

Fold Change (Fail)

I did an experiment and got the result as fold change. Experiment was repeated three times. Following is the result:

1) Control = 1 , 1, 1

2) Test = 2.2, 2.4, 2.6 (fold change)

I have several doubt over using statistical test.

a) I think I can not use parametric test because sample size is only three and normal distribution can not be determined for this.

b) All control samples have same value so standard deviation can not be calculated.

c) Can I apply any non-parametric test in this case?

Pseudoreplication

Classic “fails” in molecular biology experimentation:

- (1) The “triplicate”
- (2) The “representative graph”
- (3) P-hacking

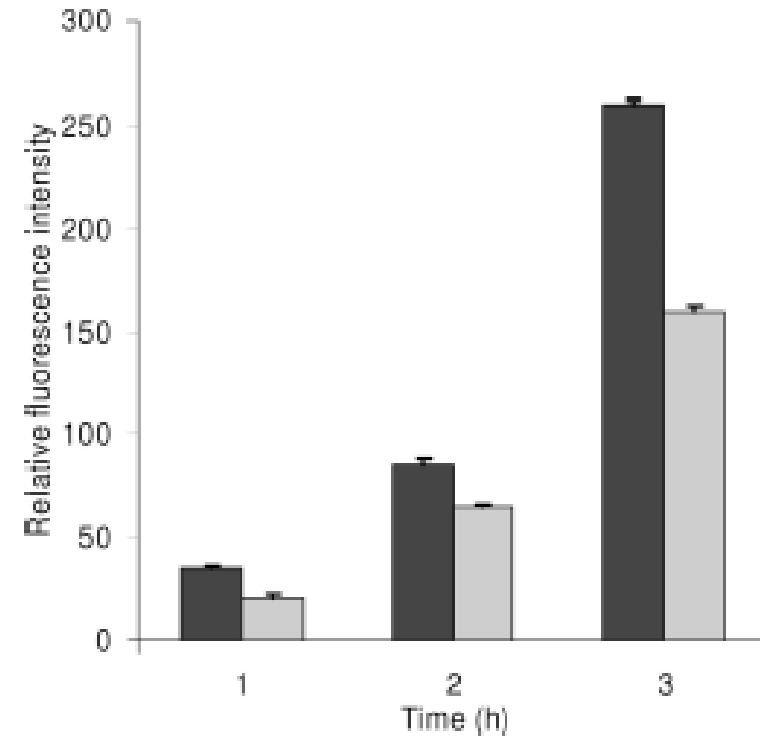


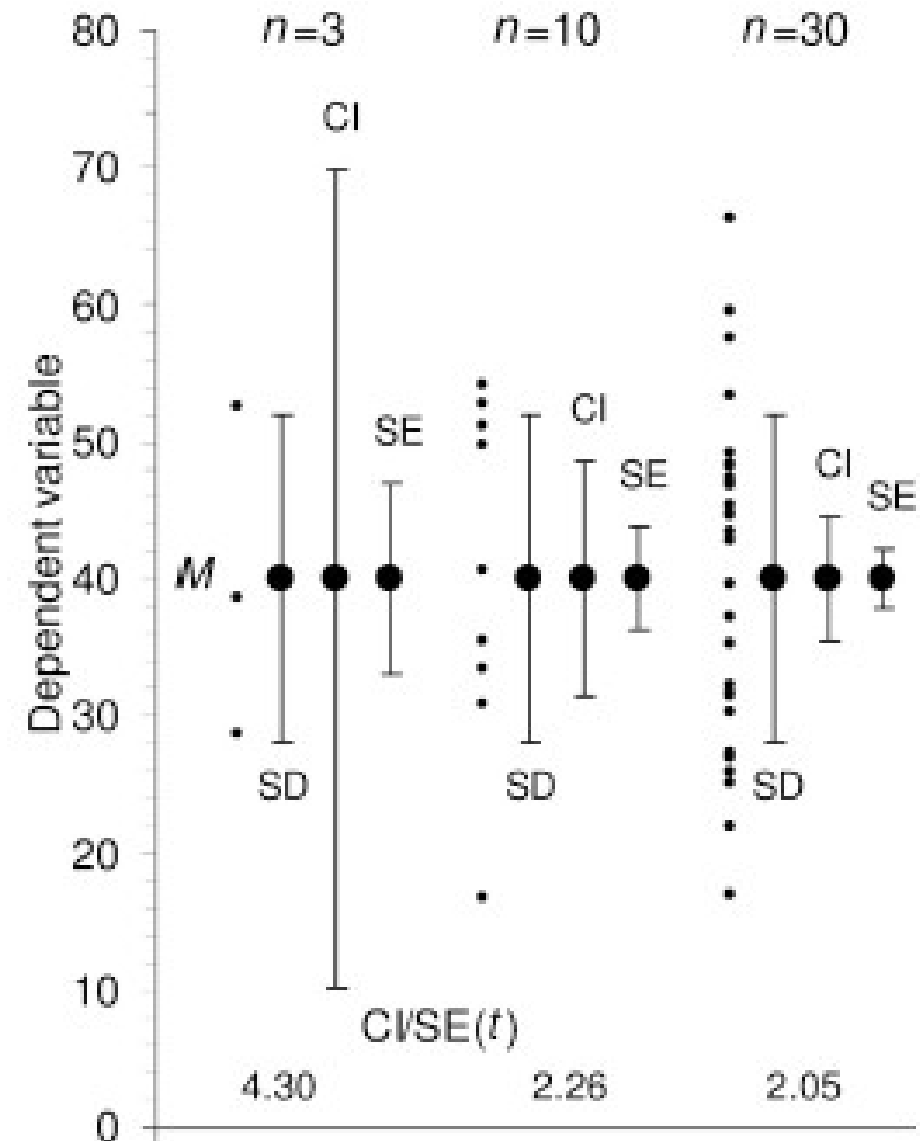
Figure 3. **Inappropriate use of error bars.** Enzyme activity for MEFs showing mean + SD from duplicate samples from one of three representative experiments. Values for wild-type vs. $-/-$ MEFs were significant for enzyme activity at the 3-h timepoint ($P < 0.0005$). This figure and its legend are typical, but illustrate inappropriate and misleading use of statistics because $n = 1$. The very low variation of the duplicate samples implies consistency of pipetting, but says nothing about whether the differences between the wild-type and $-/-$ MEFs are reproducible. In this case, the means and errors of the three experiments should have been shown.

Sample Size & Confidence Intervals

As you add data, you approach the "true mean" and your confidence of what the true mean is increases.

(Statistical consistency.)

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0086391>



BIG DATA TABLES: Metagenomics, Metabolomics, Gene Expressions patterns, Codon bias tables...etc.

Focus on diversity metrics and microbial communities.

FeatureTable[Frequency]					
	feature1	feature 2	feature 3	feature 4	feature 5
gut2	84	1	73	198	2
left.palm2	24	2	44	176	1
right.palm2	11	0	10	30	0
tongue2	0	0	25	2	0

