UniFrac

Phylogenetic Method for Comparing Microbial Communities

Lozupone, Catherine, and Rob Knight. *Appl. Environ. Microbiol.* 71.12 (2005)

Outline

- ❑ Background and motivation.
- ❑ Leading Questions.
- ❑ UniFrac – The Method.
- ❑ Results.
- ❑ Summary Conclusions and Discussion.

Outline

- ❑ **Background and motivation.**
- ❑ Leading Questions.
- ❑ UniFrac – The Method.
- ❑ Results.
- ❑ Summary Conclusions and Discussion.

Before UniFrac

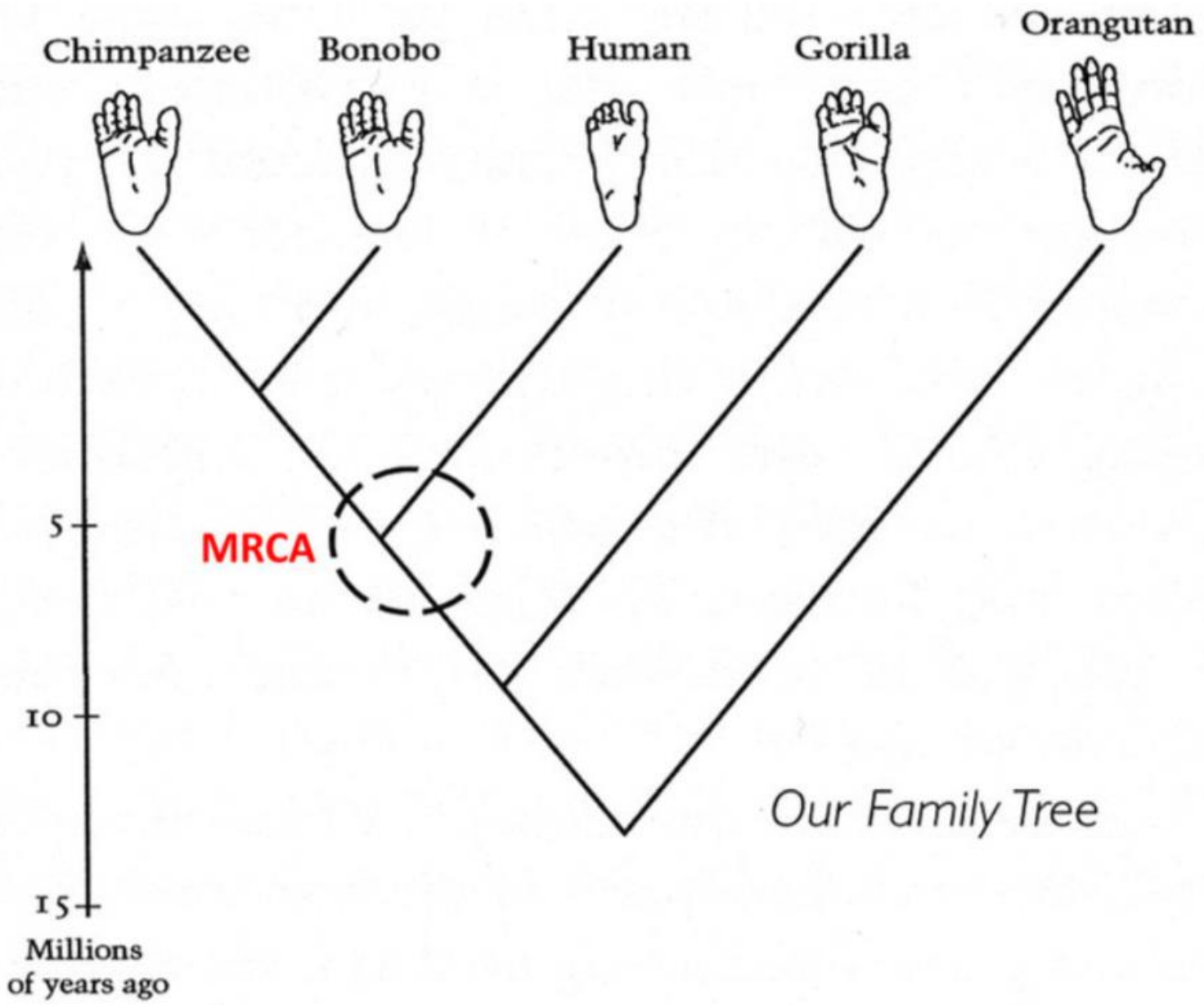
Several statistical techniques have been developed to use environmental 16S rRNA clone sequences to compare microbial communities between samples.

The problem: many of these techniques do not account for the different degrees of similarity between sequences – loss of information.

