HU CFAR QIIME Workshop
15 September 2014

www.qiime.org

John Chase
Yoshiki Vazquez-Baeza
www.caporasolab.us
https://knightlab.ucsd.edu
www.applied-bioinformatics.org

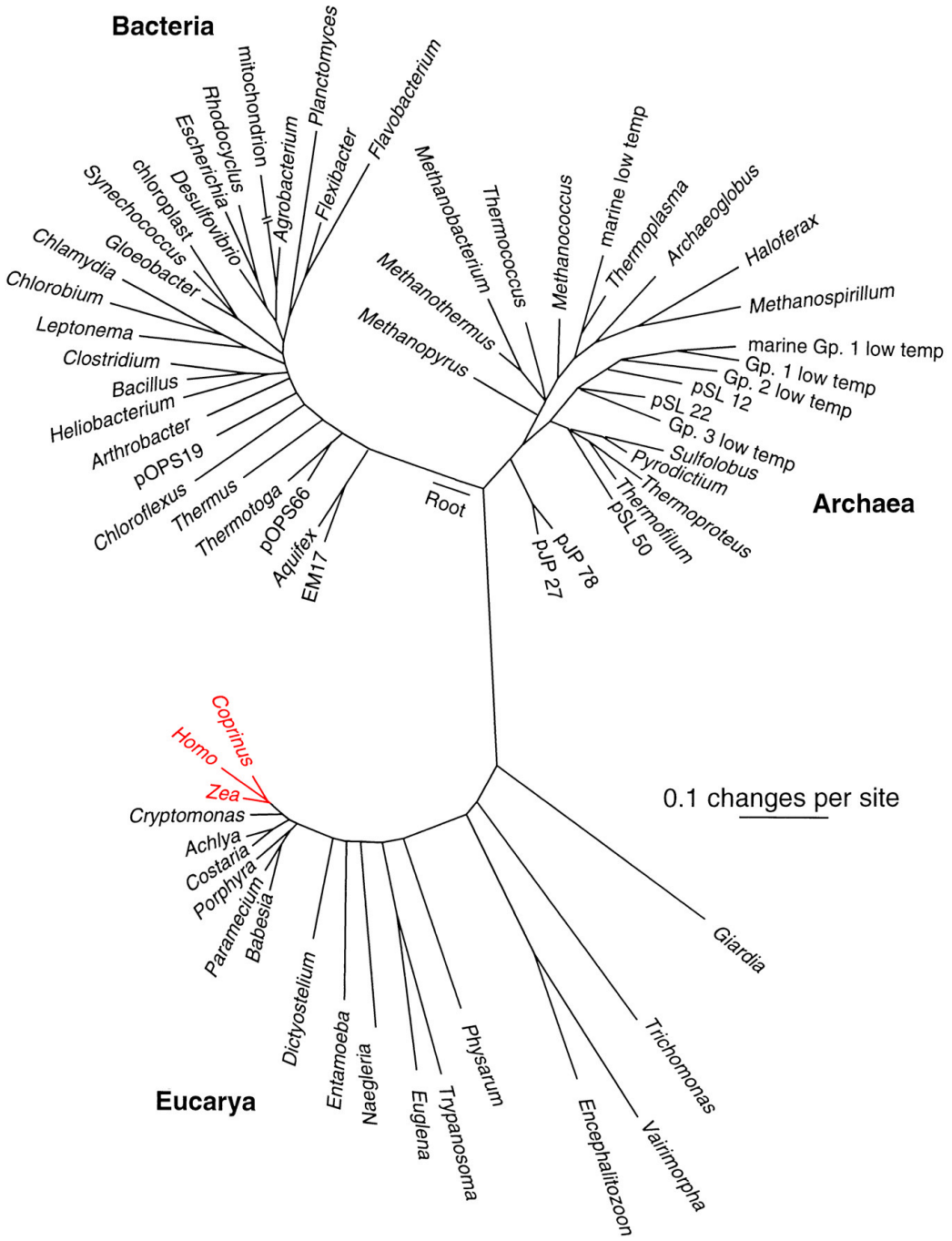
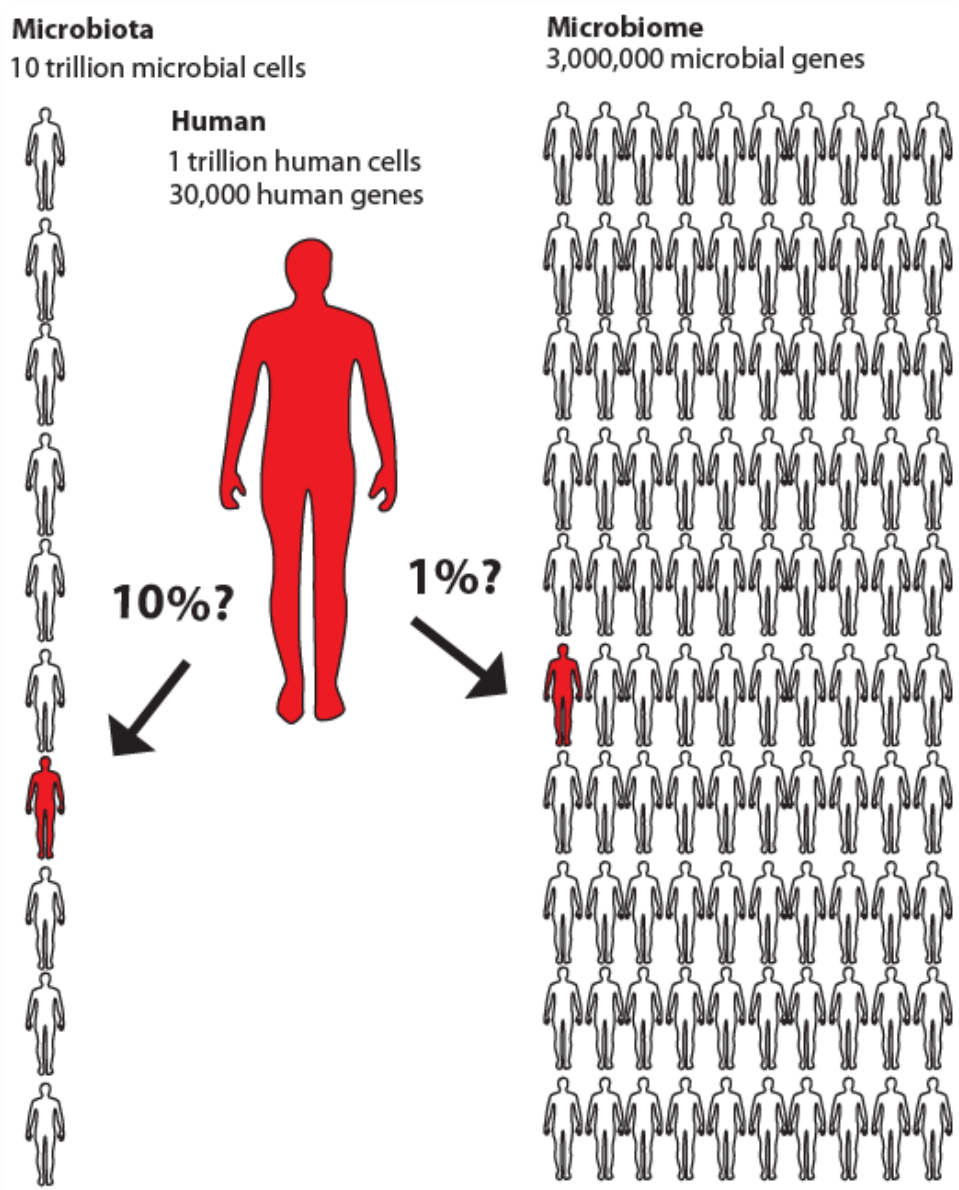
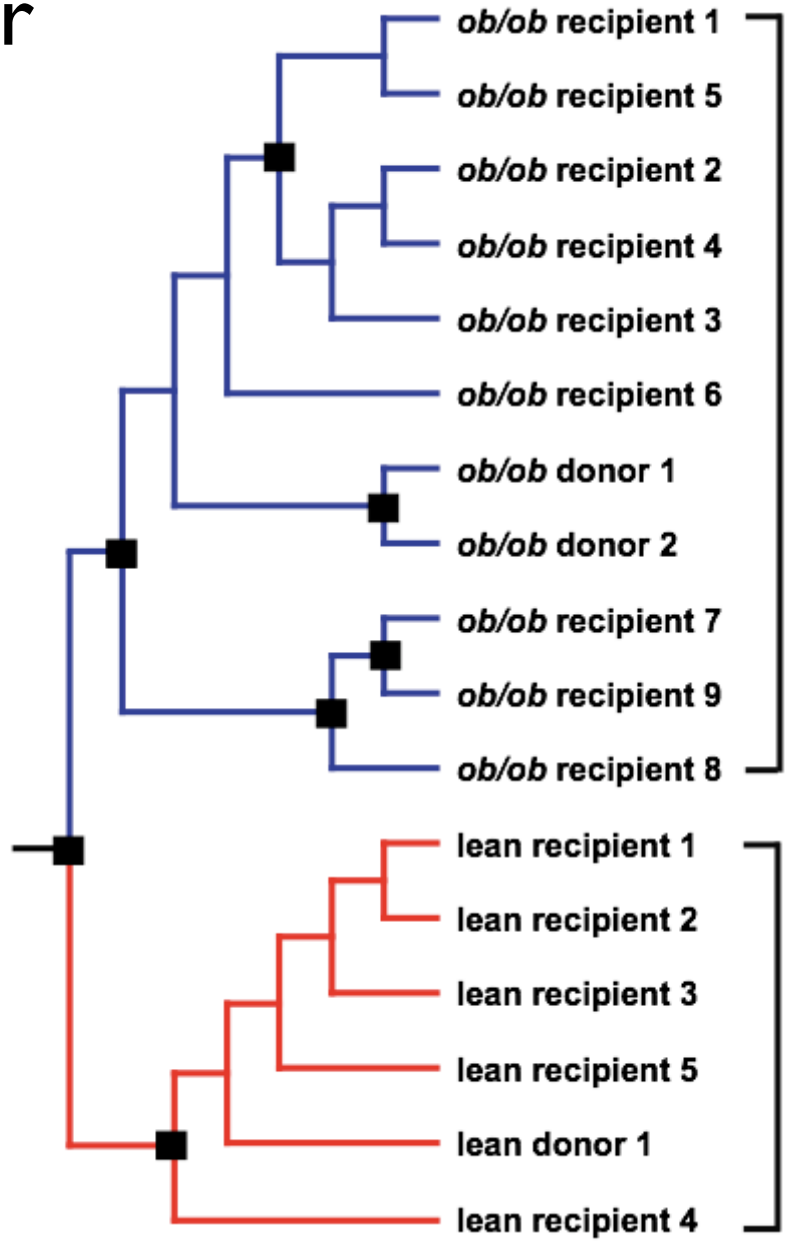


Image credit: Norman Pace



Microbial Ecology of the Gastrointestinal Tract
Annual Review of Microbiology **31**: 107–33. Savage, D. C. (1977).

Do differences in our microbiota matter?



Peter J. Turnbaugh et al., Nature 2006
An obesity-associated gut microbiome with increased capacity for energy harvest

Microbes rarely live or act alone.

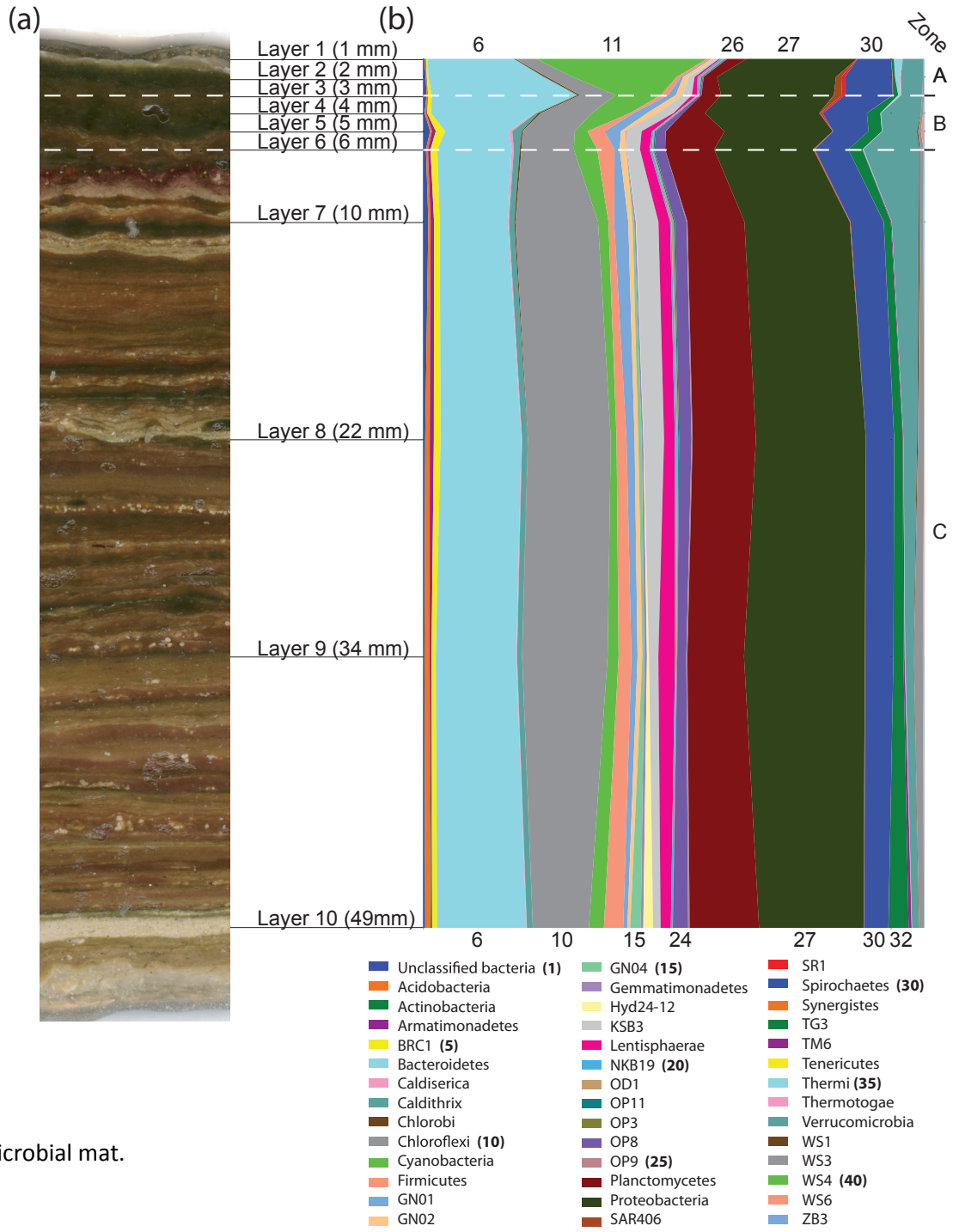


Image source:
Phylogenetic stratigraphy in the Guerrero Negro hypersaline microbial mat.
Harris, Caporaso *et al.* (2012)
International Society for Microbial Ecology Journal

Culturing microbes is hard

Back of the envelope calculation: less than 13% of bacterial species* have a representative that has been grown in culture.

Many recent advances are based on *culture-independent* approaches for studying microbial communities.

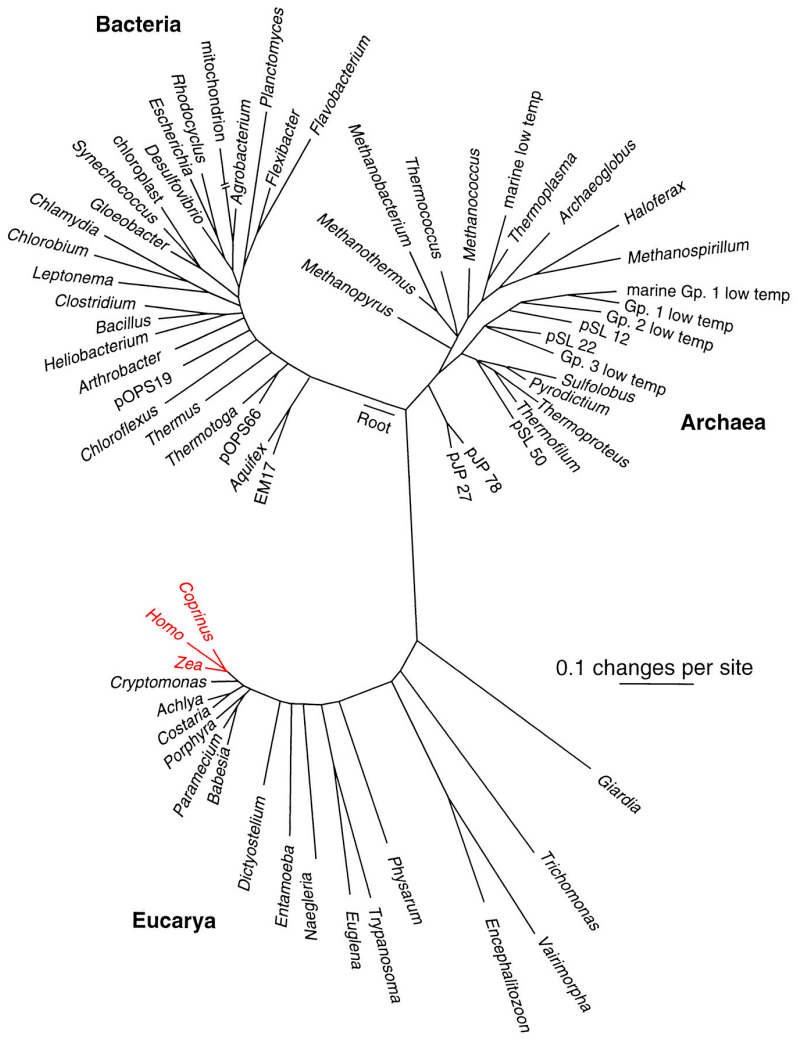
Bacillus anthracis in culture

* Defined as 97% OTUs in the Greengenes 13_5 reference database.

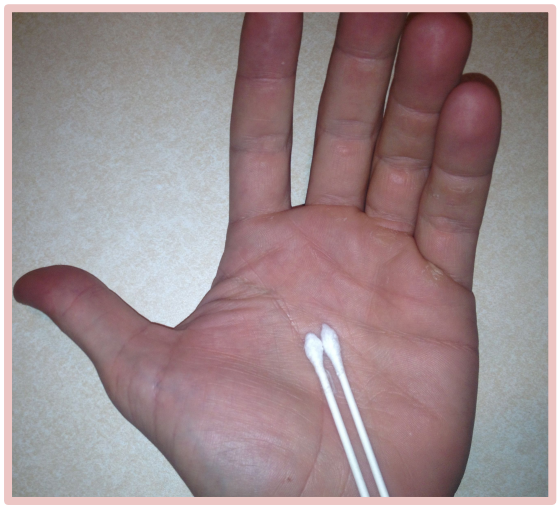
Culture-independent investigation of microbial communities

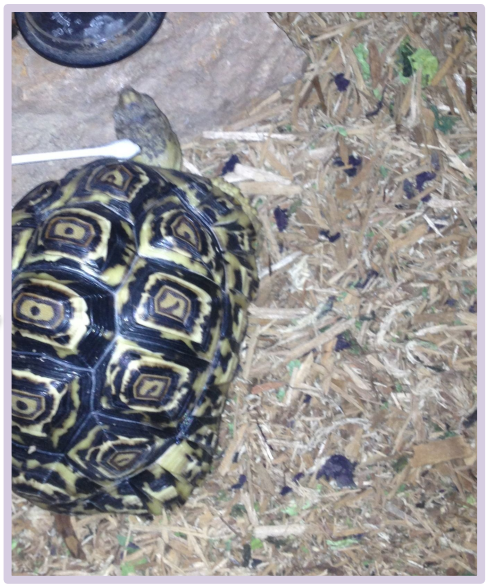
All cellular life has a shared evolutionary history, and some genes are shared by all organisms.

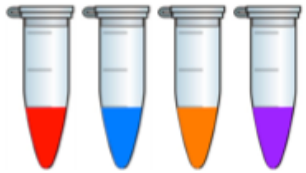
The sequence of those genes can be used as a *genetic fingerprint* for different organisms.

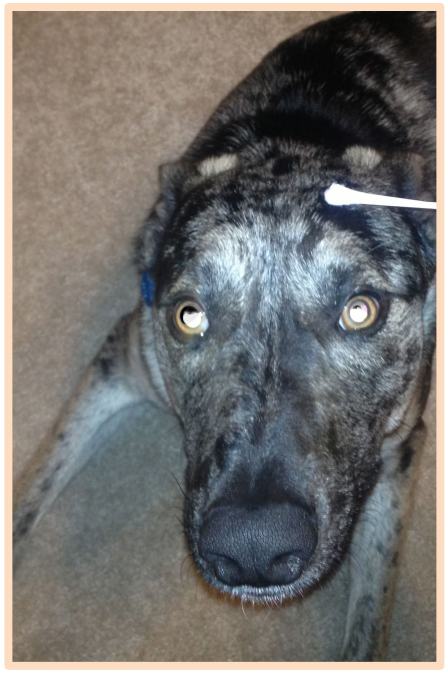


Collect samples









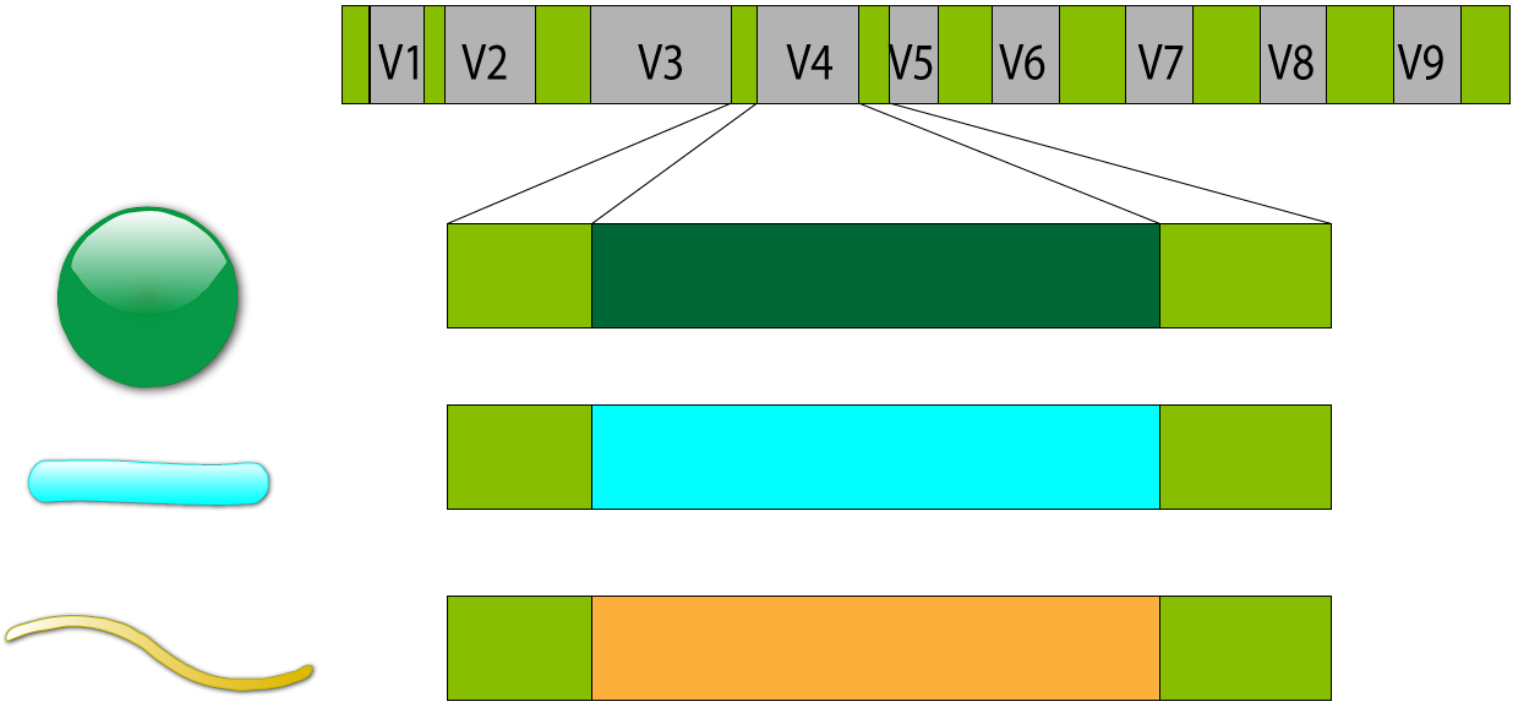
Extract DNA

(you can do this at home!)

