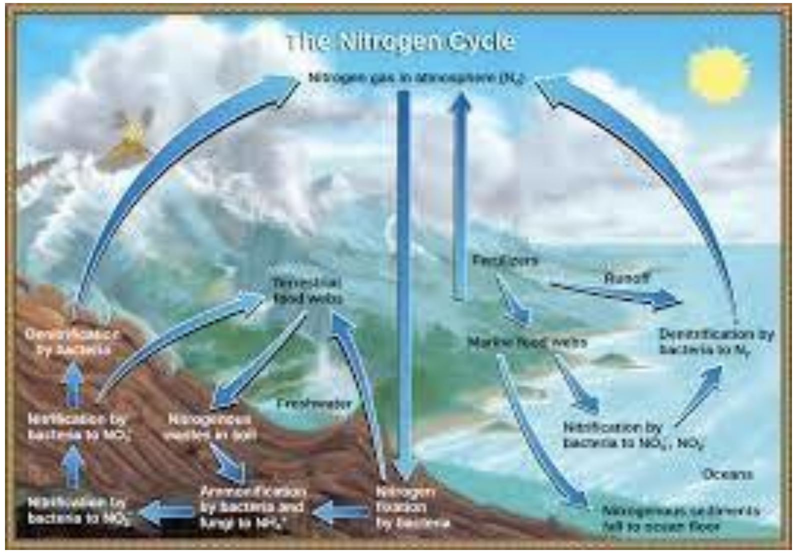
Microbial Community Profiling

Microbes were the first life forms

- First photosynthetic bacteria 3.4 BYA
- First oxygen producers emerged 2.7 BYA
- Available atmospheric oxygen 2.3 BYA
- Terrestrial plants appeared 500 MYA
- Avian flight 13 MYA
- Homo appears 10 MYA
- Human start studying microbes 400 YA

Biogeochemical cycles depend on microbes

- Proportions of elements on earth is constant
- Recycling, flux and bioavailability is the domain of microbes
- Especially nitrogen:


The Nitrogen Cycle
Nitrogen gas in atmosphere (N_2)
Fertilizers
Runoff
Terrestrial food webs
Denitrification by bacteria
Marine food webs
Denitrification by bacteria to N_2
Freshwater
Nitrification by bacteria to NO_3^-
Nitrogenous wastes in soil
Nitrification by bacteria to NO_2^-, NO_3^-
Oceans
Nitrification by bacteria to NO_2^-
Ammonification by bacteria and fungi to NH_4^+
Nitrogen fixation by bacteria
Nitrogenous sediments fall to ocean floor

- 78% of Earths atm is N2
- Required for important biological processes
- In gaseous form it is unavailable
- In fact many processes are N2 limited
- Making N2 bioavailable in a form that can be by eukaryotes depends (almost) completely on microbes

Quotes and facts

- “Microbes make up 80 percent of all biomass” -Carl Woese.
- "If you don't like bacteria, you're on the wrong planet. This is the planet of the bacteria.” -Craig Venter
- The human microbiome in our gut, mouth, skin, and elsewhere, harbors 3,000 kinds of bacteria with 3 million distinct genes.
- Most of the metabolism in the world is microbial

https://www.edge.org/response-detail/11863

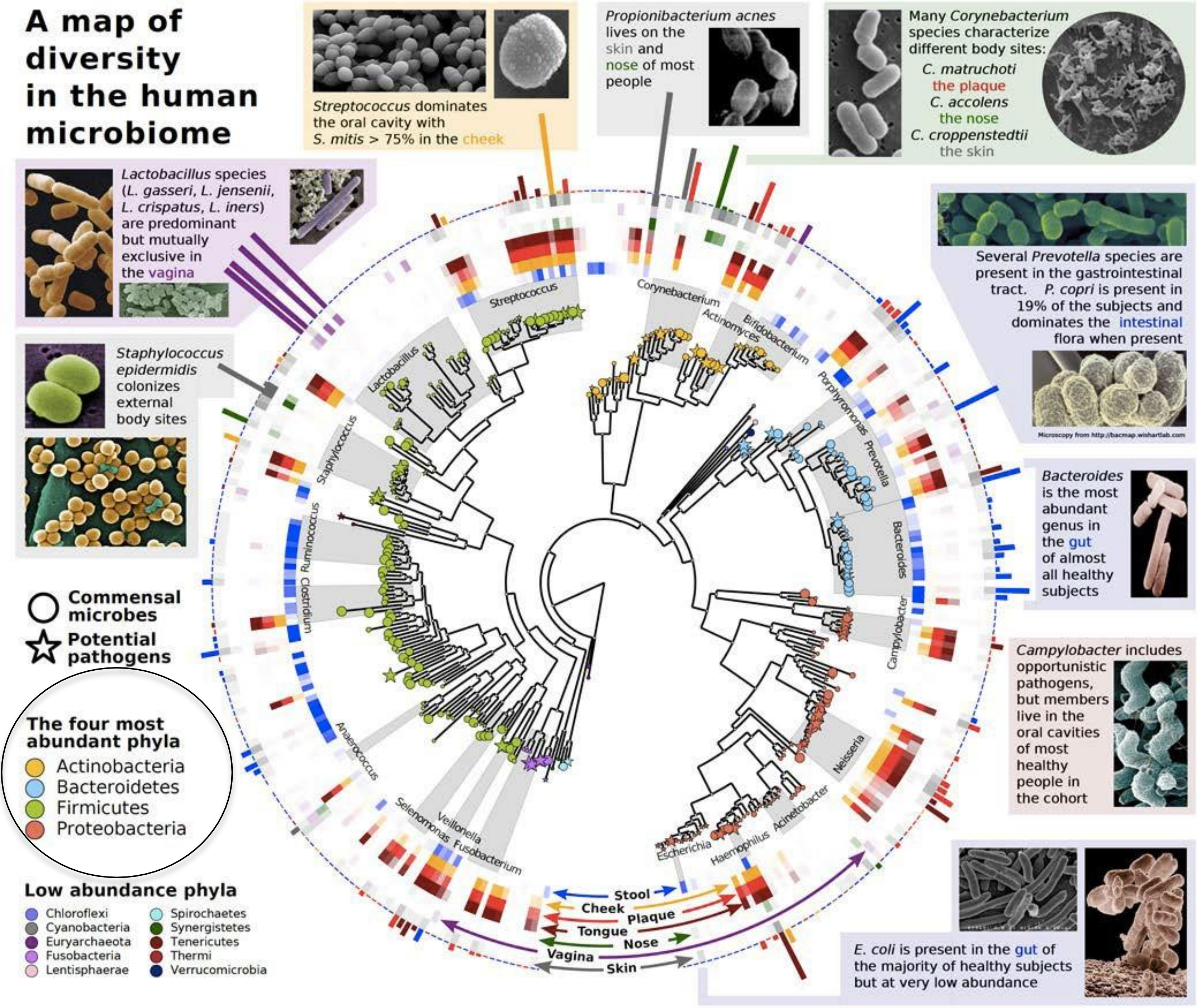
- What makes microbes so special?
  - -15ºC to 130ºC
  - 0 to 12.8 pH
  - More than 200 atm pressure
  - 4 miles deep into Earth’s crust
  - Up to 5kGy radiation

Grand Prismatic Spring – YNP – 183ºC

Validates the importance of microbes and sums up life on Earth with boundaries

Microbes are constantly trying to evolve and get deeper and deeper into the hot springs

Eukaryotes only surround this spring – cannot survive close to the hot spring



A map of diversity in the human microbiome

Streptococcus dominates the oral cavity with *S. mitis* > 75% in the cheek

Propionibacterium acnes lives on the skin and nose of most people

Many *Corynebacterium* species characterize different body sites:
C. matruchoti the plaque
C. accolens the nose
C. croppenstedtii the skin

Lactobacillus species (*L. gasseri*, *L. jensenii*, *L. crispatus*, *L. iners*) are predominant but mutually exclusive in the vagina

Staphylococcus epidermidis colonizes external body sites

Several *Prevotella* species are present in the gastrointestinal tract. *P. copri* is present in 19% of the subjects and dominates the intestinal flora when present

Microscopy from http://bacmap.wishartlab.com

Bacteroides is the most abundant genus in the gut of almost all healthy subjects

Campylobacter includes opportunistic pathogens, but members live in the oral cavities of most healthy people in the cohort

E. coli is present in the gut of the majority of healthy subjects but at very low abundance

	Human (isolated)	Microbiota
weight	~ 50-100 kg	~ 2 kg
species	1	1000-5000
cells	~ 10^{12}	10^{13} - 10^{14}
genes	25.000	>4.000.000

DOI: 10.1016/j.tig.2012.09.005

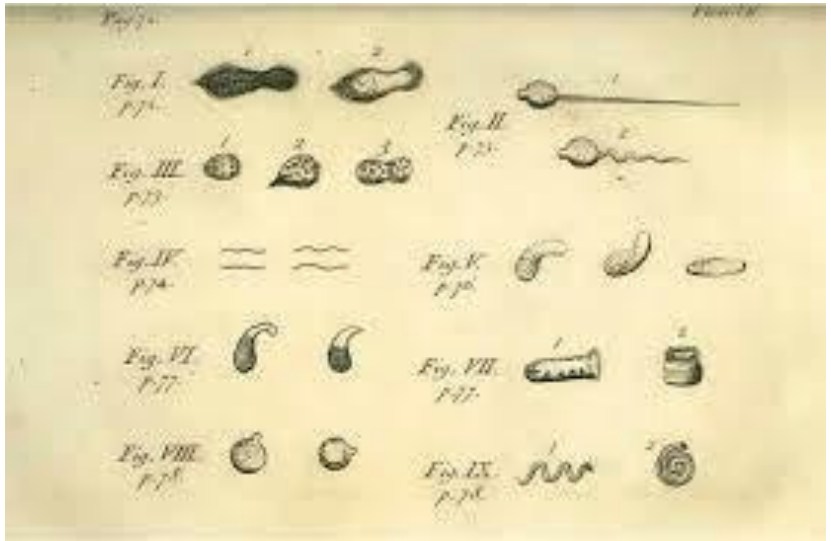
Microbial abundance

- 10^6 in 1 ml of fresh water
- 4 x 10^6 in 1 g of soil
- 10^{13-14} in a human body