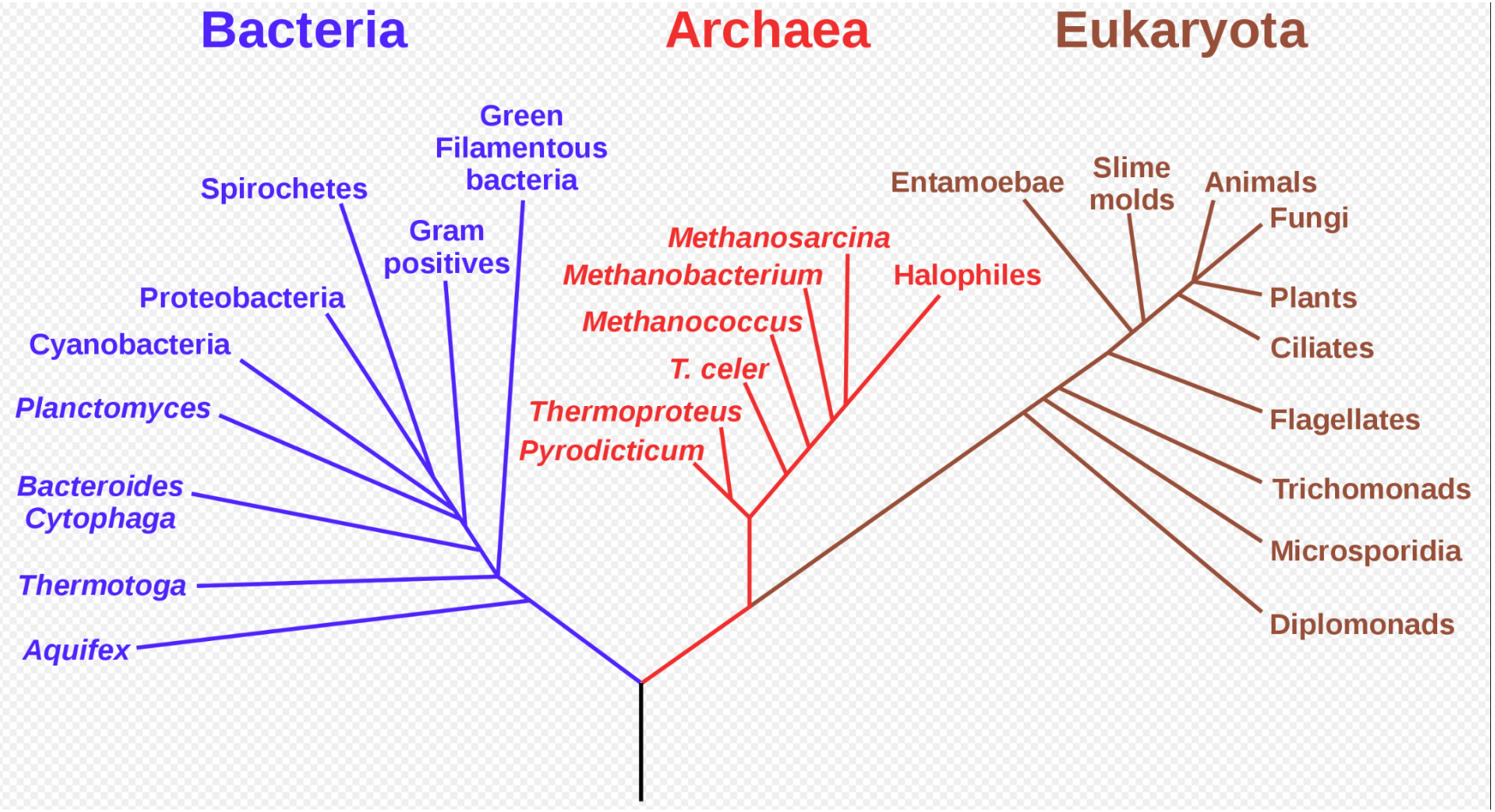
Taxonomy And Phylogenesis

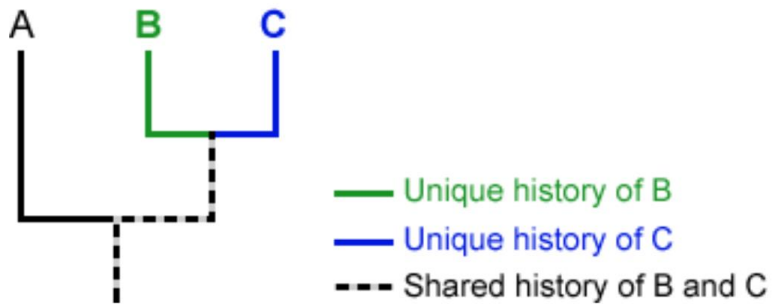
- **Taxonomy** - The science of defining and naming groups of biological organisms based on shared characteristics. Organisms are grouped together into taxa.
- **Phylogenesis** -Reconstruct the evolutionary relationship between species and taxons.

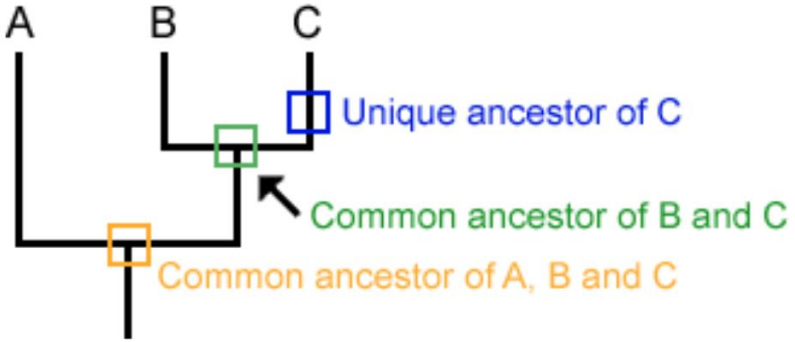
Phylogenetic Tree

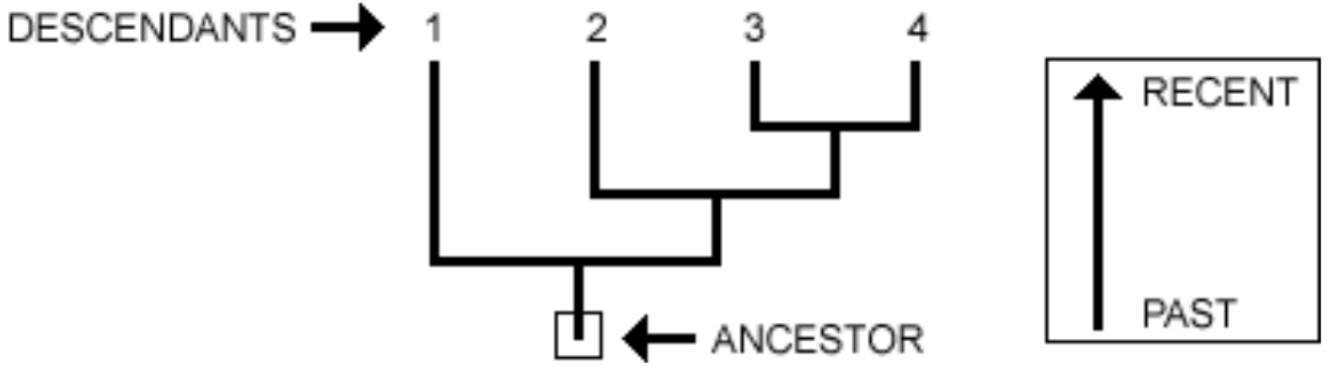


Phylogenetic Tree Of Life



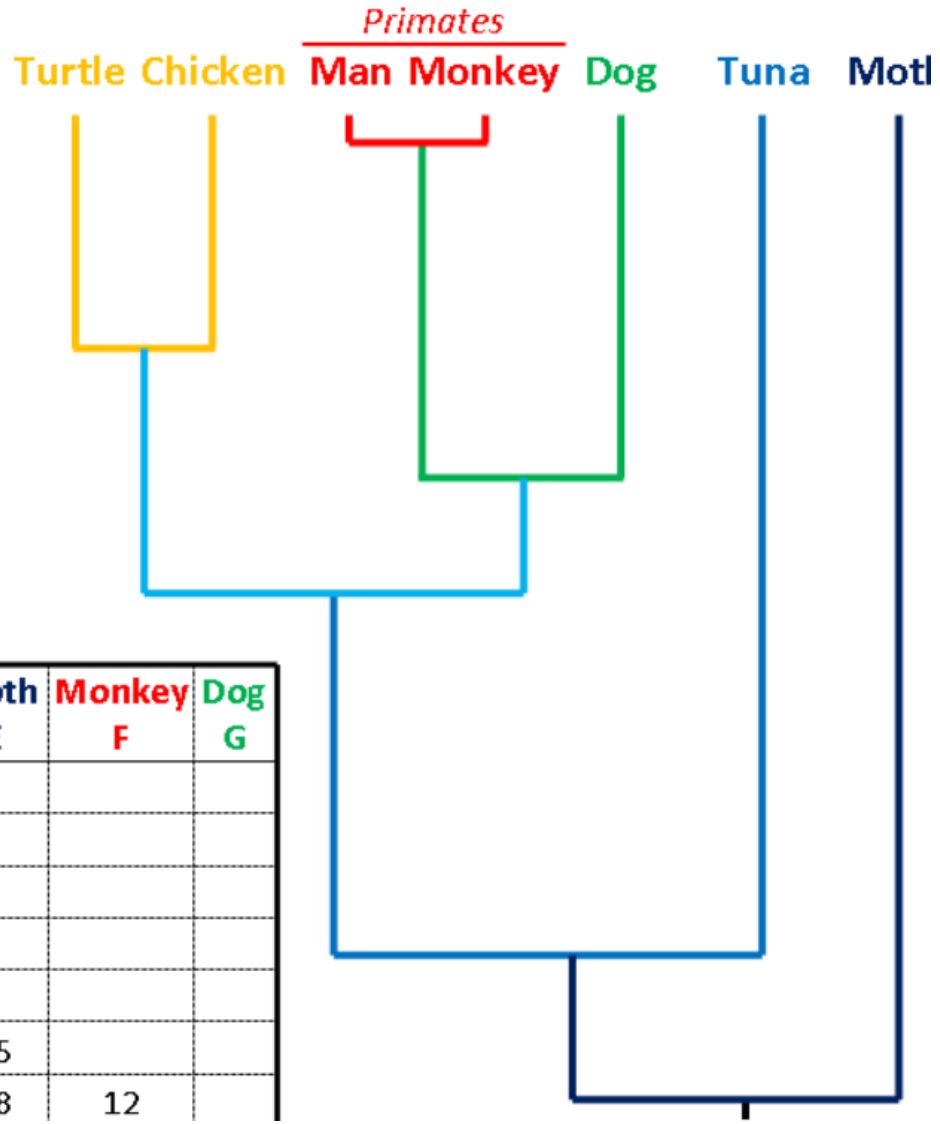

A
B
C
Unique history of B
Unique history of C
Shared history of B and C


A
B
C
Unique ancestor of C
Common ancestor of B and C
Common ancestor of A, B and C

