - Alpha-, Beta- and Gammaproteobacteria not correlated with pH
- **But**, each phylum was correlated with pH independently. That's the misinterpretation we feared.

Paper: Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale

Outline

Balance Tree to the rescue !

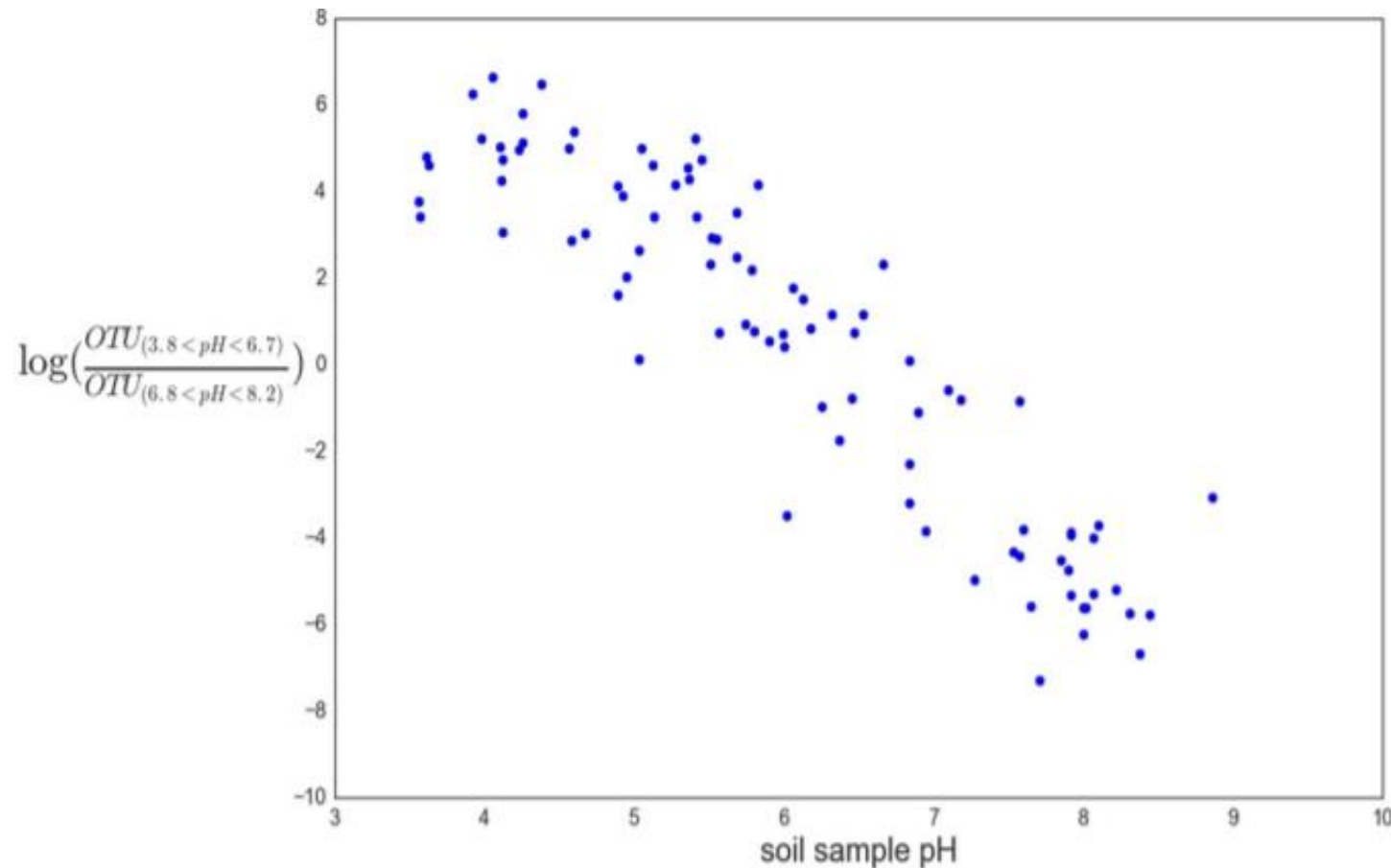
- Motivation
 - Definitions
 - Method
 - Balances
 - Balance Trees
 - ILR
 - **Results**
 - **Case study 1**
 - Case study 2
 - Implementation
 - Summary
- Order OTUs by mean pH in samples.
 - Construct a tree by hierarchal clustering.