Study methods reflect the state of science

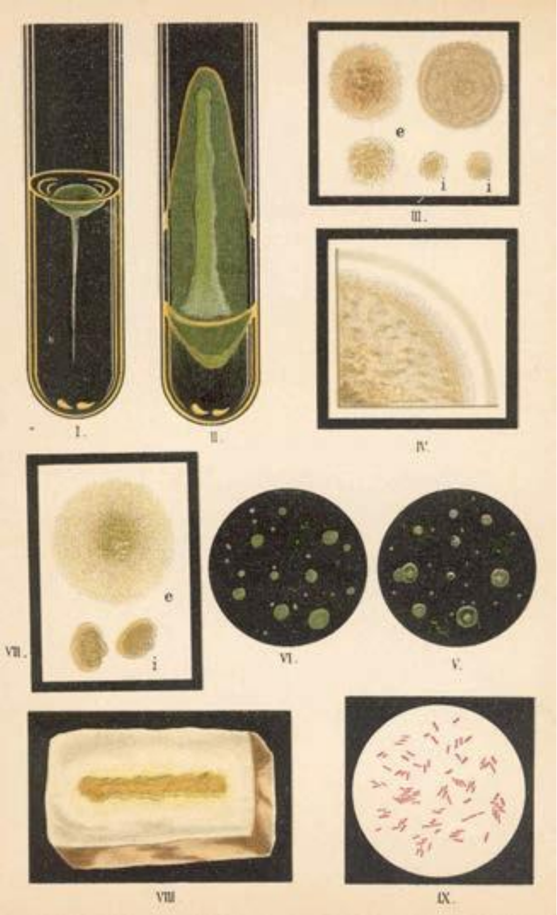
- Microscopy
- Culture
- Culture-free methods

Antonie van Leeuwenhoek (1632–1723)



- Bacteria
- Protists
- Vacuoles
- Spermatozoa
- Muscle fibers

Robert Koch (1843 –1910)



Koch’s postulates

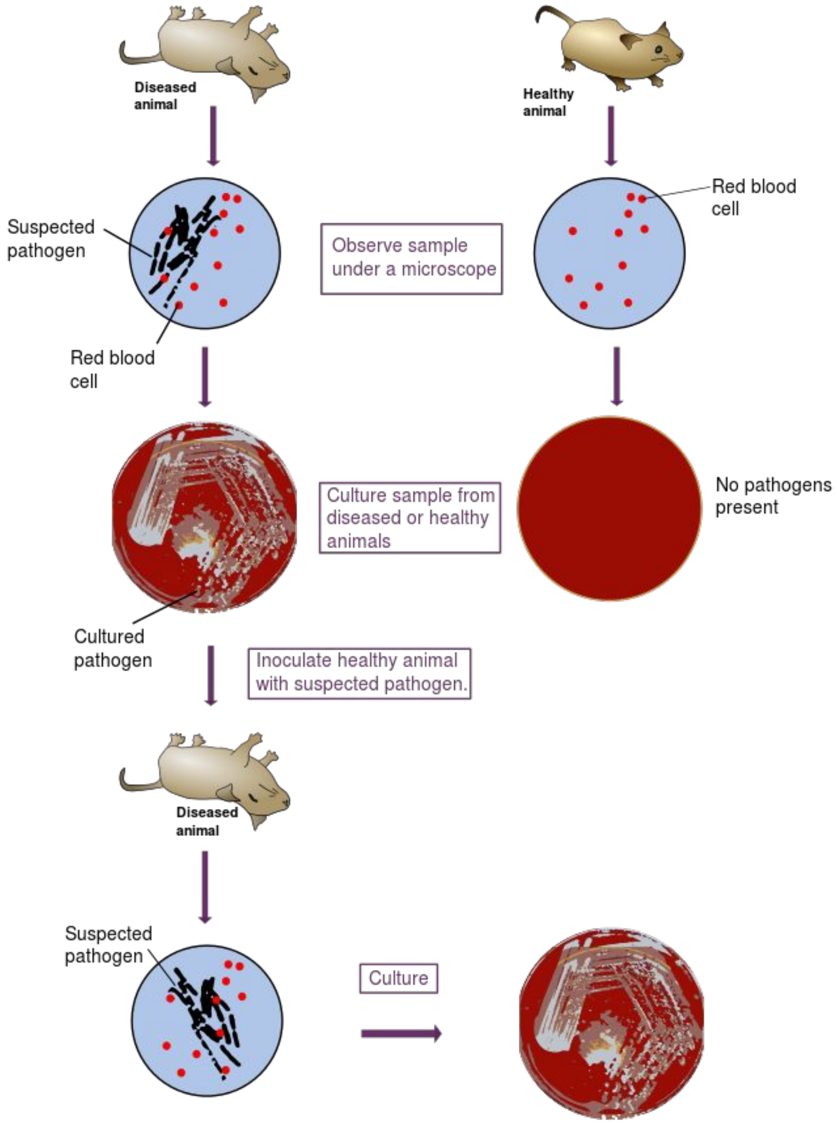
Koch's Postulates:

① The microorganism must be found in abundance in all organisms suffering from the disease, but should not be found in healthy organisms.

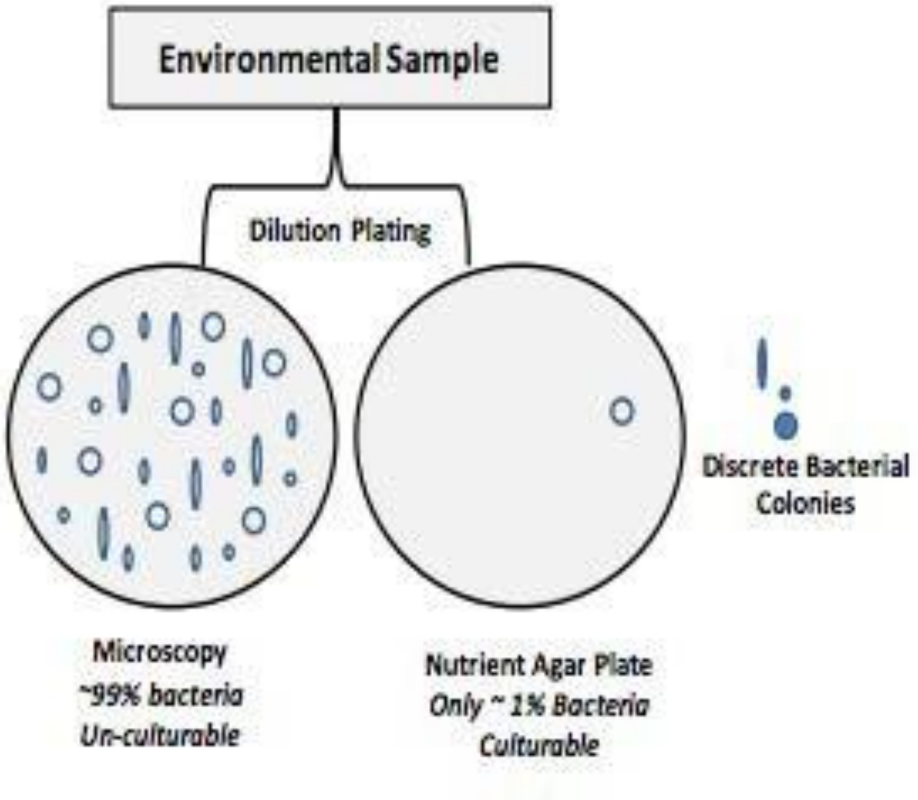
② The microorganism must be isolated from a diseased organism and grown in pure culture.

③ The cultured microorganism should cause disease when introduced into a healthy organism.

④ The microorganism must be reisolated from the inoculated, diseased experimental host and identified as being identical to the original specific causative agent.



The “plate count” anomaly



IJSR, Sept 2013

- Cultivation based cell counts are orders of magnitude lower than direct microscopic observation
- As microbiologists, we are able to cultivate only a small minority of naturally occurring microbes
- Our nucleic acid derived understanding of microbial diversity has rapidly outpaced our ability to culture new microbes

Roadmap to Culture Independent Techniques

1977: rRNA as an evolutionary marker (Woese and Fox, PNAS)

1985: Polymerase Chain Reaction (K. Mullis, Science)

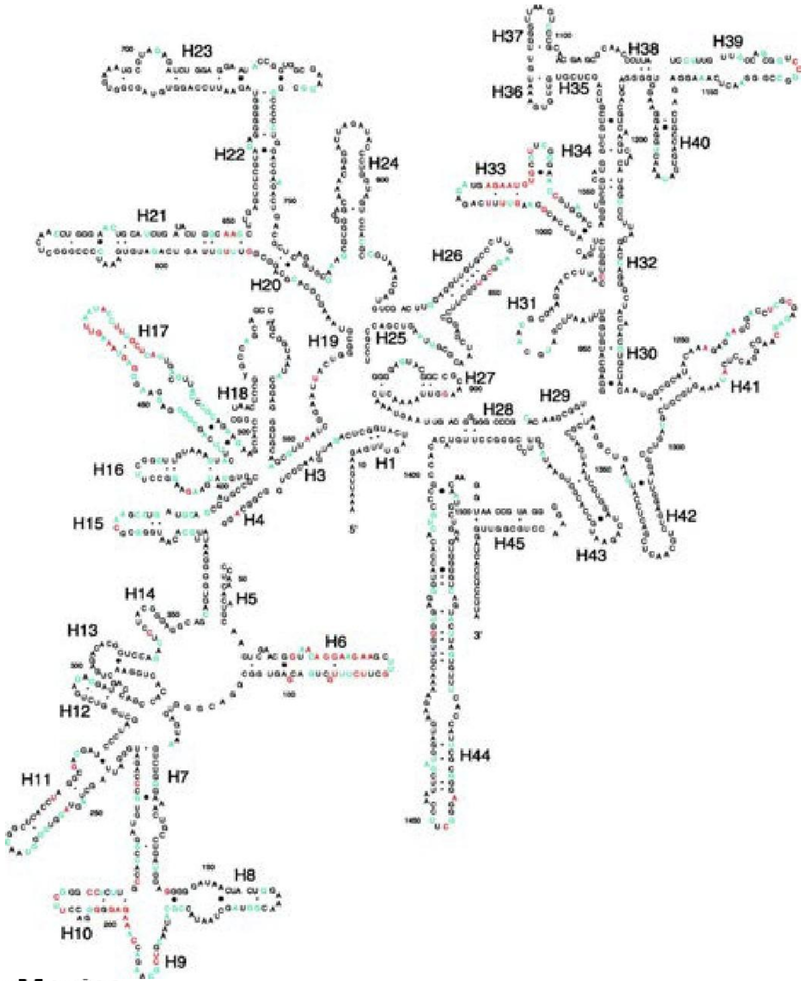
1985: "Universal Primers" for rRNA sequencing (N. Pace, PNAS)