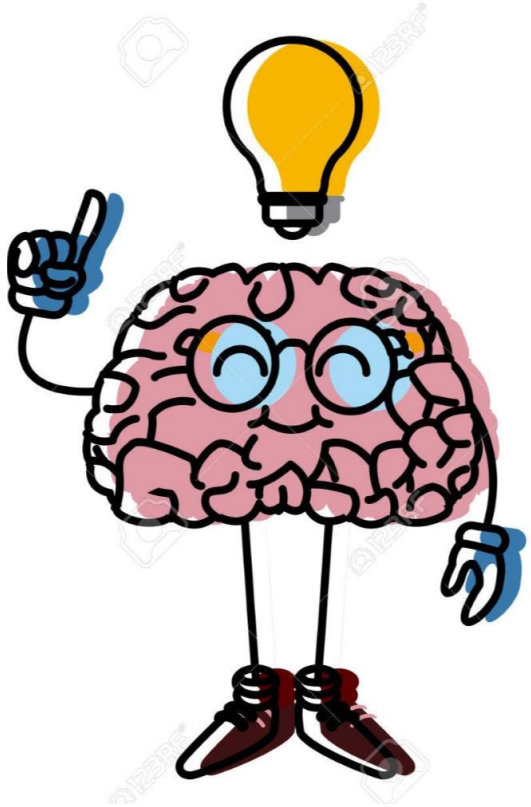

DESCENDANTS
1
2
3
4
ANCESTOR
RECENT
PAST

UPGMA

- Clustering method that gets a distance matrix as input and returns phylogenetic tree.



	Turtle A	Man B	Tuna C	Chicken D	Moth E	Monkey F	Dog G
Turtle							
Man	19						
Tuna	27	31					
Chicken	8	18	26				
Moth	33	36	41	31			
Monkey	18	1	32	17	35		
Dog	13	13	29	14	28	12	



Phylogenetic distance measures can provide far more power, exploit the degree of divergence between different sequences.

Usage of Phylogenesis

- Two phylogenetic approaches that asses whether communities differ significantly in composition have been developed not long before the paper was published.

But:

  - These methods have only been applied to determining whether samples are significantly different and have not been used to compare many samples simultaneously
  - They don’t use the branch length information.

Outline

- ✓ Background and motivation.
- ❑ **Leading Questions**.
- ❑ UniFrac – The Method.
- ❑ Results.
- ❑ Summary Conclusions and Discussion.

Leading questions in the research that UniFrac can answer

- **How does culturing affect similarities between samples?**

Examination if cultured samples appeared more similar to uncultured samples from the same environment or to each other?

- **How Cosmopolitan are bacterial lineage?**

Comparing samples from the same environment but from different geographical origin.

Leading questions – cont.

- **Are marine ice, sediment and seawater three distinct homogenous habitats?**

These three habitats treated in the literature as distinct habitat types.

The Authors test whether these habitat types harbor consistent bacterial communities that differ from one another.

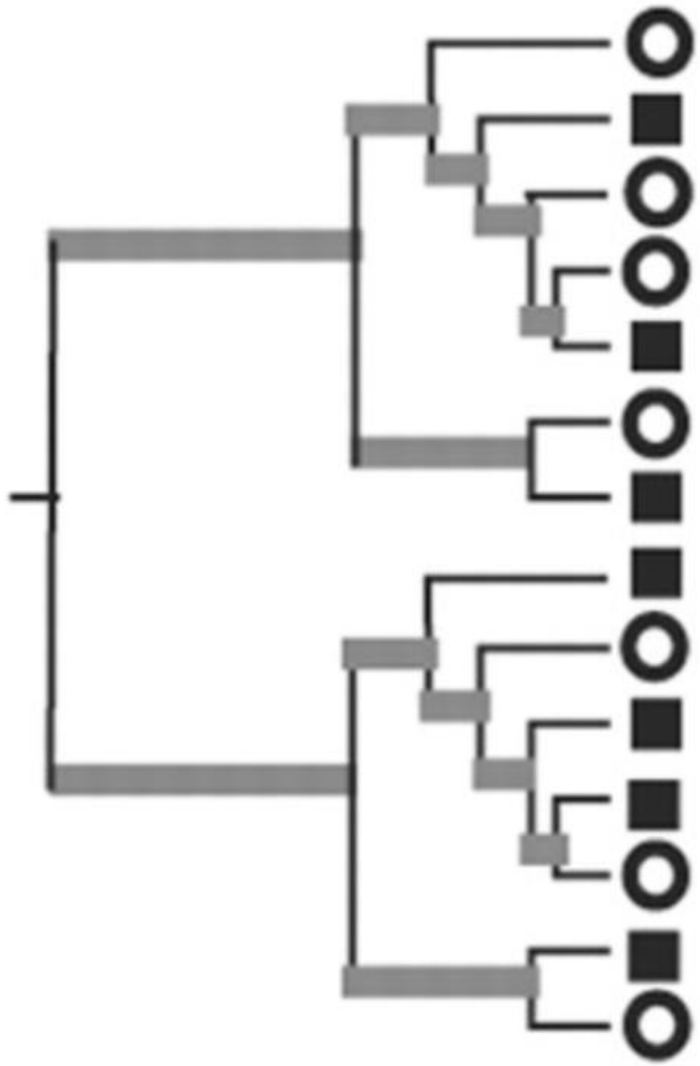
Outline

- ✓ Background and motivation.
- ✓ Leading Questions.
- ❑ **UniFrac – The Method.**
- ❑ Results.
- ❑ Summary Conclusions and Discussion.

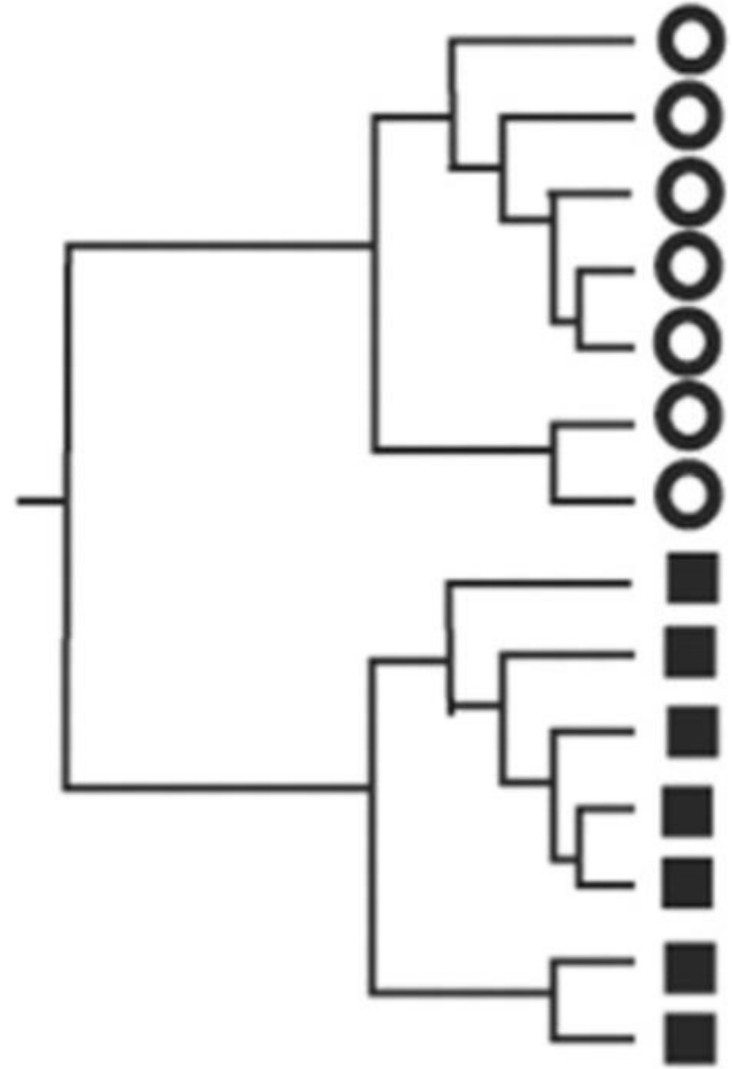
The UniFrac (unique fraction) Metric

- Measures the phylogenetic distance between sets of taxa in a phylogenetic tree as the fraction of the branch length of the tree that leads to descendants from either on environment or the other but not both.
- rRNA is used purely as a phylogenetic marker.