	Total frequency
gut2	358
left.palm2	247
right.palm2	51
tongue2	27

Comparing microbial communities

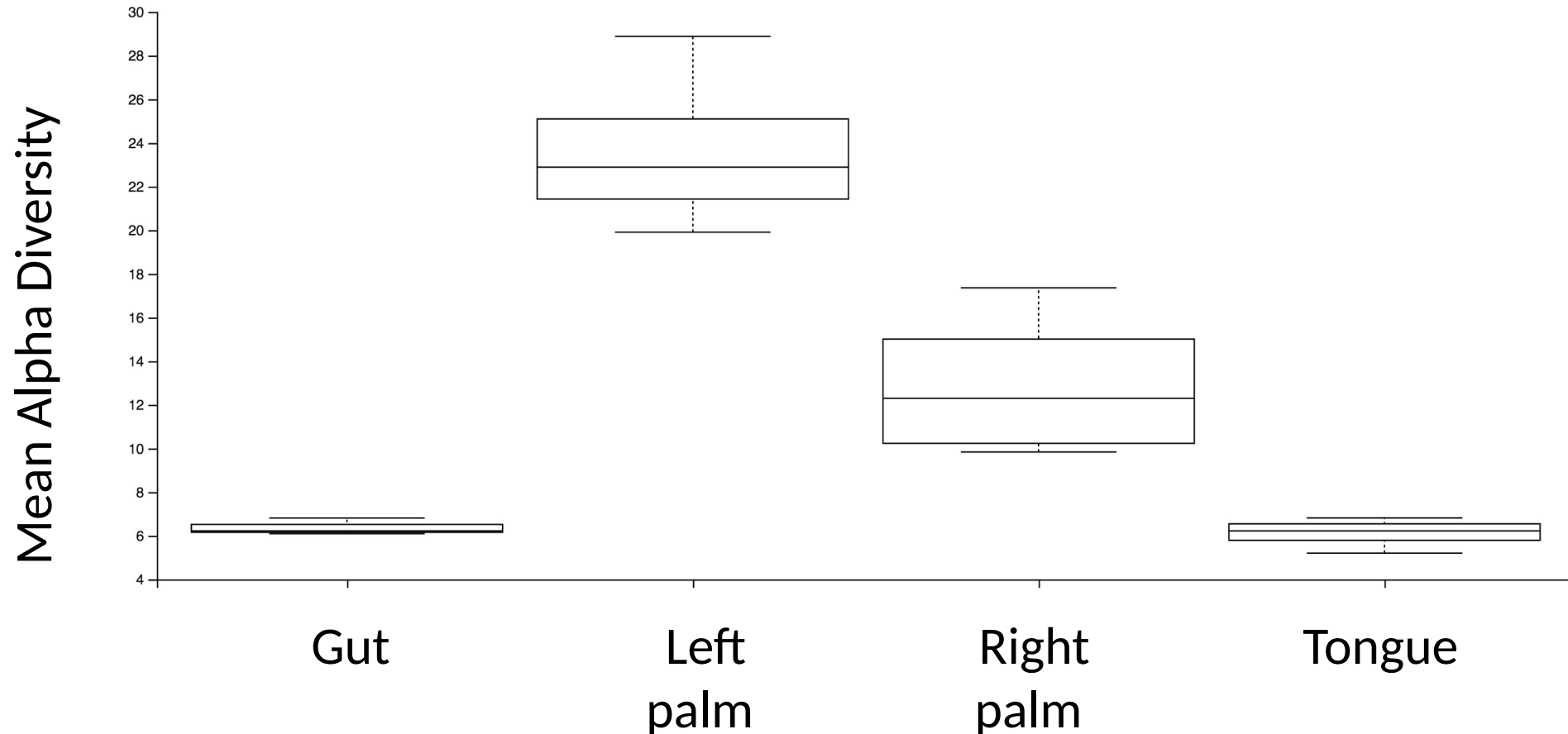
How many “species” are there? *Alpha diversity (richness, evenness, or both).*

How similar are pairs of samples? *Beta diversity*

Who is there? *Taxonomic profiling, differential abundance testing.*

Alpha diversity comparison

- visually with distribution comparison plots (discrete data) or scatter plots (continuous data)
- statistically with Kruskal-Wallis (discrete data) or Spearman correlation (continuous data)



Alpha diversity metrics

operate on a single sample (i.e., within sample diversity).

Beta diversity metrics

operate on pairs of samples (i.e., between sample diversity).

Qualitative diversity metrics only consider the presence/absence of features.

Quantitative diversity metrics consider abundance of features

ALPHA DIVERSITY

Observed OTUs (or Observed Species):
a qualitative, non-phylogenetic, alpha diversity metric

FeatureTable[Frequency]					
	feature1	feature 2	feature 3	feature 4	feature 5
gut3	25	30	15	0	0
left.palm3	0	17	33	25	0



SampleData[AlphaDiversity]	
	Observed OTUs
gut3	TOTAL?
left.palm3	TOTAL?

Count the number of different features in a sample.

ALPHA DIVERSITY

Observed OTUs (or Observed Species):
a qualitative, non-phylogenetic, alpha diversity metric

FeatureTable[Frequency]					
	feature1	feature 2	feature 3	feature 4	feature 5
gut3	25	30	15	0	0
left.palm3	0	17	33	25	0



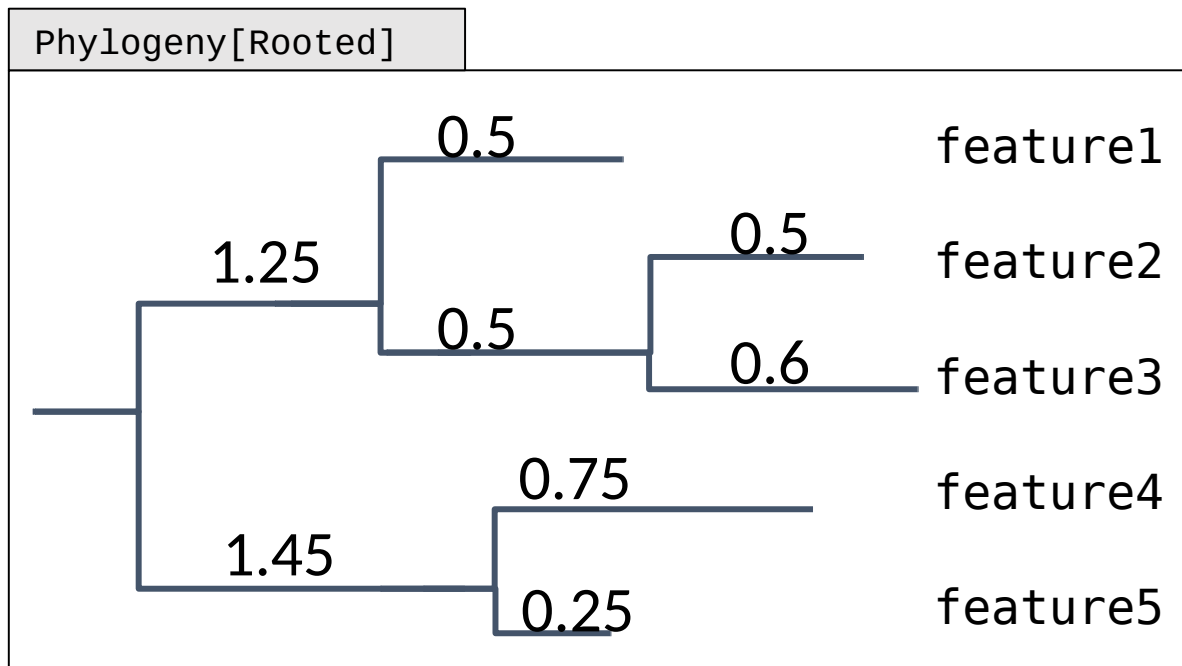
SampleData[AlphaDiversity]	
	Observed OTUs
gut3	3
left.palm3	3

Count the number of different features in a sample.

Why incorporate phylogeny in a diversity metric?