1989: PCR amplification of 16S rRNA gene (Bottger, FEMS Microbiol)

Early 1990's: Curation and hosting of RDP (rRNA database)

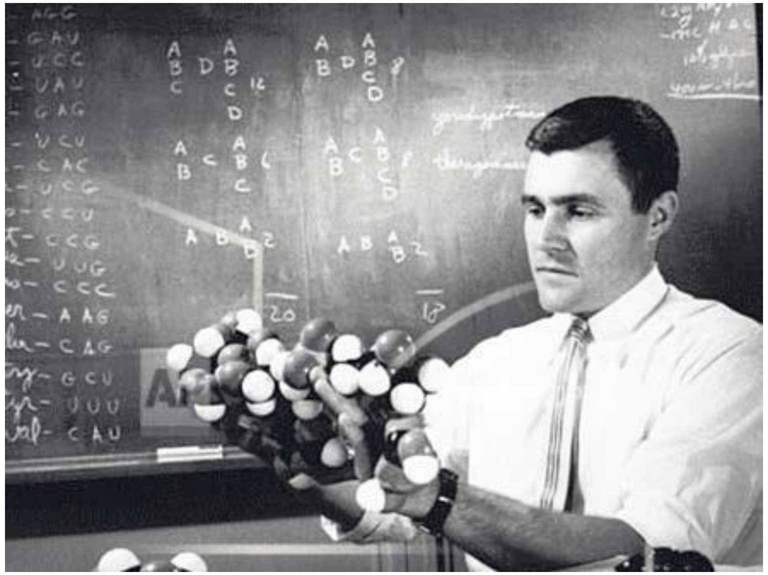
2001: Term 'microbiome' coined by Lederberg and McCray

16S rRNA as an evolutionary chronometer

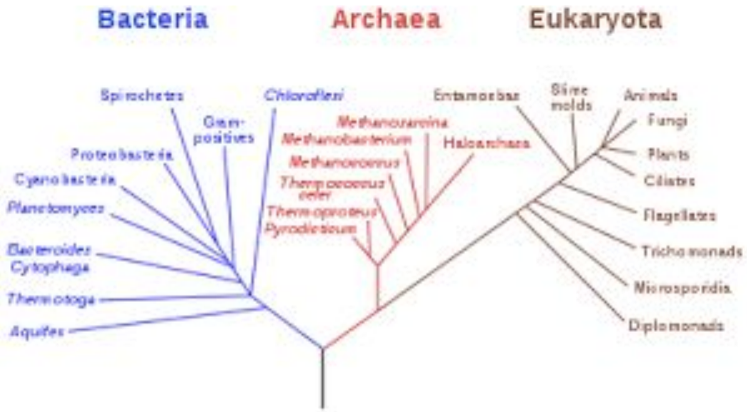


- Ubiquitous and ancient – present in all known life
- Functionally constant wrt translation and secondary structure
- Evolves very slowly – mutations are extremely rare
- Large enough to extract information for evolutionary inference
- Limited exchange – limited examples of rRNA gene sharing between organisms

DOI: 10.1128/AEM.01177-06

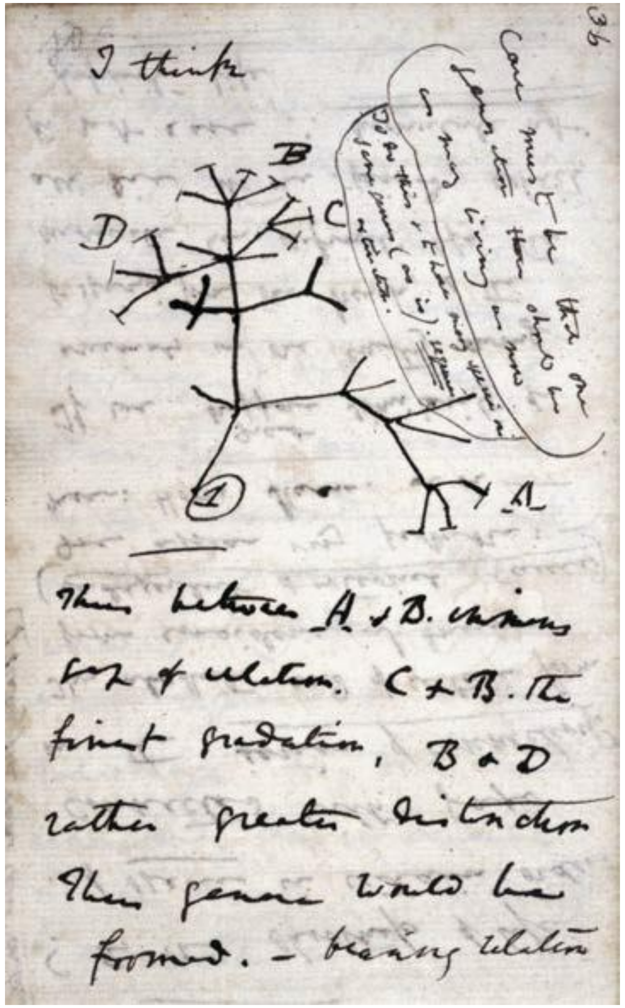


Carl Woese, 1977



DOI: 10.1016/s0960-9822(02)70664-7

Charles Darwin in 1837

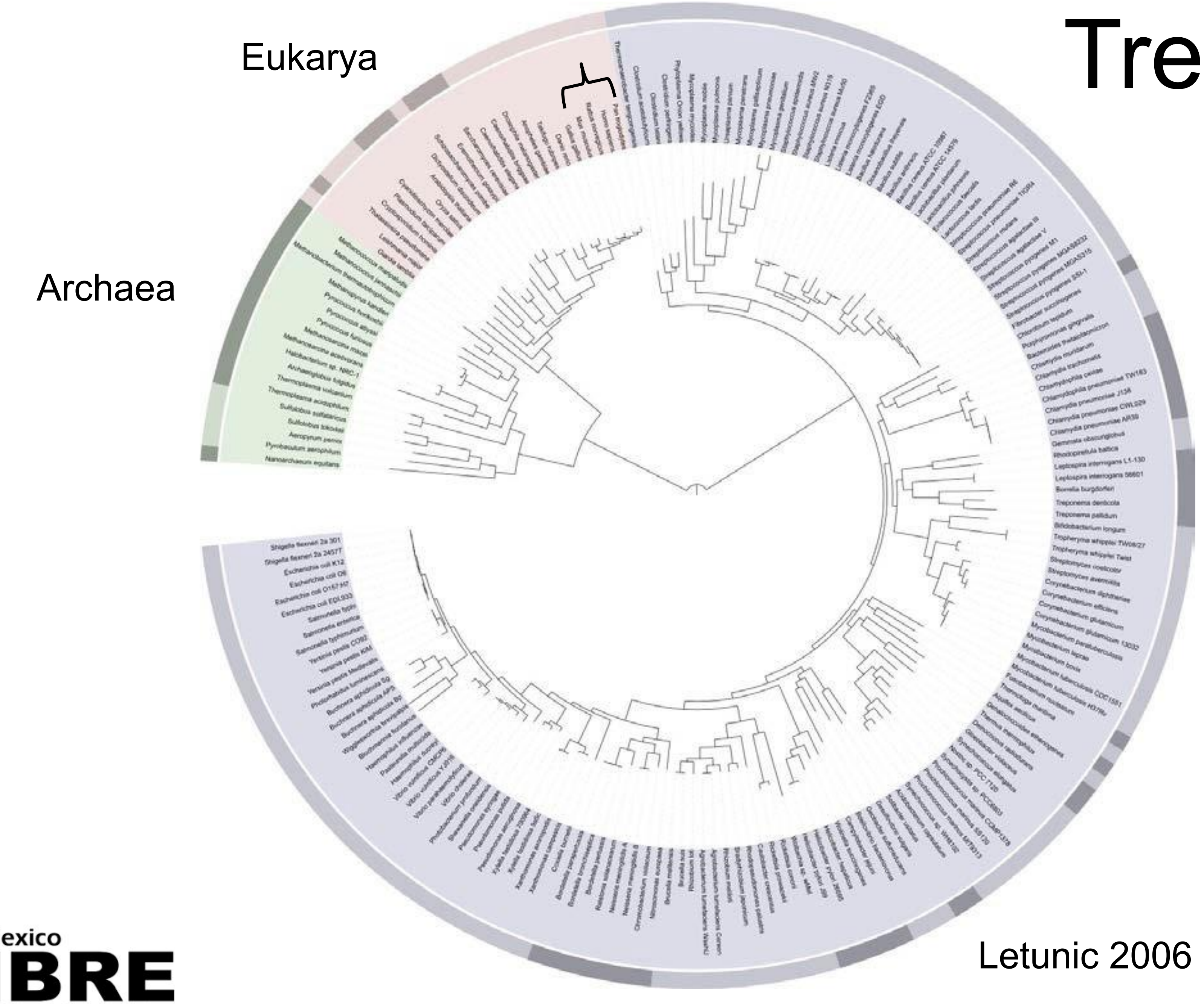

36
I think
B
C
D
1
A

Case must be that one generation then should be as many living as now. To do this & to have many species in same genus (as is) requires extinction.