A.

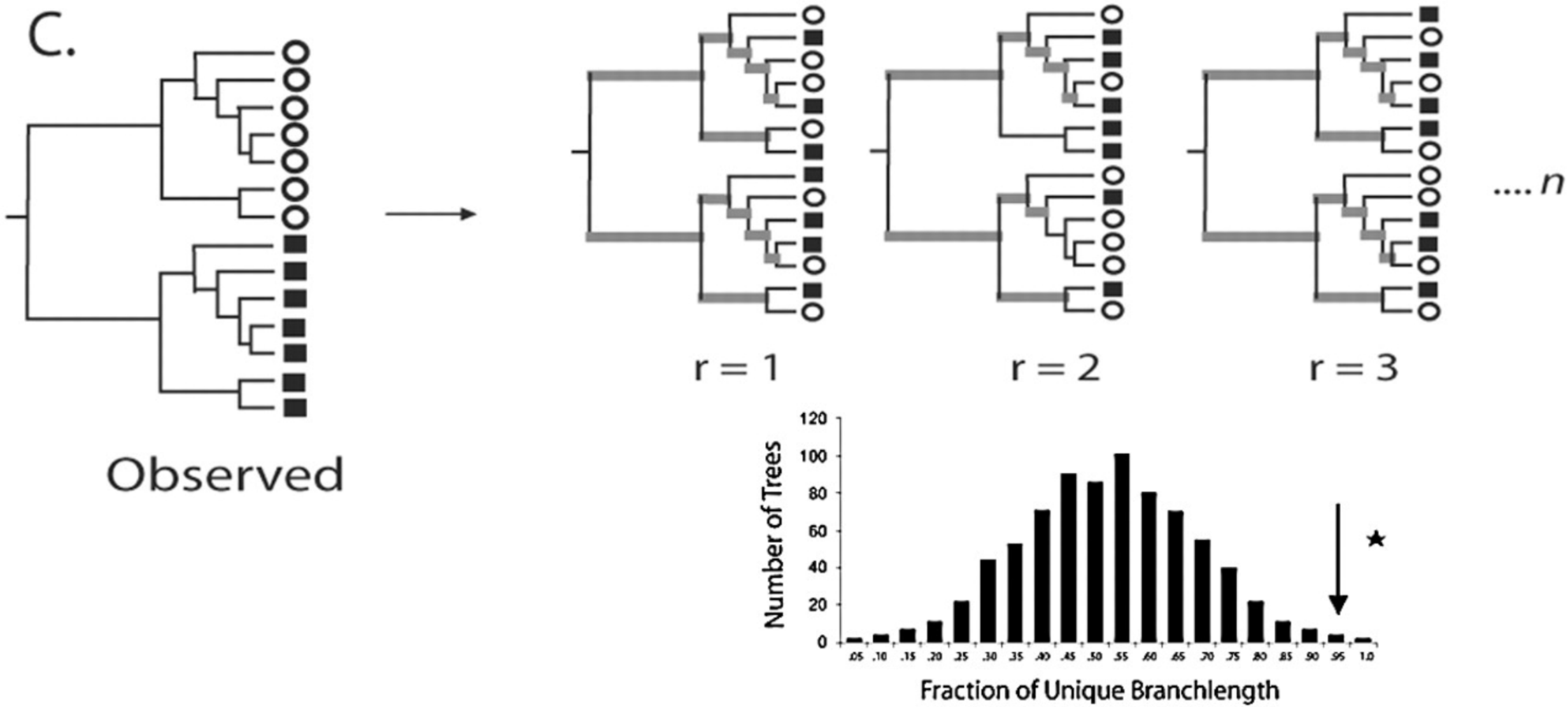


B.

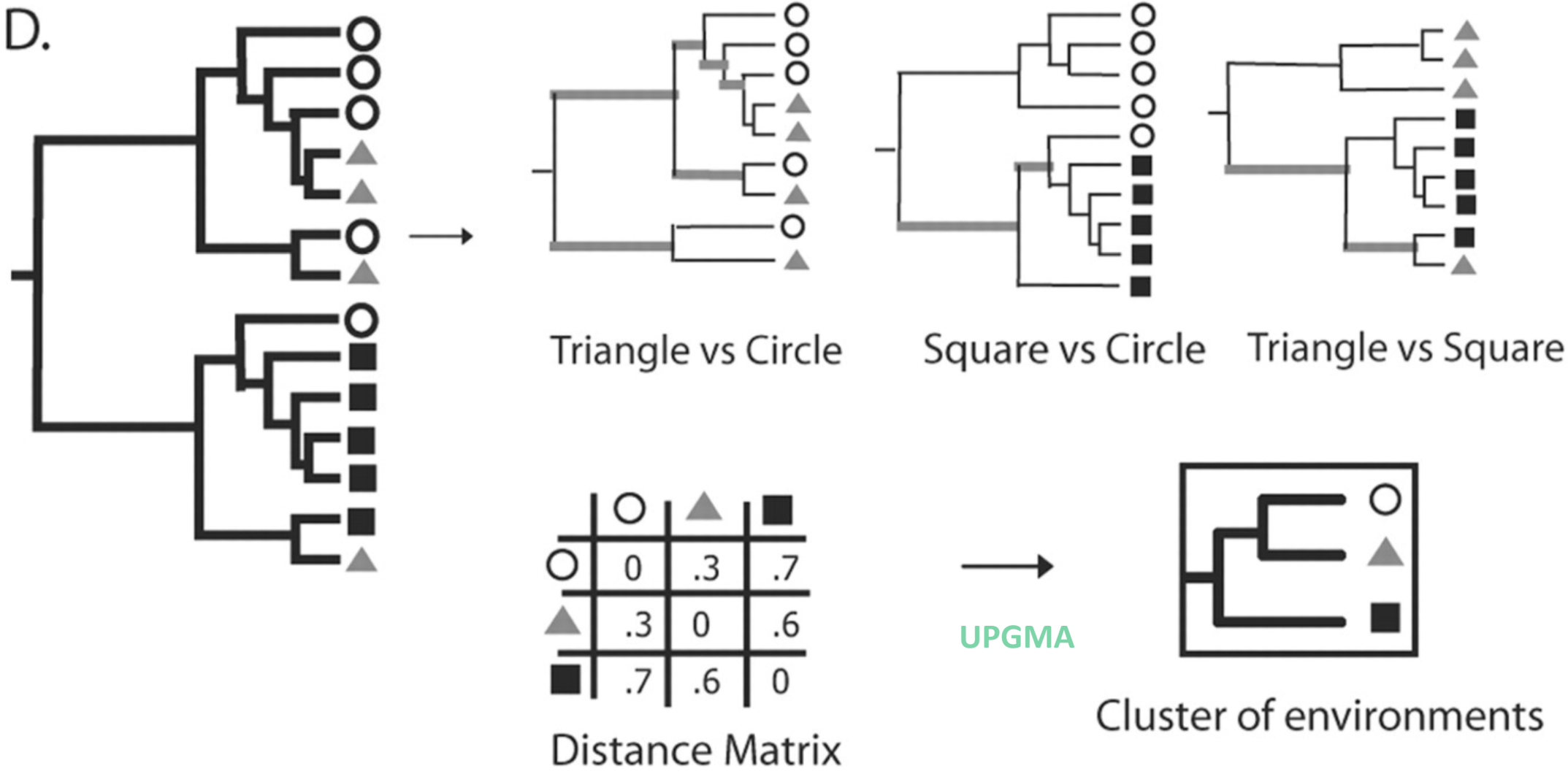


Monte Carlo Simulation

- UniFrac can be used to determine whether two communities differ significantly by using Monte Carlo simulation.
- Two communities are considered different if the fraction of the tree unique to one environment is greater than would be expected by chance.