Outline

- Motivation
- Definitions
- Method
 - Balances
 - Balance Trees
 - ILR
- **Results**
 - **Case study 1**
 - Case study 2
- Implementation
- Summary

The tree's root balance



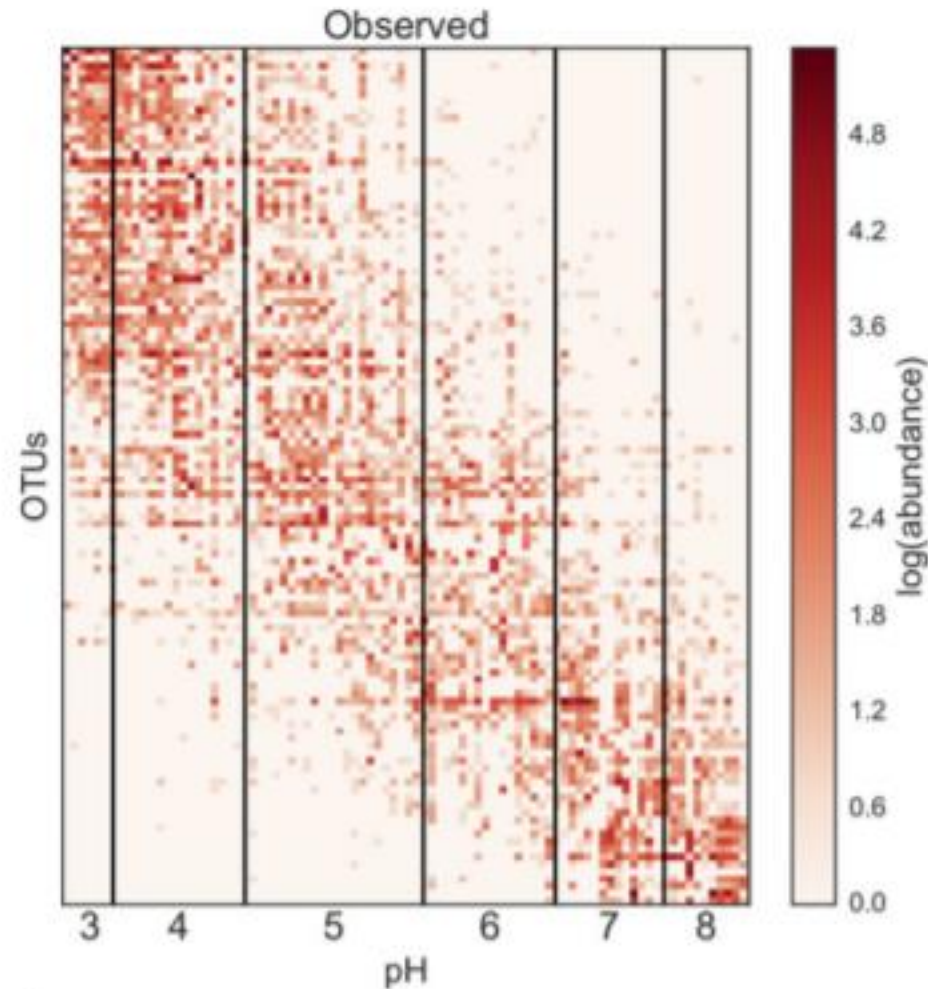
We observe a nice linear trend !