Image source and instructions:
http://learn.genetics.utah.edu/content/labs/extraction/howto/

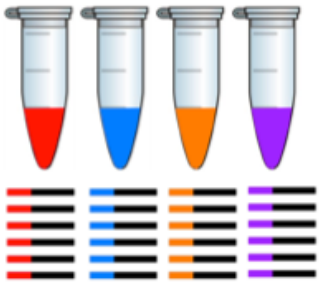
Isolate the *small subunit ribosomal RNA* gene to "fingerprint" different microbial organisms. (PCR)

Why 16S rRNA?



Sequence the rRNA from all samples on a “high-throughput” DNA sequencer

Per-sample rRNA

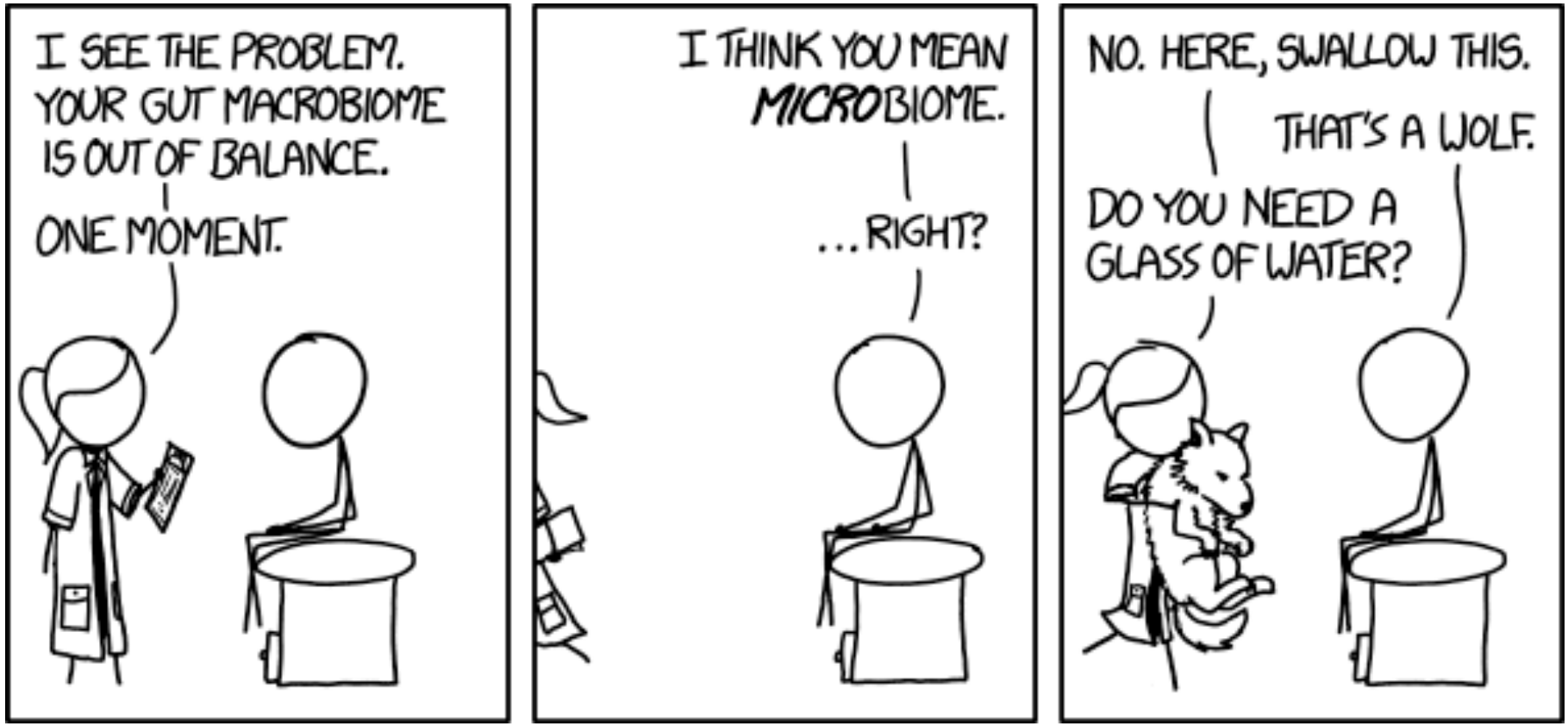


Pool samples and sequence

>GCACCTGAGGACAGGCATGAGGAA…
>GCACCTGAGGACAGGGGAGGAGGA…
>TCACATGAACCTAGGCAGGACGAA…
>CTACCGGAGGACAGGCATGAGGAT…
>TCACATGAACCTAGGCAGGAGGAA…
>GCACCTGAGGACACGCAGGACGAC…
>CTACCGGAGGACAGGCAGGAGGAA…
>CTACCGGAGGACACACAGGAGGAA…
>GAACCTTCACATAGGCAGGAGGAT…
>TCACATGAACCTAGGGGCAAGGAA…
>GCACCTGAGGACAGGCAGGAGGAA…

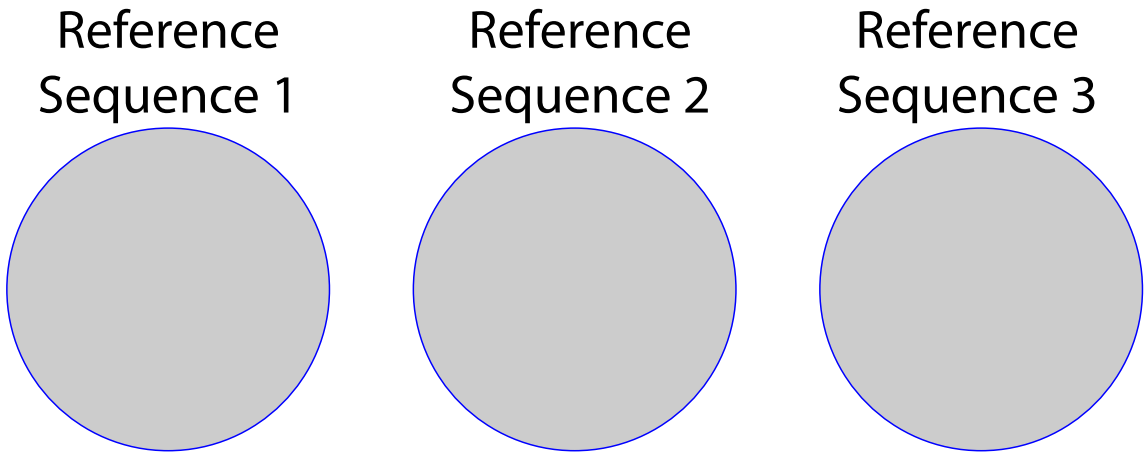
Micah Hamady, et al., Nature Methods, 2008.
Error-correcting barcodes for pyrosequencing hundreds of samples in multiplex.

Analyzing Microbial Diversity



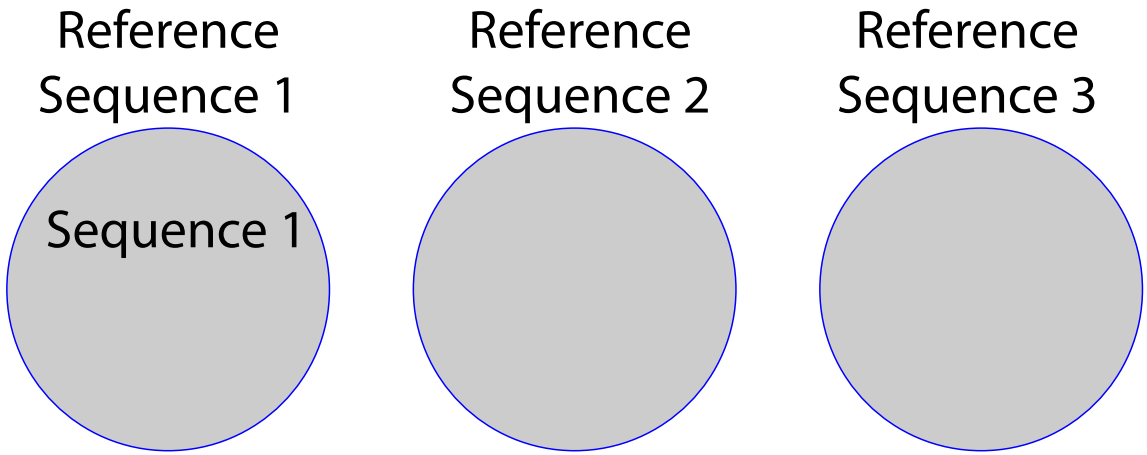
Sequences are clustered into Operational Taxonomic Units (OTUs)

Similar sequences are grouped together



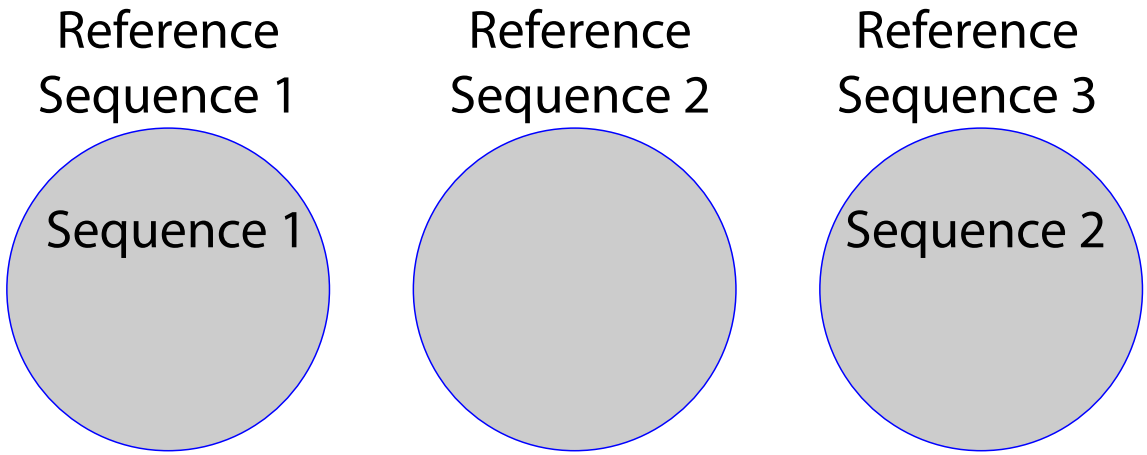
Sequence 1
Sequence 2
Seqeunce 3
Sequence 4
Sequence 5
Sequence 6

Similar sequences are grouped together



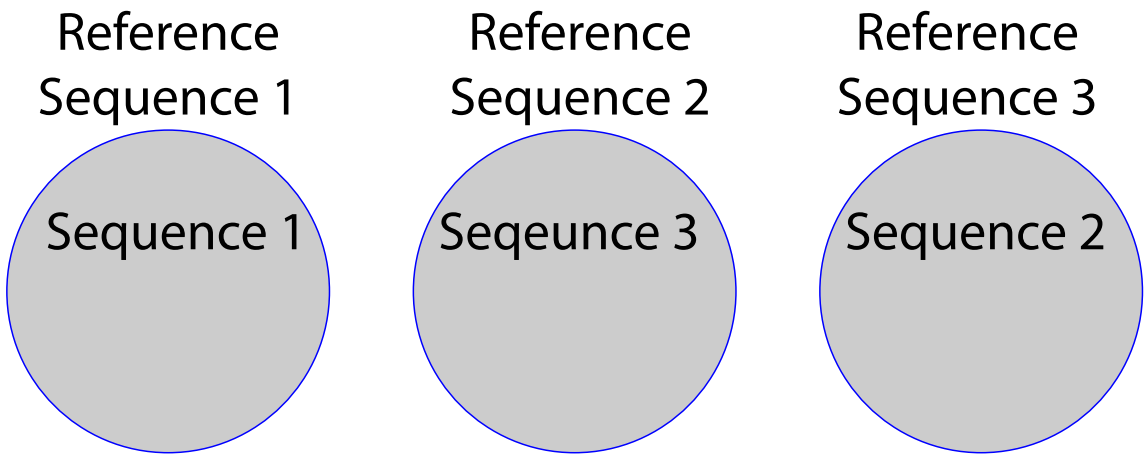
Sequence 2
Seqeunce 3
Sequence 4
Sequence 5
Sequence 6

Similar sequences are grouped together



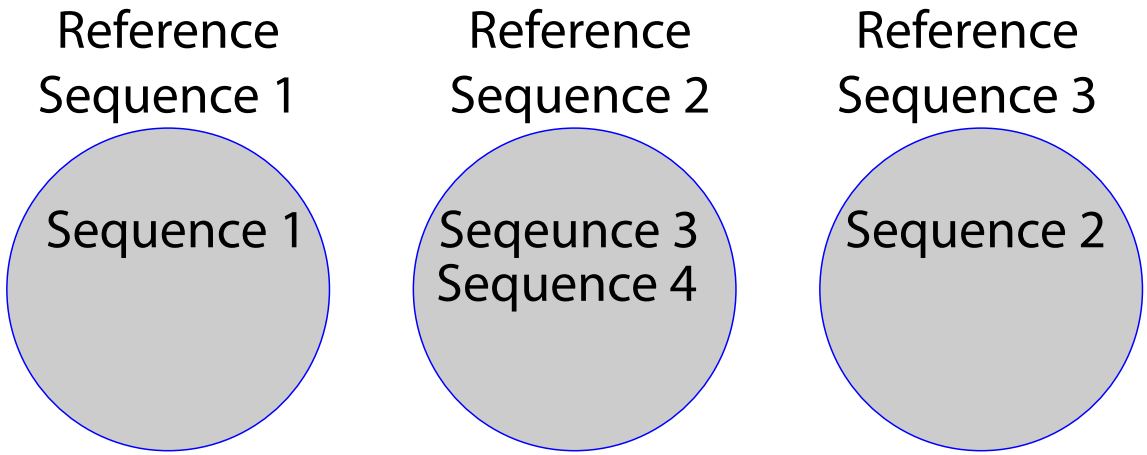
Seqeunce 3
Sequence 4
Sequence 5
Sequence 6

Similar sequences are grouped together



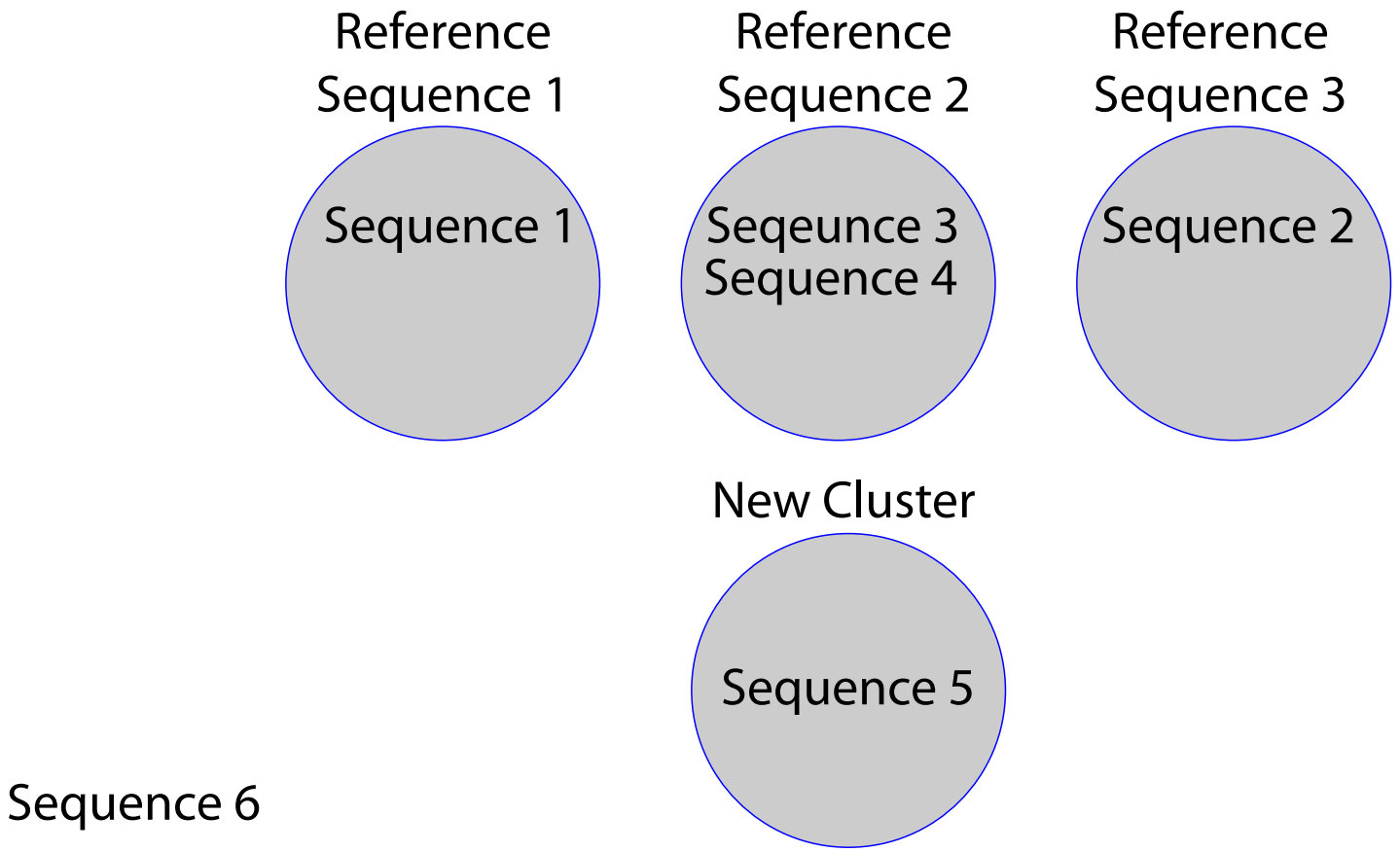
Sequence 4
Sequence 5
Sequence 6

Similar sequences are grouped together

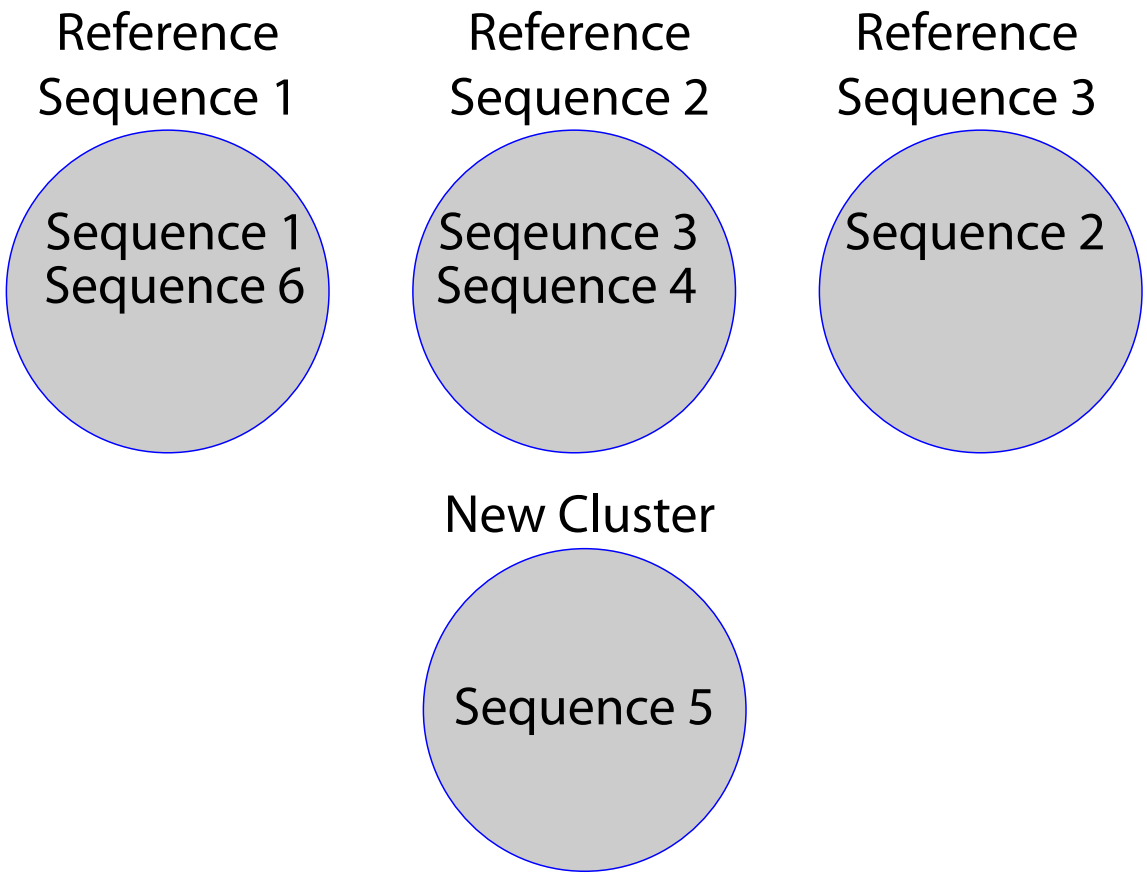


Sequence 5
Sequence 6

If a sequence does not match the database it is assigned to a new cluster



This continues until all sequences are assigned



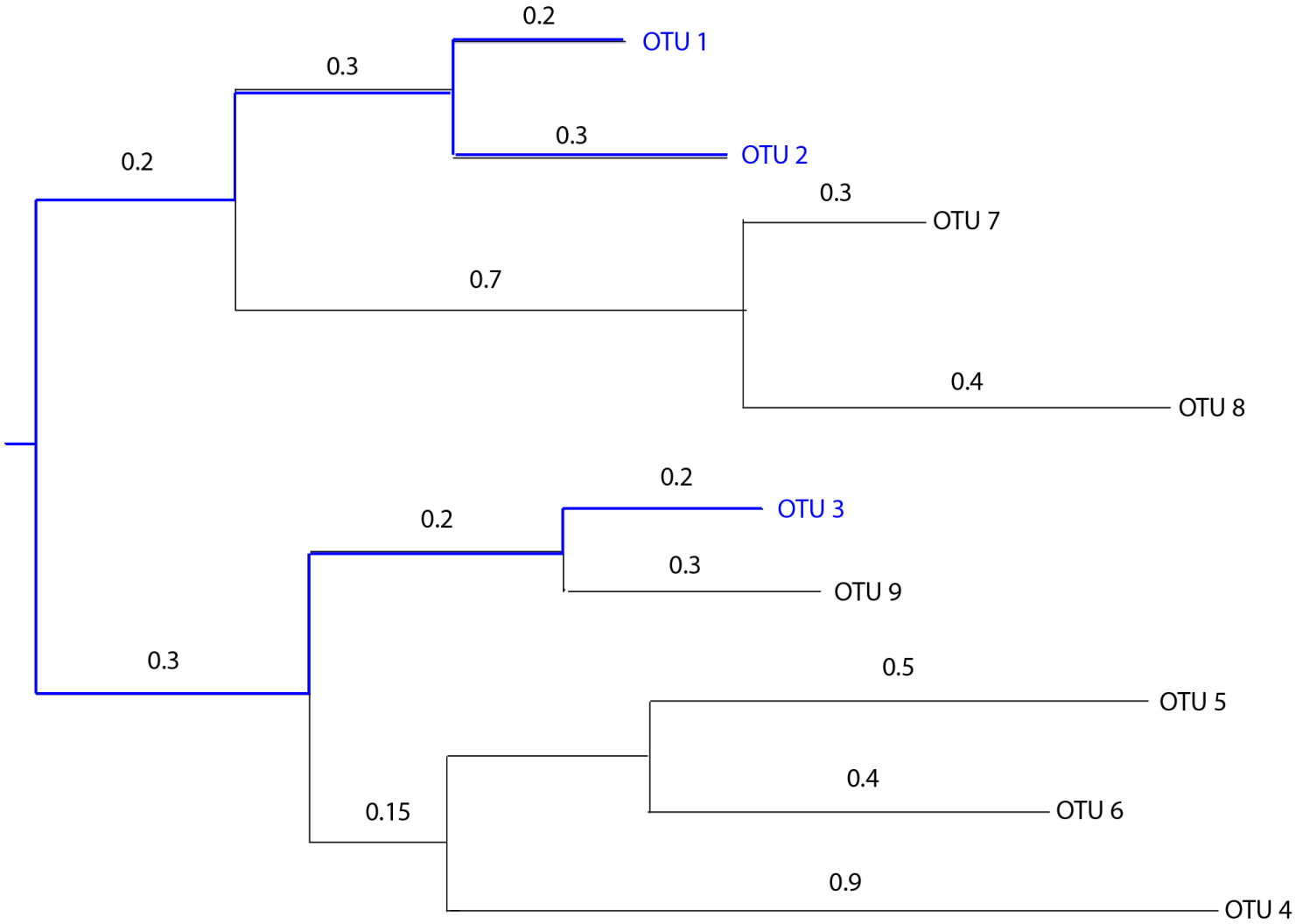
A representative sequence is chosen from each cluster and an OUT table is made