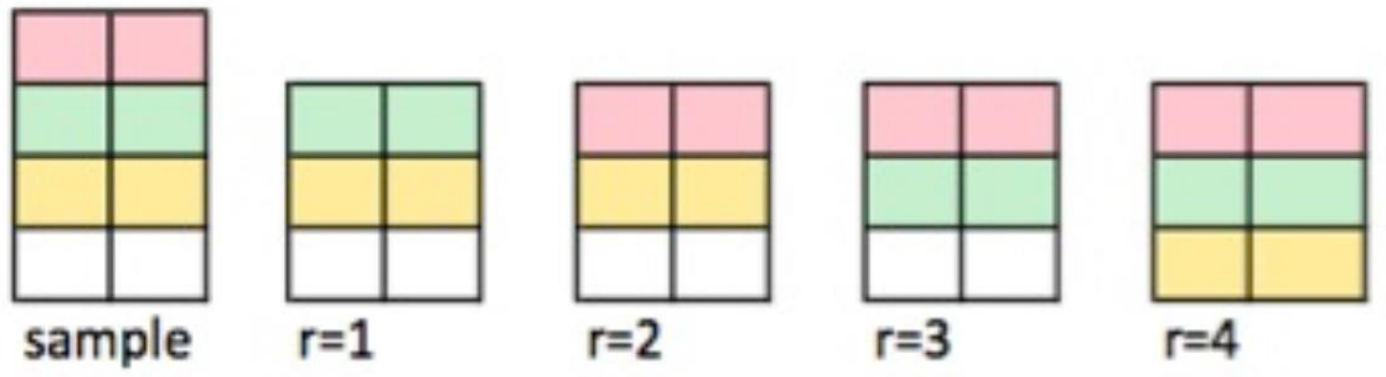

C.
Observed
r = 1
r = 2
r = 3
.... n
Number of Trees
120
100
80
60
40
20
0
.05 .10 .15 .20 .25 .30 .35 .40 .45 .50 .55 .60 .65 .70 .75 .80 .85 .90 .95 1.0
Fraction of Unique Branchlength

Producing distance matrix with UniFrac



Jackknifing

- After applying UniFrac to get distance matrix and UPGMA to get phylogenetic tree, Jackknifing can be applied to assess confidence in the node of UPGMA.
- the jackknife is a resampling technique especially useful for variance and bias estimation.



Outline

- ✓ Background and motivation.
- ✓ Leading Questions.
- ✓ UniFrac – The Method.
- ❑ **Results.**
- ❑ Conclusions and Discussion.

Data

Environment type: S, marine sediment; W, water; I, ice.

Geographic location: R, Arctic; N, Antarctic; T, temarate; P tropical.

Derived from: U, environmental clones; C, cultured isolated.

Sample[a]	Reference	No. of sequences
SRU1	38	79
STU2	25	33
SNU3	4	36
SNC4	4	31
SNU5	5	101
SNU6	5	146
SNU7	5	231
WRU8	2	87
WTC9	8	36
WTU10	22	75
WTC11	22	21
WTU12	1	544
WPU13	11	17
WPU14	11	40
INC15	3	58
IRU16	6	62
IRC17	6	109
INU18	6	20
INC19	6	87
INU20	7	75

UniFrac P values

- Based on comparisons to 1000 trees.
- Results are listed only if the P value ≥ 0.05.

Sample	Compared sample(s) (*P* value)
SRU1	SNU3 (0.118), STU2 (0.111)
STU2	SNU3 (0.201), SRU1 (0.111), SNU5 (0.066), SNU6 (0.107)
SNU3	SNU6 (0.802), SNU7 (0.070), SRU1 (0.118), STU2 (0.201)
SNC4	WTC11 (0.105), SNU5 (0.053)
SNU5	SNC4 (0.053), STU2 (0.066)
SNU6	SNU7 (0.394), SNU3 (0.802), STU2 (0.107)
SNU7	SNU6 (0.394), SNU3 (0.070)
WRU8	
WTC9	WTC11 (0.639), INU20 (0.076), INC19 (0.155), INU18 (0.097)
WTU10	
WTC11	SNC4 (0.105), WTC9 (0.639), INC19 (0.055)
WTU12	
WPU13	WPU14 (0.238)
WPU14	WPU13 (0.238)
INC15	
IRU16	INU18 (0.257)
IRC17	
INU18	WTC9 (0.097), INU20 (0.055), INC19 (0.233), IRC17 (0.257)
INC19	WTC11 (0.055), WTC9 (0.155), INU18 (0.233)
INU20	WTC9 (0.076), INU18 (0.055)

UPGMA cluster of marine samples

- The number of sequences that represent each environment is indicated next to the sample name.