Outline

- Motivation
- Definitions
- Method
 - Balances
 - Balance Trees
 - ILR
- **Results**
 - **Case study 1**
 - Case study 2
- Implementation
- Summary

Abundance “heatmap”

OTUs are sorted
by mean pH



- Motivation
- Definitions
- Method
 - Balances
 - Balance Trees
 - ILR
- **Results**
 - **Case study 1**
 - Case study 2
- Implementation
- Summary

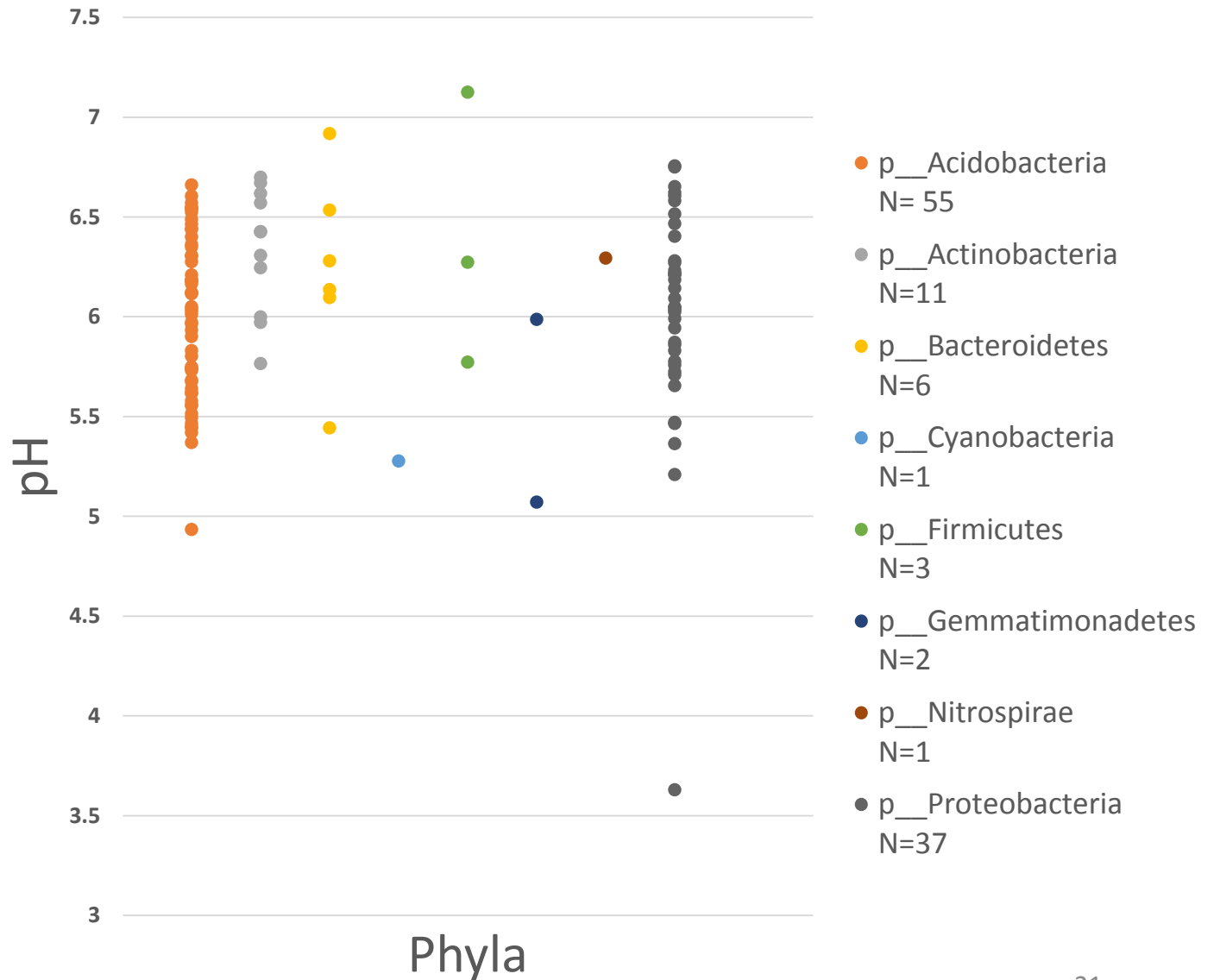
Conclusions

- OTUs tend to be observed in very specific pH ranges.
- This ties to concept from niche theory: OTUs are suited to live in a specific environment, and when placed outside of it are outcompeted.
- These patterns are missed if looked only at phyla level. The phyla mentioned in the original paper were distributed all over the pH.