	Sample 1	**Sample 2**	**Sample 3**	**Sample 4**
OTU 1	0	0	4	6
OTU 2	1	7	0	1
OTU 3	4	0	4	2

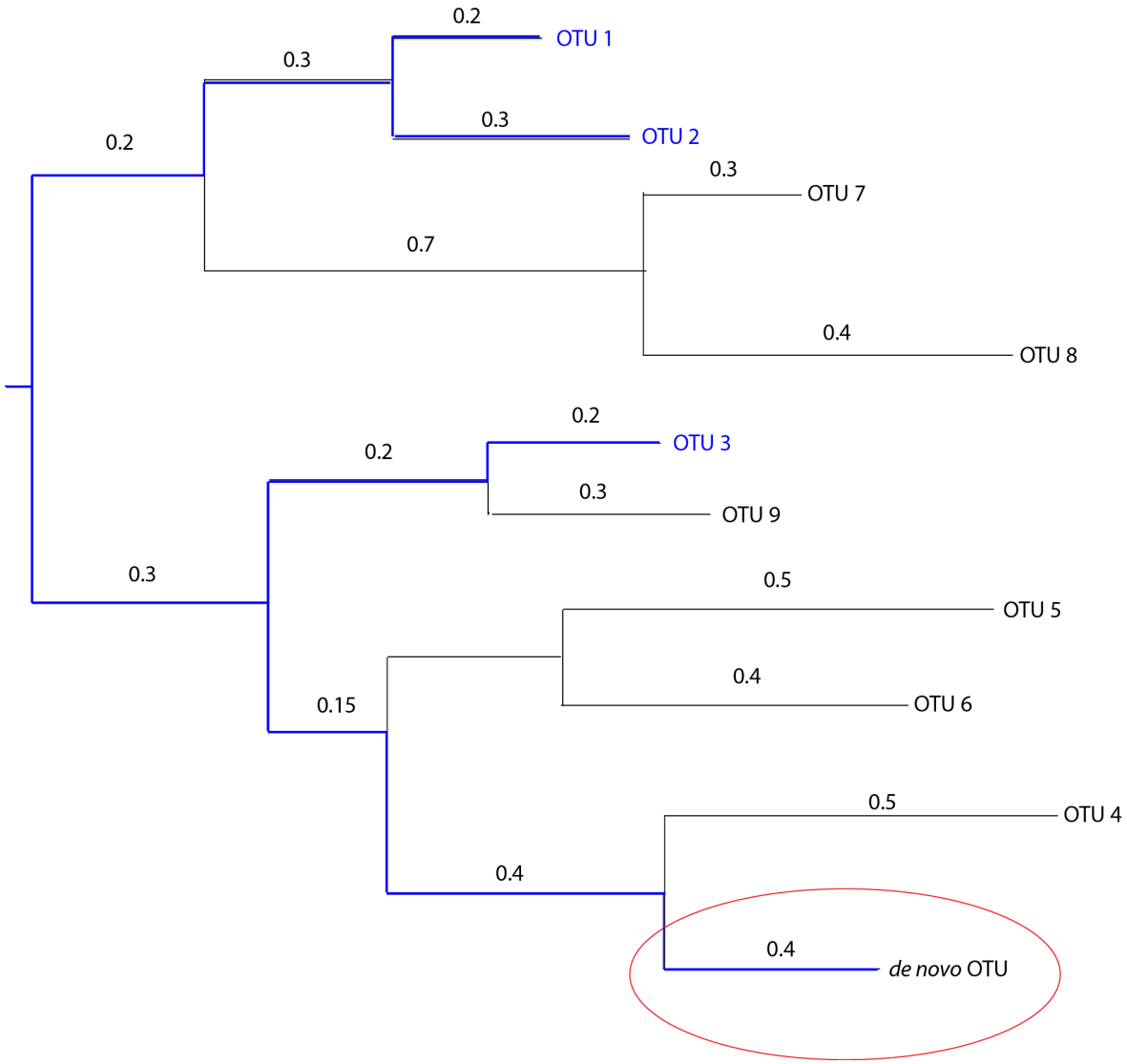
Metadata is recorded for each sample that is taken

	BodySite	**Date**	**Sex**
Sample1	gut	01/10/1970	male
Sample2	palm	01/10/1970	male
Sample3	gut	01/25/1970	female
Sample4	palm	01/25/1970	female

A table may be created from an existing phylogenetic tree



New "*de novo*" OTUs can be added to a tree, however phylogenetic analyses should not be performed with a tree that does not contain all OTUs



Comparing microbial communities

Who is there?

How many unique “species” are there?

How similar (or different are pairs of samples?

How do we compare the communities (Samples)?

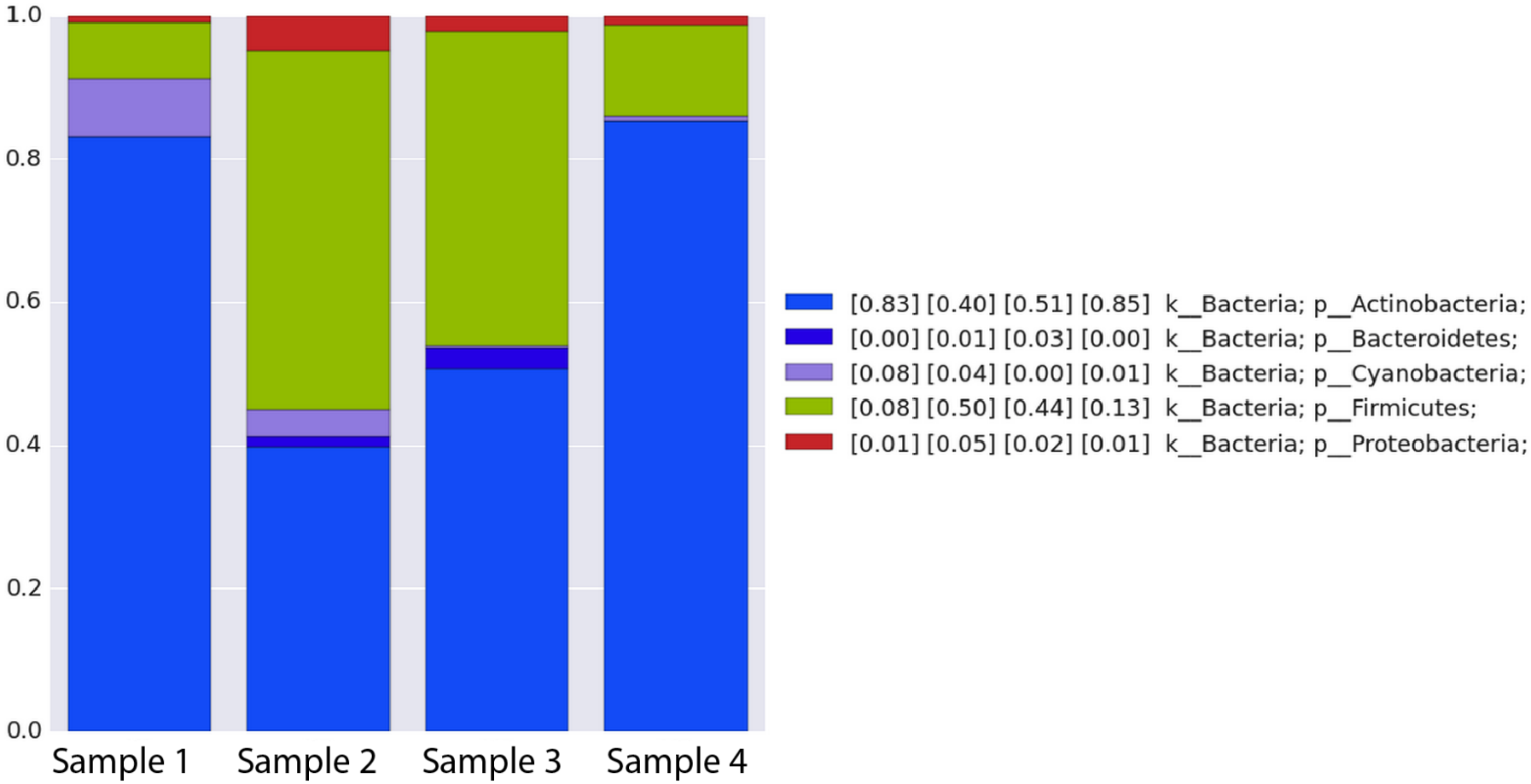
- Taxonomy Summary
- Alpha Diversity (within sample diversity)
- Beta Diversity (between sample diversity)

Some Terminology

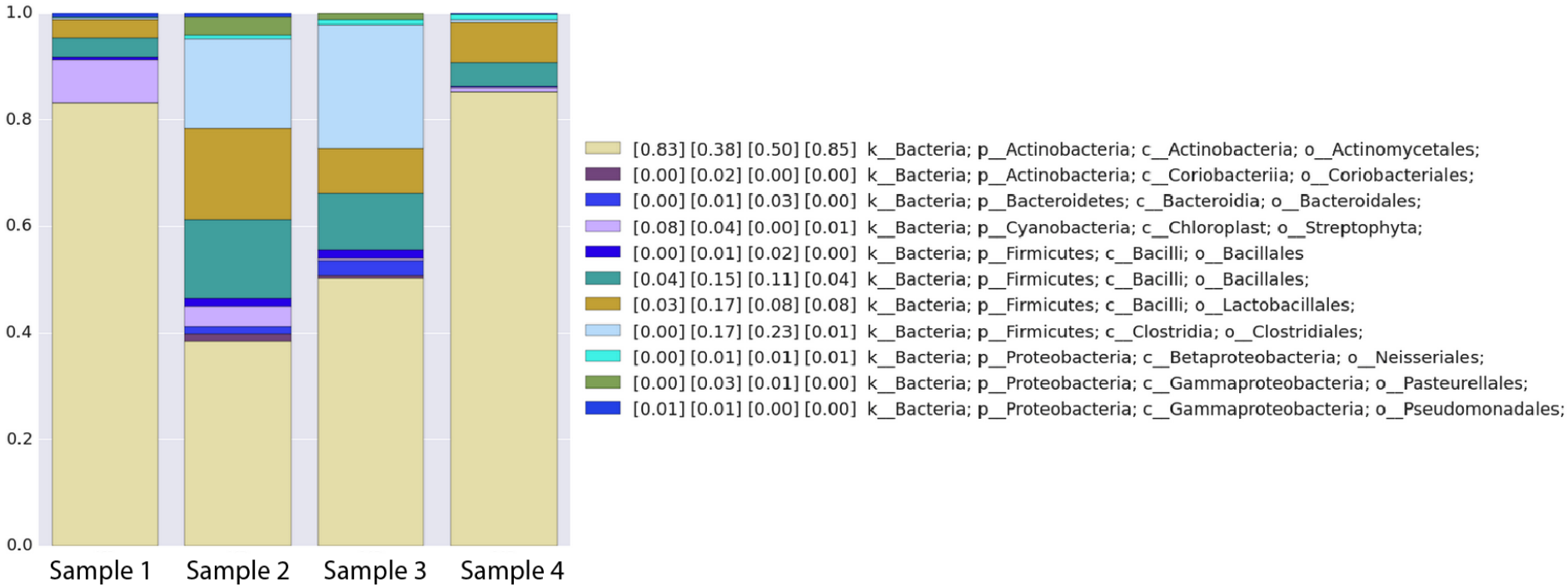
- Qualitative metric - Presence absence of an organism
- Quantitative metric - Takes into account the abundances of different organisms
- Non-phylogenetic metric - Treats all OTUs as being equally related
- Phylogenetic metric - Incorporates evolutionary relationships between the OTUs

Taxonomy Summary

Phylum level differences by sample

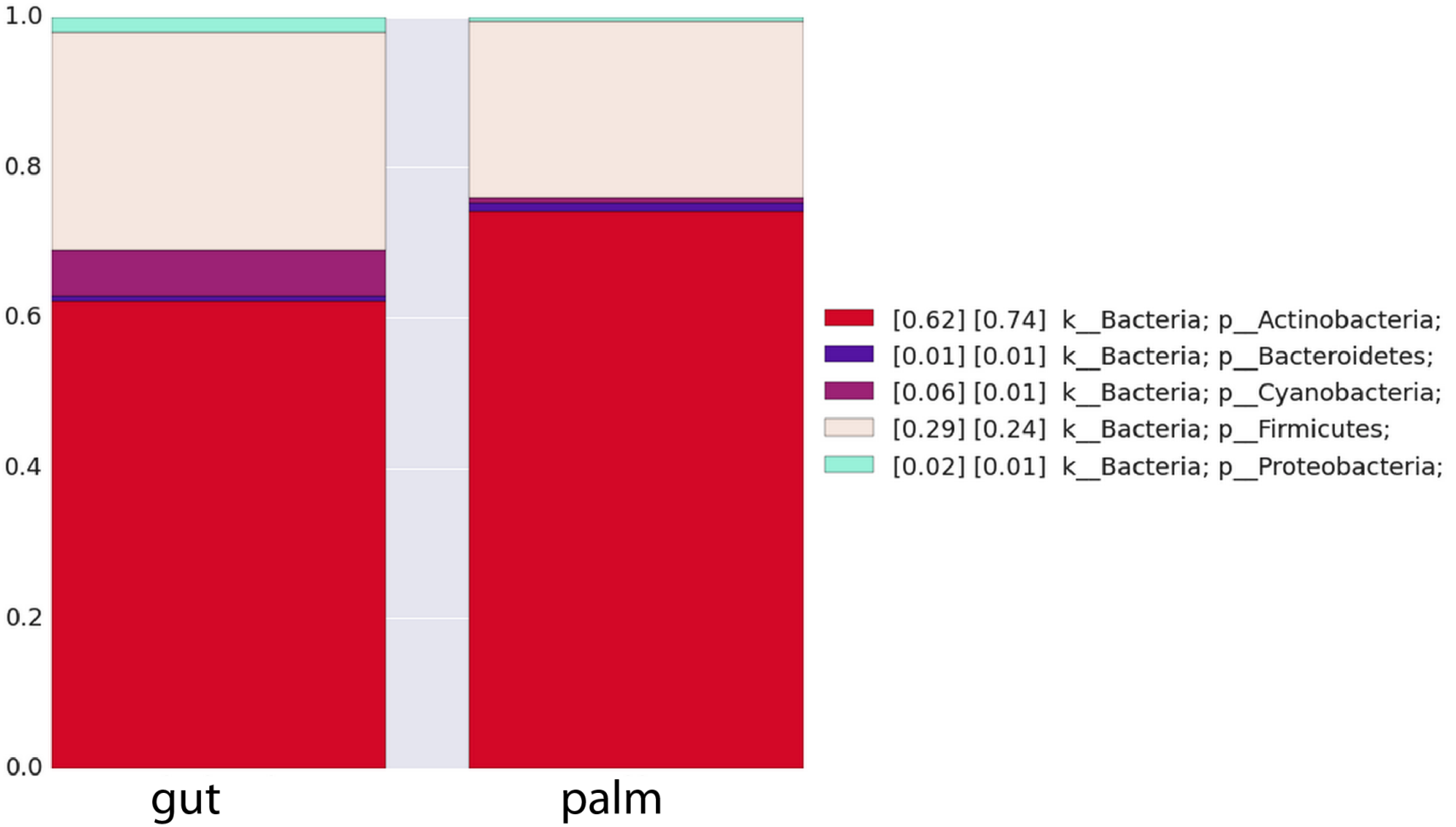


Order level differences by sample



Samples can be combined to look for category differences

(not real data!)



Are relative abundances of OTUs statistically different between groups?

- Kruskal-Wallis (non-parametric)
  - Groups must contain > 5 samples
  - Commonly used for marker gene surveys
- Anova (parametric)
  - Equal variance
  - Normality of residuals
  - Independence
  - Usually violated by marker gene surveys

- *Many others*

Kruskal-Wallis test for OTU differences

OTU	Test Stat	P-value	Bonferroni Corrected p-value	Gut mean	Palm mean
k__Bacteria; p__Firmicutes	10.86	0.001	0.01	1.12E-10	23.751
k__Bacteria; p__Bacteroidetes	5.07	0.005	0.05	0.003	7.767
k__Bacteria; p__Proteobacteria	2	0.078	0.078	1.151	1.616

Alpha Diversity
(within sample diversity)

Observed Species: a non-phylogenetic qualitative metric

Sample A

Pseudomonas aeruginosa
Pseudomonas argentinensis
Pseudomonas flavescens

Sample B

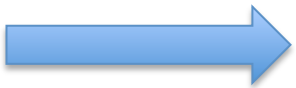
Pseudomonas aeruginosa
Pseudomonas argentinensis
Escherichia coli

Sample C

Pseudomonas aeruginosa
Giardia lamblia
Methanobrevibacter smithii

Observed species

Sample A 3
Sample B 3
Sample C 3



Conclusion

A = B = C

	Sample 1	**Sample 2**	**Sample 3**	**Sample 4**
OTU 1	0	0	4	6
OTU 2	1	7	0	1
OTU 3	4	0	4	2

Alpha diversity (Observed Species)

Sample A

Pseudomonas aeruginosa
Pseudomonas argentinensis
Pseudomonas flavescens

Sample B

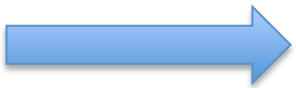
Pseudomonas aeruginosa
Pseudomonas argentinensis
Escherichia coli

Sample C

Pseudomonas aeruginosa
Giardia lamblia
Methanobrevibacter smithii

Observed species

Sample A 3
Sample B 3
Sample C 3



Conclusion

A = B = C

Phylogenetic Diversity: a qualitative phylogenetic metric

Sample A

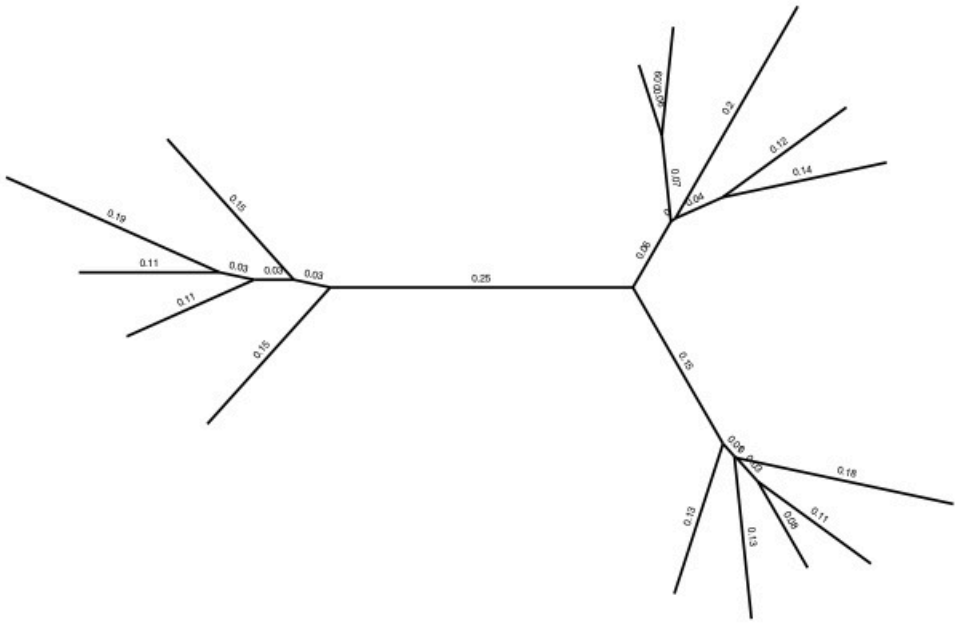
Pseudomonas aeruginosa
Pseudomonas argentinensis
Pseudomonas flavescens