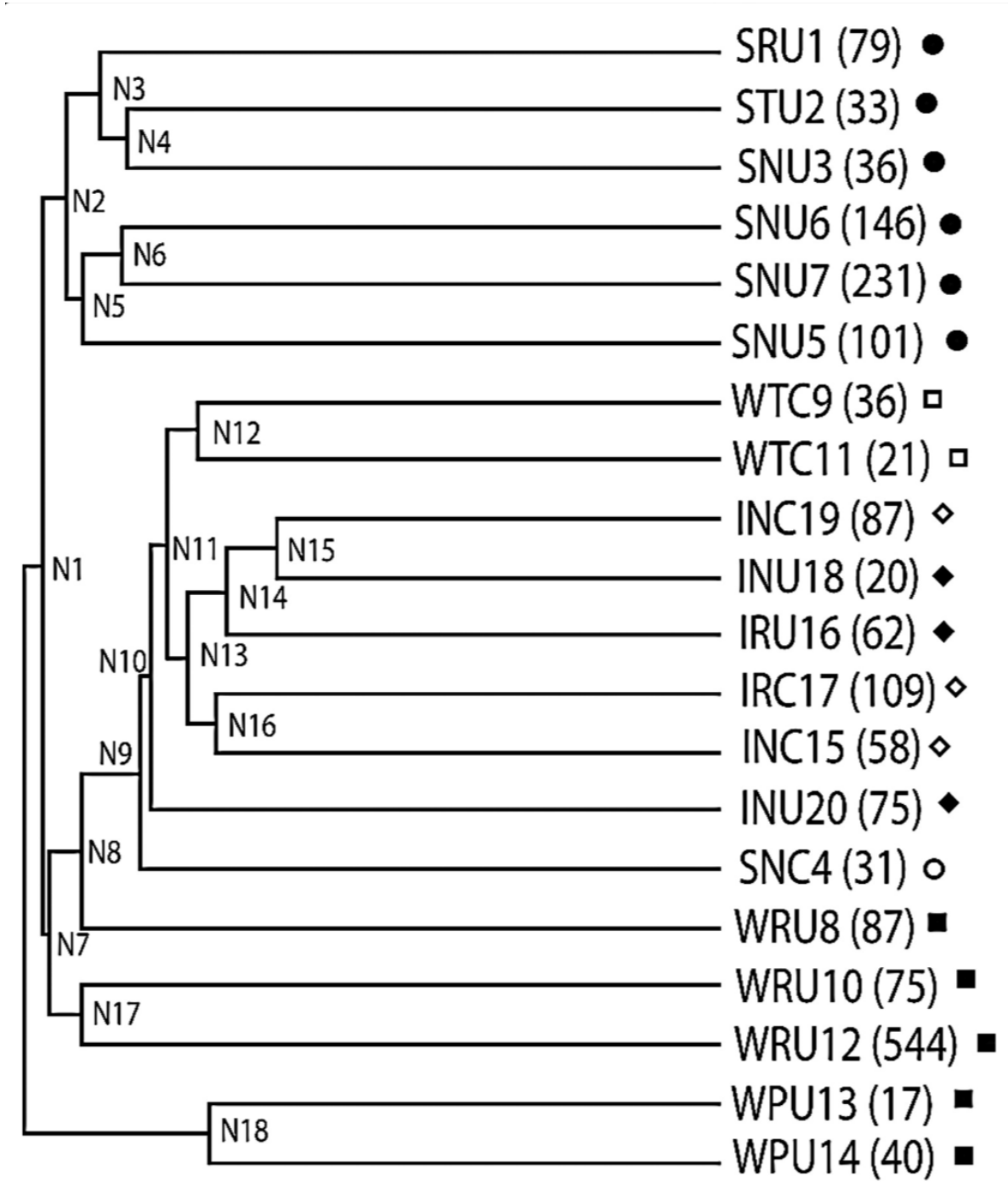


TABLE 3. UPGMA jackknifing results

Node	% of trials with node[a]					
	17	20	31	36	40	58
N1	3	14	31	27	12	NA
N2	8	1	29	33	48	63
N3	1	8	7	11	NA	NA
N4	14	16	11	NA	NA	NA
N5	1	0	0	1	27	37
N6	27	36	57	67	53	63
N7	23	23	36	44	52	66
N8	22	17	17	39	31	37
N9	52	58	64	NA	NA	NA
N10	8	16	79	96	94	100
N11	6	12	40	46	NA	NA
N12	13	31	NA	NA	NA	NA
N13	16	38	41	38	64	79
N14	34	50	29	23	12	6
N15	69	77	NA	NA	NA	NA
N16	18	40	27	28	28	21
N17	24	35	43	46	37	50
N18	97	NA	NA	NA	NA	NA

The numbers show % of trails that the node occurred in when each environment was represented by only 17,20,31,36,40,58 seq.

Leading questions in the research that UniFrac can answer

- **How**

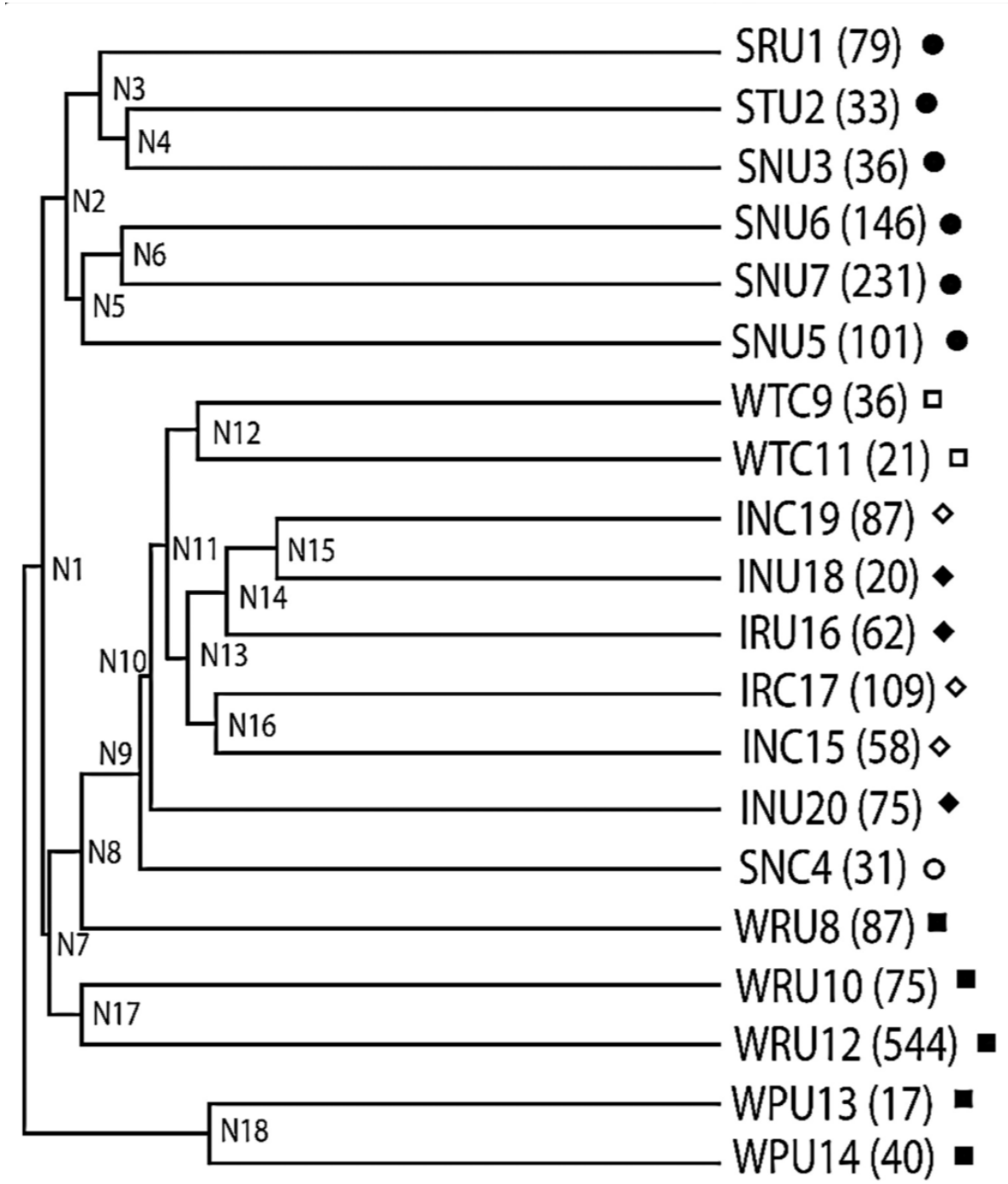
Exam ncultured

samp

Samples from cultured isolates resemble each other rather than uncultured samples from the same environment.

UPGMA cluster of marine samples

- The number of sequences that represent each environment is indicated next to the sample name.



Leading questions – Cont.