FeatureTable[Frequency]					
	feature1	feature2	feature3	feature4	feature5
gut3	25	30	15	0	0
left.palm3	0	17	33	25	0

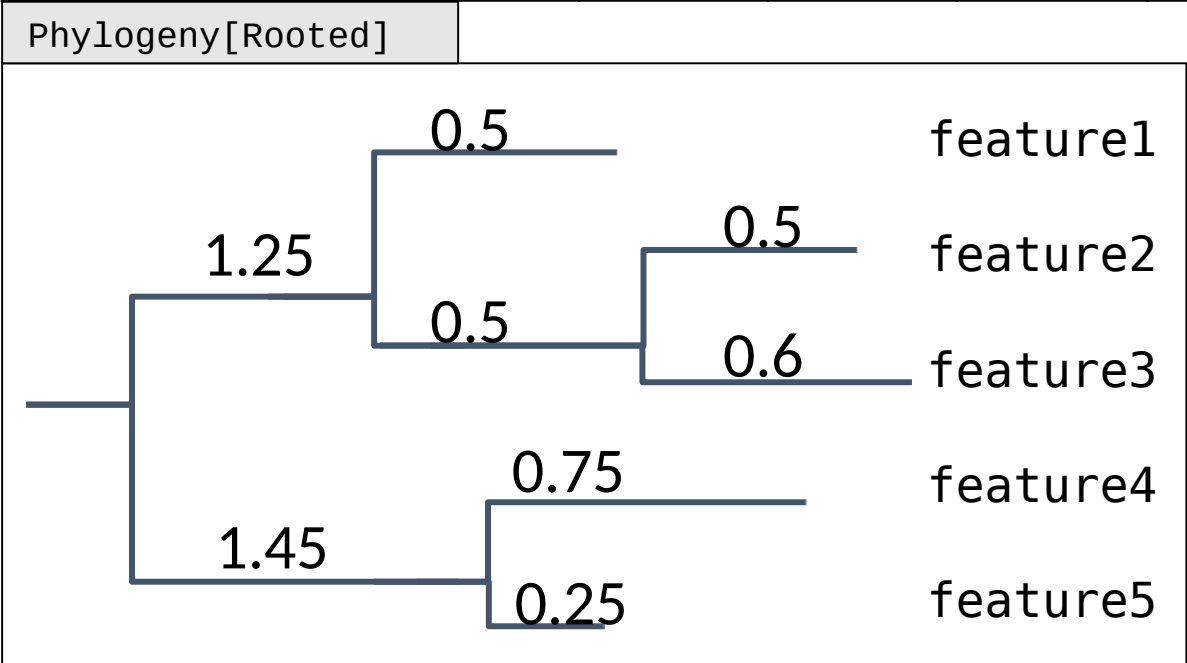
SampleData[AlphaDiversity]	
	Observed OTUs
gut3	3
left.palm3	3



FeatureData[Taxonomy]	
	Domain
feature1	Bacteria
feature2	Bacteria
feature3	Bacteria
feature4	Archaea
feature5	Archaea

Faith's Phylogenetic Diversity (PD): a qualitative, phylogenetic, alpha diversity metric

FeatureTable[Frequency]					
	feature1	feature 2	feature 3	feature 4	feature 5
gut3	25	30	15	0	0
left.palm3	0	17	33	25	0



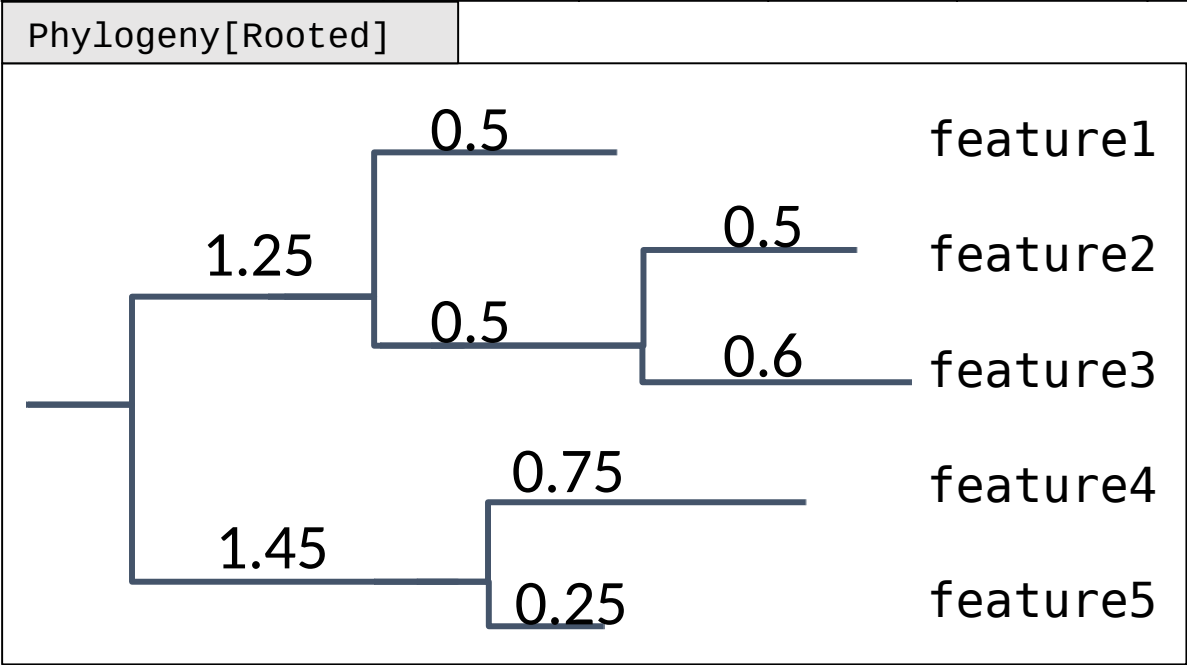
SampleData[AlphaDiversity]	
	Faith's PD
gut3	TOTAL?
left.palm3	TOTAL?

Sum of branch length covered by a sample.

Faith DP (1992) Conservation evaluation and phylogenetic diversity. Biological Conservation. 61:1-10.

Faith's Phylogenetic Diversity (PD): a qualitative, phylogenetic, alpha diversity metric

FeatureTable[Frequency]					
	feature1	feature 2	feature 3	feature 4	feature 5
gut3	25	30	15	0	0
left.palm3	0	17	33	25	0



SampleData[AlphaDiversity]	
	Faith's PD
gut3	3.35
left.palm3	5.05

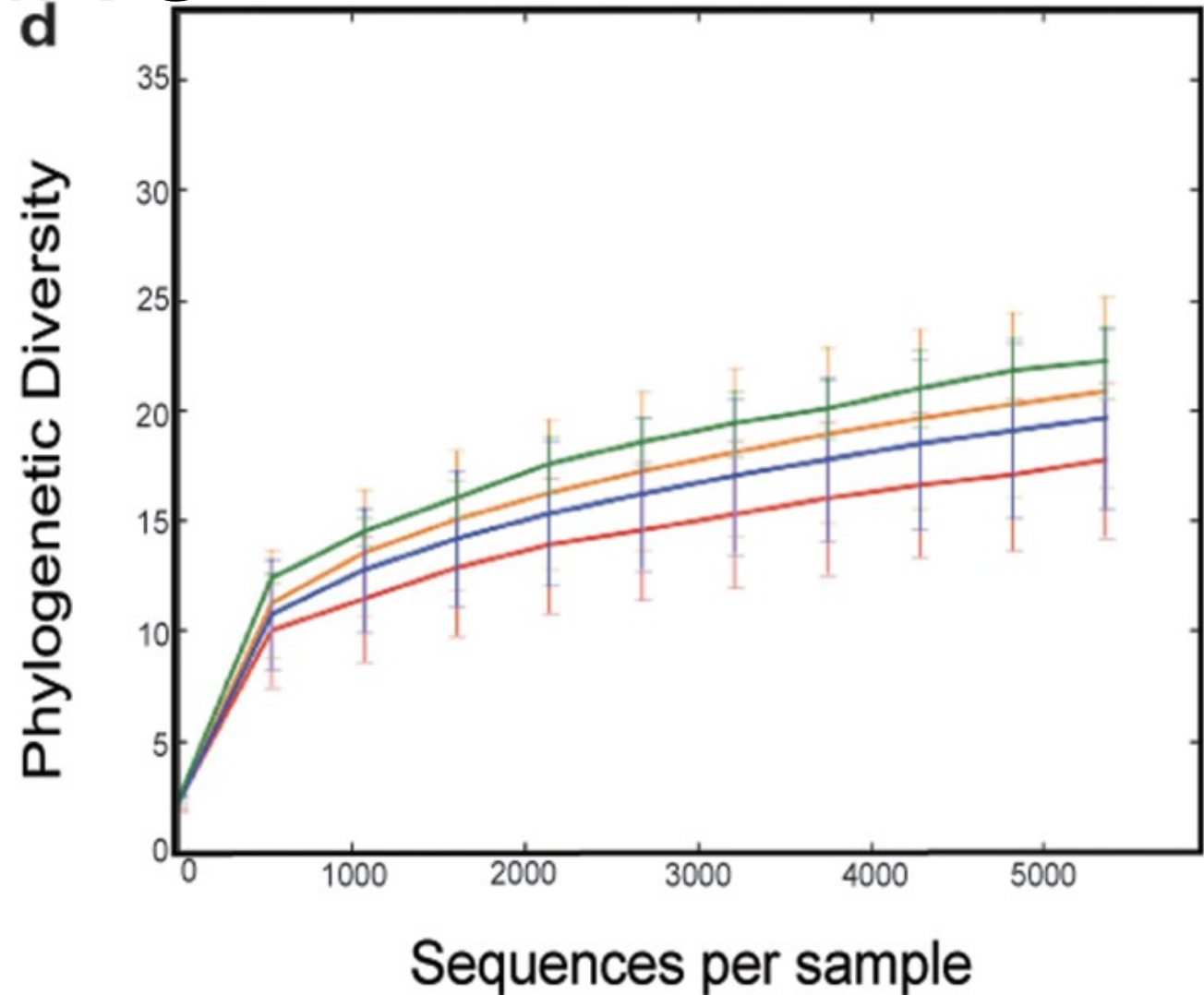
Sum of branch length covered by a sample.