Outline

- Motivation
- Definitions
- Method
 - Balances
 - Balance Trees
 - ILR
- **Results**
 - **Case study 1**
 - Case study 2
- Implementation
- Summary

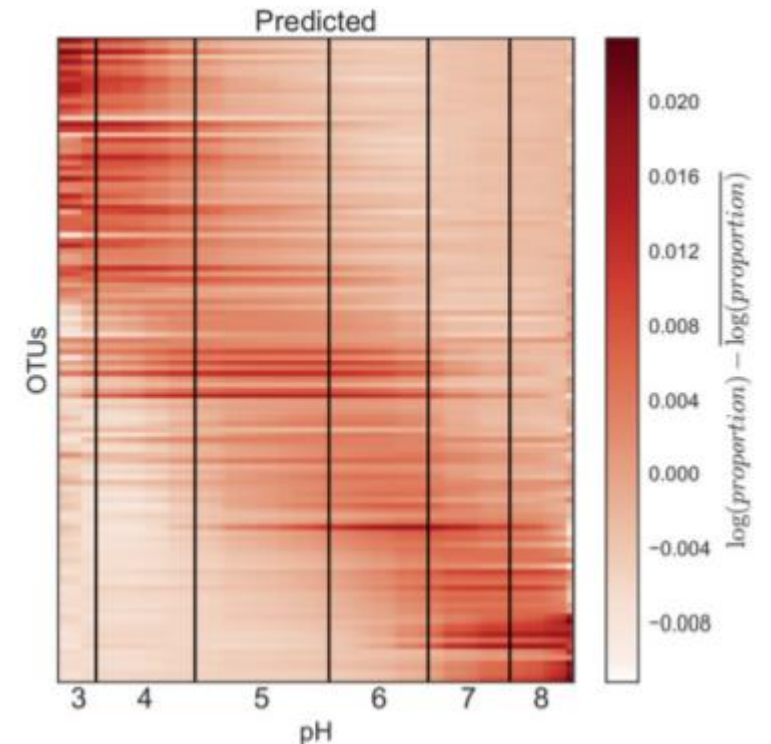
Conclusions



Predictive model

- Motivation
- Definitions
- Method
 - Balances
 - Balance Trees
 - ILR
- **Results**
 - **Case study 1**
 - Case study 2
- Implementation
- Summary

- Next, we'll build a predictive model that will predict a microbial community based on pH levels.
 - Fit a quartic to each balance (node) with Least Squares.
 - Map predicted balances back to proportions with inverse ILR
 - The overall trend is predicted
 - $R^2 = 0.35$



Introduction – Case study 2

- Motivation
- Definitions
- Method
 - Balances
 - Balance Trees
 - ILR
- **Results**
 - Case study 1
 - **Case study 2**
- Implementation
- Summary

- Lung sputum samples collected from 16 cystic fibrosis (CF) patients. Samples were placed into separate tubes and pH was adjusted from 5 to 8.5 at intervals of 0.5
- It was observed that patients had different microbiomes
- It was hypothesized that there were two subcommunities, one thrives at high pH and one at low pH.

Paper: A Winogradsky-based culture system shows an association between microbial fermentation and cystic fibrosis exacerbation.

- Motivation
- Definitions
- Method
 - Balances
 - Balance Trees
 - ILR
- **Results**
 - Case study 1
 - **Case study 2**
- Implementation
- Summary

Introduction – Case study 2

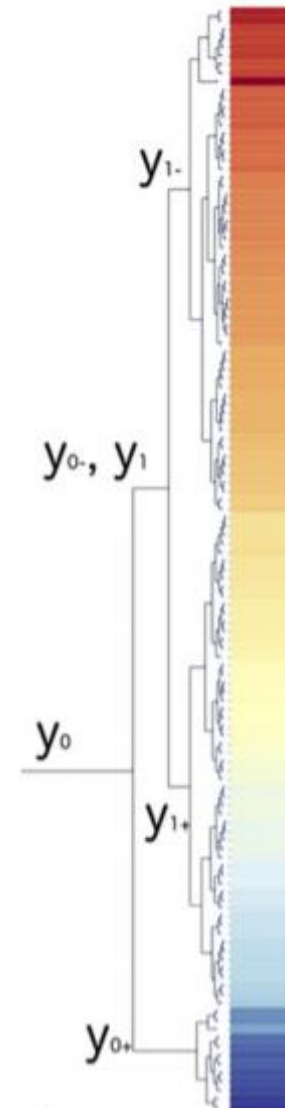
- **But**, it could not be proven !
- Since patients had different OTUs, it was difficult to test the effect of pH on any OTU.