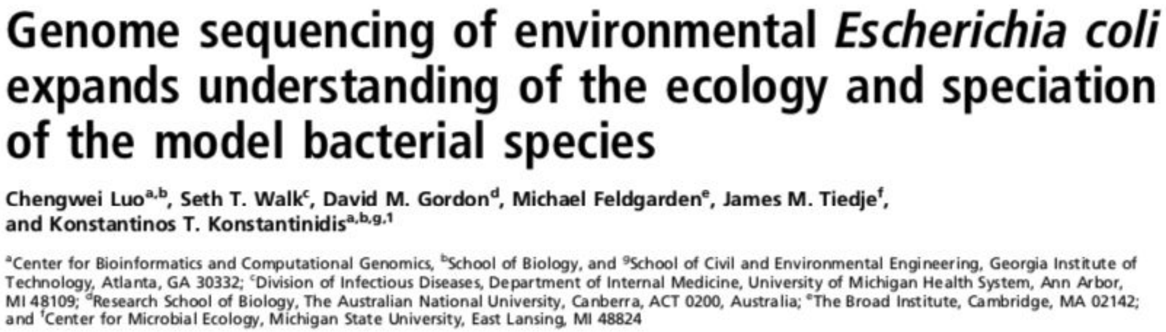
Thus between A & B. immense gap of relation. C & B. the finest gradation, B & D rather greater distinction. Thus genera would be formed. — bearing relation

- Introduced the idea of descent from a common ancestor
- Not just a hierarchical relationship

Tree of Life



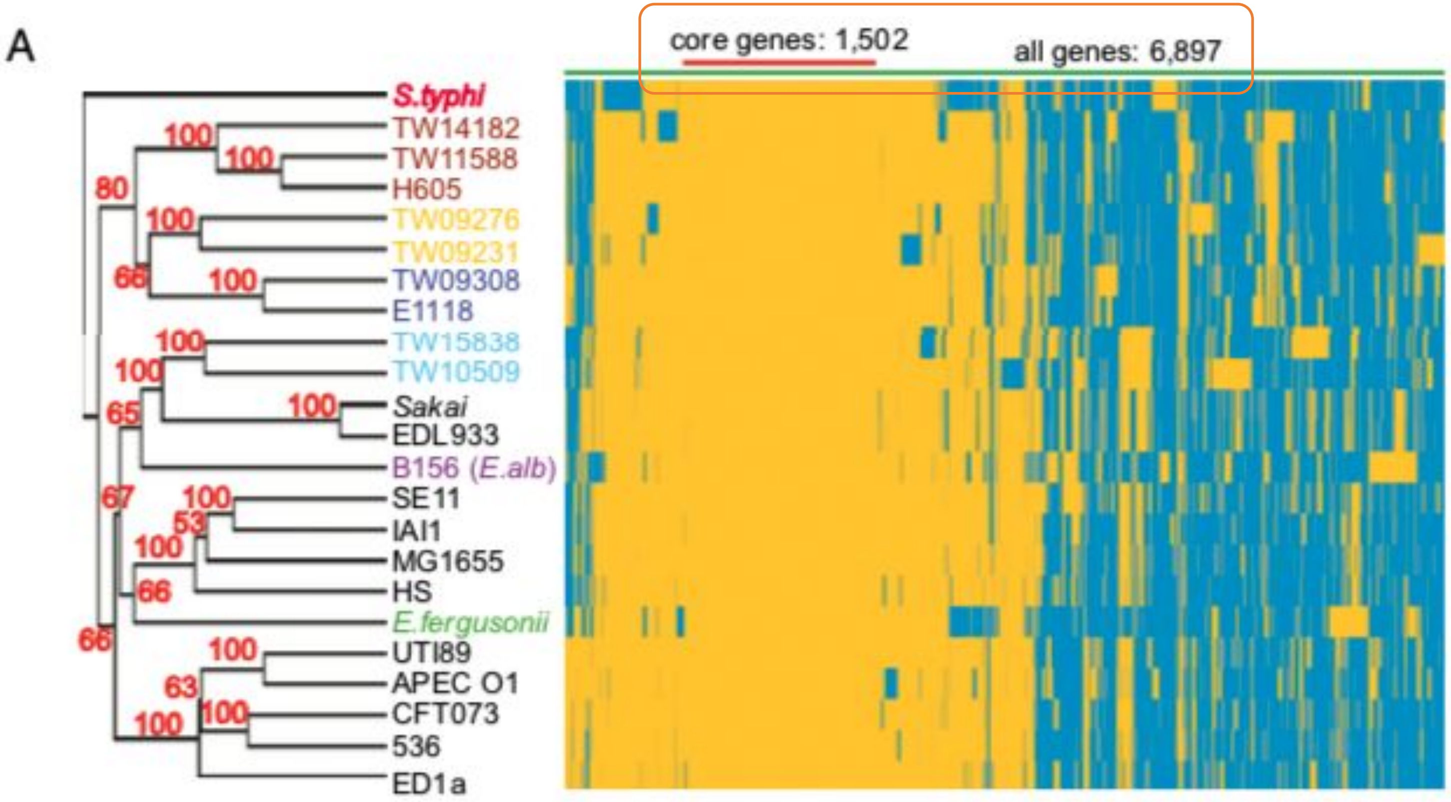
Letunic 2006



Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species

Chengwei Luo[a,b], Seth T. Walk[c], David M. Gordon[d], Michael Feldgarden[e], James M. Tiedje[f], and Konstantinos T. Konstantinidis[a,b,g,1]

[a]Center for Bioinformatics and Computational Genomics, [b]School of Biology, and [g]School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332; [c]Division of Infectious Diseases, Department of Internal Medicine, University of Michigan Health System, Ann Arbor, MI 48109; [d]Research School of Biology, The Australian National University, Canberra, ACT 0200, Australia; [e]The Broad Institute, Cambridge, MA 02142; and [f]Center for Microbial Ecology, Michigan State University, East Lansing, MI 48824



www.pnas.org/cgi/doi/10.1073/pnas.1015622108

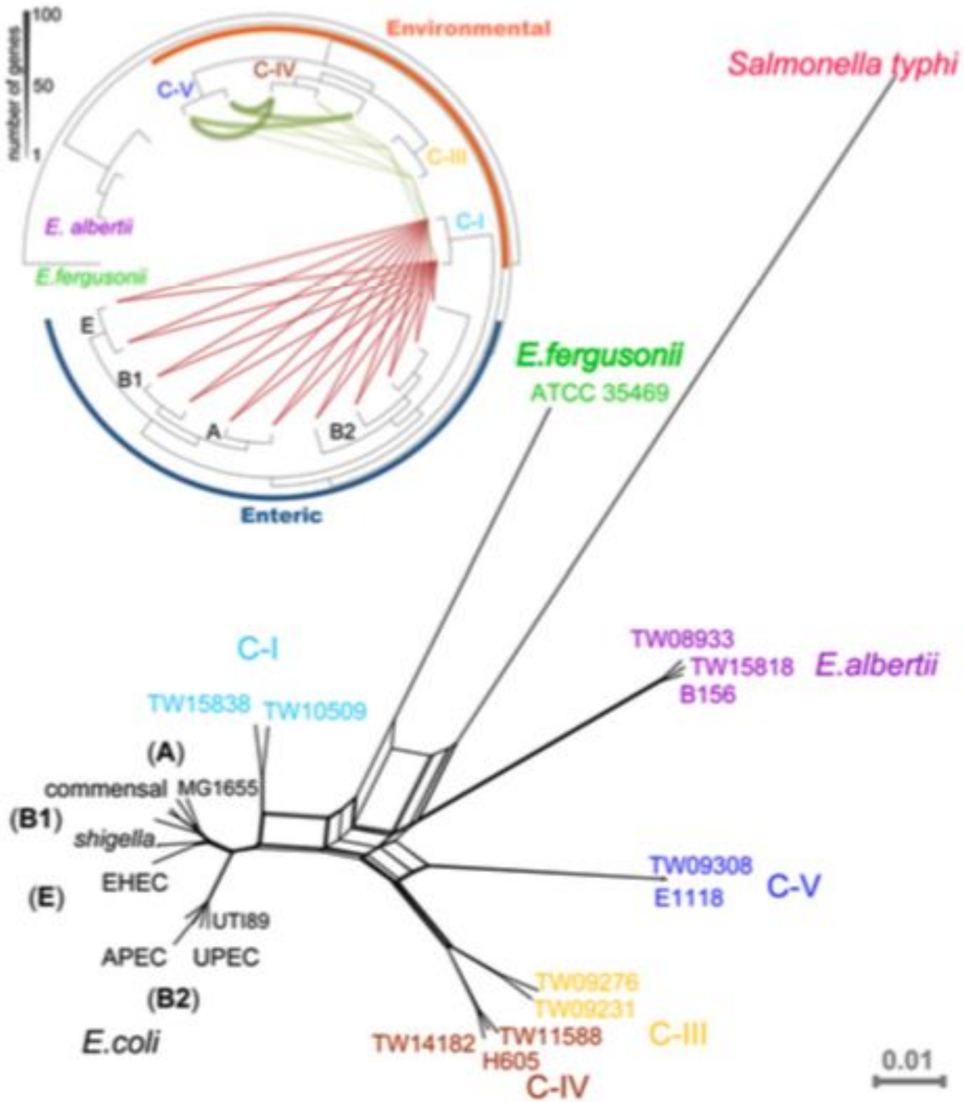
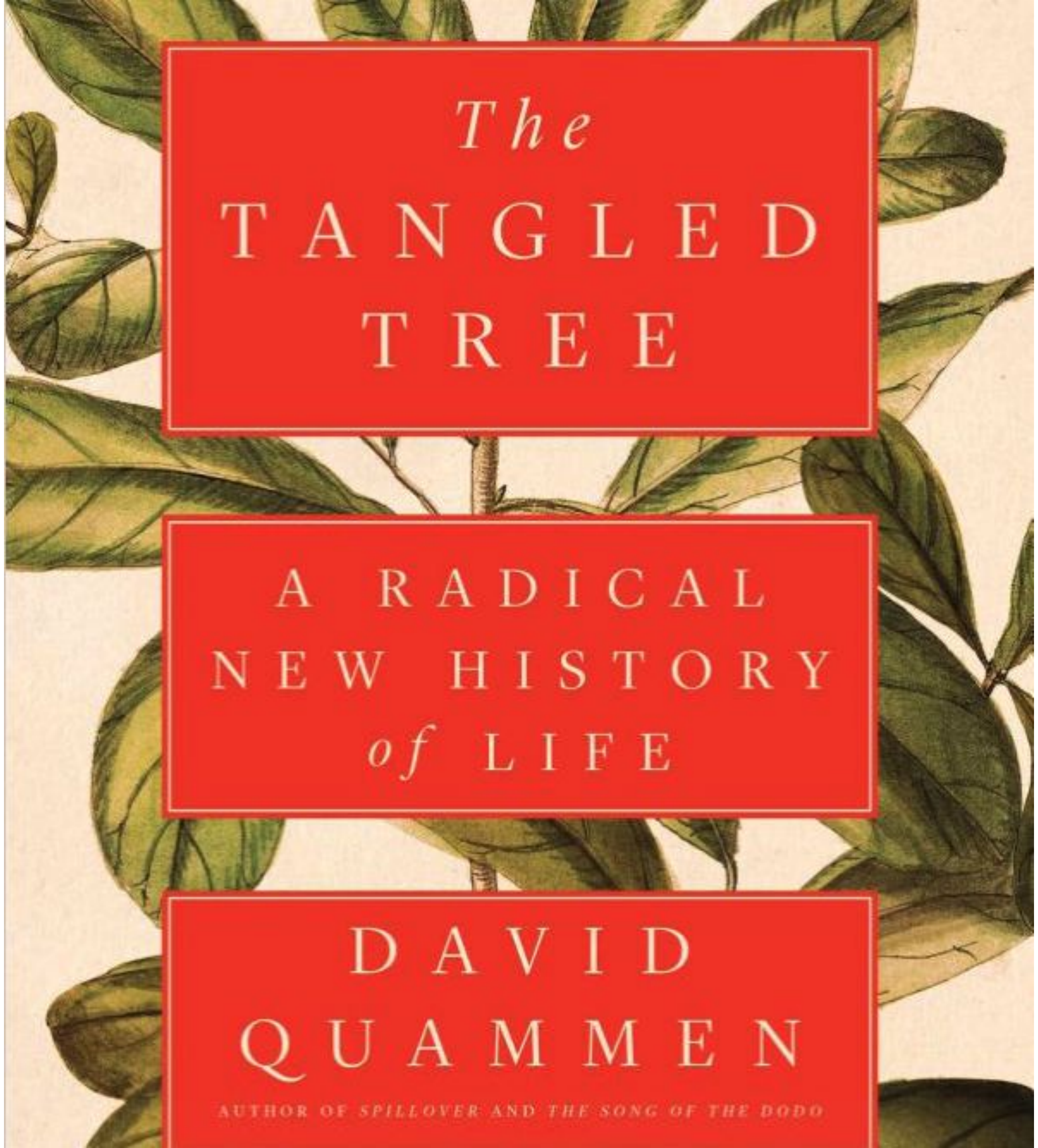