Sample B

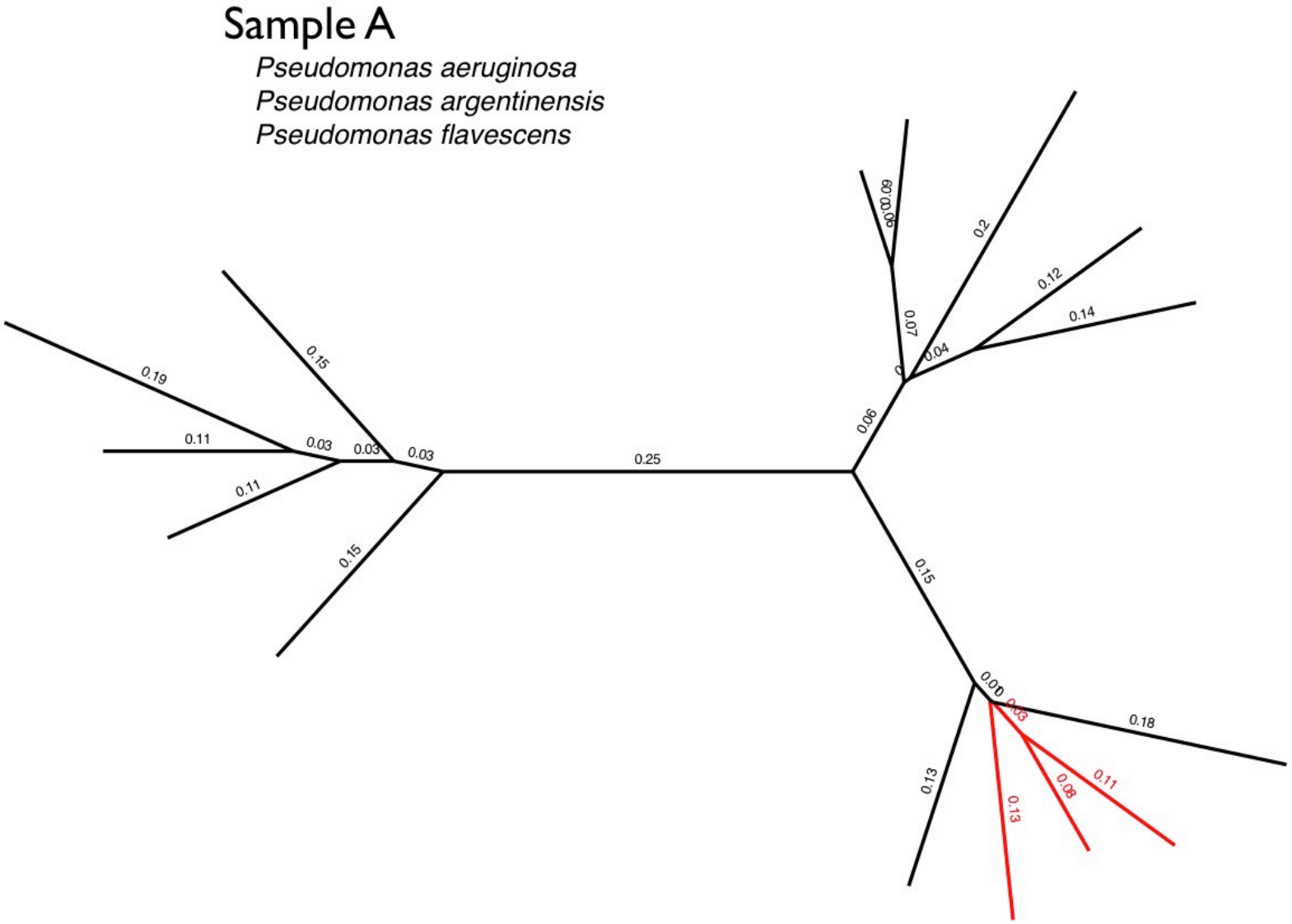
Pseudomonas aeruginosa
Pseudomonas argentinensis
Escherichia coli

Sample C

Pseudomonas aeruginosa
Giardia lamblia
Methanobrevibacter smithii



Alpha diversity

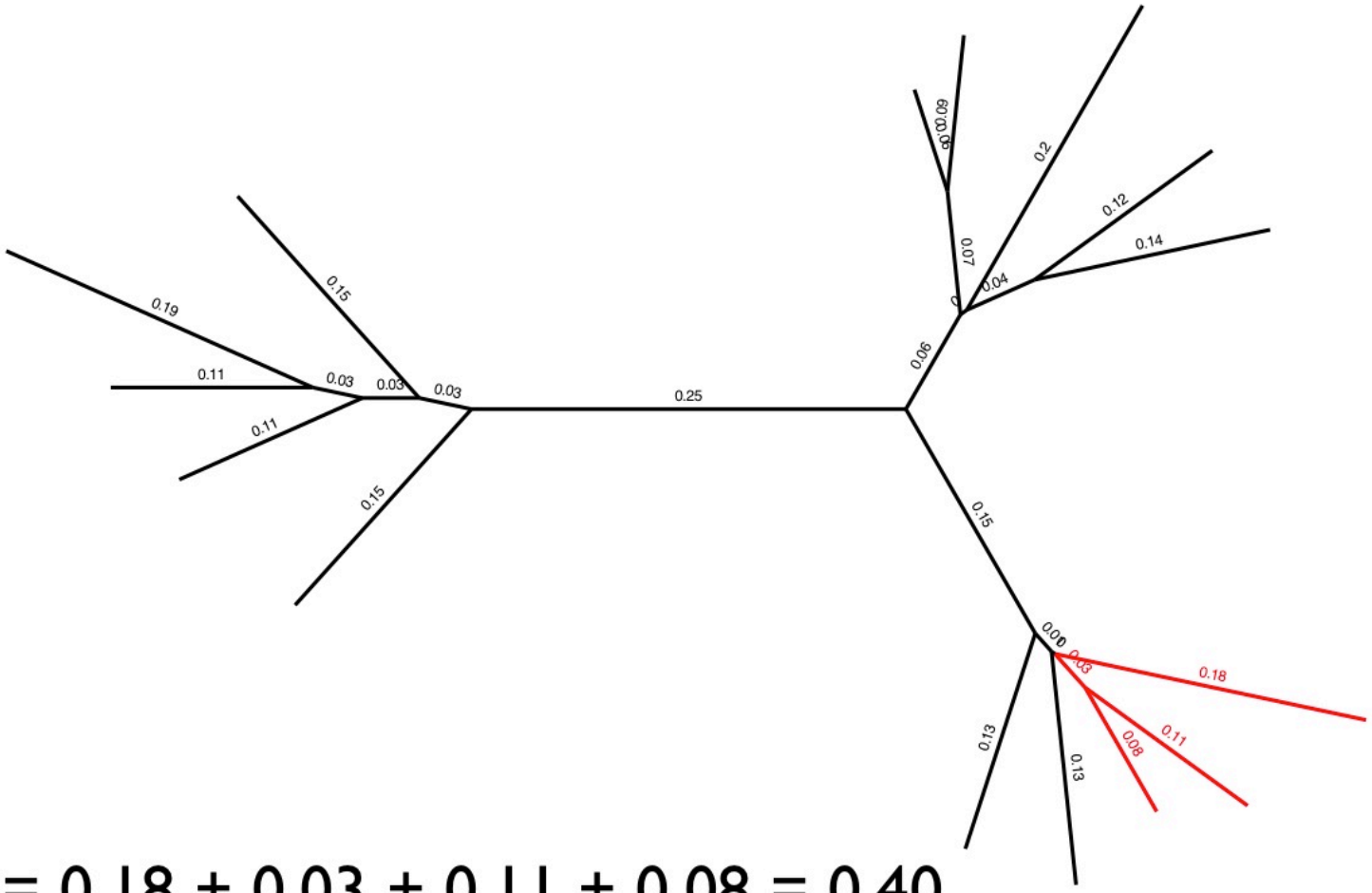


PD = 0.13 + 0.03 + 0.11 + 0.08 = 0.35

Alpha diversity

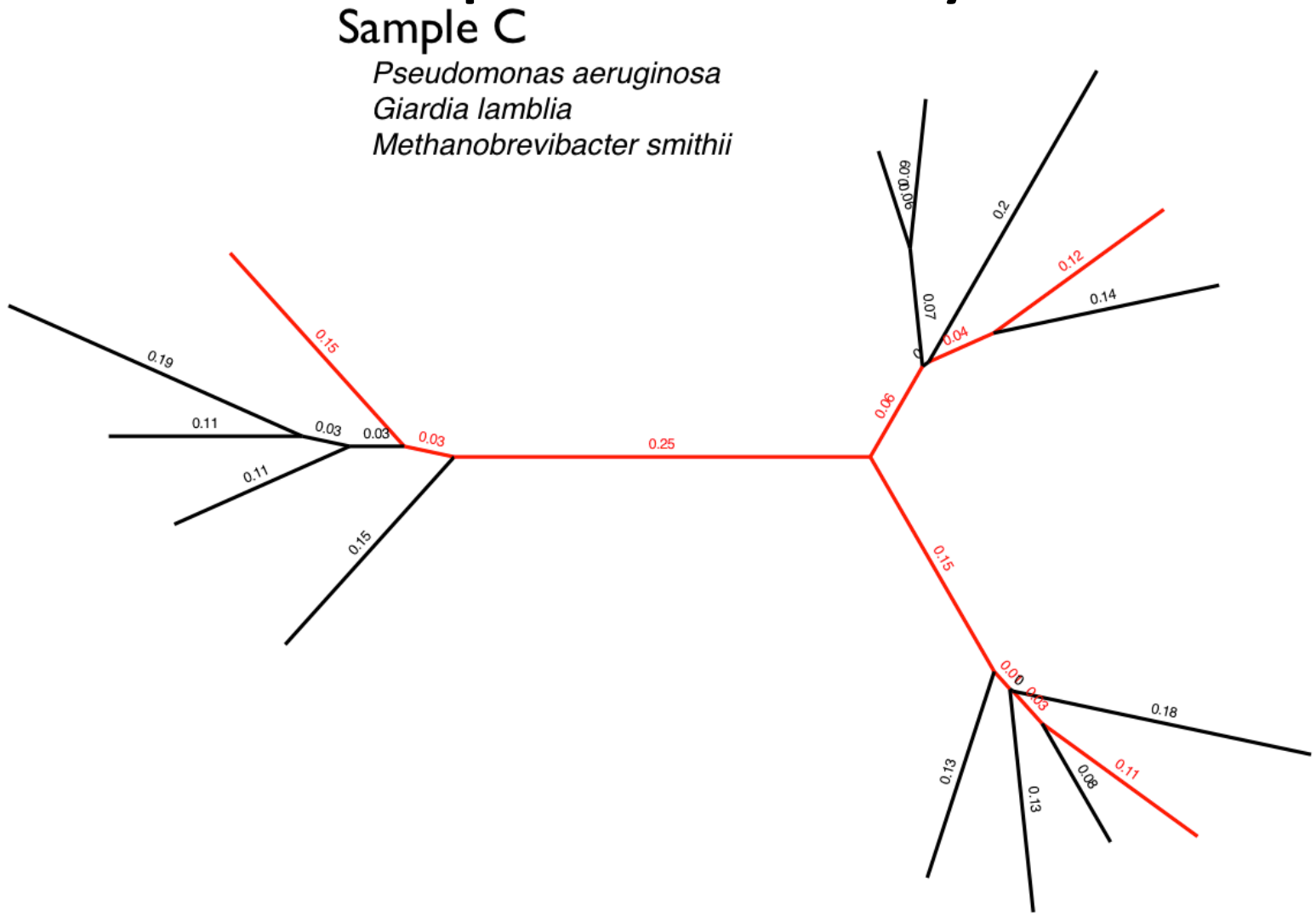
Sample B

Pseudomonas aeruginosa
Pseudomonas argentinensis
Escherichia coli



PD = 0.18 + 0.03 + 0.11 + 0.08 = 0.40

Alpha diversity



PD = 0.15+0.03+0.25+0.06+0.04+0.12+0.15+0.01+0.03+0.11= 0.95

Alpha diversity

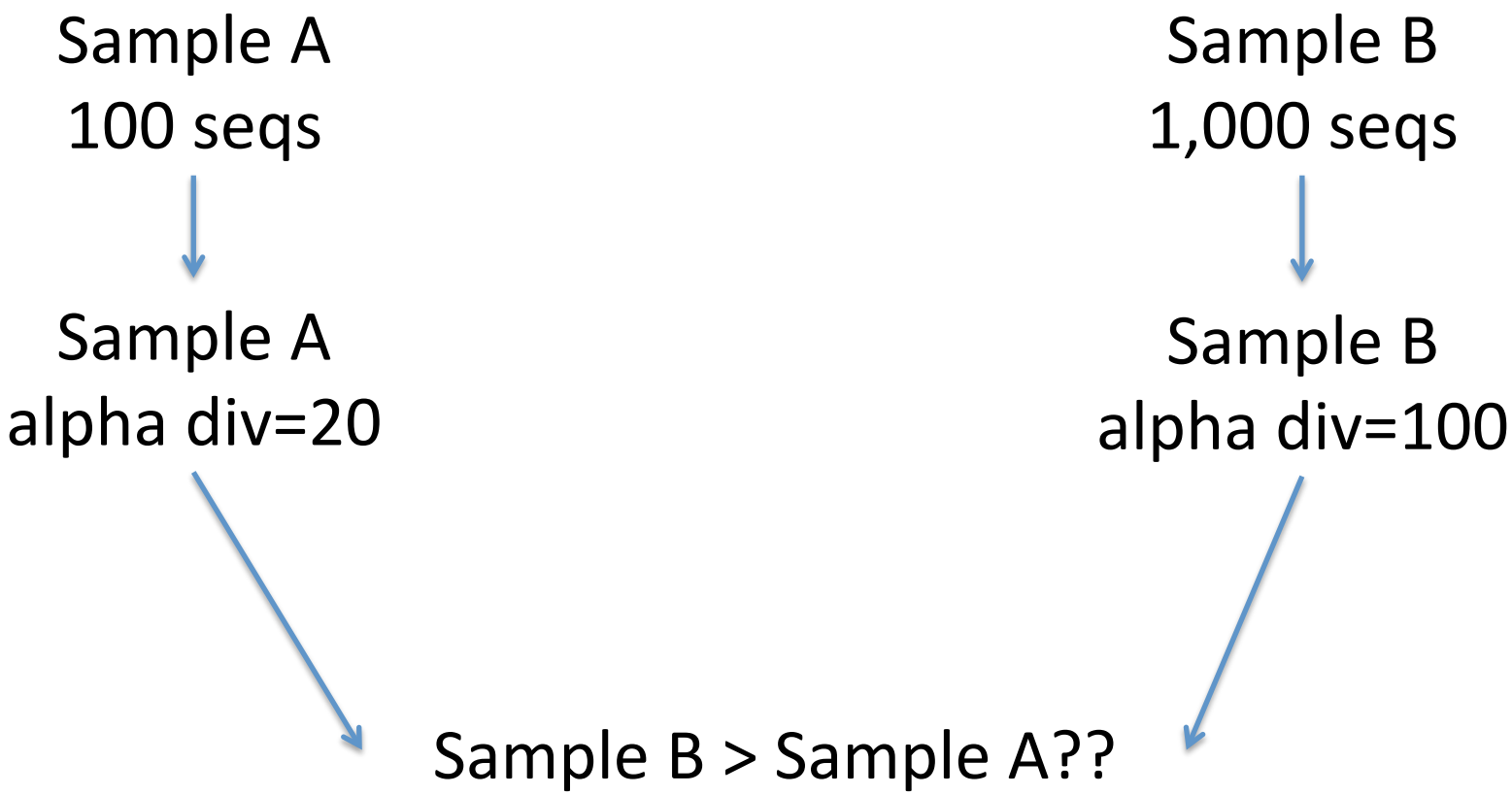
Sample A	**Sample B**	**Sample C**
Pseudomonas aeruginosa	Pseudomonas aeruginosa	Pseudomonas aeruginosa
Pseudomonas argentinensis	Pseudomonas argentinensis	Giardia lamblia
Pseudomonas flavescens	Escherichia coli	Methanobrevibacter smithii

PD = 0.35 < PD = 0.40 < PD = 0.95

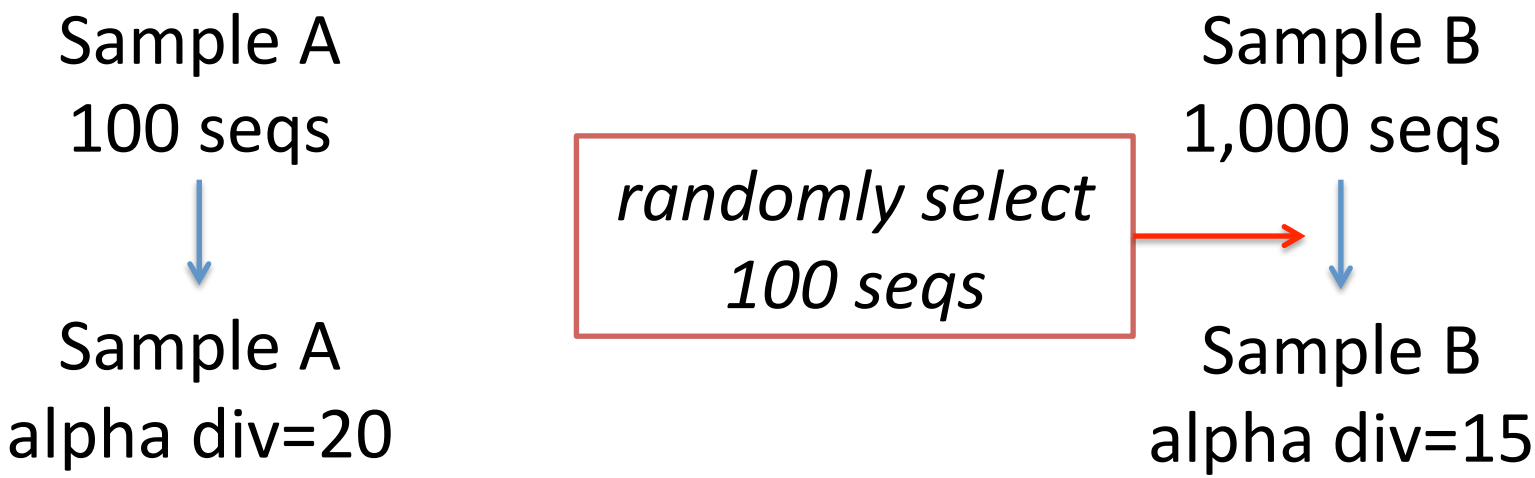
Sample C is more diverse than sample B,
which is more diverse than sample A

Alpha rarefaction

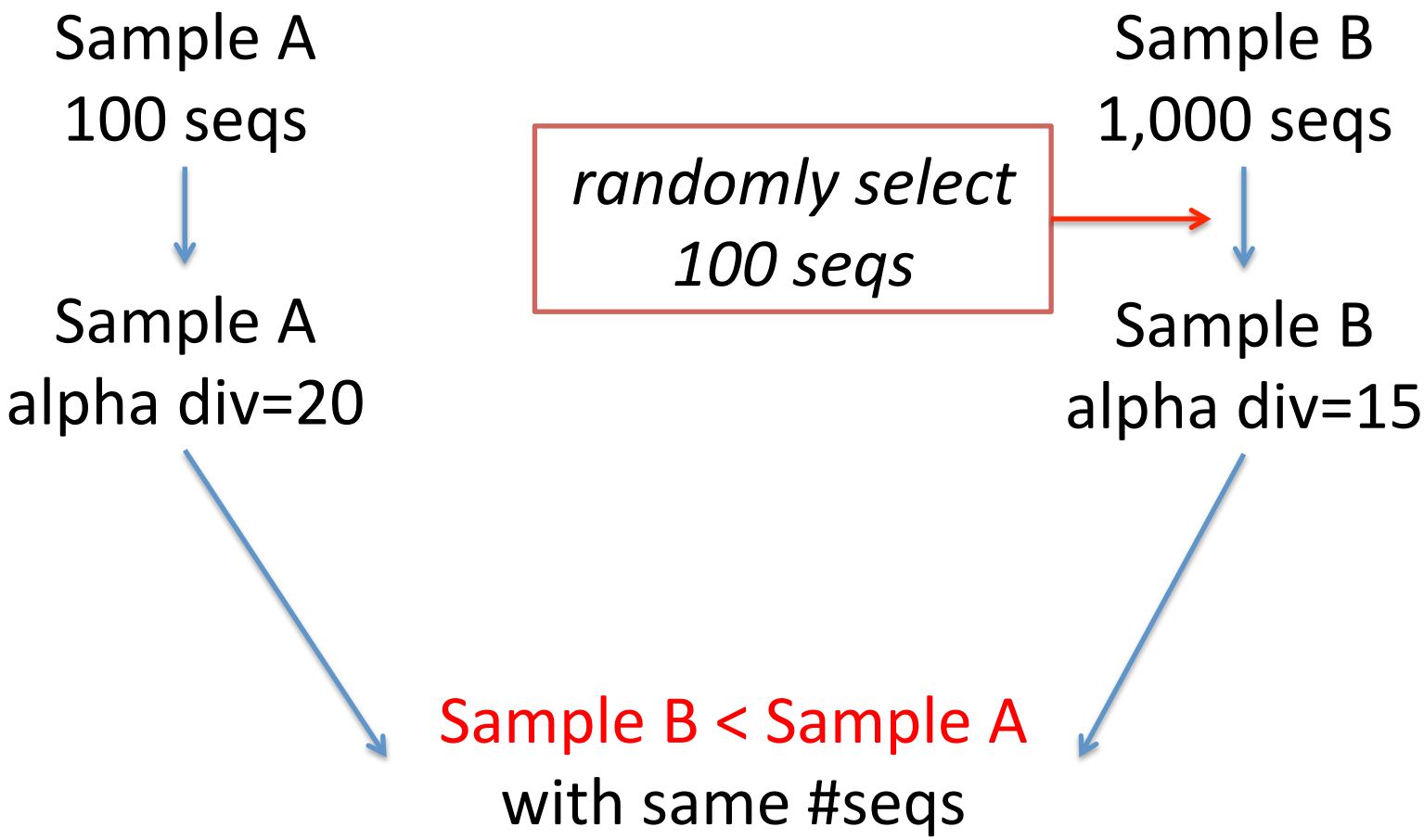
Alpha rarefaction



Alpha rarefaction

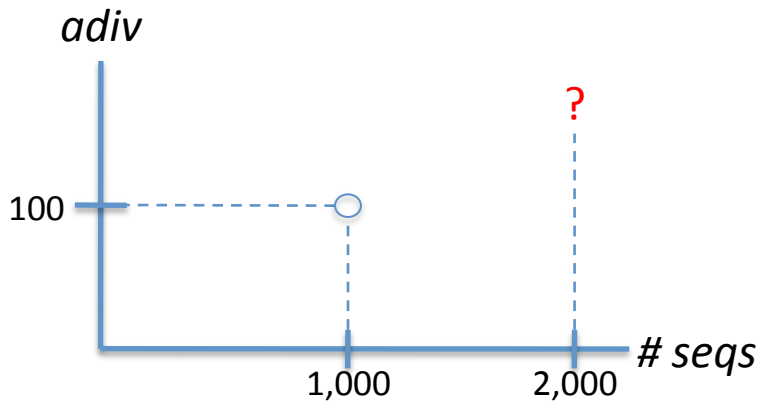


Alpha rarefaction



Multiple alpha rarefaction

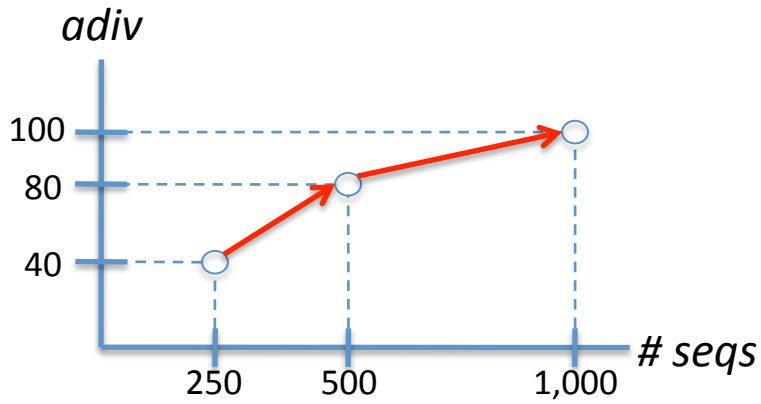
Sample A
Alpha div = 100
with 1,000 seqs



What if we had 2,000 seqs?