Outline

- Motivation
- Definitions
- Method
 - Balances
 - Balance Trees
 - ILR
- **Results**
 - Case study 1
 - **Case study 2**
- Implementation
- Summary

Balance Tree to the rescue !

- Construct a balance tree in the exact same manner.

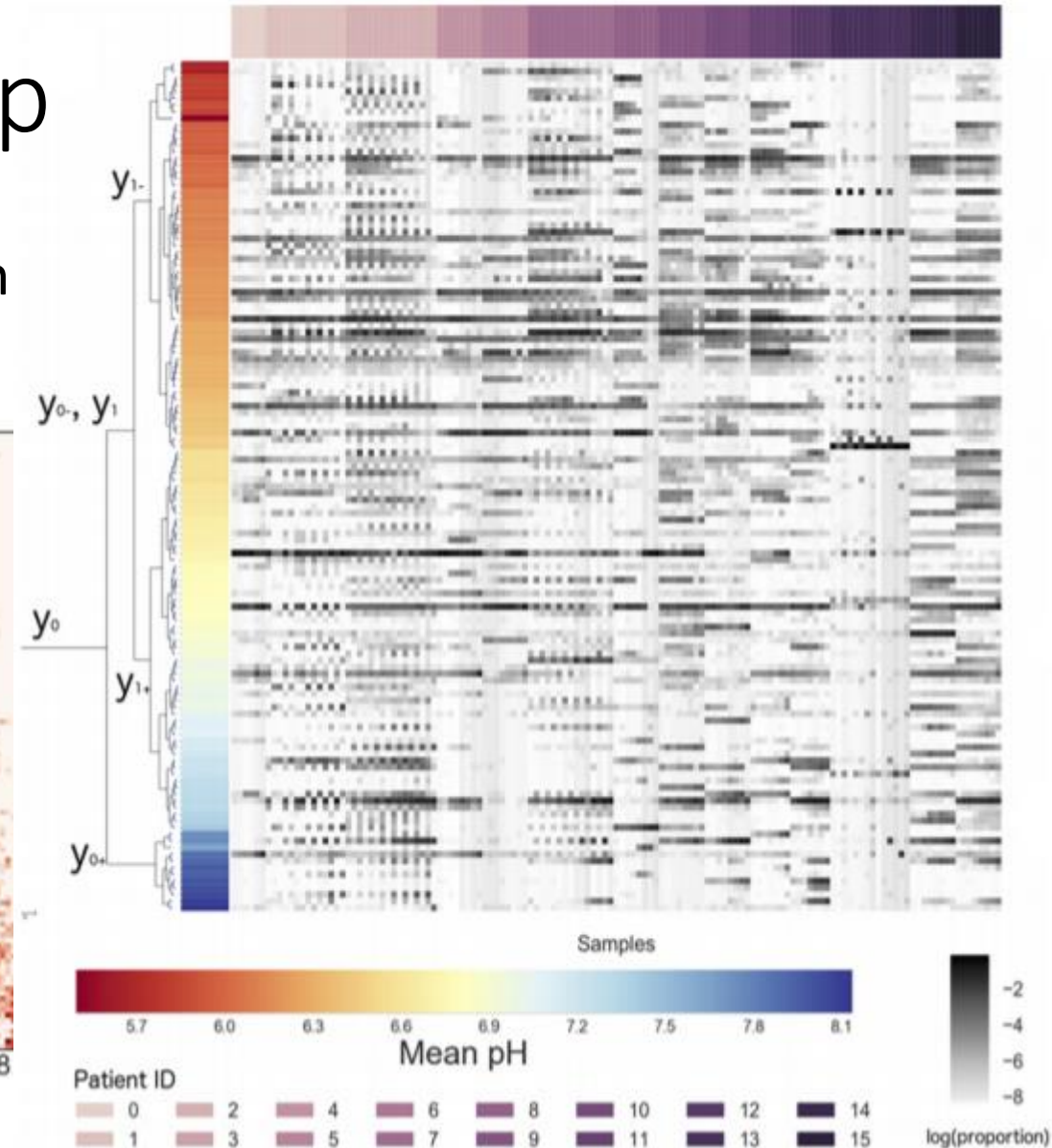
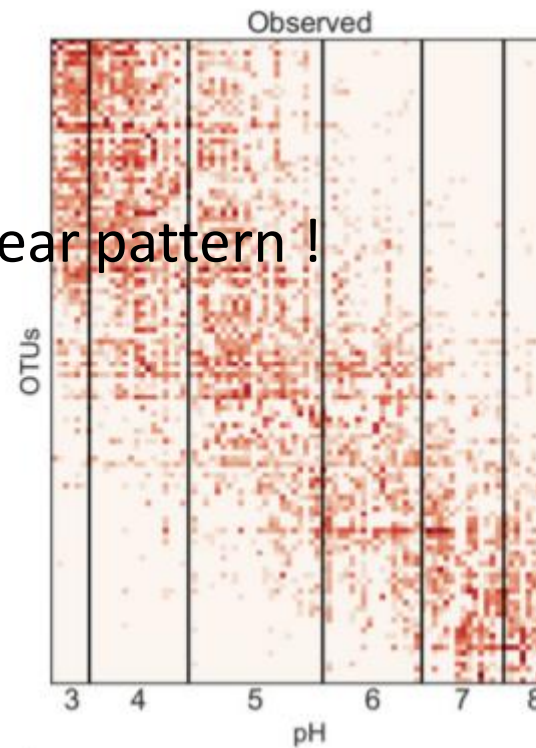