Fig. 1. Whole-genome phylogeny of the *Escherichia* genomes used in the study. The phylogenetic network shown was constructed with the SplitsTree software (27), using as input the concatenated alignment of 1,910 single-copy core genes. (*Inset*) The graph represents the amount of recent horizontal transfer of core genes between the genomes of the clades. The thickness of the line is proportional to the number of genes transferred (scale at upper left in figure).

www.pnas.org/cgi/doi/10.1073/pnas.1015622108



The
TANGLED
TREE

A RADICAL
NEW HISTORY
of LIFE

DAVID
QUAMMEN

AUTHOR OF *SPILLOVER* AND *THE SONG OF THE DODO*

We need microbial classification in order to study microbes, but please be aware of the limitations of how we think about species and the tree of life.

Doolittle, W.F., 2009. Eradicating typological thinking in prokaryotic systematics and evolution. Cold Spring Harb Symp Quant Biol 74, 197–204. https://doi.org/10.1101/sqb.2009.74.002

Dagan, T., Martin, W., 2006. The tree of one percent. Genome Biol 7, 118. https://doi.org/10.1186/gb-2006-7-10-118

What is a microbiome?

- Totality of microbes in a defined environment, and their intricate interactions with each other and the surrounding environment
- Microbes seldom work alone
- Monoculture is extremely rare outside of lab and in some infections
- A microbiome is a mixed population of different microbial species
- Most microbial activities are performed by complex communities of microorganisms
- Mixed community is the norm

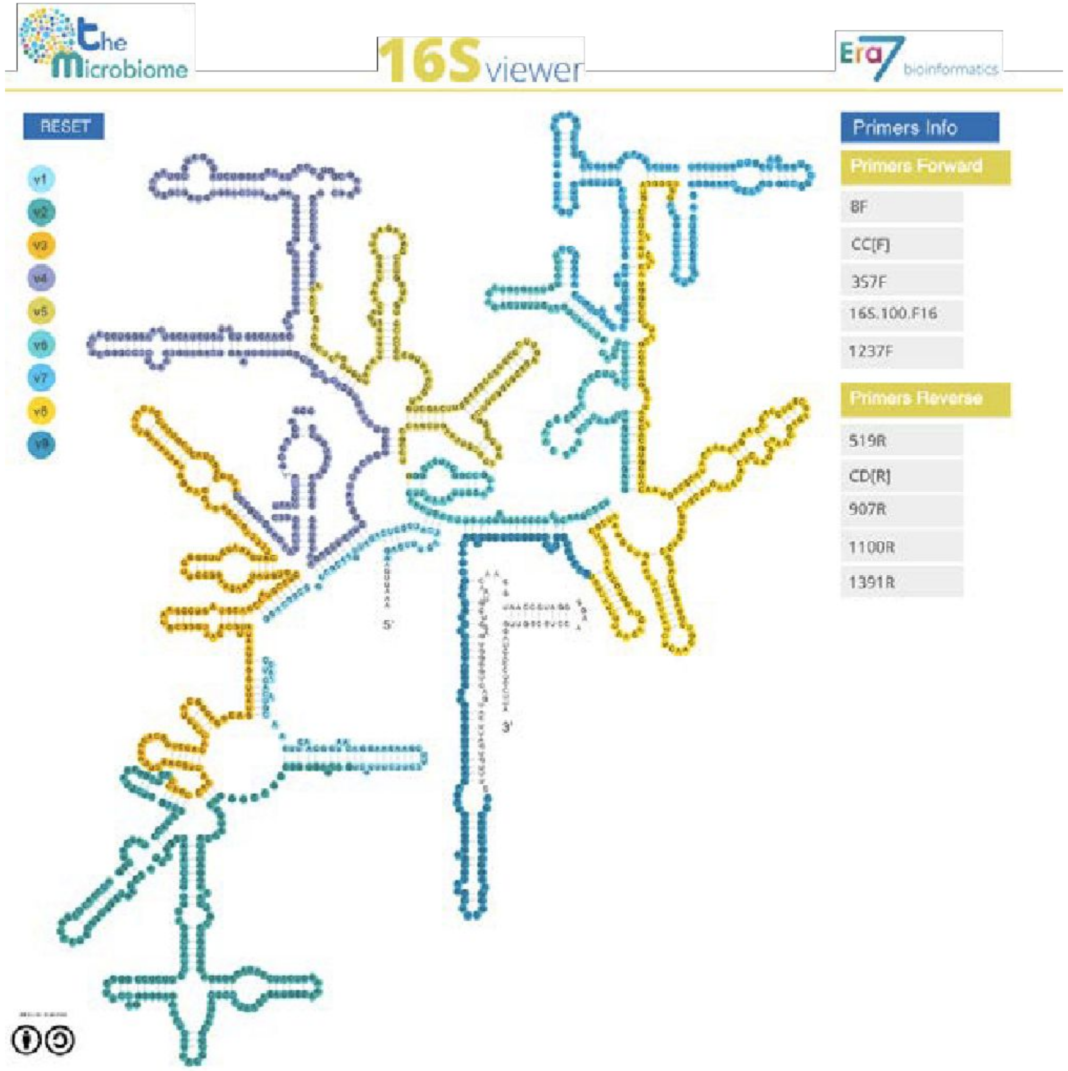
Why study the microbiome?

- Microbes modulate and maintain the atmosphere
- Critical elemental cycles (carbon, nitrogen, sulfur, iron,…)
- Bioredmediation
- Microbes keep us healthy
- Protection from pathogens
- Absorption/production of nutrients in the gut
- Role in chronic diseases (obesity, Crohn’s/IBD, arthritis…)
- Microbes support plant growth and suppress plant disease
- Crop productivity/protection/stress

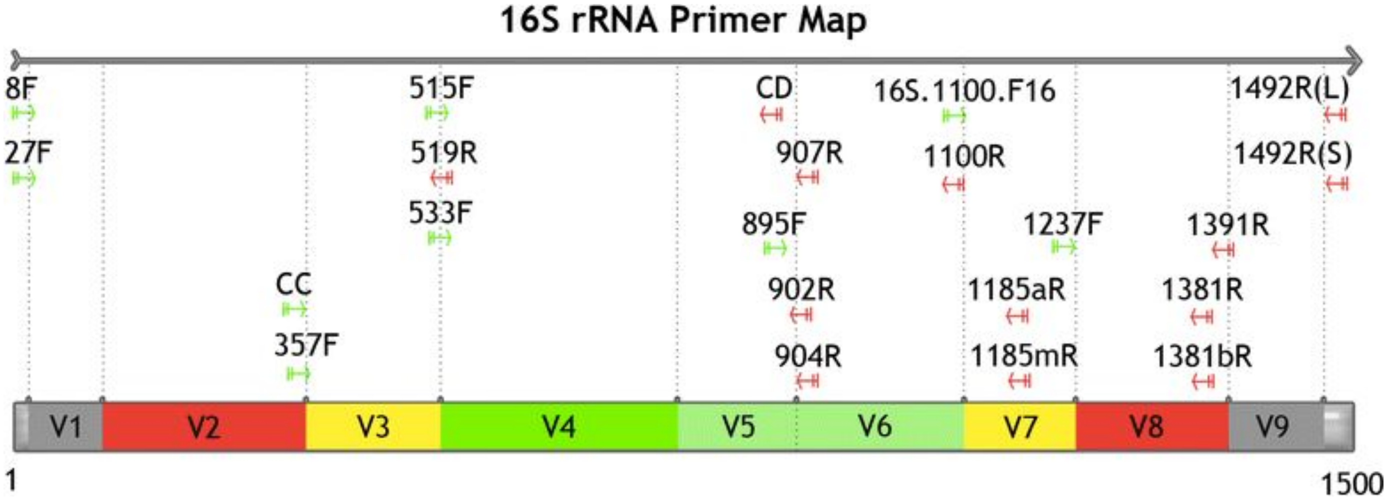
Why is microbiome research new?