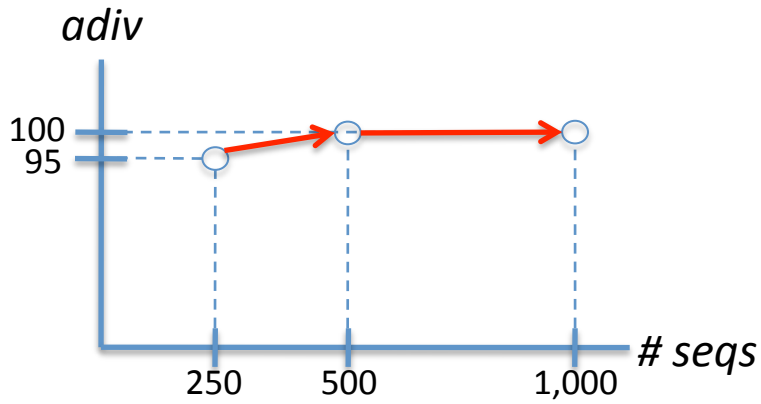
Repeatedly calculate alpha div
at decreasing number of seqs

Multiple alpha rarefaction



Higher sequencing effort might result
in higher observed diversity

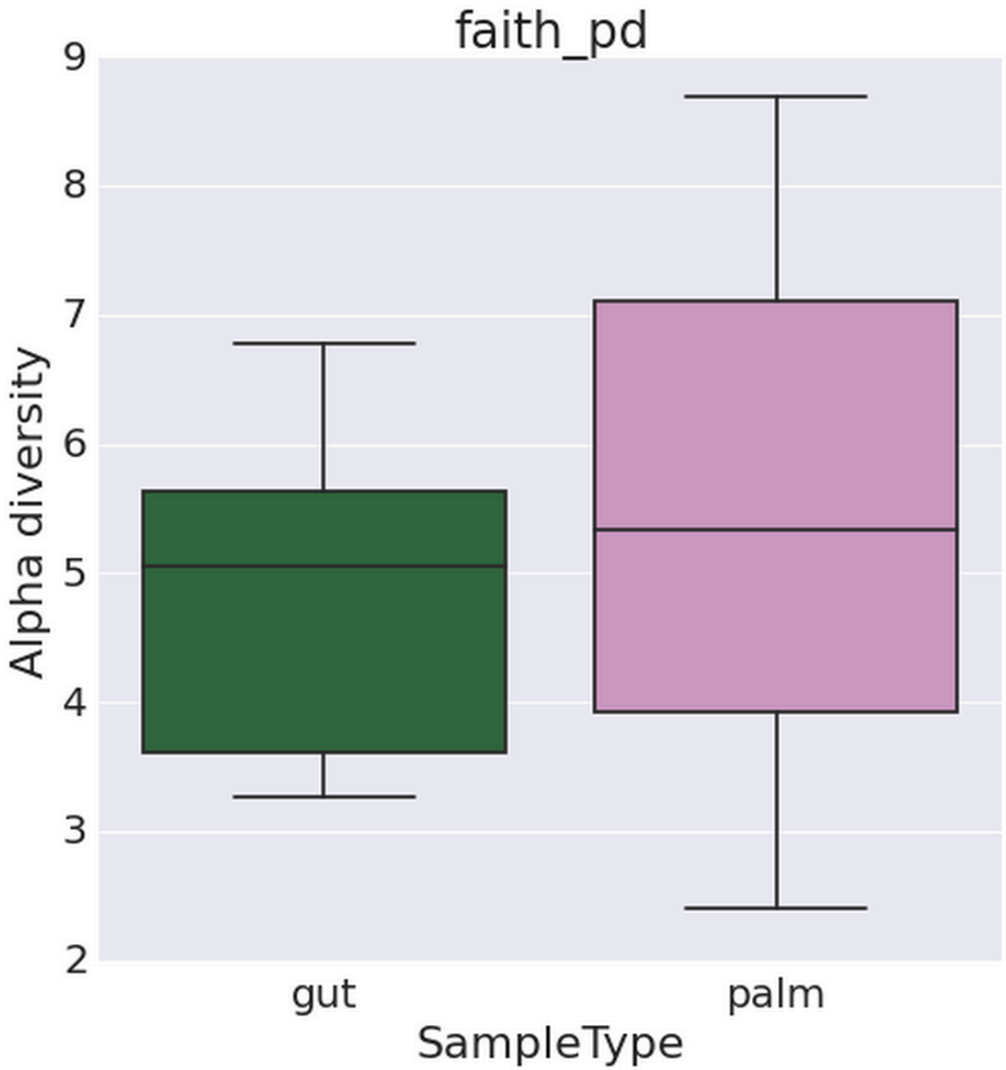
Multiple alpha rarefaction



Higher sequencing effort will probably not add to observed diversity

- Rarefaction results in a loss of data
- Can be difficult to determine optimal rarefaction level
- Rarefaction increases error
- Not always the best option
  - Small libraries/sample size

Are skin samples more diverse than gut?



Comparing alpha diversity between categories

- T-test
  - Parametric
    - Not usually appropriate
  - Non-parametric
    - Mont carlo permutations
    - More commonly used

Beta Diversity
(between sample diversity)

Alpha Diversity

Sample A

Wolbachia melophagi
Lactobacillus delbruecki
Pseudomonas flavescens

Sample C

Pseudomonas aeruginosa
Giardia lamblia
Methanobrevibacter smithii

Sample B

Pseudomonas aeruginosa
Giardia lamblia
Escherichia coli

Alpha Diversity Conclusion

A = B = C

Sample A

Wolbachia melophagi
Lactobacillus delbruecki
Pseudomonas flavescens

Sample C

Pseudomonas aeruginosa
Giardia lamblia
Methanobrevibacter smithii

Sample B

Pseudomonas aeruginosa
Giardia lamblia
Escherichia coli

Bray-Curtis dissimilarity

Non-phylogenetic, quantitative method

$$BC_{jk} = \frac{\sum_i |X_{ij} - X_{ik}|}{\sum_i (X_{ij} + X_{ik})}$$

BC = dissimilarity
Sample
j = jth sample
k = kth sample
i = ith observation
X = Value

Bray-Curtis dissimilarity

$$BC_{jk} = \frac{\sum_i |X_{ij} - X_{ik}|}{\sum_i (X_{ij} + X_{ik})}$$

To calculate distance from Sample 1 to Sample 2

	Sample 1	**Sample 2**	**Sample 3**	**Sample 4**
OTU 1	0	0	4	6
OTU 2	1	7	0	1
OTU 3	4	0	4	2

Numerator:
$|0 + 0| = 0$

Bray-Curtis dissimilarity

$$BC_{jk} = \frac{\sum_i |X_{ij} - X_{ik}|}{\sum_i (X_{ij} + X_{ik})}$$

To calculate distance from Sample 1 to Sample 2

	Sample 1	**Sample 2**	**Sample 3**	**Sample 4**
OTU 1	0	0	4	6
OTU 2	1	7	0	1
OTU 3	4	0	4	2

Numerator:
$|0 - 0| = 0$
$|1 - 7| = 6$

Bray-Curtis dissimilarity

$$BC_{jk} = \frac{\sum_i |X_{ij} - X_{ik}|}{\sum_i (X_{ij} + X_{ik})}$$

To calculate distance from Sample 1 to Sample 2

	Sample 1	**Sample 2**	**Sample 3**	**Sample 4**
OTU 1	0	0	4	6
OTU 2	1	7	0	1
OTU 3	4	0	4	2

Numerator
$|0 - 0| = 0$
$|1 - 7| = 6$
$|4 - 0| = 4$

Bray-Curtis dissimilarity

$$BC_{jk} = \frac{\sum_i |X_{ij} - X_{ik}|}{\sum_i (X_{ij} + X_{ik})}$$

To calculate distance from Sample 1 to Sample 2

	Sample 1	**Sample 2**	**Sample 3**	**Sample 4**
OTU 1	0	0	4	6
OTU 2	1	7	0	1
OTU 3	4	0	4	2

Numerator:
$|0 - 0| = 0$
$|1 - 7| = 6$
$|4 - 0| = 4$

Denominator
$0 + 0 = 0$

Bray-Curtis dissimilarity

$$BC_{jk} = \frac{\sum_i |X_{ij} - X_{ik}|}{\sum_i (X_{ij} + X_{ik})}$$

To calculate distance from Sample 1 to Sample 2

	Sample 1	**Sample 2**	**Sample 3**	**Sample 4**
OTU 1	0	0	4	6
OTU 2	1	7	0	1
OTU 3	4	0	4	2

Numerator:
$|0 - 0| = 0$
$|1 - 7| = 6$
$|4 - 0| = 4$

Denominator:
$0 + 0 = 0$
$1 + 7 = 8$

Bray-Curtis dissimilarity

$$BC_{jk} = \frac{\sum_i |X_{ij} - X_{ik}|}{\sum_i (X_{ij} + X_{ik})}$$

To calculate distance from Sample 1 to Sample 2

	Sample 1	**Sample 2**	**Sample 3**	**Sample 4**
OTU 1	0	0	4	6
OTU 2	1	7	0	1
OTU 3	4	0	4	2

Numerator:
$|0 - 0| = 0$
$|1 - 7| = 6$
$|4 - 0| = 4$

Denominator
$0 + 0 = 0$
$1 + 7 = 8$
$4 + 0 = 4$

Bray-Curtis dissimilarity

$$BC_{jk} = \frac{\sum_i |X_{ij} - X_{ik}|}{\sum_i (X_{ij} + X_{ik})}$$

To calculate distance from Sample 1 to Sample 2

	Sample 1	**Sample 2**	**Sample 3**	**Sample 4**
OTU 1	0	0	4	6
OTU 2	1	7	0	1
OTU 3	4	0	4	2

Numerator:
$|0 - 0| = 0$
$|1 - 7| = 6$
$|4 - 0| = 4$

Denominator
$0 + 0 = 0$
$1 + 7 = 8$
$4 + 0 = 4$

BC = 10/12

Unweighted Unifrac
A phylogenetic qualitative metric

$$U_{AB} = \frac{unique}{observed}$$

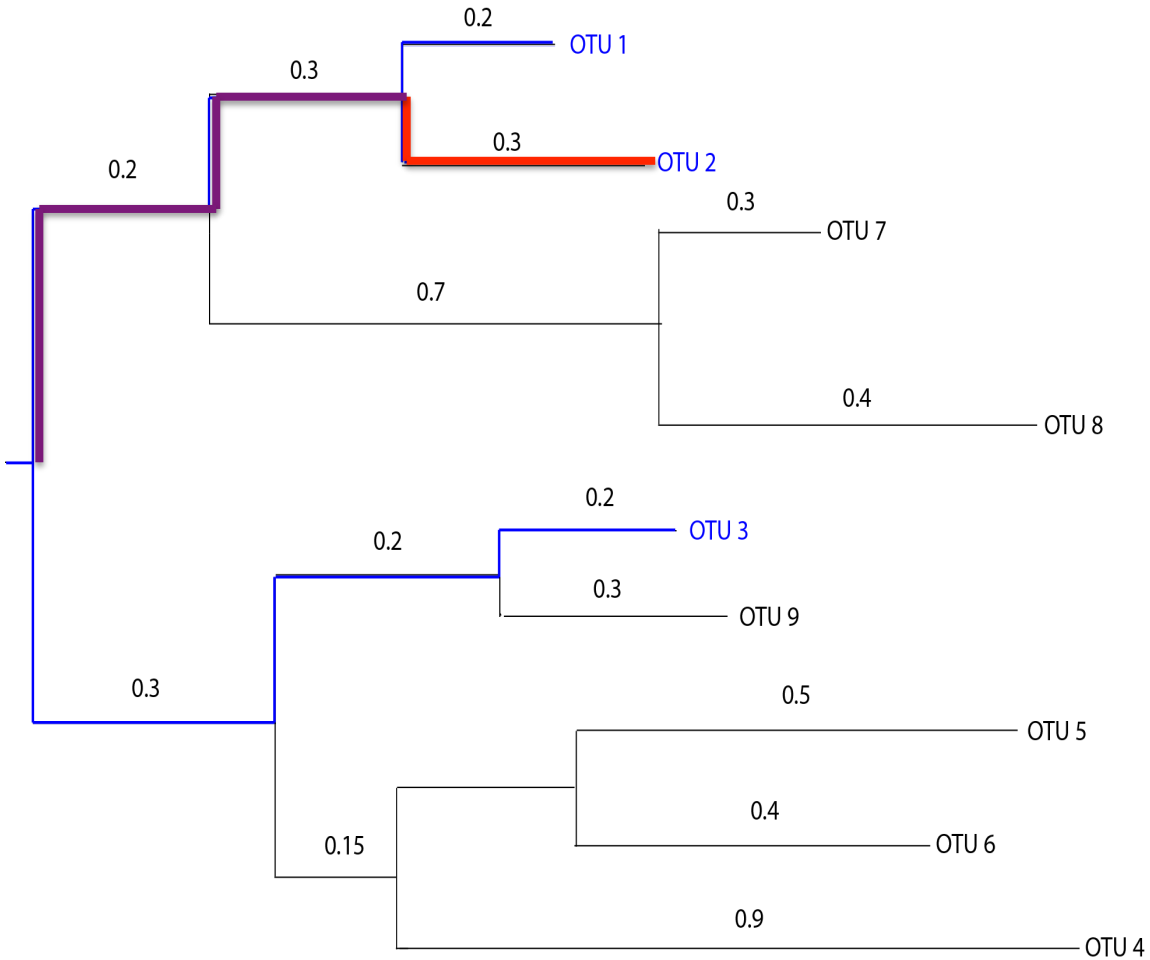
Uniqe: the sum of the branch lengths that only lead to an observations found in one of samples being compared

Observed: the sum of the branch lengths that lead to an observation in either sample