Outline

Abundance heatmap

Remember the heatmap graph from previous case ?

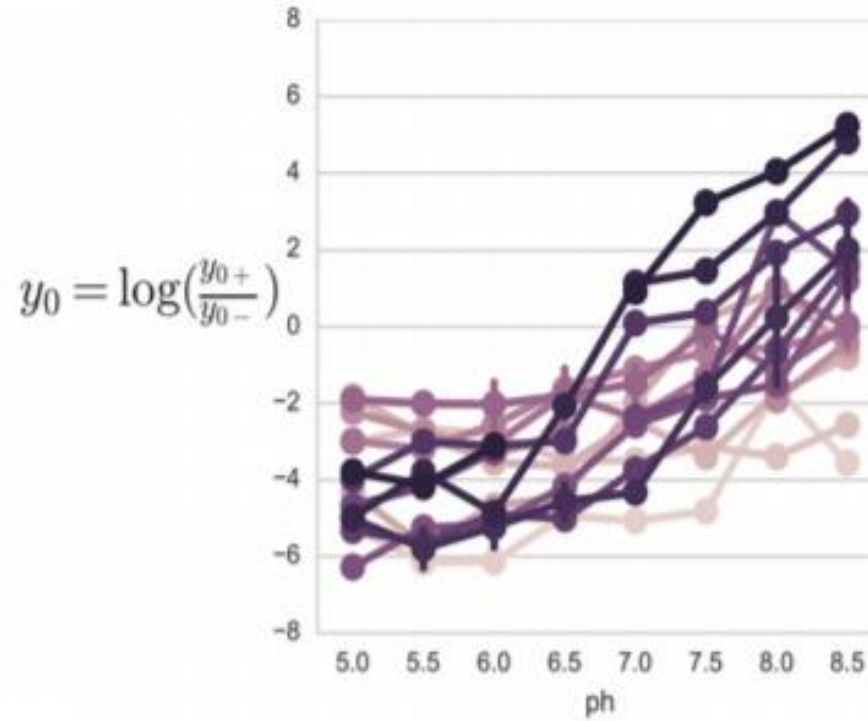
This time, no clear pattern !