Faith DP (1992) Conservation evaluation and phylogenetic diversity. Biological Conservation. 61:1-10.

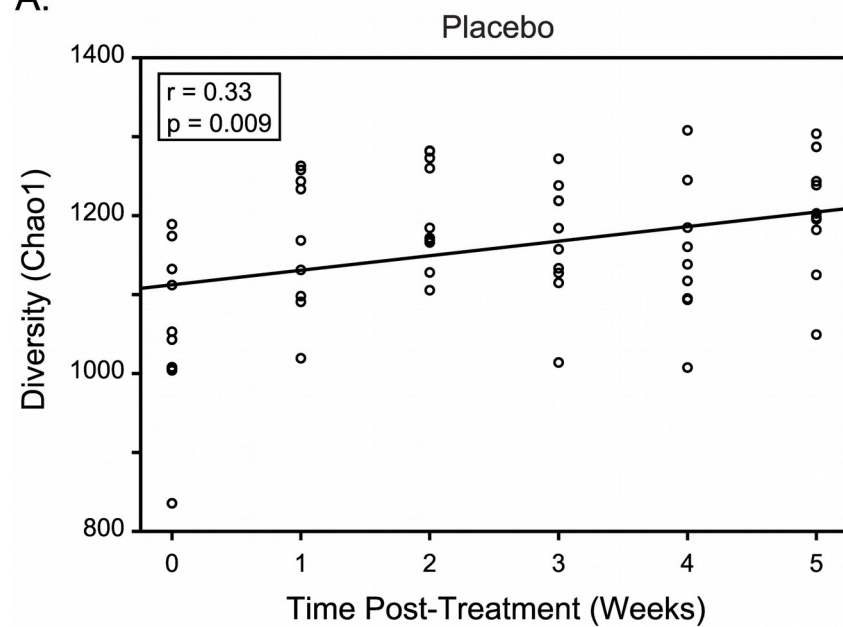
Rarefaction Curve

Q1) What if I have different sampling or sequencing for each sample?

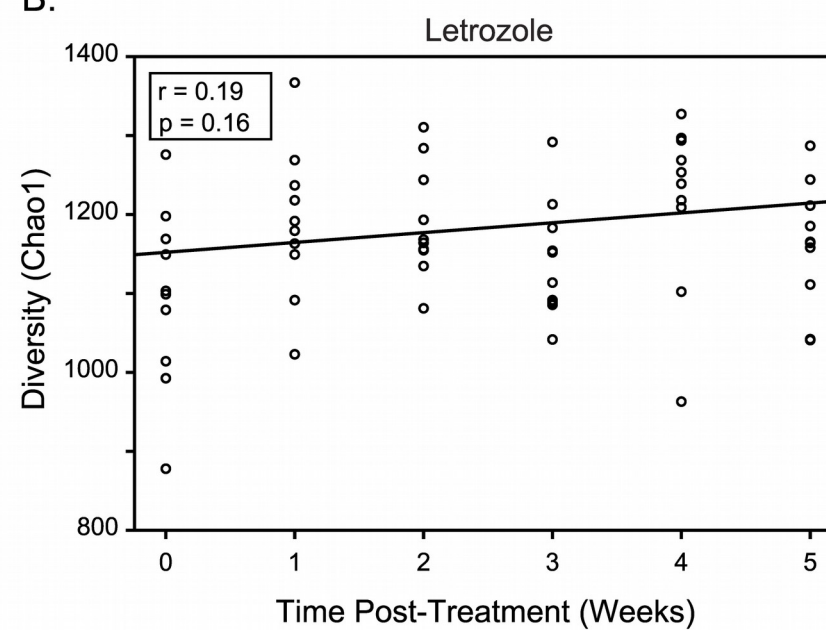
Q2) What happens if I change the sampling levels?



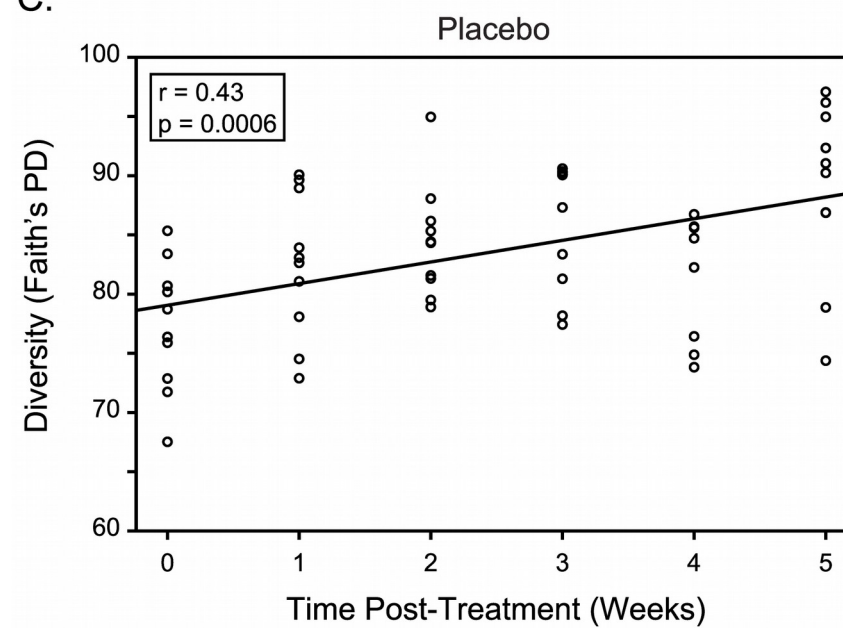
A.



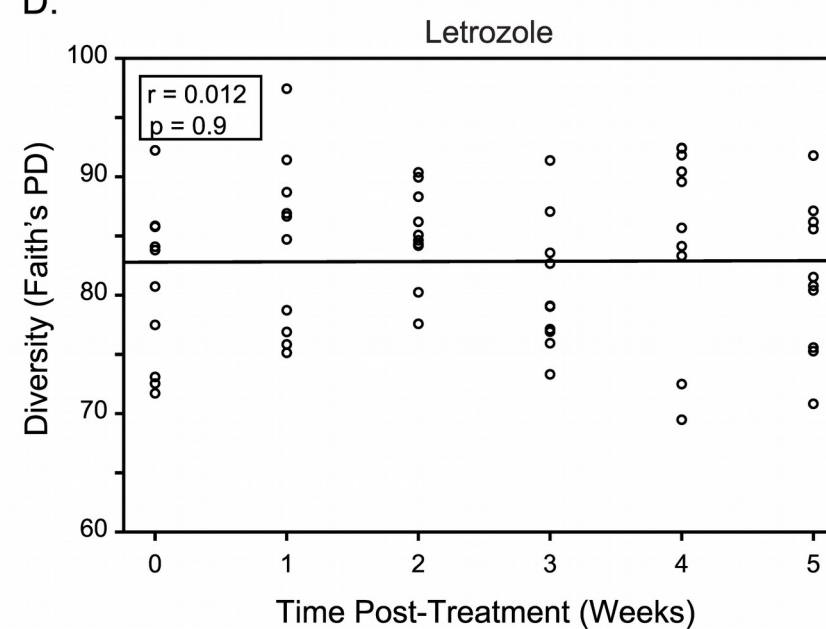
B.