- **How**

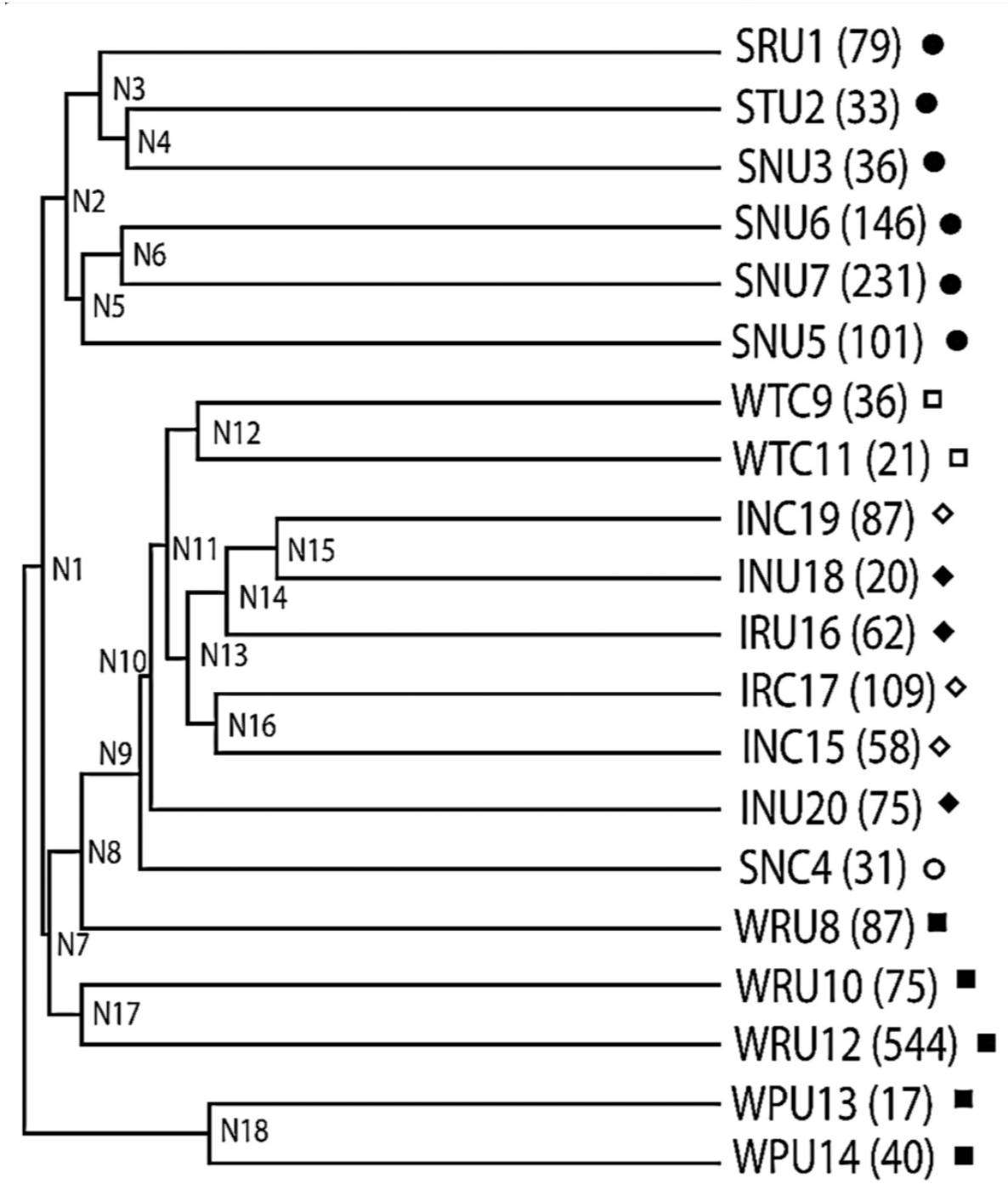
Comp ifferent

geogra

Geography plays a minor role in structing communities compared to the environment type.

UPGMA cluster of marine samples

- The number of sequences that represent each environment is indicated next to the sample name.



Leading questions – cont.

- **Are marine ice, sediment and seawater three distinct homogenous habitats?**

Uncultured bacterial communities in sediment and ice form distinct cluster but communities in seawater samples do not

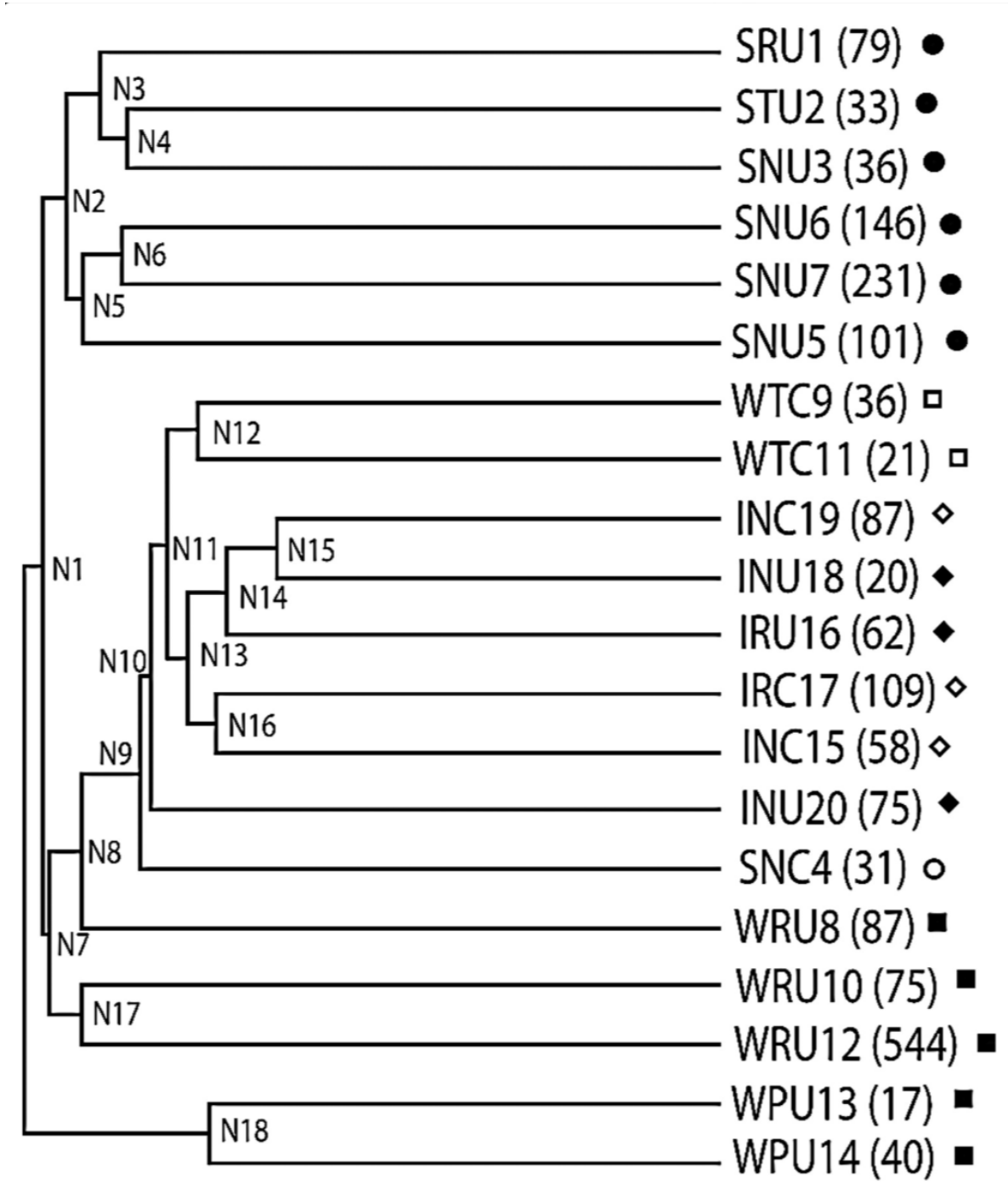
Thes tat types.

The A ent

bacterial communities that differ from one another.

UPGMA cluster of marine samples

- The number of sequences that represent each environment is indicated next to the sample name.