- Bias for microbes (especially pathogens) that are cultivable
  - Culture-based methods do not detect majority of microbes
  - Only pathogens are easily detected
  - And most microbes are not pathogens
- Availability of tools
  - Discovery of culture independent techniques
  - Amplicon sequencing and DNA sequencing

16S rRNA hypervariable regions



Microbiome.com



BMC Bioinf, 2016

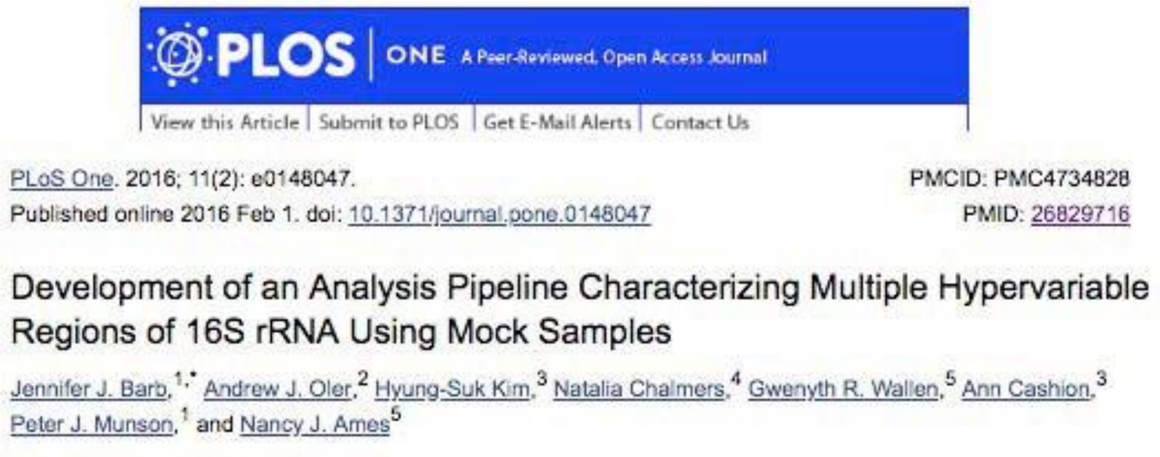
Illustration of different hypervariable regions of 16S rRNA

Choice of variable segment

- V2, V3 and V6 contain maximum nucleotide heterogeneity
- V6 is the shortest hypervariable region with the maximum sequence heterogeneity
- V1 is best target for distinguishing pathogenic S aureus
- V2 and V3 are excellent targets for speciation among Staph and Strep pathogens as well as Clostridium and Neisseria species
- V2 especially useful for speciation of Mycobacterium sp. and detection of E coli O157:H7
- V3 useful for speciation of Haemophilus sp
- V6 best target for probe based PCR assays to identify CDC select agents

DOI: 10.1016/j.mimet.2007.02.005

Selection of primers and region of 16S gene influence microbial profile



PLOS | ONE A Peer-Reviewed, Open Access Journal
View this Article | Submit to PLOS | Get E-Mail Alerts | Contact Us

PLoS One. 2016; 11(2): e0148047.
Published online 2016 Feb 1. doi: 10.1371/journal.pone.0148047

PMCID: PMC4734828
PMID: 26829716

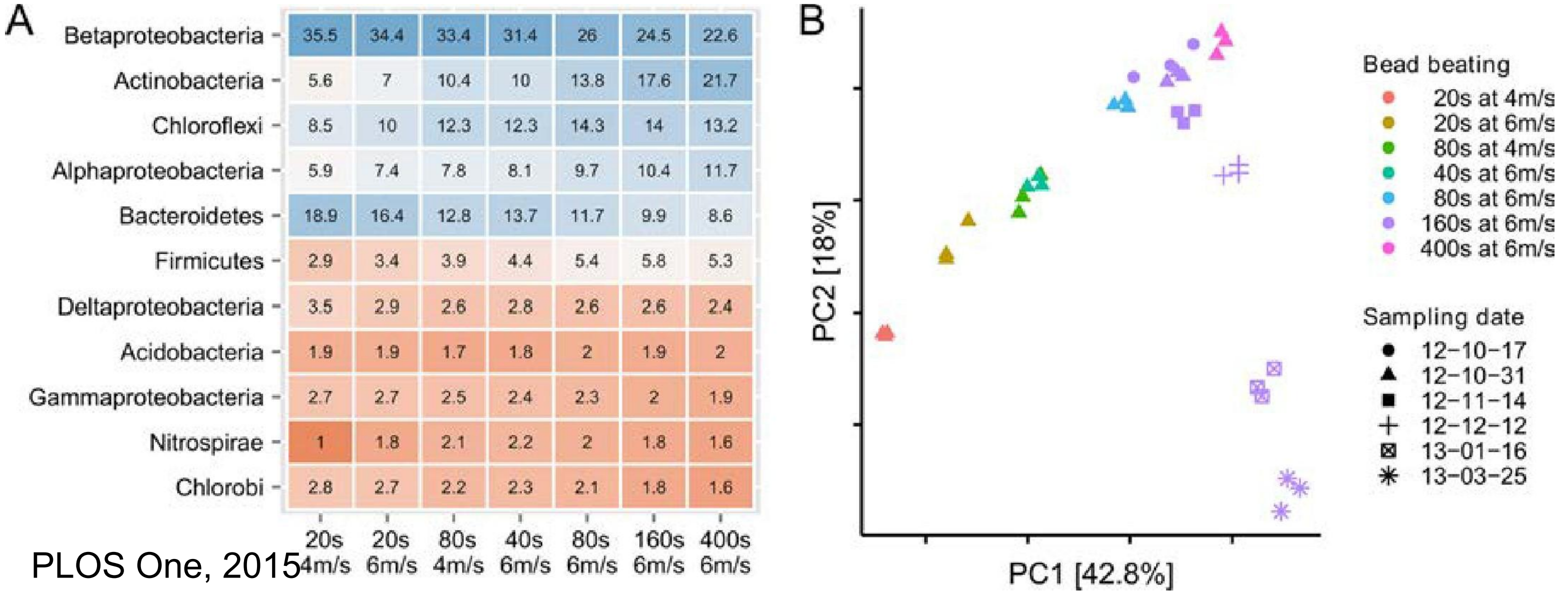
Development of an Analysis Pipeline Characterizing Multiple Hypervariable Regions of 16S rRNA Using Mock Samples

Jennifer J. Barb,[1,*] Andrew J. Oler,[2] Hyung-Suk Kim,[3] Natalia Chalmers,[4] Gwenyth R. Wallen,[5] Ann Cashion,[3] Peter J. Munson,[1] and Nancy J. Ames[5]

V2, V4, V6-V7 regions produced consistent results

DNA extraction protocol

- Effect of mechanical lysis methods for extraction
- Presence of inhibitors such as organic matter, humic acid, bile salts, polysaccharides
- DNA yield post extraction and reproducibility



A

	20s 4m/s	20s 6m/s	80s 4m/s	40s 6m/s	80s 6m/s	160s 6m/s	400s 6m/s
Betaproteobacteria	35.5	34.4	33.4	31.4	26	24.5	22.6
Actinobacteria	5.6	7	10.4	10	13.8	17.6	21.7
Chloroflexi	8.5	10	12.3	12.3	14.3	14	13.2
Alphaproteobacteria	5.9	7.4	7.8	8.1	9.7	10.4	11.7
Bacteroidetes	18.9	16.4	12.8	13.7	11.7	9.9	8.6
Firmicutes	2.9	3.4	3.9	4.4	5.4	5.8	5.3
Deltaproteobacteria	3.5	2.9	2.6	2.8	2.6	2.6	2.4
Acidobacteria	1.9	1.9	1.7	1.8	2	1.9	2
Gammaproteobacteria	2.7	2.7	2.5	2.4	2.3	2	1.9
Nitrospirae	1	1.8	2.1	2.2	2	1.8	1.6
Chlorobi	2.8	2.7	2.2	2.3	2.1	1.8	1.6

PLOS One, 2015

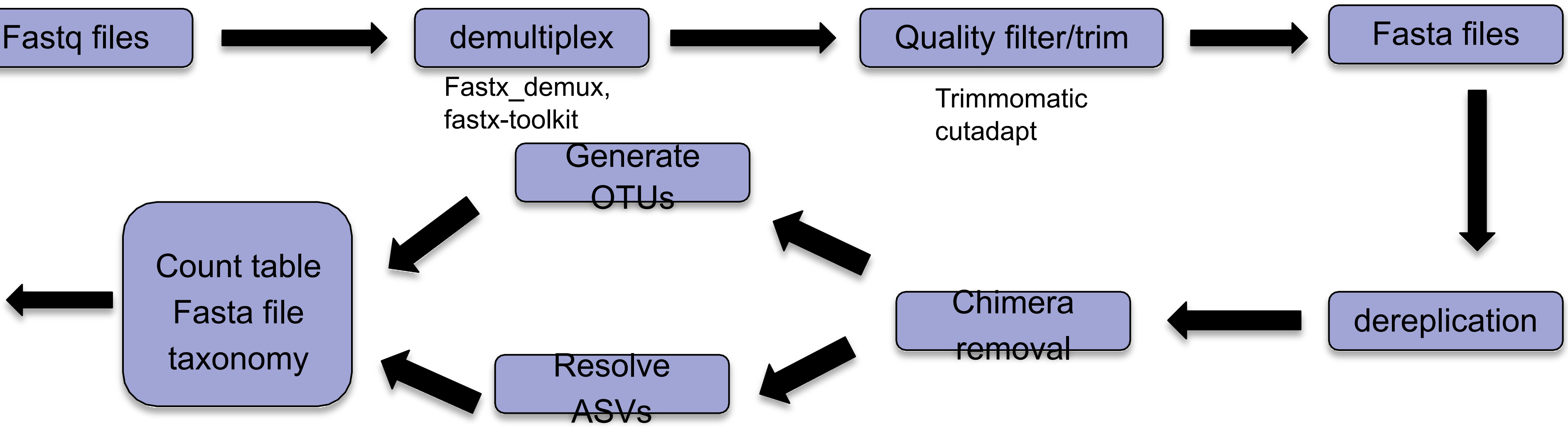
Effect of bead beating was larger than sampling time over 5 months