C.

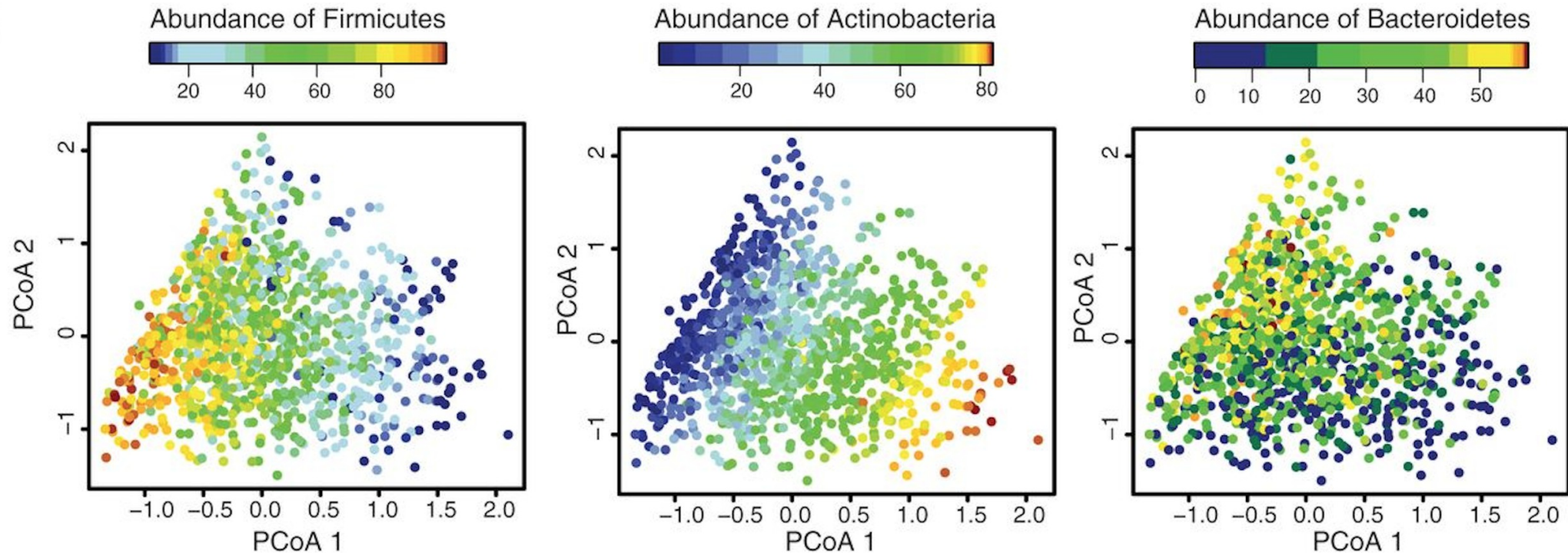


D.



Beta diversity comparison

- visually with ordination plots (e.g., NMDS or PCoA; discrete or continuous data) or distribution box plots (discrete data)
- statistically with Kruskal-Wallis (discrete data), Mantel correlation (continuous data) or BEST (continuous data)



Jaccard distance:

a qualitative, non-phylogenetic beta diversity metric

$$J(A, B) = 1 - \frac{A \cap B}{A \cup B}$$

FeatureTable[Frequency]					
	feature1	feature 2	feature 3	feature 4	feature 5
gut1	42	0	37	99	1
left.palm1	12	1	22	88	0
right.palm1	25	3	23	86	0
tongue1	0	0	87	12	0

DistanceMatrix				
	gut1	left.palm1	right.palm1	tongue1
gut1				
left.palm1				
right.palm1				
tongue1				

Jaccard distance:

a qualitative, non-phylogenetic beta diversity metric

$$J(A, B) = 1 - \frac{A \cap B}{A \cup B}$$

FeatureTable[Frequency]					
	feature1	feature 2	feature 3	feature 4	feature 5
gut1	42	0	37	99	1
left.palm1	12	1	22	88	0
right.palm1	25	3	23	86	0
tongue1	0	0	87	12	0

DistanceMatrix				
	gut1	left.palm1	right.palm1	tongue1
gut1	0.0			
left.palm1	VAL?	0.0		
right.palm1			0.0	
tongue1				0.0

Jaccard distance:

a qualitative, non-phylogenetic beta diversity metric

$$J(A, B) = 1 - \frac{A \cap B}{A \cup B}$$

FeatureTable[Frequency]					
	feature1	feature2	feature3	feature4	feature5
gut1	42	0	37	99	1
left.palm1	12	1	22	88	0
right.palm1	25	3	23	86	0
tongue1	0	0	87	12	0

DistanceMatrix				
	gut1	left.palm1	right.palm1	tongue1
gut1	0.0			
left.palm1	0.4	0.0		
right.palm1	0.4	0.0	0.0	
tongue1	0.5	0.5	0.5	0.0

Jaccard distance:

a qualitative, non-phylogenetic beta diversity metric

$$J(A, B) = 1 - \frac{A \cap B}{A \cup B}$$

FeatureTable[Frequency]					
	feature1	feature2	feature3	feature4	feature5
gut1	42	0	37	99	1
left.palm1	12	1	22	88	0
right.palm1	25	3	23	86	0
tongue1	0	0	87	12	0

DistanceMatrix				
	gut1	left.palm1	right.palm1	tongue1
gut1	0.0	0.4	0.4	0.5
left.palm1	0.4	0.0	0.0	0.5
right.palm1	0.4	0.0	0.0	0.5
tongue1	0.5	0.5	0.5	0.0

Bray-Curtis distance:

a quantitative, non-phylogenetic beta diversity metric

FeatureTable[Frequency]					
	feature1	feature2	feature3	feature4	feature5
gut1	42	0	37	99	1
left.palm1	12	1	22	88	0
right.palm1	25	3	23	86	0
tongue1	0	0	87	12	0

$$BC(A, B) = \frac{\sum_i |X_{iA} - X_{iB}|}{\sum_i (X_{iA} + X_{iB})}$$

X_{iA} : frequency of feature i in sample A

DistanceMatrix				
	gut1	left.palm1	right.palm1	tongue1
gut1	0.0			
left.palm1		0.0		
right.palm1			0.0	
tongue1	VAL?			0.0

Bray-Curtis distance:

a quantitative, non-phylogenetic beta diversity metric

FeatureTable[Frequency]					
	feature1	feature2	feature3	feature4	feature5
gut1	42	0	37	99	1
left.palm1	12	1	22	88	0
right.palm1	25	3	23	86	0
tongue1	0	0	87	12	0

$$BC(A, B) = \frac{\sum_i |X_{iA} - X_{iB}|}{\sum_i (X_{iA} + X_{iB})}$$

X_{iA} : frequency of feature i in sample A

DistanceMatrix				
	gut1	left.palm1	right.palm1	tongue1
gut1	0.0			
left.palm1	0.19	0.0		
right.palm1	0.15	0.07	0.0	
tongue1	0.65	0.69	0.70	0.0

Bray-Curtis distance:

a quantitative, non-phylogenetic beta diversity metric

FeatureTable[Frequency]					
	feature1	feature 2	feature 3	feature 4	feature 5
gut1	42	0	37	99	1
left.palm1	12	1	22	88	0
right.palm1	25	3	23	86	0
tongue1	0	0	87	12	0