	Sample 1	Sample 2	Sample 3	Sample 4
OTU 1	0	0	4	6
OTU 2	1	7	0	1
OTU 3	4	0	4	2

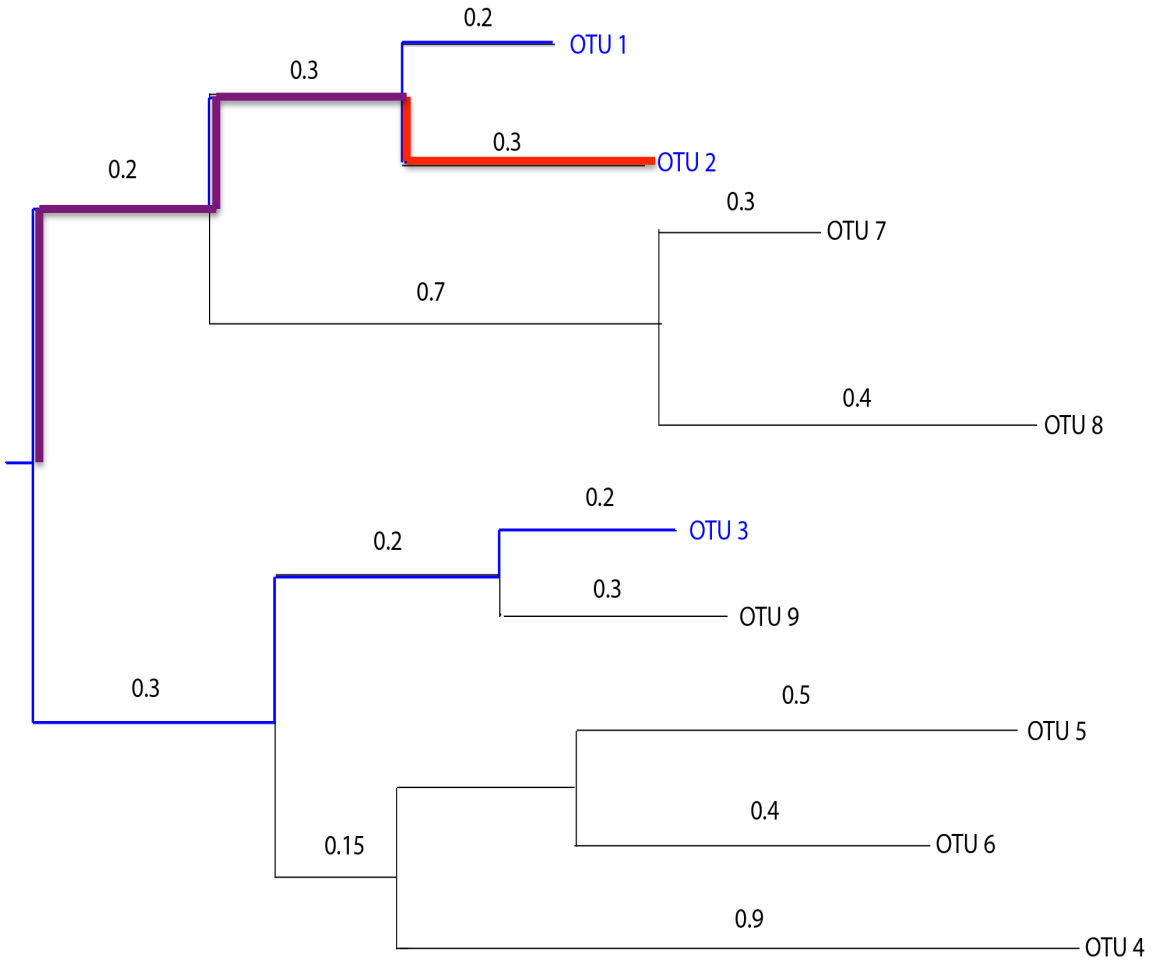
$$U_{AB} = \frac{unique}{observed}$$

Sample 2 to Sample 3 Distance



	Sample 1	Sample 2	Sample 3	Sample 4
OTU 1	0	0	4	6
OTU 2	1	7	0	1
OTU 3	4	0	4	2

$$U_{AB} = \frac{unique}{observed}$$

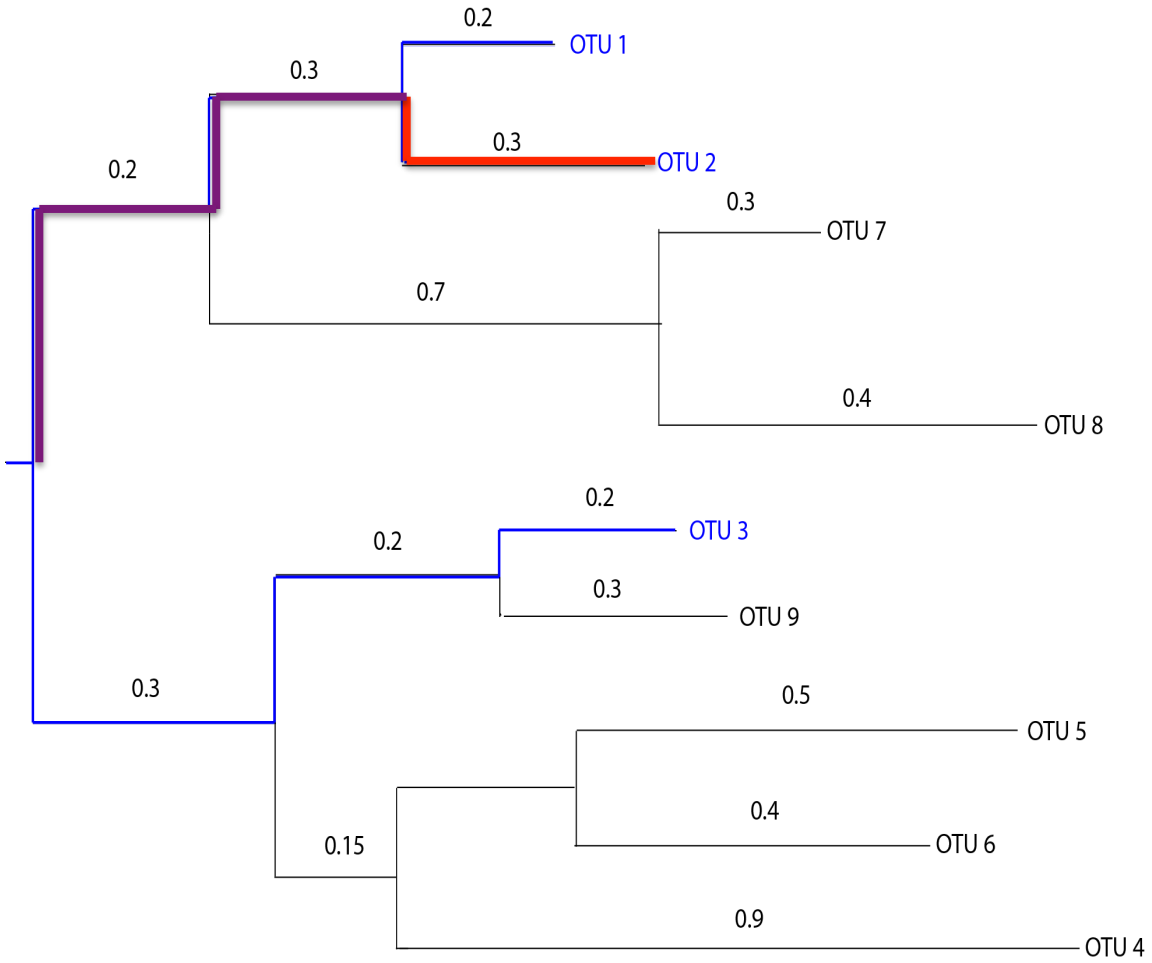


Sample 2 to Sample 3 Distance

Unique = 0.2 + 0.3 + 0.3 + 0.2 + 0.2 + 0.2

	Sample 1	Sample 2	Sample 3	Sample 4
OTU 1	0	0	4	6
OTU 2	1	7	0	1
OTU 3	4	0	4	2

$$U_{AB} = \frac{unique}{observed}$$

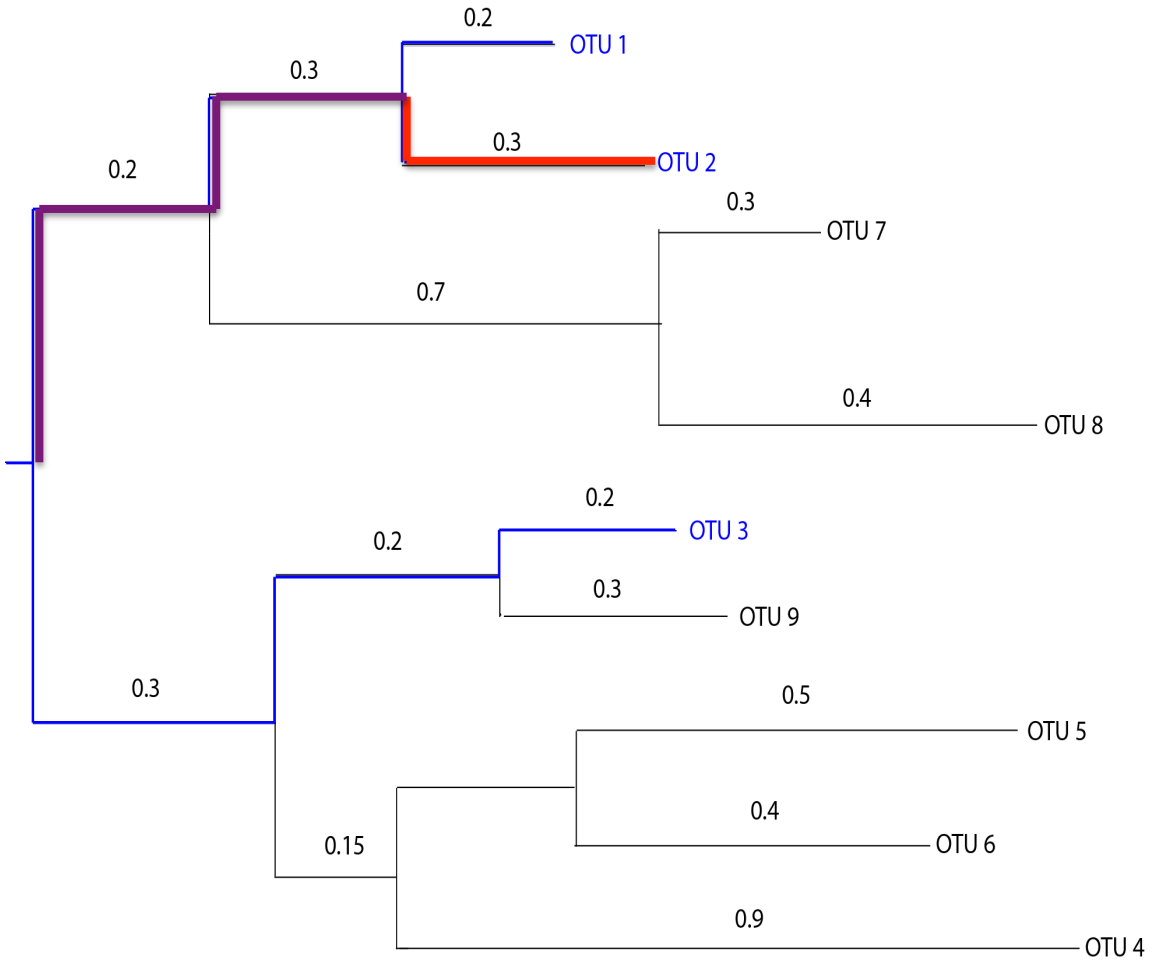


Sample 2 to Sample 3 Distance

Unique = 0.2 + 0.3 + 0.3 + 0.2 + 0.2 + 0.2

Observed = 0.2 + 0.3 + 0.2 + 0.3 + 0.3 + 0.2 + 0.2 + 0.2

	Sample 1	Sample 2	Sample 3	Sample 4
OTU 1	0	0	4	6
OTU 2	1	7	0	1
OTU 3	4	0	4	2



$$U_{AB} = \frac{unique}{observed}$$

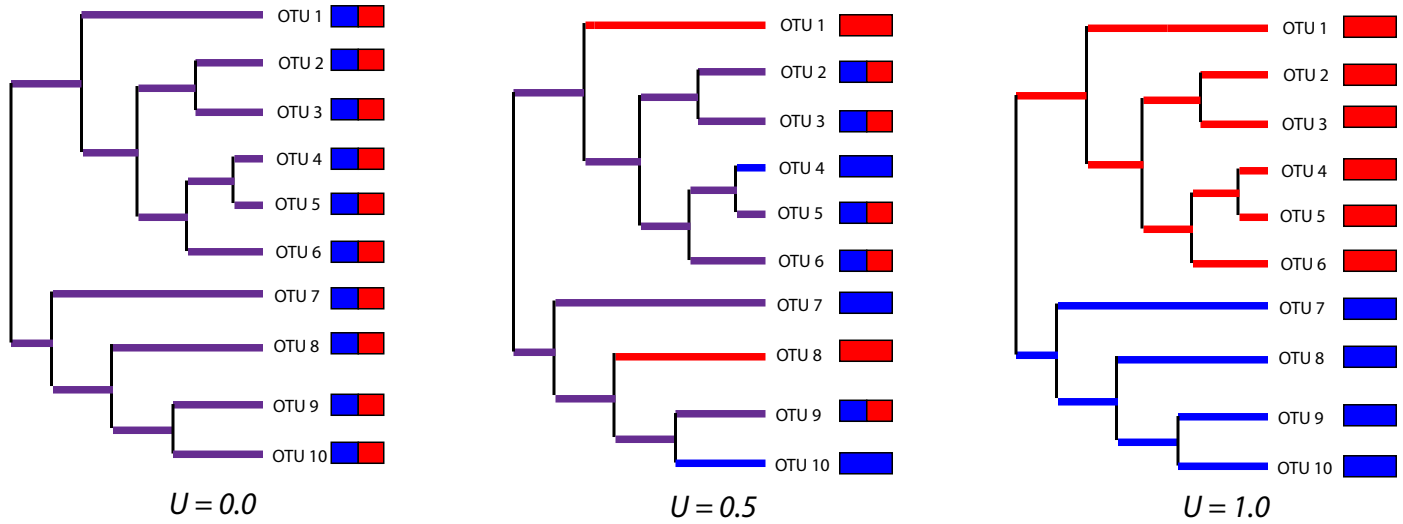
Sample 2 to Sample 3 Distance

Unique = 0.2 + 0.3 + 0.3 + 0.2 + 0.2 + 0.2

Observed = 0.2 + 0.3 + 0.2 + 0.3 + 0.3 + 0.2 + 0.2 + 0.2

U_{AB} = 1.4/1.9

Unweighted Unifrac: a phylogenetic measure of the dissimilarity of microbial communities



$$U_{AB} = \frac{unique}{observed}$$

where:

$unique$: the unique branch length, or branch length that only leads to OTU(s) observed in sample A or sample B

$observed$: the total branch length observed in either sample A or sample B

Slide Credit: Greg Caporaso

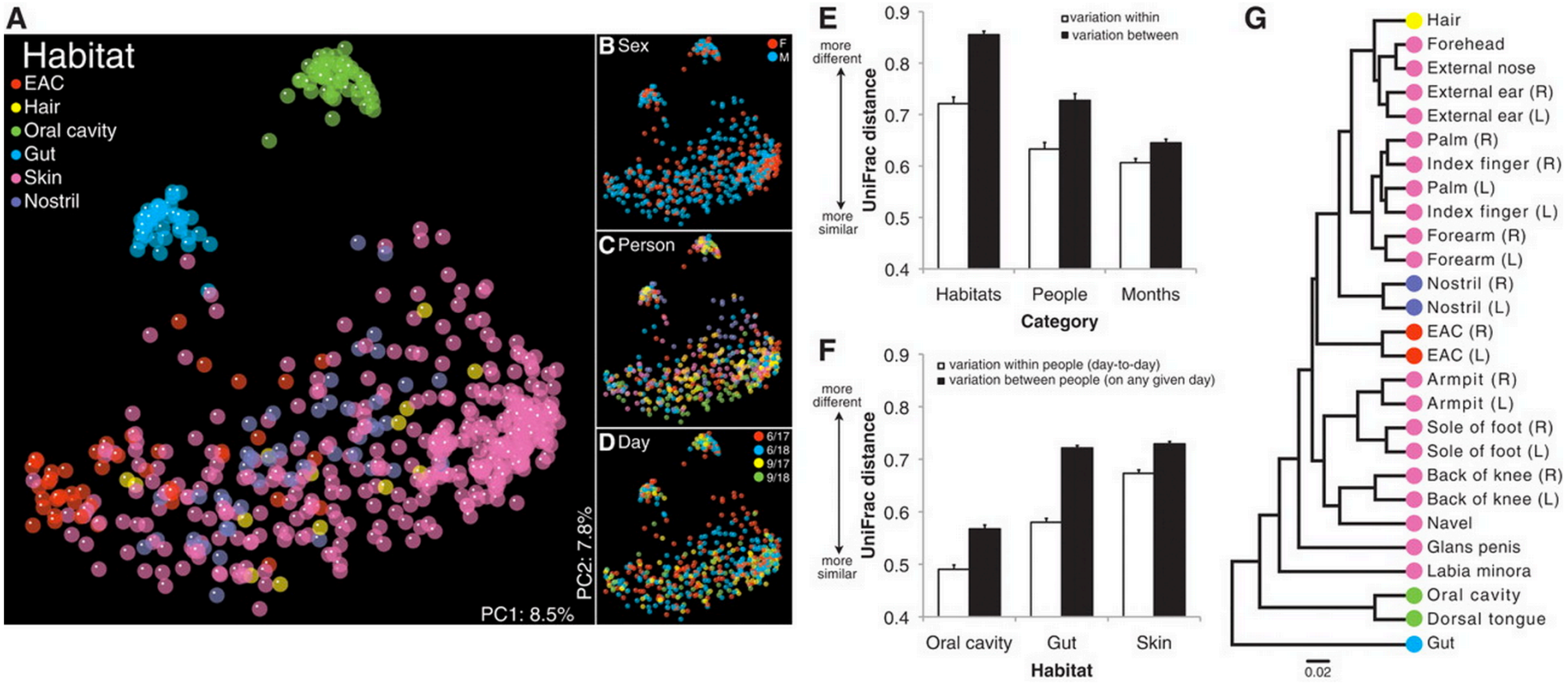
Result of beta diversity is a distance matrix

	Sample 1	**Sample 2**	**Sample 3**	**Sample 4**	**Sample 5**	**Sample 6**
Sample 1	0	0.35	0.83	0.83	0.9	0.9
Sample 2	0.35	0	0.86	0.85	0.92	0.91
Sample 3	0.83	0.7	0	0.25	0.88	0.87
Sample 4	0.83	0.1	0.4	0	0.88	0.88
Sample 5	0.9	0.92	0.88	0.88	0	0.5
Sample 6	0.9	0.91	0.87	0.88	0.5	0

unweighted_unifrac_dm.txt

	Stillton4R2	Stillton4R3	Stillton4R1	HCanyon3R3	HCanyon3R2	HCanyon3R1	Halls9R2	HCanyon2R2	HCanyon2R3	HCanyon2R1	Halls9R1	Stillton10R3	HCanyon0R1	HCanyon0R2	HCanyon0R3	HCa
Stillton4R2	0.0	0.382273294624	0.391416675288	0.560309484808	0.553938232028	0.566136031815	0.557134987546	0.531719852875	0.53655901824	0.567041909667	0.574831935502	0.474244845718				
Stillton4R3	0.382273294624	0.0	0.394399899497	0.586303332083	0.583969320358	0.589363298118	0.588066262733	0.560923487354	0.57767776062	0.601572335069	0.609758664376	0.479910046195				
Stillton4R1	0.391416675288	0.394399899497	0.0	0.580628929798	0.583767102601	0.584463513906	0.573828216898	0.550892933103	0.566056393448	0.584611276938	0.607144746402	0.469209424282				
HCanyon3R3	0.560309484808	0.586303332083	0.580628929798	0.0	0.354600316745	0.345731612479	0.458450145834	0.412488628393	0.385127601086	0.384525100123	0.46823652766	0.585402372425				
HCanyon3R2	0.553938232028	0.583969320358	0.583767102601	0.354600316745	0.0	0.36256949665	0.462707750398	0.414528760004	0.385453766442	0.380087766632	0.46109426257	0.574488221279				
HCanyon3R1	0.566136031815	0.589363298118	0.584463513906	0.345731612479	0.36256949665	0.0	0.452351806146	0.415483559719	0.398796424875	0.381981707704	0.485596583752	0.580926122772				
Halls9R2	0.557134987546	0.588066262733	0.573828216898	0.458450145834	0.462707750398	0.452351806146	0.0	0.447883445295	0.429943464459	0.409064513124	0.344264504725	0.561110815583				
HCanyon2R2	0.531719852875	0.560923487354	0.550892933103	0.412488628393	0.414528760004	0.415483559719	0.447883445295	0.0	0.404179520995	0.388727659604	0.468345488157	0.549328940024				
HCanyon2R3	0.53655901824	0.57767776062	0.566056393448	0.385127601086	0.385453766442	0.398796424875	0.429943464459	0.404179520995	0.0	0.361444528983	0.439367245599	0.566330894525				
HCanyon2R1	0.567041909667	0.601572335069	0.584611276938	0.384525100123	0.380087766632	0.381981707704	0.409064513124	0.388727659604	0.361444528983	0.0	0.454633280595	0.582635621853				
Halls9R1	0.574831935502	0.609758664376	0.607144746402	0.46823652766	0.46109426257	0.485596583752	0.344264504725	0.468345488157	0.439367245599	0.454633280595	0.0	0.578254150137				
Stillton10R3	0.474244845718	0.479910046195	0.469209424282	0.585402372425	0.574488221279	0.580926122772	0.561110815583	0.549328940024	0.566330894525	0.582635621853	0.578254150137					
HCanyon0R1	0.776189708931	0.779315452075	0.781074100354	0.69071344042	0.692644612288	0.683086513009	0.727520276229	0.709578414217	0.711736423222	0.707912554104	0.745189206089	0.7				
HCanyon0R2	0.797241309582	0.795032626325	0.797809039119	0.713804444578	0.714717508733	0.697490417772	0.75335887058	0.723508997775	0.728837315629	0.722257776622	0.761081079305	0.7				
HCanyon0R3	0.784124036272	0.791711071574	0.795527048923	0.693476495058	0.699412891734	0.688360016409	0.732489014053	0.713925200799	0.708354096579	0.708462144382	0.747778043238	0.7				
HCanyon7R3	0.539194289149	0.563745794923	0.557216919336	0.457709184042	0.467073058252	0.473888308307	0.42997132225	0.455576263999	0.448029116277	0.441566044862	0.444788573194	0.5				
HCanyon7R2	0.665877547192	0.683833494738	0.674471865383	0.55308292175	0.563316168524	0.562611936812	0.546938312206	0.558741865535	0.562533096516	0.534453165883	0.558141564634	0.6				
HCanyon7R1	0.554857668962	0.578652540309	0.570402839551	0.475303182122	0.474309745536	0.462587318718	0.43076590681	0.463222699451	0.452021582021	0.42664191335	0.44552755863	0.5				
HCanyon1R1	0.746066596617	0.754645795641	0.759778994683	0.654674045733	0.651121732227	0.640930093945	0.703134531793	0.67682938985	0.675957137688	0.676955919301	0.714539249748	0.7				
HCanyon1R3	0.751522854919	0.764074726397	0.767226018039	0.641176135477	0.65043468158	0.629924888988	0.702785457767	0.675191090098	0.668258446703	0.66343689429	0.713256418398	0.7				
HCanyon1R2	0.748585453295	0.753698061339	0.759152313976	0.666322115311	0.666043304584	0.654075859985	0.711448710816	0.680563093126	0.679964026606	0.678341781539	0.728364004942	0.7				
HCanyon10R2	0.56835786215	0.584149104488	0.584183423897	0.466055758676	0.464168854592	0.470680372129	0.458866733827	0.4814190318	0.460896823727	0.442543616102	0.4619867362	0.5				
HCanyon10R3	0.570798042641	0.588843077275	0.591942476159	0.50320590177	0.498220269993	0.503094520399	0.460532212318	0.496962587514	0.490816424808	0.497919293873	0.463277414447	0.5				
HCanyon10R1	0.505926071682	0.528347698837	0.529350507225	0.472413250039	0.479553556468	0.478832889376	0.45591841873	0.458027517992	0.428733419477	0.456719821741	0.457712834343	0.5				
HCanyon11R3	0.582831143875	0.594413158276	0.60681715155	0.463621293703	0.45842280462	0.486656215192	0.430493752745	0.47439497384	0.445815433526	0.432013882772	0.437061115715	0.5				
HCanyon11R2	0.52855255543	0.55774407831	0.555177188694	0.432763917311	0.429877059678	0.440567014981	0.41332040453	0.426696806769	0.398447088329	0.396059530935	0.433778641456	0.5				
HCanyon11R1	0.555250897098	0.579991048797	0.575126874355	0.453486254954	0.44588780311	0.456712334902	0.419232496891	0.446896791255	0.430325765526	0.428966840449	0.431167398777	0.5				
HCanyon6R2	0.568527068157	0.58513246508	0.583868790939	0.430174126838	0.419369519662	0.432708499988	0.416828113391	0.445815626173	0.426928785146	0.410839210322	0.422526407071	0.5				
HCanyon6R3	0.566346578975	0.587780110737	0.586940661236	0.438634245182	0.425883025257	0.449519180577	0.424862337935	0.447100684715	0.425434475646	0.426218627557	0.424693270252	0.5				
HCanyon6R1	0.583972679946	0.586778207344	0.581534515565	0.434912987895	0.424307243199	0.441765941119	0.415236907762	0.428454635151	0.428113957425	0.406954252133	0.422536839493	0.5				
HCanyon5R1	0.558910457622	0.57470972671	0.562821643814	0.440681773256	0.451596106791	0.444216261284	0.424754580111	0.435763316753	0.417737787339	0.416631599644	0.455877139948	0.5				
HCanyon5R3	0.559987492698	0.582019682	0.584250873735	0.450336948189	0.426574240559	0.448524323812	0.439745925494	0.440643510478	0.417613041188	0.413563495656	0.464820696561	0.5				
HCanyon5R2	0.577951660858	0.593568671618	0.596389572645	0.422545765978	0.409515246412	0.434403209926	0.429247944328	0.453870629905	0.431590973836	0.41366014112	0.436175614058	0.5				
HCanyon4R1	0.516189922282	0.555158871988	0.544024708263	0.416449154884	0.420465768689	0.437380878857	0.46015301609	0.459537309642	0.425192263132	0.439771880651	0.474121489459	0.5				
HCanyon4R2	0.601827528628	0.632734373589	0.628005145571	0.414666304996	0.417716489314	0.443390188238	0.497188855337	0.473766452031	0.447416361784	0.454050692681	0.502442274303	0.6				
HCanyon4R3	0.624113459554	0.653406187405	0.649892943176	0.438848993112	0.452700670776	0.45666604112	0.506388612937	0.493725468949	0.479954328128	0.462327913958	0.510888415471	0.6				
HCanyon12R1	0.513376891954	0.547194076841	0.519066521533	0.471737648119	0.464638931219	0.461841527684	0.430362902794	0.439277983407	0.419105789781	0.41992903626	0.466361101825	0.5				
HCanyon12R2	0.552892298156	0.560516044793	0.549257756132	0.504271040435	0.504749255647	0.517681811493	0.453014756694	0.480827849426	0.485996535935	0.465436661211	0.5					
HCanyon12R3	0.553185666616	0.585465060164	0.585665024134	0.439712101711	0.431336995109	0.463612059381	0.409326408365	0.435088785406	0.414920465304	0.407840817122	0.422739541765	0.5				
HCanyon8R1	0.552014316498	0.559504417239	0.562836483618	0.483602818412	0.470838841272	0.485053828711	0.446654297414	0.465192835869	0.464892486774	0.441733317446	0.466837938413	0.5				
HCanyon8R2	0.547616166832	0.576958110793	0.578527645361	0.458907582912	0.452778768072	0.456800424019	0.420176245854	0.449481849055	0.430019204227	0.413022684081	0.432050808518	0.5				
HCanyon8R3	0.549015567252	0.573444281355	0.580500957664	0.446205572204	0.446999609932	0.453171155433	0.4327591563	0.439870648001	0.425623148959	0.408822227859	0.446705079363	0.5				
Halls8R2	0.535707795107	0.569114464296	0.574510728774	0.46528819939	0.452053418068	0.460087502347	0.379587194176	0.458489263366	0.42118148129	0.440297005744	0.366881097176	0.5				
Stillton10R1	0.462296292442	0.468444771913	0.470923059335	0.575228561838	0.585038958921	0.5846812234	0.56445075535	0.557710461145	0.559334675708	0.5881972363	0.584771138758					
Stillton10R2	0.538798221769	0.543866219051	0.557970880727	0.543811509839	0.541537645813	0.56301479339	0.486181550953	0.53366160907	0.52362238711	0.535021530589	0.480837812931					
Halls8R1	0.614648837651	0.6401217906	0.644104393777	0.493196498338	0.492370158337	0.486233469981	0.445575362279	0.508540989086	0.484457264025	0.473964462468	0.458044156786	0.6				
Halls6R1	0.484233643835	0.518598404668	0.519082685954	0.476689255656	0.489545391911	0.479649954411	0.484643817074	0.479596366267	0.472472195267	0.47020063594	0.505956019295	0.5				
Halls6R2	0.496757718878	0.528784811375	0.521274731439	0.45457903264	0.434119427081	0.447597701249	0.457381539248	0.442946410151	0.453605919538	0.450370943322	0.476499991041	0.5				
Halls4R2	0.509683393195	0.495879096417	0.488951752223	0.609282443564	0.612813488861	0.603872514737	0.603436206122	0.590101531162	0.603855889606	0.609643606143	0.639977381353	0.5				
Stillton5R2	0.552954619991	0.583608736216	0.583498177592	0.471532456639	0.464725650565	0.479581685887	0.521034266478	0.507044011114	0.481356798038	0.487411548649	0.523238604314	0.5				
Halls4R1	0.549133039964	0.534016529239	0.54216719333	0.597412160187	0.596104351076	0.583900421014	0.611452811208	0.594261248789	0.610285284262	0.62138814384	0.619094392421	0.5				
Halls3R1	0.473322285448	0.512621151343	0.492765444818	0.486663734375	0.479618250755	0.49005497706	0.495762448414	0.481055233469	0.485571399694	0.495623098076	0.512078468475	0.5				
Halls3R2	0.477906252536	0.504664103964	0.493403206238	0.505477512498	0.485003185584	0.489265134199	0.506592335264	0.495688280035	0.482524313016	0.488678891103	0.531945821721	0.5				
Stillton1R2	0.74000698752	0.73918413177	0.752036339232	0.661324275604	0.663374157184	0.659034469034	0.713789910279	0.684872131566	0.686043819813	0.691158070649	0.727674238476	0.7				
Stillton1R1	0.740478436173	0.749933432793	0.756329963502	0.66650820781	0.6690775674	0.658929067552	0.718833771904	0.685754060798	0.673445656096	0.689569094328	0.717503361858	0.7				
Stillton2R1	0.550373678768	0.590091691031	0.587820892951	0.47932468001	0.495155380615	0.492382688781	0.545201016787	0.509560123521	0.475440520725	0.493748269889	0.554215167119	0.6				
Stillton2R2	0.484475701033	0.524107691975	0.511649195406	0.579401294383	0.56801243576	0.576872243674	0.58870881228	0.553944702967	0.540477801825	0.571888742292	0.60609853984	0.5				
Stillton2R3	0.474431262772	0.519246404275	0.499359209064	0.57485592581	0.556873882461	0.566375658845	0.577858048887	0.555184929823	0.525423272042	0.555979805507	0.600952560439	0.5				
Stillton3R1	0.432052935213	0.454275671487	0.444726425606	0.549799848291	0.533275339629	0.535991712539	0.549881648961	0.515665559359	0.526949949975	0.541831963172	0.581231050668	0.5				
Stillton3R3	0.416032757783	0.462195063102	0.466371835339	0.507434721309	0.489038123044	0.512316406392	0.520518662126	0.493960260433	0.46471863963	0.492944864273	0.531358800178	0.5				

Pairwise distances between samples are the basis of most microbiome surveys



Bacterial Community Variation in Human Body Habitats Across Space and Time.
Costello et al. *Science* (2009)

Polar Ordination

$$a = \frac{D^2 + D1^2 - D2^2}{2 \times D}$$

D = is distance between the endpoints

D1 is distance between the current sample and endpoint 1

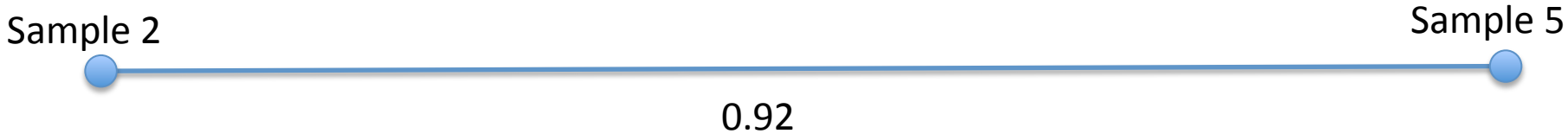
D2 is distance between sample and endpoint 2.