A. Percentage read abundance of the 11 most abundant phyla as a result of bead beating intensity

B. PCA of samples with different bead beating intensities vs. samples taken at different dates

Overview of generic amplicon workflow

Sequencing facility



Fastq files

demultiplex

Fastx_demux,
fastx-toolkit

Quality filter/trim

Trimmomatic
cutadapt

Fasta files

dereplication

Chimera removal

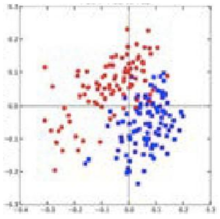
Generate OTUs

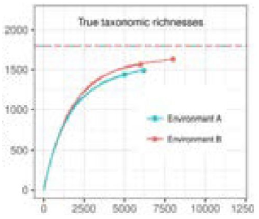
Resolve ASVs

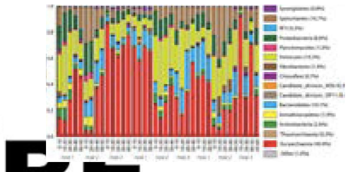
Count table
Fasta file
taxonomy

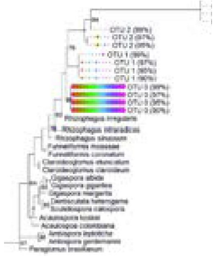
	Sample1	Sample B	…
Seq_1	0	400	…
Seq_2	305	350	…
Seq_3	200	1	…
…	…	…	…

Analysis









Alpha diversity
Beta diveristy
Taxonomic summaries
Phylogeny

For this workshop

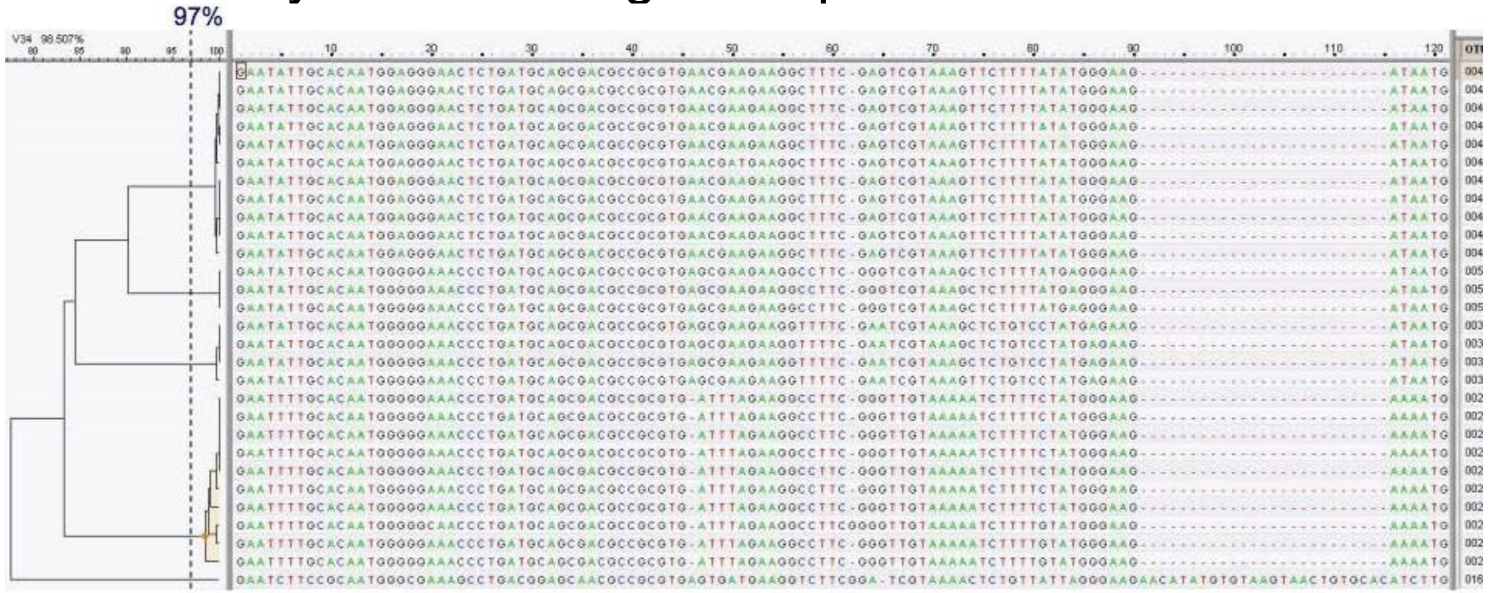
QIIME2

New Mexico INBRE
IDeA Networks of Biomedical Research Excellence

NCGR
National Center for Genome Resources

Clustering

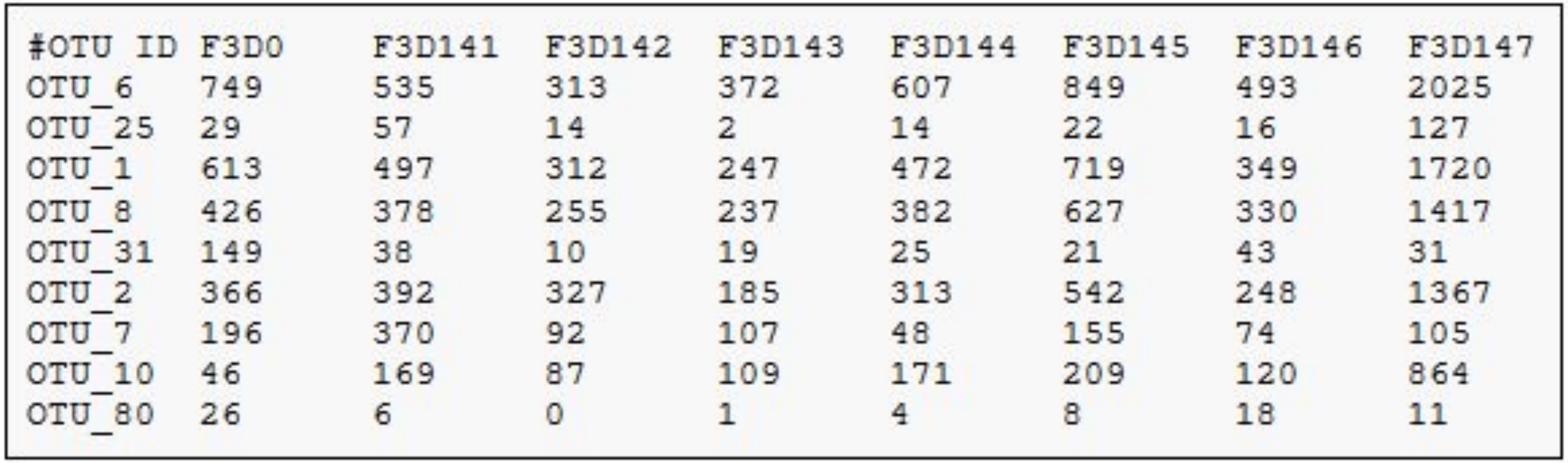
- Analysis of 16S rRNA relies on clustering of related sequences at a particular level of identity and counting the representatives of each cluster



- Some level of sequence divergence should be allowed – 95% (genus-level, partial 16S gene), 97% (species-level) or 99% typical similarity cutoffs used in practice and the resulting cluster of nearly identical tags (assumedly identical genomes) is referred to as an OTU (Operational Taxonomic Unit)

Create OTU tables

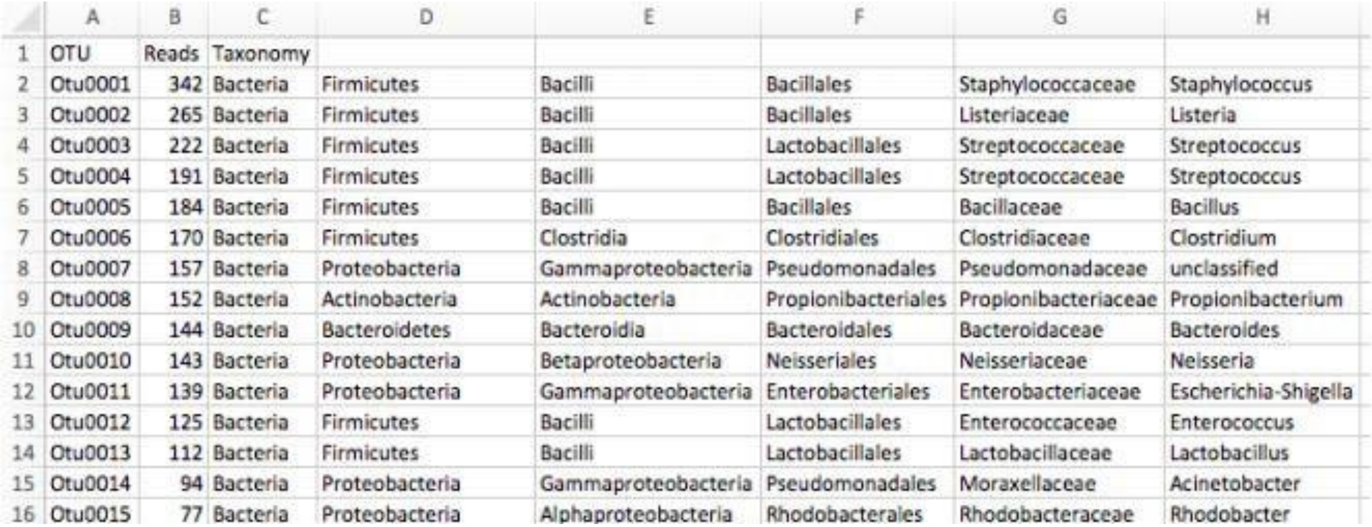
- OTU table is a matrix that gives the number of reads per sample per OTU



#OTU ID	F3D0	F3D141	F3D142	F3D143	F3D144	F3D145	F3D146	F3D147
OTU_6	749	535	313	372	607	849	493	2025
OTU_25	29	57	14	2	14	22	16	127
OTU_1	613	497	312	247	472	719	349	1720
OTU_8	426	378	255	237	382	627	330	1417
OTU_31	149	38	10	19	25	21	43	31
OTU_2	366	392	327	185	313	542	248	1367
OTU_7	196	370	92	107	48	155	74	105
OTU_10	46	169	87	109	171	209	120	864
OTU_80	26	6	0	1	4	8	18	11

Bin OTUs into Taxonomy (assign taxonomy)

- Accuracy of assigning taxonomy depends on the reference database chosen
  - Ribosomal Database Project
  - GreenGenes
  - SILVA
- Accuracy depends on the completeness of databases



	A	B	C	D	E	F	G	H
1	OTU	Reads	Taxonomy					
2	Otu0001	342	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus
3	Otu0002	265	Bacteria	Firmicutes	Bacilli	Bacillales	Listeriaceae	Listeria
4	Otu0003	222	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus
5	Otu0004	191	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus
6	Otu0005	184	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus
7	Otu0006	170	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium
8	Otu0007	157	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	unclassified
9	Otu0008	152	Bacteria	Actinobacteria	Actinobacteria	Propionibacteriales	Propionibacteriaceae	Propionibacterium
10	Otu0009	144	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
11	Otu0010	143	Bacteria	Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae	Neisseria
12	Otu0011	139	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Escherichia-Shigella
13	Otu0012	125	Bacteria	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Enterococcus
14	Otu0013	112	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus
15	Otu0014	94	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter
16	Otu0015	77	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Rhodobacter

So what is an OTU, anyway?