PCA

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components

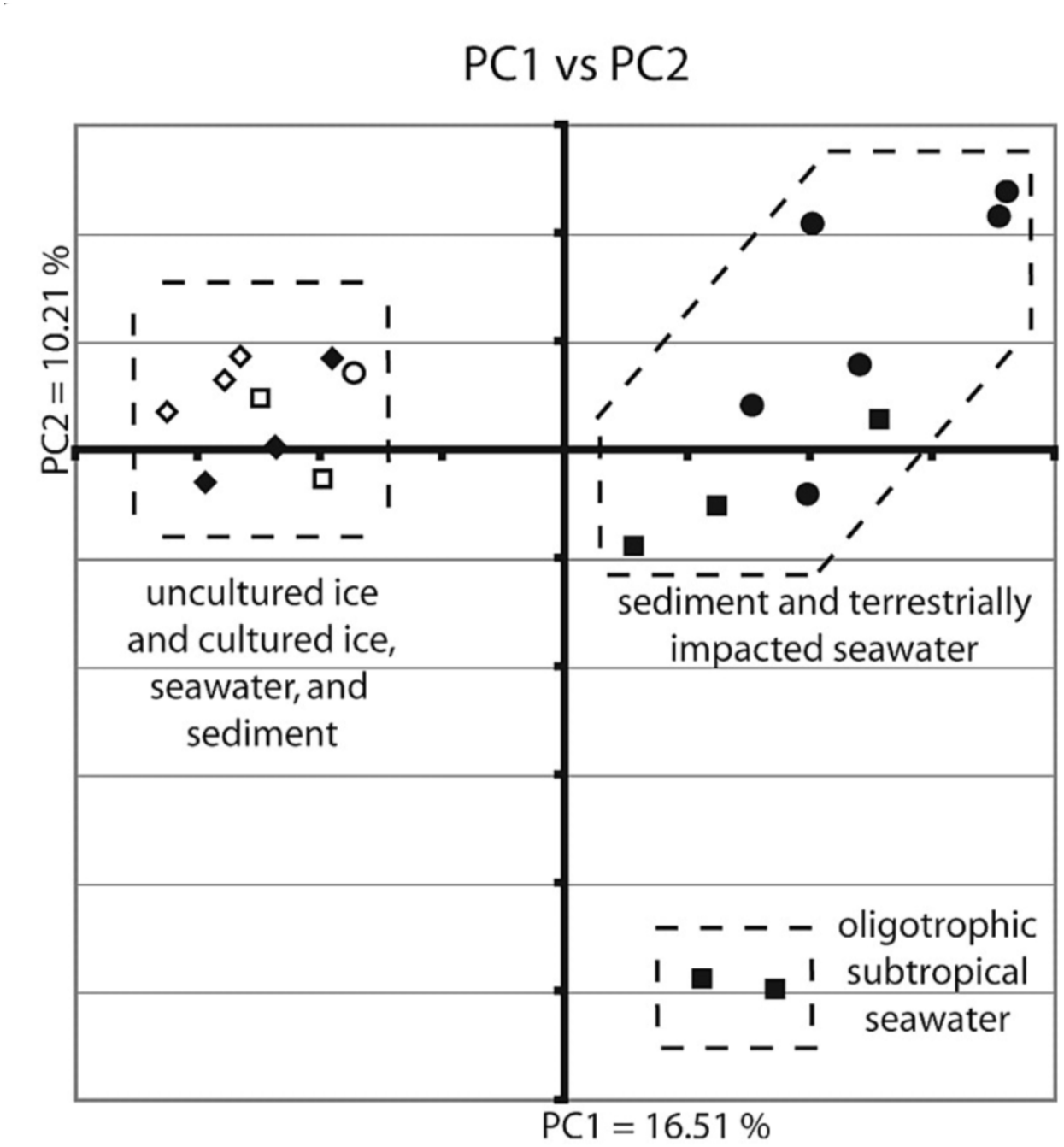
Ice

Seawater

Sediment

(Filled) – Uncultured

(Open) – Cultured



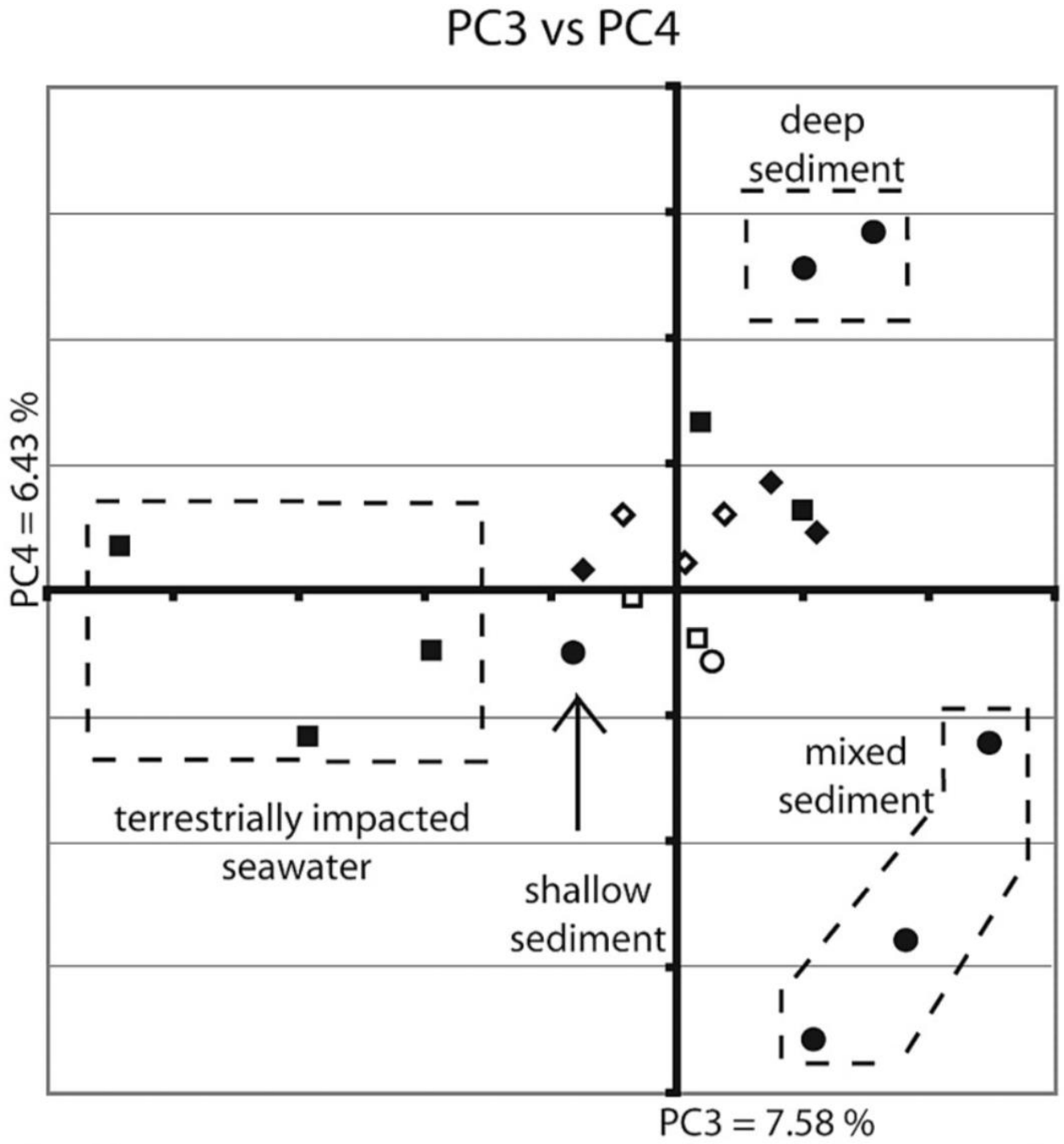
Ice

Seawater

Sediment

(Filled) – Uncultured

(Open) – Cultured



Outline

- ✓ Background and motivation.
- ✓ Leading Questions.
- ✓ UniFrac – The Method.
- ✓ Results.
- ❑ **Summary Conclusions and Discussion.**

Summary and Conclusions

- Phylogenetic information can be very useful to understand microbial communities.
- In contrast to previous methods, UniFrac utilize the phylogenetic information and also can compare many environments simultaneously.
- The ability of UniFrac to integrate sequence data from many diverse studies make it suitable for large-scale comparations between environments.
- Promising potential to shed light on biological factors that structure microbial communities.

Summary and Conclusions - Cont.

- Very known and popular method.

UniFrac: a New Phylogenetic Method for Comparing ...

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1317376/ ▾ לדף המתורגם

מאת C Lozupone - 2005 - צוטט על ידי 4136 - מאמרים בנושא זה

This method, **UniFrac**, measures the phylogenetic distance between sets of taxa in a .... While three of the sediment **papers** reported sequences from multiple ...

- Open source
- A lot of discussions and related papers published after this paper.

Discussion

- The method is not considering the number of samples per community. Maybe this information can be useful to get more accurate results?
- The number of sequences per sample is rather low. Is it enough for the conclusions that presented in the paper?