Outline

The tree's root balance

- We can still observe niche differentiation across pH gradient.

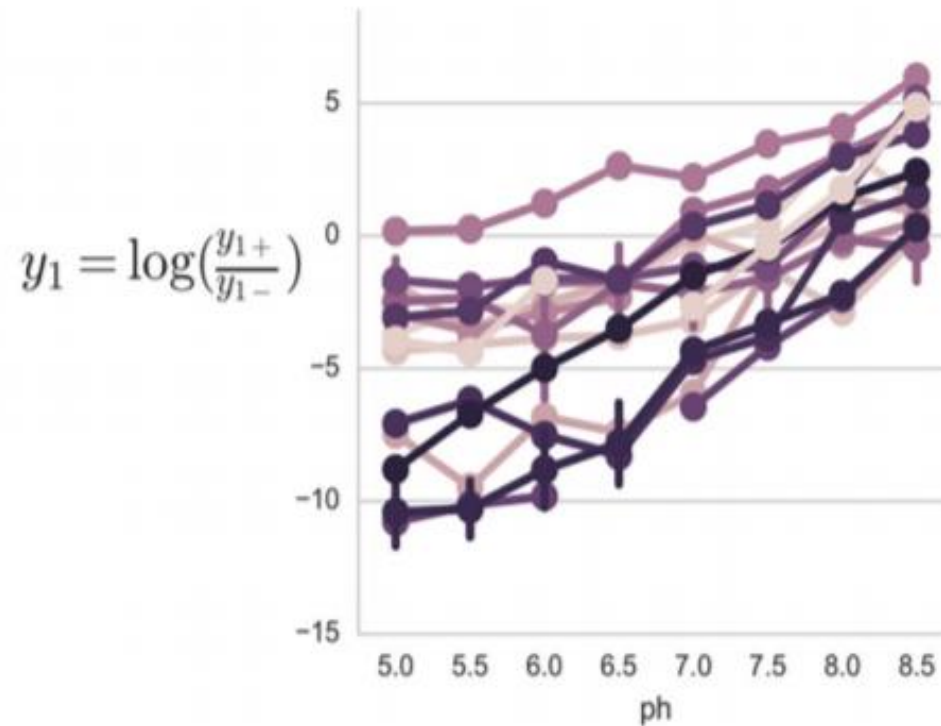


$y_{0+} \rightarrow 7.6 < \text{mean pH} < 8.12,$
 $y_{0-} \rightarrow 5.4 < \text{mean pH} < 7.4,$

Outline

Balance of subcommunity (depth=1)

- We can still observe niche differentiation across pH gradient.



$y_{1+} \rightarrow 6.5 < \text{mean pH} < 7.4$

$y_{1-} \rightarrow 5.4 < \text{mean pH} < 6.4$

- Motivation
- Definitions
- Method
 - Balances
 - Balance Trees
 - ILR
- **Results**
 - Case study 1
 - **Case study 2**
- Implementation
- Summary

Conclusion

- We conclude that although CF patients had different lung sputum microbiome, they contain OTUs that behave the same with respect to pH.
- Grouping OTUs using a meaningful metric is beneficial.

Outline

Implementation

- Motivation
- Definitions
- Method
 - Balances
 - Balance Trees
 - ILR
- Results
 - Case study 1
 - Case study 2
- **Implementation**
- Summary

- Open Source, <https://github.com/biocore/gneiss>
- Implemented as Python package. Setup as easy as *'pip install'* .
- Awesome documentation and tutorials !



R

Python

- Motivation
- Definitions
- Method
 - Balances
 - Balance Trees
 - ILR
- Results
 - Case study 1
 - Case study 2
- Implementation
- **Summary**

Summary

- Compositional data has its limitations.
- Balances are a robust concept to sidestep those limitations, and infer meaningful properties of subcommunities.
- Standard statistical techniques could be applied to balances.

Outline

- Motivation
- Definitions
- Method
 - Balances
 - Balance Trees
 - ILR
- Results
 - Case study 1
 - Case study 2
- Implementation
- Summary

Discussion

Discussion will be presented together with next lecture

Thanks !