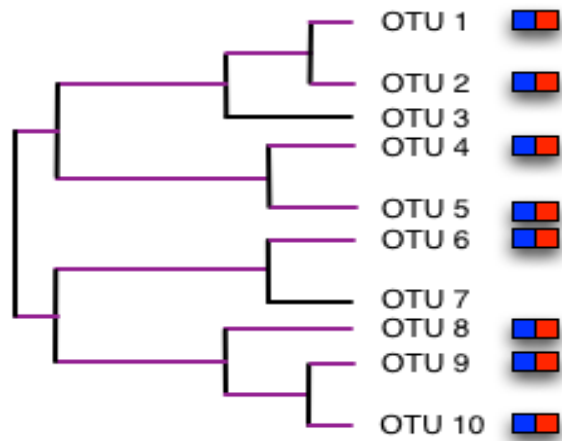
$$BC(A, B) = \frac{\sum_i |X_{iA} - X_{iB}|}{\sum_i (X_{iA} + X_{iB})}$$

X_{iA} : frequency of feature i in sample A

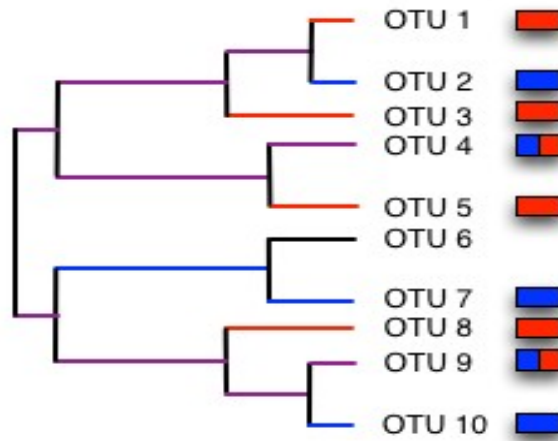
DistanceMatrix				
	gut1	left.palm1	right.palm1	tongue1
gut1	0.0	0.19	0.15	0.65
left.palm1	0.19	0.0	0.07	0.69
right.palm1	0.15	0.07	0.0	0.70
tongue1	0.65	0.69	0.70	0.0

Unweighted UniFrac distance:

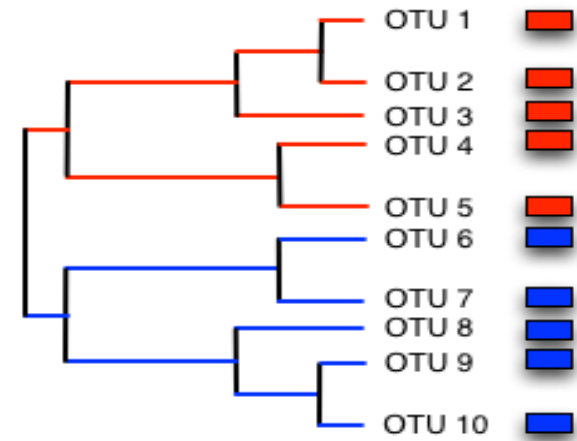
a qualitative, phylogenetic beta diversity metric



$UU =$
0.0



$UU \approx$
0.5



$UU = 1.0$

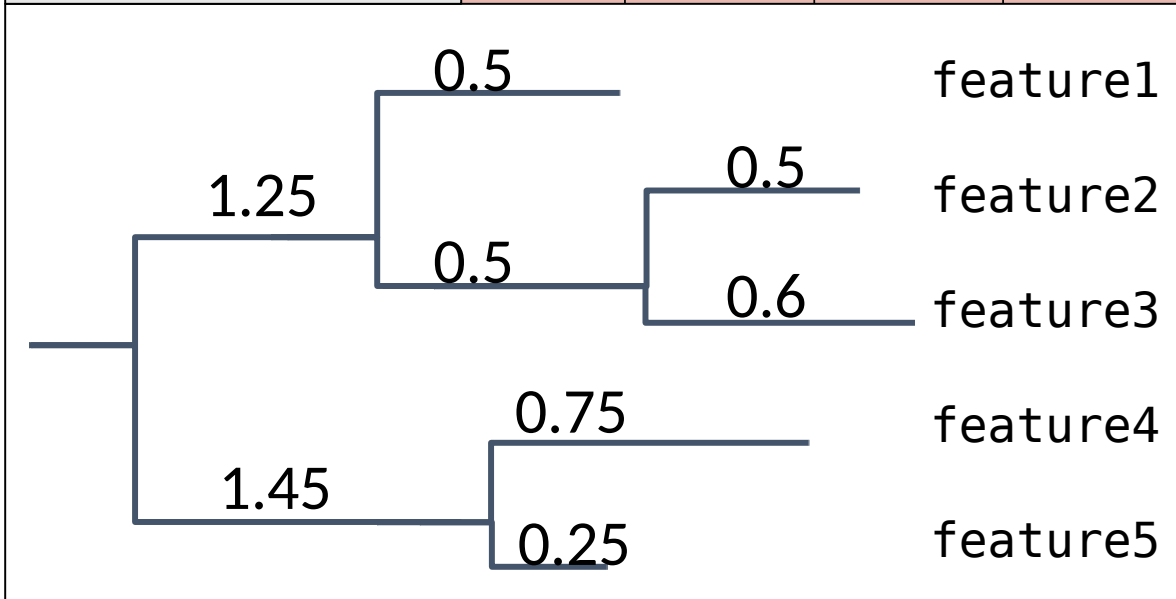
$$UU(A, B) = \frac{\text{sum of unique branch length}}{\text{sum of observed branch length}}$$

Unweighted UniFrac distance:

a qualitative, phylogenetic beta diversity metric

FeatureTable[Frequency]					
	feature1	feature 2	feature 3	feature 4	feature 5
gut1	42	0	37	99	1
left.palm1	12	1	22	88	0
Phylogeny[Rooted]	25	3	23	86	0

$$UU(A, B) = \frac{\text{sum of unique branch length}}{\text{sum of observed branch length}}$$



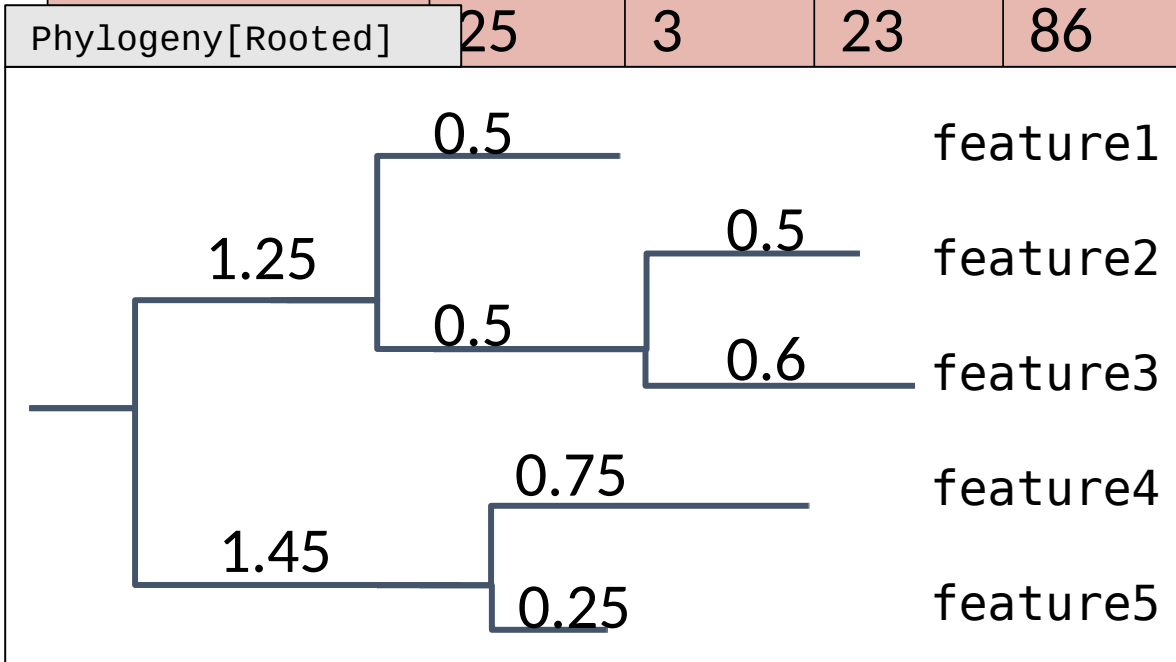
DistanceMatrix				
	gut1	left.palm1	right.palm1	tongue1
gut1	0.0			
left.palm1	VAL?	0.0		
right.palm1			0.0	
tongue1				0.0