Step 1. Identify the largest distance in the distance matrix.

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
Sample 1	0	0.35	0.83	0.83	0.9	0.9
Sample 2	0.35	0	0.86	0.85	0.92	0.91
Sample 3	0.83	0.7	0	0.25	0.88	0.87
Sample 4	0.83	0.1	0.4	0	0.88	0.88
Sample 5	0.9	0.92	0.88	0.88	0	0.5
Sample 6	0.9	0.91	0.87	0.88	0.5	0

Step 2. Define a line with the two samples contributing to that distance as follows:

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
Sample 1	0	0.35	0.83	0.83	0.9	0.9
Sample 2	0.35	0	0.86	0.85	0.92	0.91
Sample 3	0.83	0.7	0	0.25	0.88	0.87
Sample 4	0.83	0.1	0.4	0	0.88	0.88
Sample 5	0.9	0.92	0.88	0.88	0	0.5
Sample 6	0.9	0.91	0.87	0.88	0.5	0



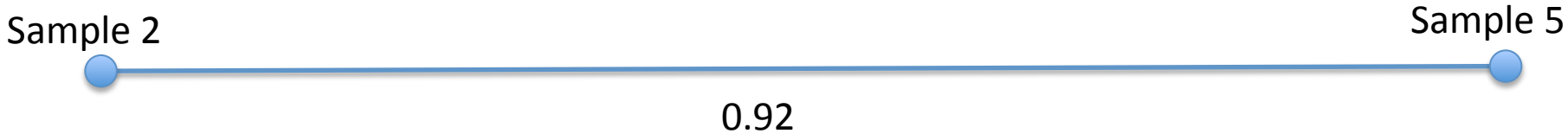
Step 3. Compute the location of each other sample on that axis as follows:

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
Sample 1	0	0.35	0.83	0.83	0.9	0.9
Sample 2	0.35	0	0.86	0.85	0.92	0.91
Sample 3	0.83	0.7	0	0.25	0.88	0.87
Sample 4	0.83	0.1	0.4	0	0.88	0.88
Sample 5	0.9	0.92	0.88	0.88	0	0.5
Sample 6	0.9	0.91	0.87	0.88	0.5	0

$$a = \frac{D^2 + D1^2 - D2^2}{2 \times D}$$

Sample 1:

$$\frac{0.92^2 + 0.35^2 - 0.9^2}{2 \times 0.92} = 0.86$$



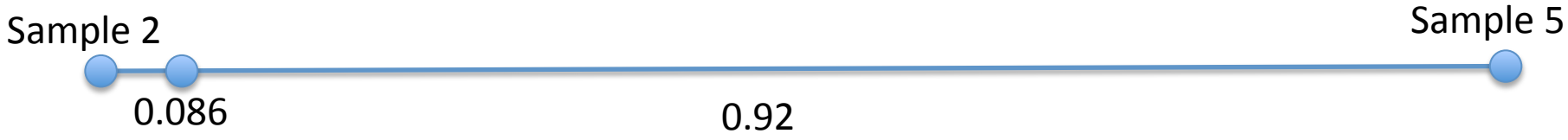
Step 3. Compute the location of each other sample on that axis as follows:

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
Sample 1	0	0.35	0.83	0.83	0.9	0.9
Sample 2	0.35	0	0.86	0.85	0.92	0.91
Sample 3	0.83	0.7	0	0.25	0.88	0.87
Sample 4	0.83	0.1	0.4	0	0.88	0.88
Sample 5	0.9	0.92	0.88	0.88	0	0.5
Sample 6	0.9	0.91	0.87	0.88	0.5	0

$$a = \frac{D^2 + D1^2 - D2^2}{2 \times D}$$

Sample 1:

$$\frac{0.92^2 + 0.35^2 - 0.9^2}{2 \times 0.92} = 0.086$$



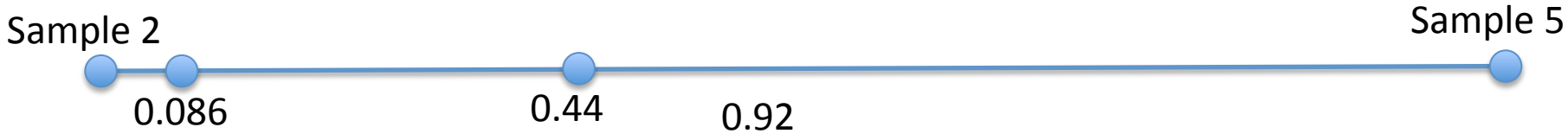
Step 3. Compute the location of each other sample on that axis as follows:

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
Sample 1	0	0.35	0.83	0.83	0.9	0.9
Sample 2	0.35	0	0.86	0.85	0.92	0.91
Sample 3	0.83	0.7	0	0.25	0.88	0.87
Sample 4	0.83	0.1	0.4	0	0.88	0.88
Sample 5	0.9	0.92	0.88	0.88	0	0.5
Sample 6	0.9	0.91	0.87	0.88	0.5	0

$$a = \frac{D^2 + D1^2 - D2^2}{2 \times D}$$

Sample 2:

$$\frac{0.92^2 + 0.86^2 - 0.88^2}{2 \times 0.92} = 0.44$$

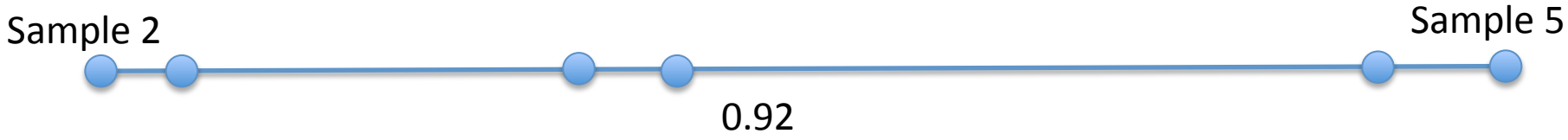


Step 3. Compute the location of each other sample on that axis as follows:

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
Sample 1	0	0.35	0.83	0.83	0.9	0.9
Sample 2	0.35	0	0.86	0.85	0.92	0.91
Sample 3	0.83	0.7	0	0.25	0.88	0.87
Sample 4	0.83	0.1	0.4	0	0.88	0.88
Sample 5	0.9	0.92	0.88	0.88	0	0.5
Sample 6	0.9	0.91	0.87	0.88	0.5	0

$$a = \frac{D^2 + D1^2 - D2^2}{2 \times D}$$

Repeat for all samples



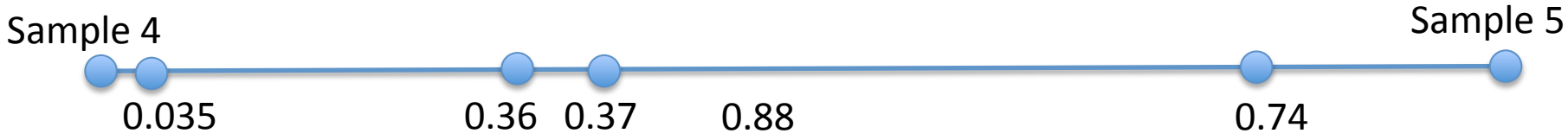
Step 4**.** Find the next largest distance that could be used to define an *uncorrelated axis*. This step can be labor-intensive to do by hand - usually you would compute all of the axes, along with correlation scores.

Step 5. Compute the location of each other sample on that axis as follows:

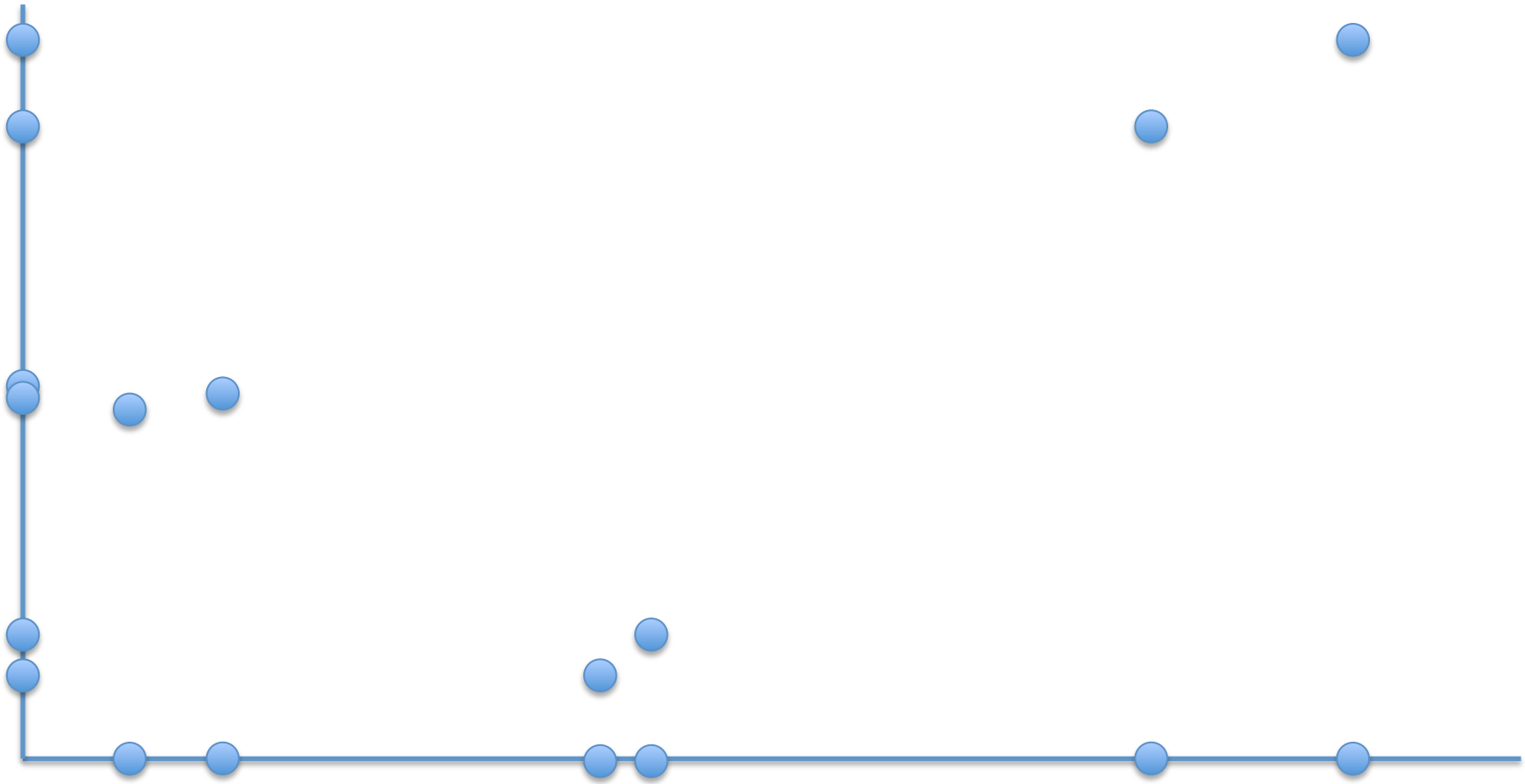
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
Sample 1	0	0.35	0.83	0.83	0.9	0.9
Sample 2	0.35	0	0.86	0.85	0.92	0.91
Sample 3	0.83	0.7	0	0.25	0.88	0.87
Sample 4	0.83	0.1	0.4	0	0.88	0.88
Sample 5	0.9	0.92	0.88	0.88	0	0.5
Sample 6	0.9	0.91	0.87	0.88	0.5	0

$$a = \frac{D^2 + D1^2 - D2^2}{2 \times D}$$

Repeat for all samples

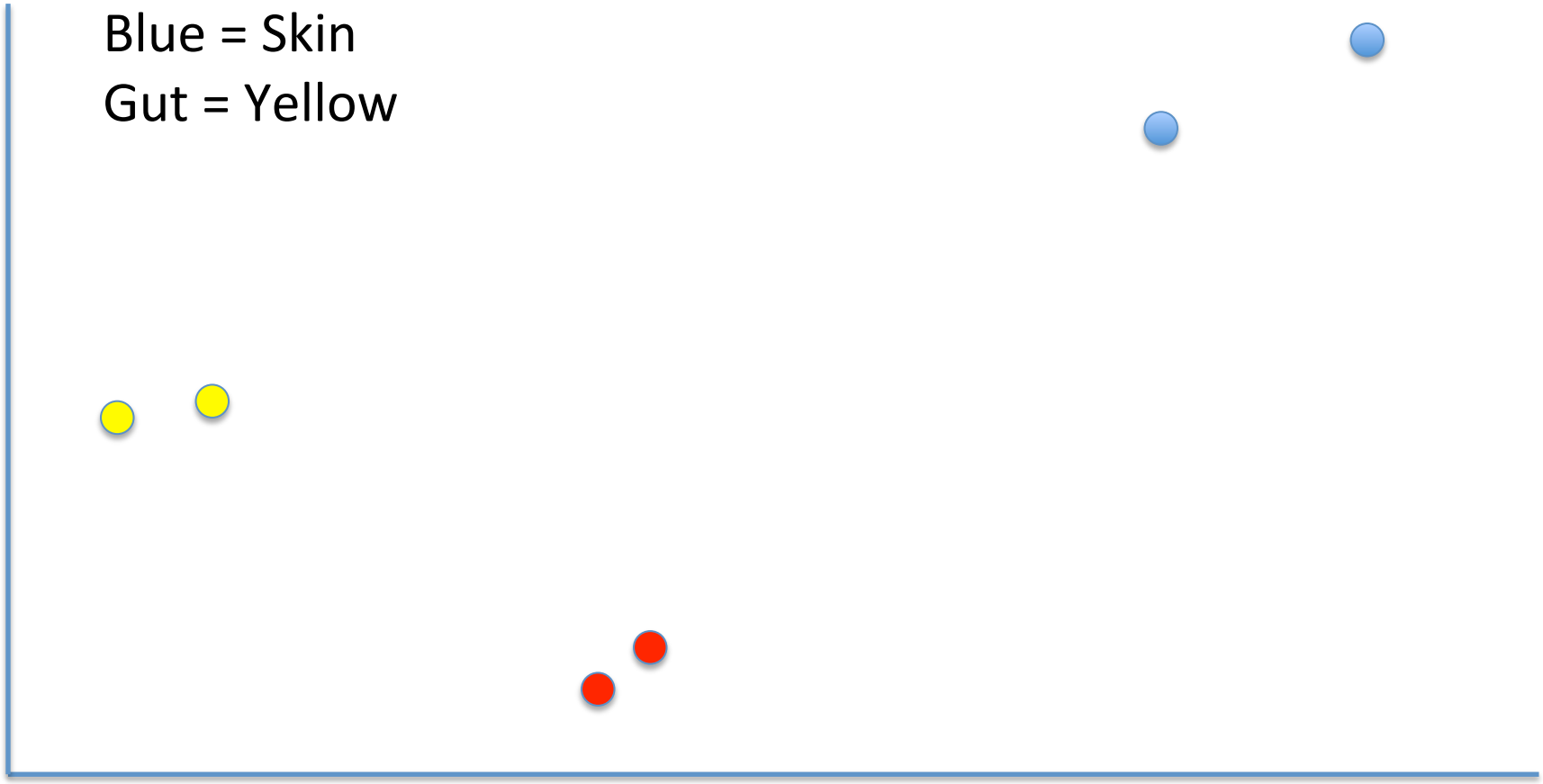


Step 6. Plot pairs of points:



Step 7. Color by metadata category :

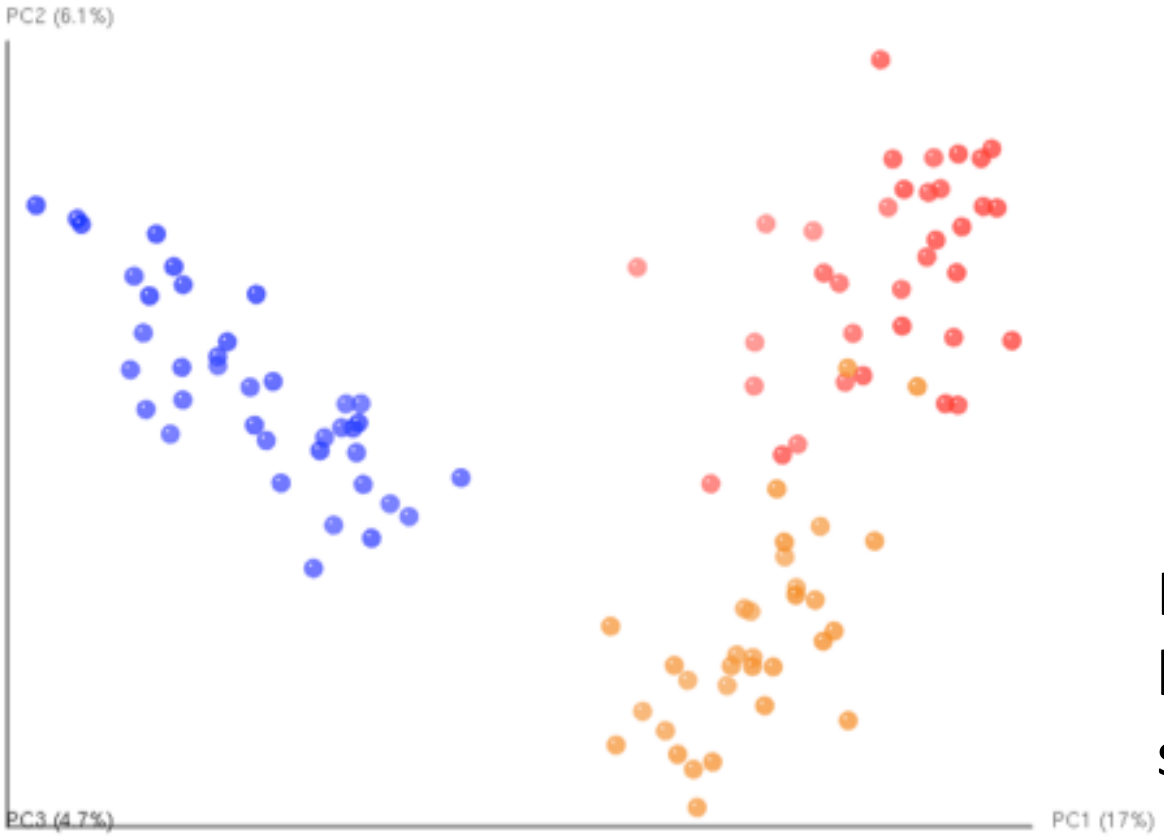
Red = Tongue
Blue = Skin
Gut = Yellow



- Anosim
  - Permutation
  - Compares within and between distances
  - R statistic -1 to 1. 0 indicates random grouping
  - P-values can be unreliable with a large sample size
  - Rank based
- Permanova
  - Works with any distances (as opposed to manova)
  - Permutations
- T-test
  - Not likely appropriate for a

Variation in sampling depth also needs to be controlled for beta diversity!

Variation in sampling depth is an important consideration

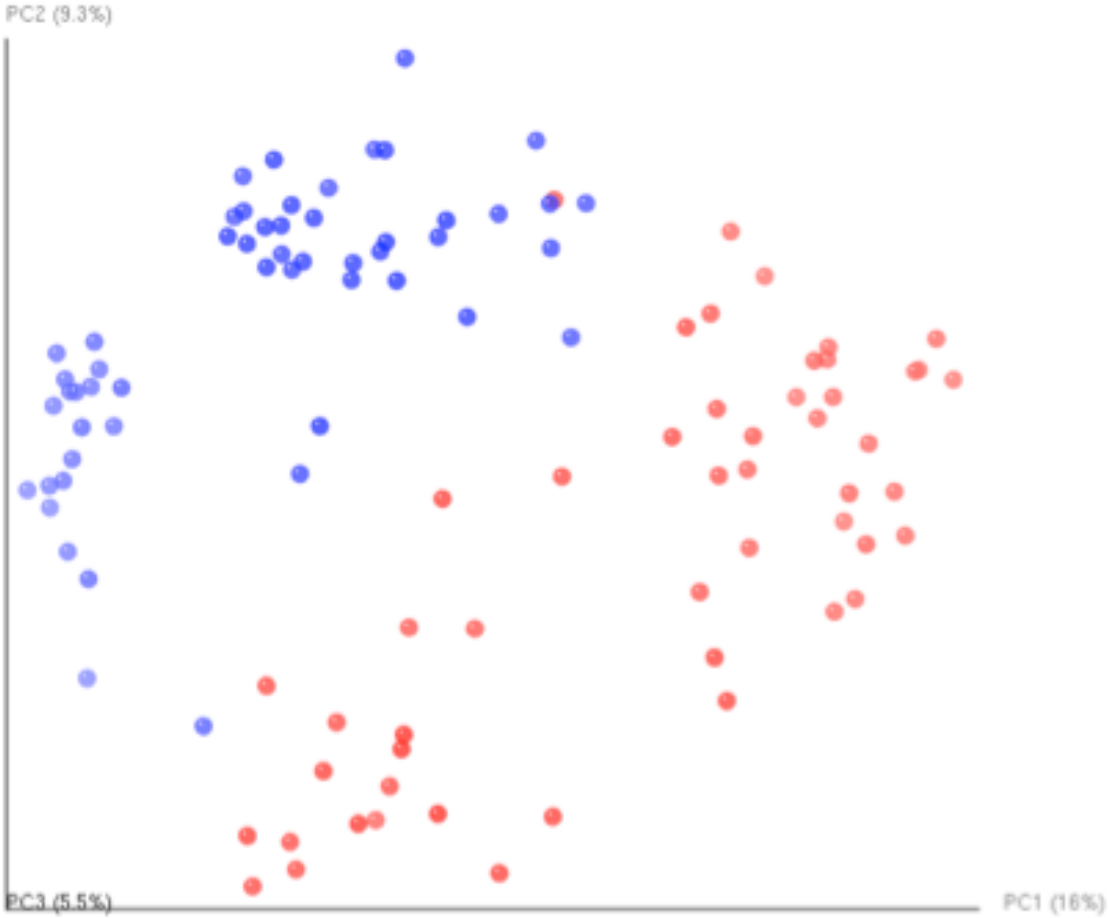


Human skin, colored by individual, at 500 sequence/sample

Image/analysis credit: Justin Kuczynski

Data reference:
Forensic identification using skin bacterial communities. Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R. Proc Natl Acad Sci U S A. 2010 Apr 6;107(14):6477-81.