Unweighted UniFrac distance:

a qualitative, phylogenetic beta diversity metric

FeatureTable[Frequency]					
	feature1	feature 2	feature 3	feature 4	feature 5
gut1	42	0	37	99	1
left.palm1	12	1	22	88	0
right.palm1	25	3	23	86	0
tongue1	0	0	0	0	0

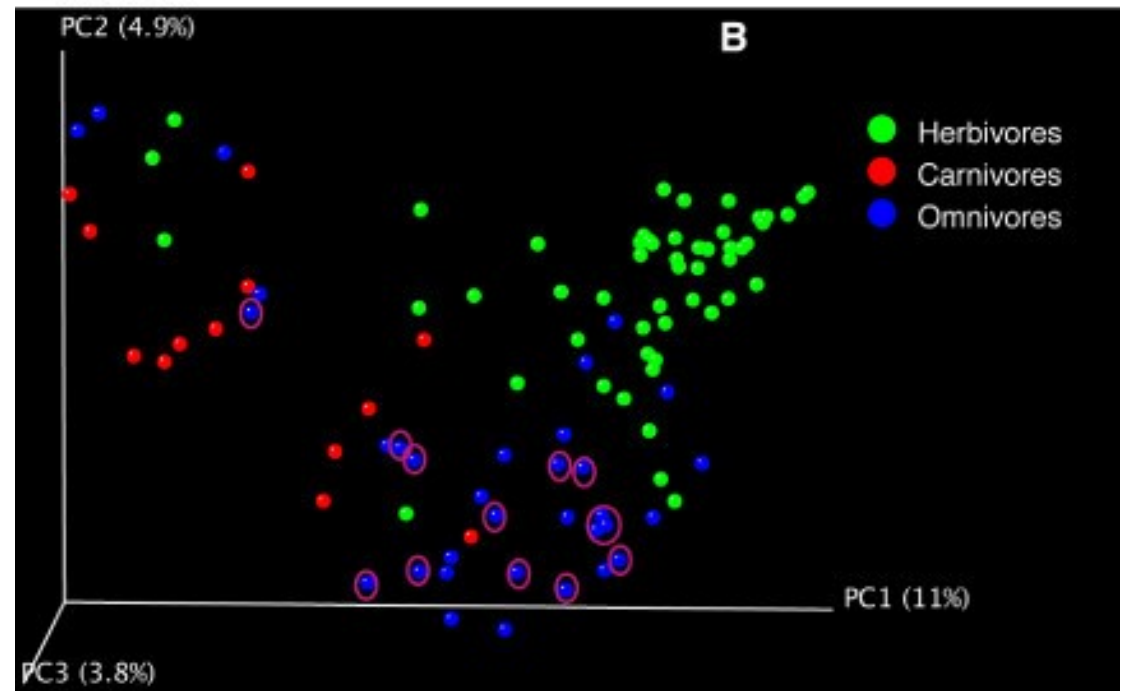
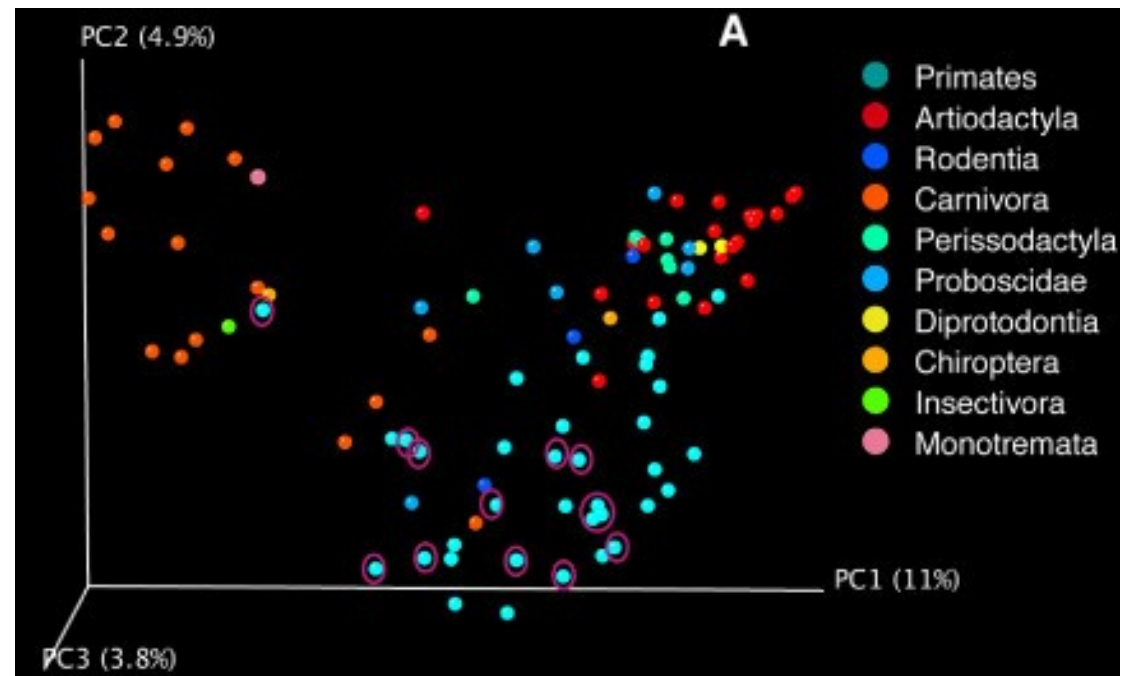
$$UU(A, B) = \frac{\text{sum of unique branch length}}{\text{sum of observed branch length}}$$