Variation in sampling depth is an important consideration



Human skin, colored by sampling depth, at either 50 (blue) or 500 (red) sequences/sample

Image/analysis credit: Justin Kuczynski

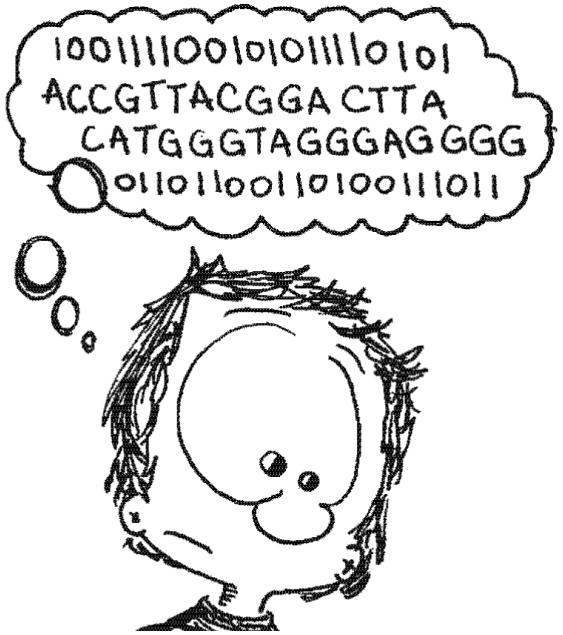
Data reference:
Forensic identification using skin bacterial communities. Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R. Proc Natl Acad Sci U S A. 2010 Apr 6;107(14):6477-81.

An Introduction To Applied Bioinformatics

An **I**ntroduction to **A**pplied **B**ioinformatics (or IAB) is a **free, open source interactive text** that introduces readers to core concepts of bioinformatics in the context of their implementation and application.

- To learn more about who should read IAB or how to read IAB, see *Reading IAB*.
- For an overview of the content, see the *IAB Table of Contents*.
- If you're interested in writing or editing content or code for IAB, see CONTRIBUTING.md.

http://readiab.org

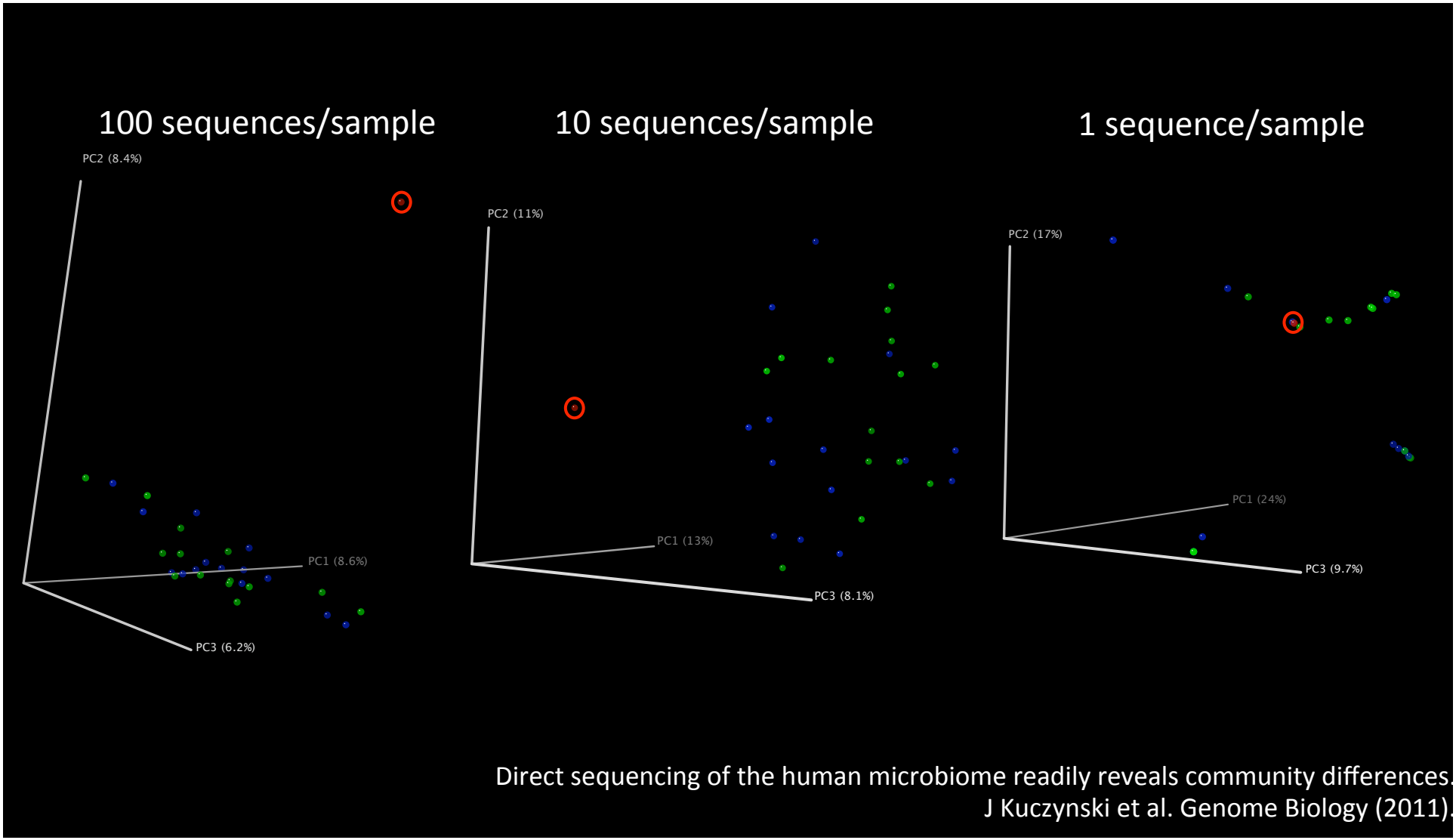


How deep is deep enough?

It depends on the question…

- Differences between community types: not many sequences.
- Rare biosphere: more (but be careful about sequencing noise!)

How deep is deep enough?



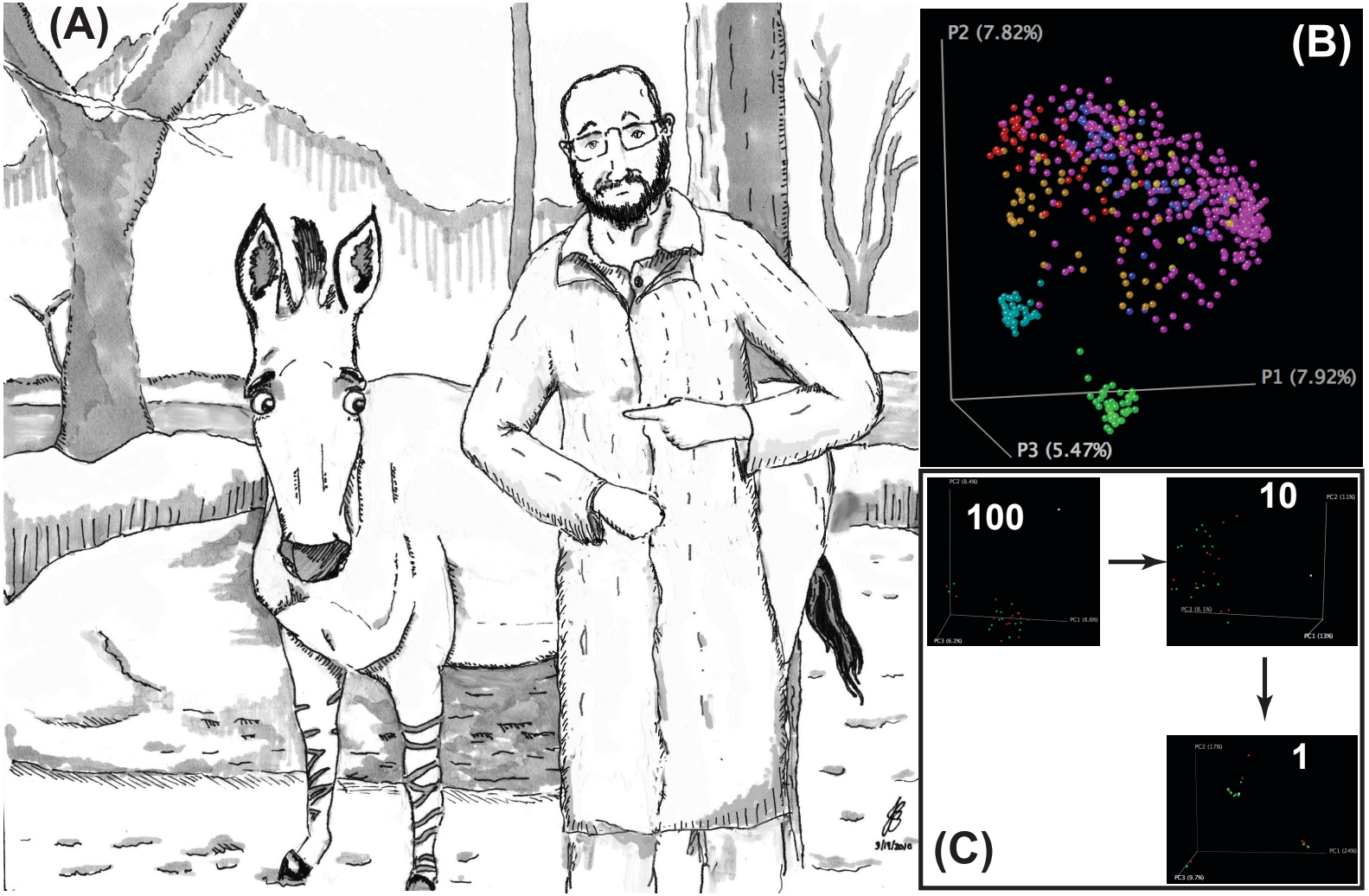
Direct sequencing of the human microbiome readily reveals community differences.
J Kuczynski et al. Genome Biology (2011).

Genome Biol. 2010;11(5):210. Epub 2010 May 5.

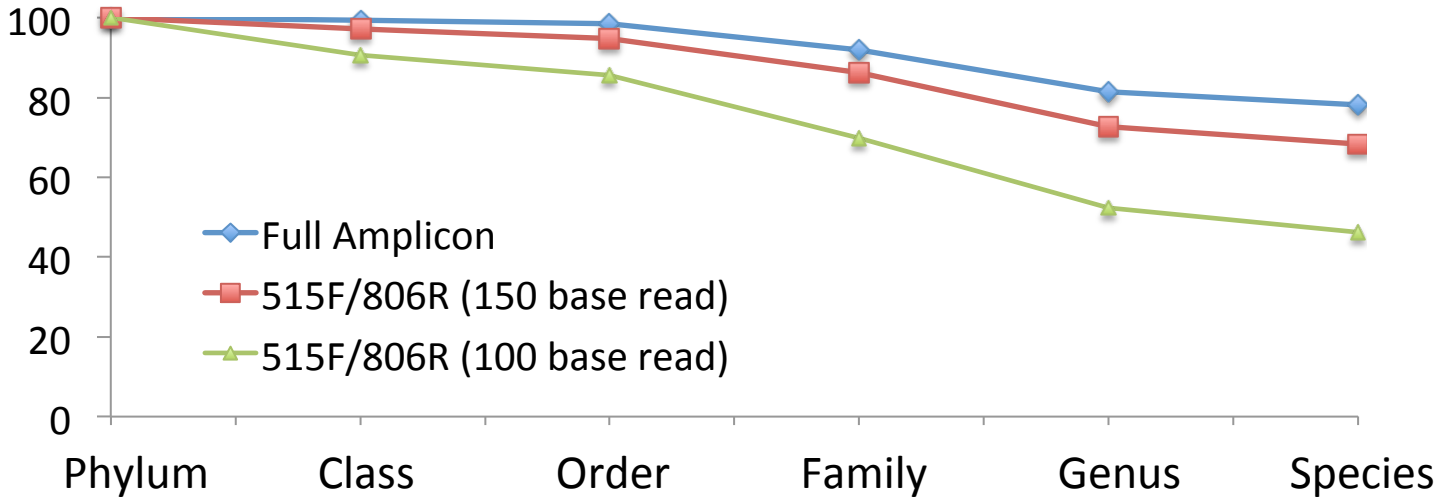
Direct sequencing of the human microbiome readily reveals community differences.

Kuczynski J, Costello EK, Nemergut DR, Zaneveld J, Lauber CL, Knights D, Koren O, Fierer N, Kelley ST, Ley RE, Gordon JI, Knight R.

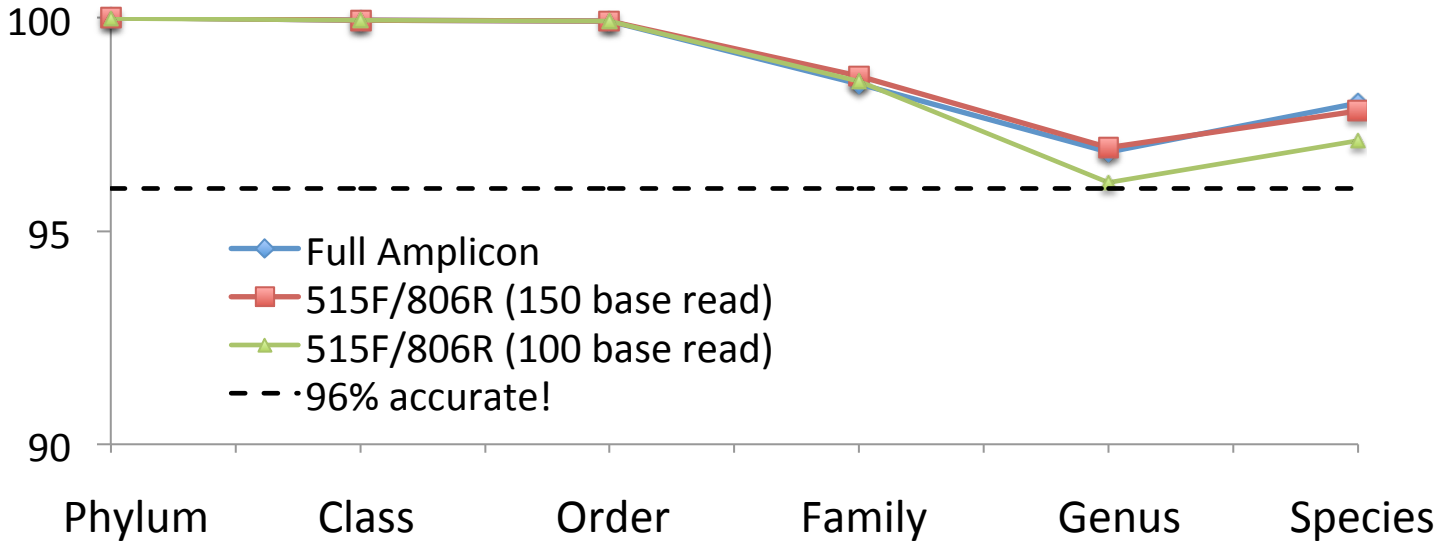
Figure 1



Fraction of Greengenes *simulated reads* classified by taxonomic level using the RDP Classifier (80% confidence)

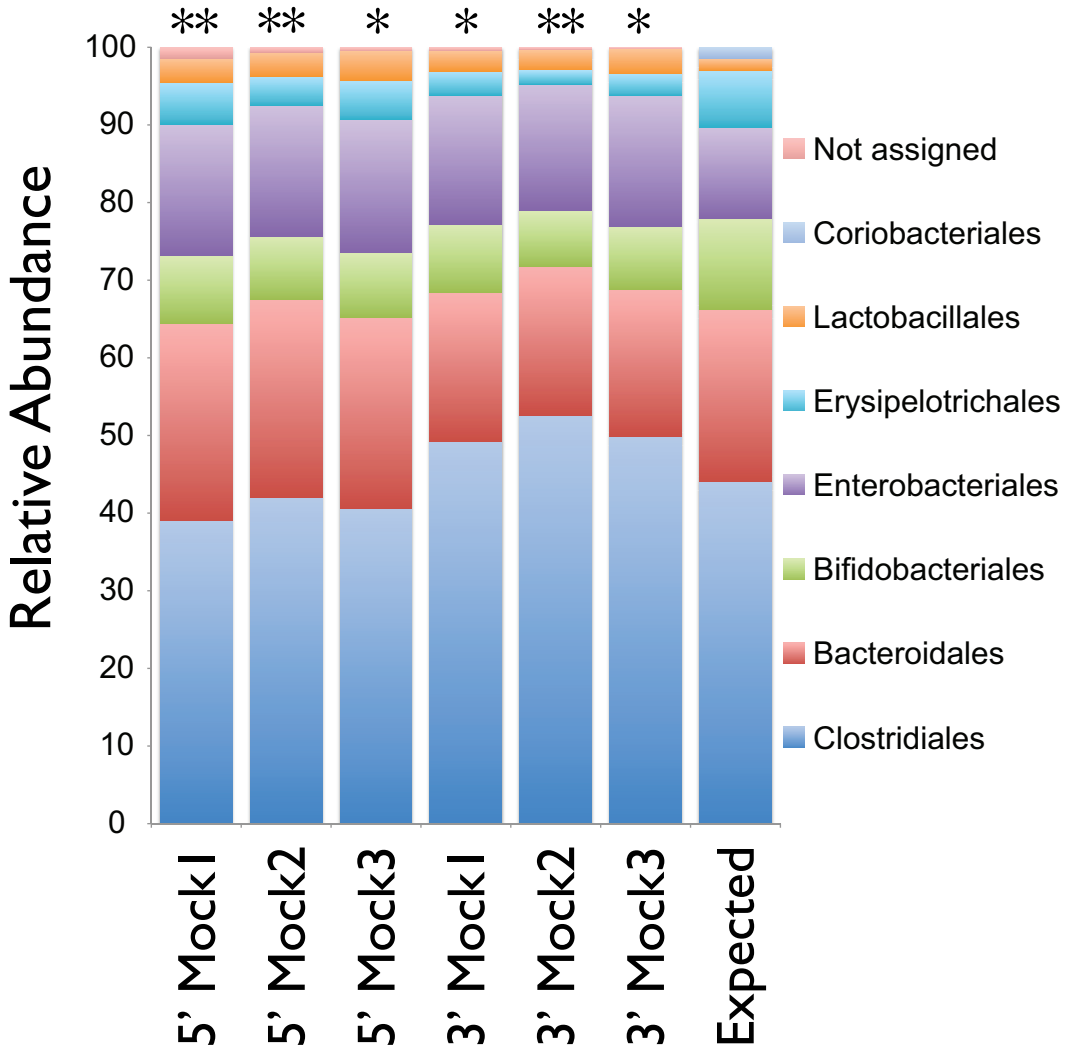


Accuracy of classified reads



Acknowledgement: Tony Walters

Can accurate taxonomy assignments be achieved?



Order-level taxonomy assignments

G-test (goodness of fit)

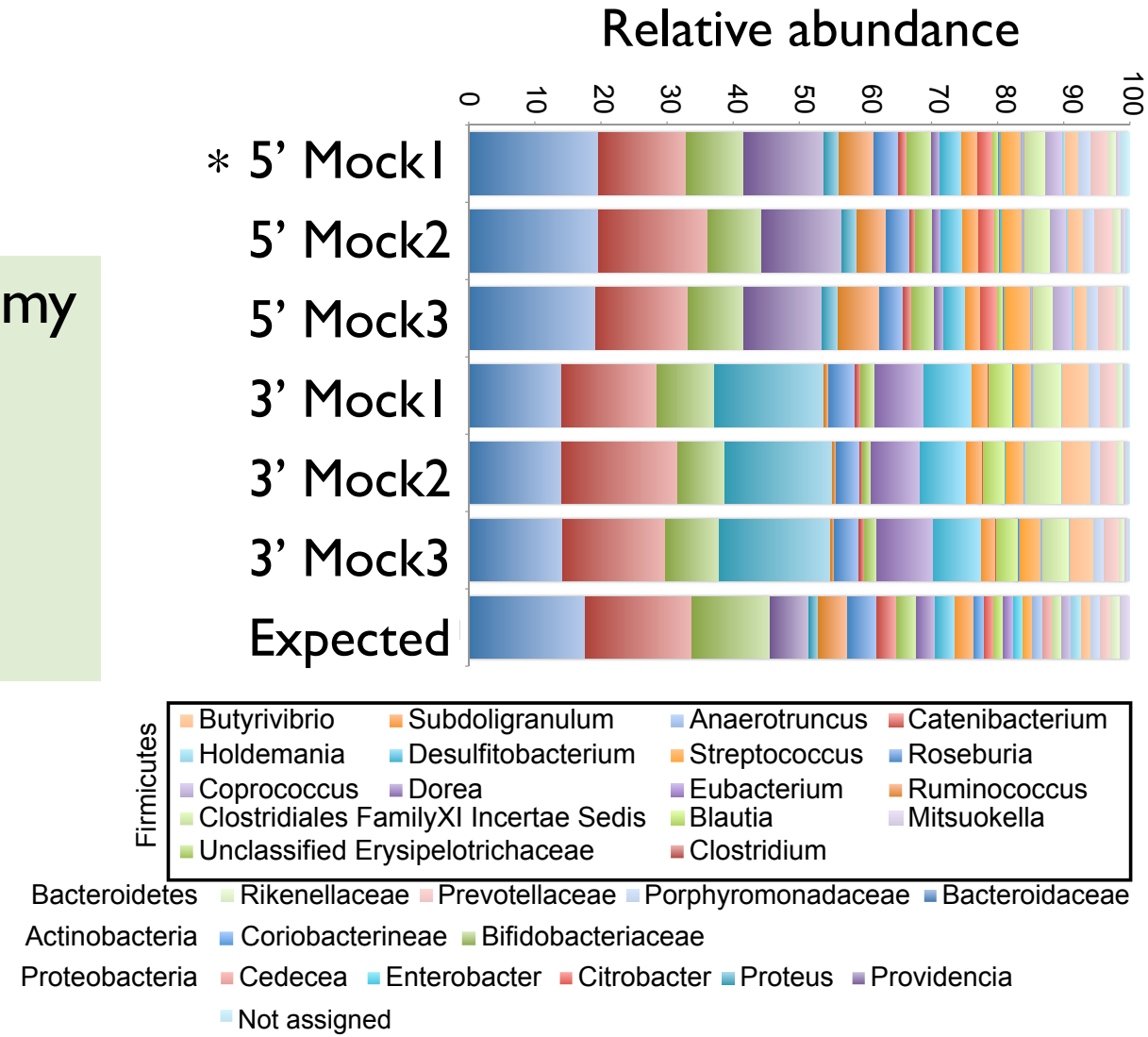
** $p < 0.01$

* $p < 0.05$

Can accurate taxonomy assignments be achieved?

Genus-level taxonomy assignments

G-test (goodness of fit)
** $p < 0.01$
* $p < 0.05$



Slides compiled by:

Greg Caporaso
John Chase
Jose Clemente
Antonio Gonzalez Peña
Rob Knight
Cathy Lozupone
Daniel McDonald
Jai Rideout
Yoshiki Vázquez Baeza

www.qiime.org