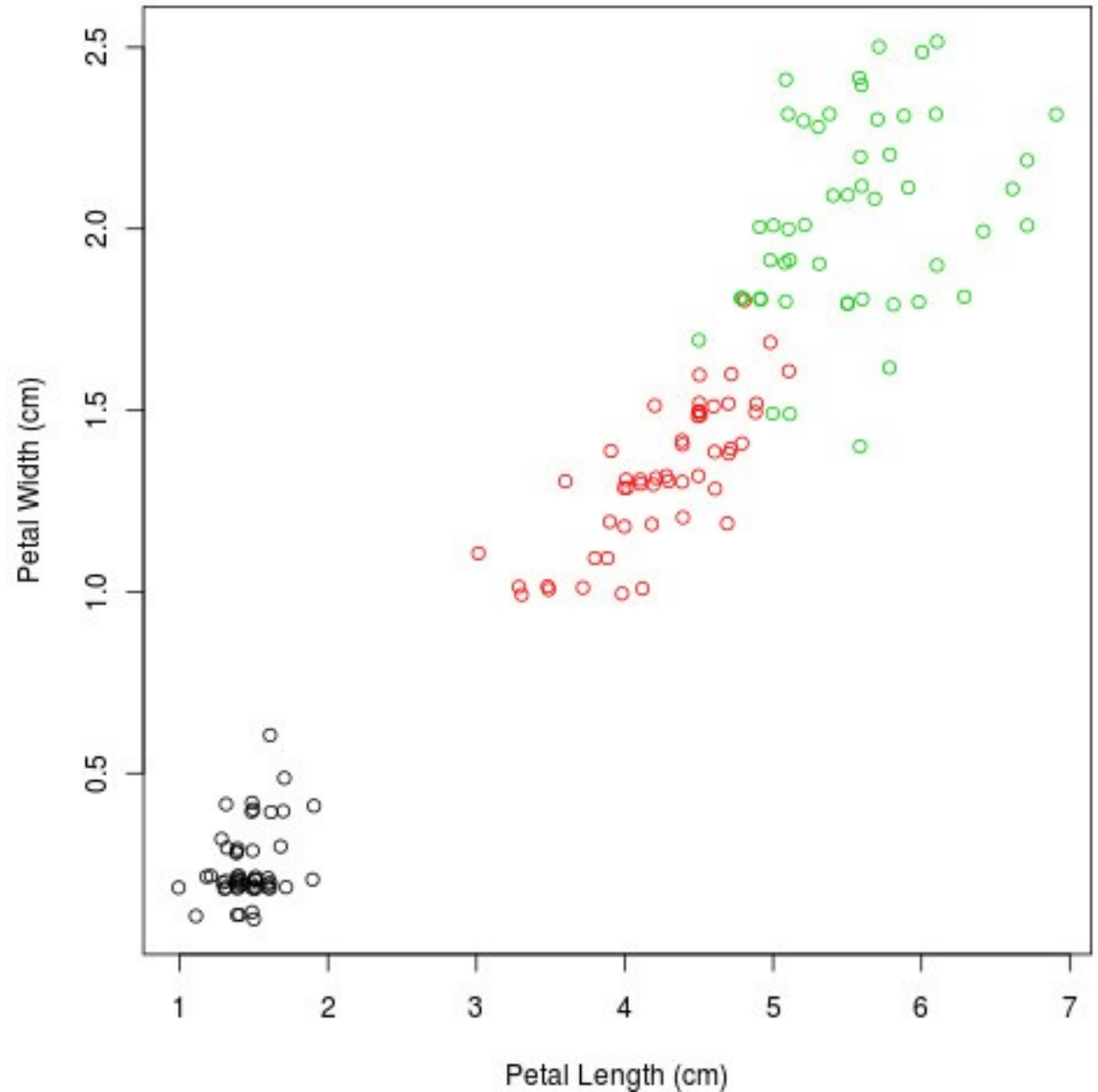
DistanceMatrix				
	gut1	left.palm1	right.palm1	tongue1
gut1	0.0			
left.palm1	0.13	0.0		
right.palm1	0.13	0.0	0.0	
tongue1	0.14	0.18	0.18	0.0

[illegible]

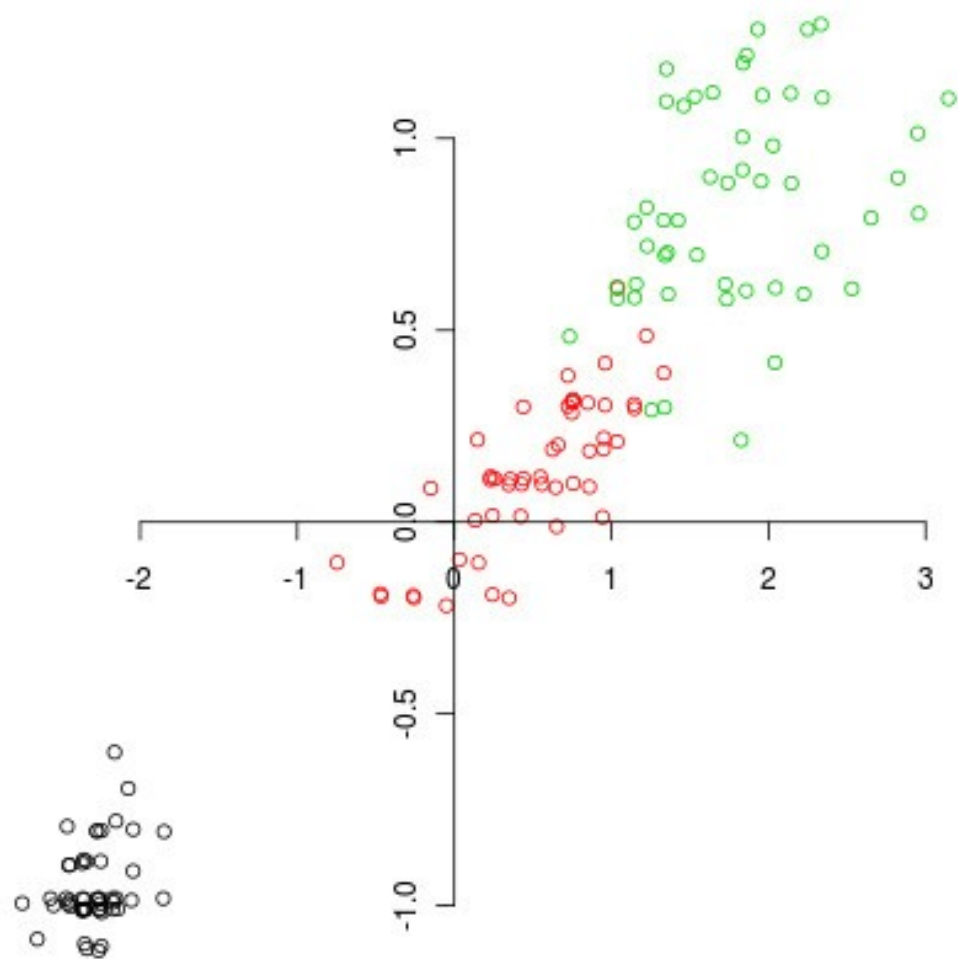
Principal Coordinates Analysis: Example



PCA and PCoA explained

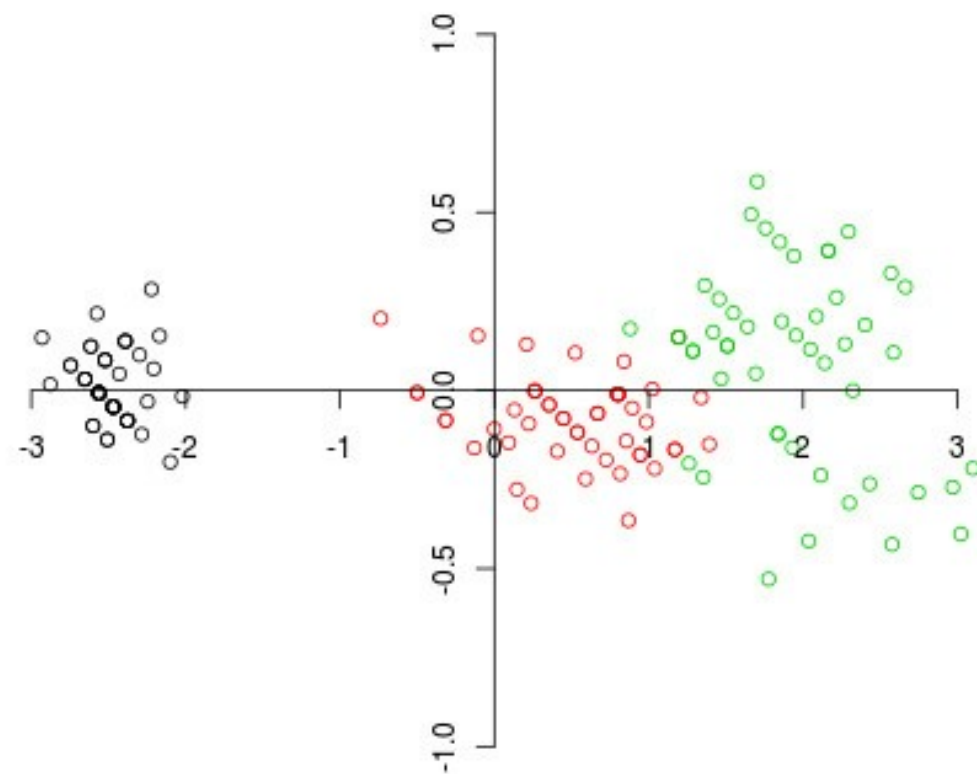


Centred Petal Width (cm)



Centred Petal Length (cm)

Second Dimension



First Dimension

Principal Coordinates Analysis: Example