The Operational Taxonomic Unit

- PCR the 16s rRNA region
- Sequence the product
- Cluster the reads at 97% (or other) identity
- Each cluster is an OTU
- Count the reads in each OTU
- Those are your abundances

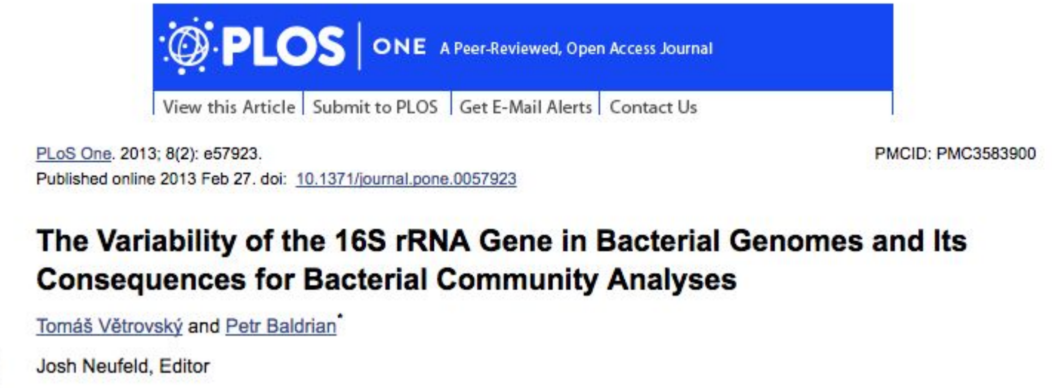
BUT…

- You want to know abundances of actual taxa
- Qiime session will show you how

Problems

- Samples bacteria (and some archaea) only
- Primers may not be universal
- Databases are not complete
- Limited resolution
- The OTU problem
- The copy number problem

The copy number problem


PLOS | ONE A Peer-Reviewed, Open Access Journal
View this Article | Submit to PLOS | Get E-Mail Alerts | Contact Us

PLoS One. 2013; 8(2): e57923.
Published online 2013 Feb 27. doi: 10.1371/journal.pone.0057923

PMCID: PMC3583900

The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses

Tomáš Větrovský and Petr Baldrian*

Josh Neufeld, Editor

- Sampled 1,690 bacterial genomes
- 1-16 16s rRNA gene copies per genome
- Sequences can differ within a genome
- Many species have identical copies

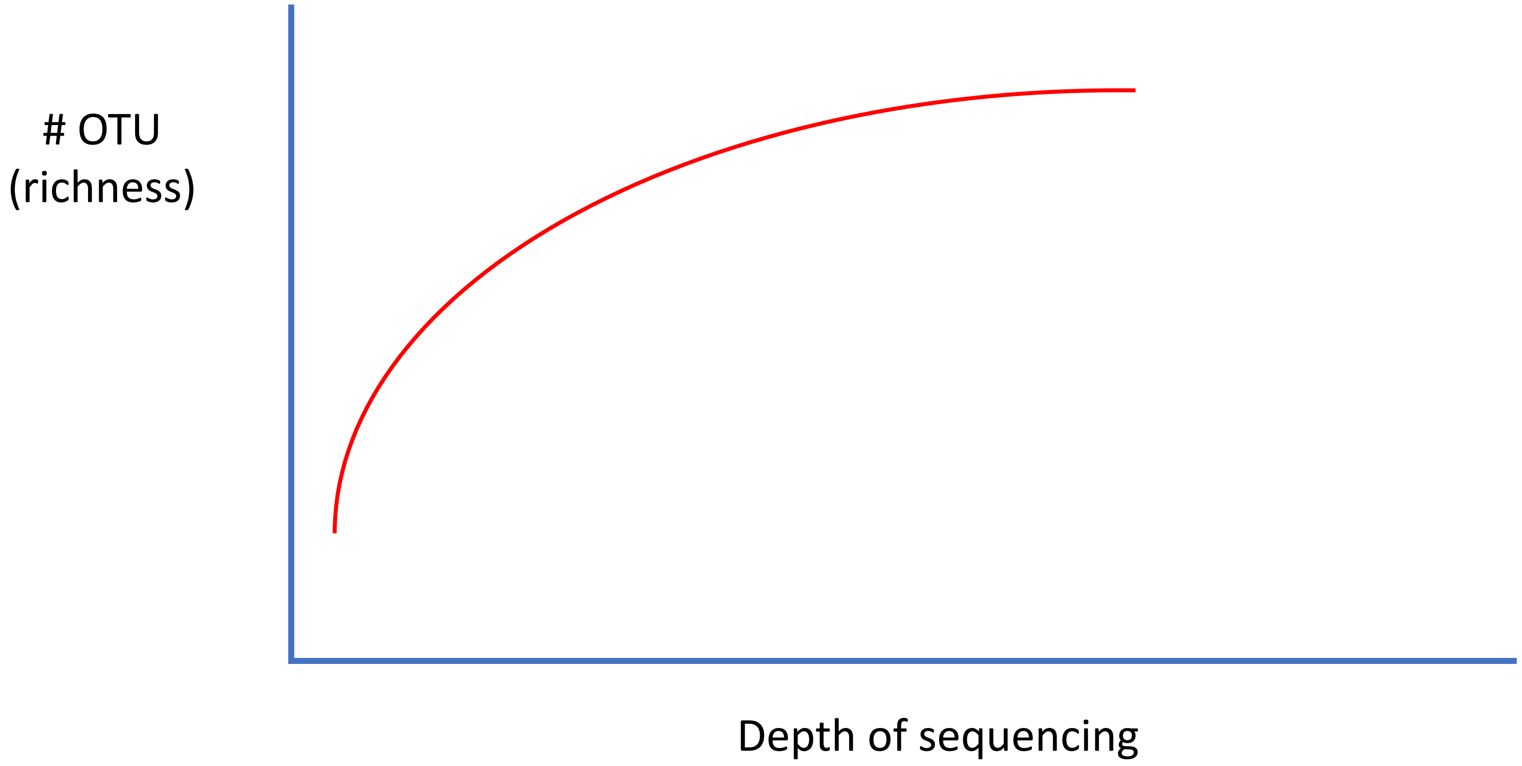
This means

- A species may have >1 OTU
- Many species may belong to the same OTU

Some terminology

- Alpha diversity
  - Number of OTUs in a sample
  - Their relative abundance

How to count the uncountable?



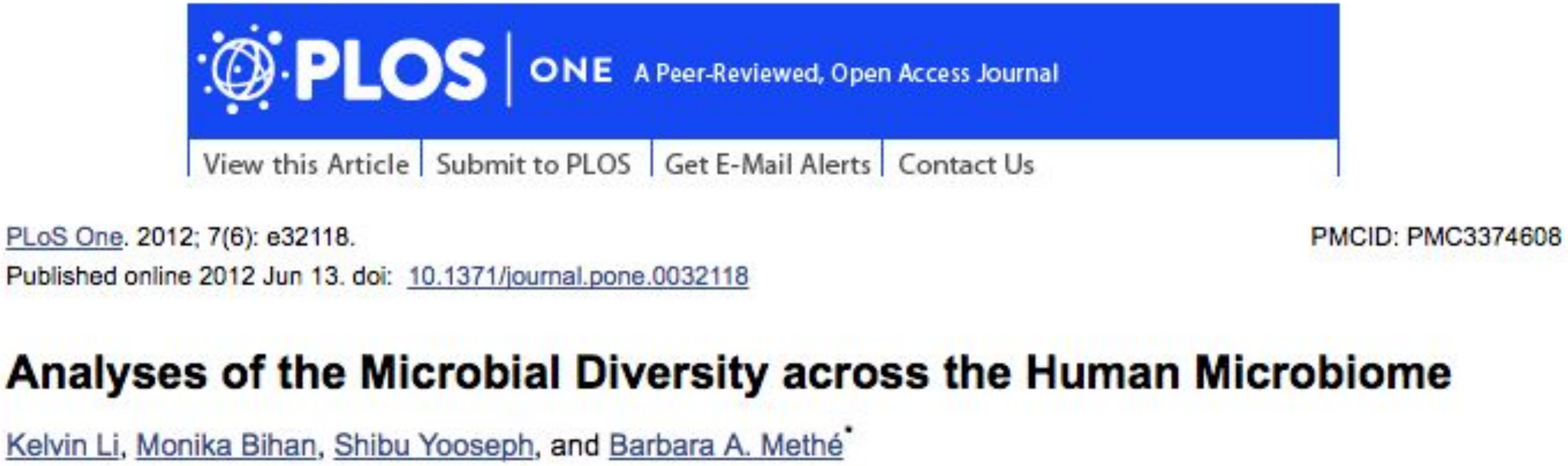
DOI: 10.1128/aem.67.10.4399-4406.2001

Richness estimates

- Chao1 index:

$$S_{EST} = S_{OBS} + \frac{N_1^2}{2N_2}$$

Richness estimates



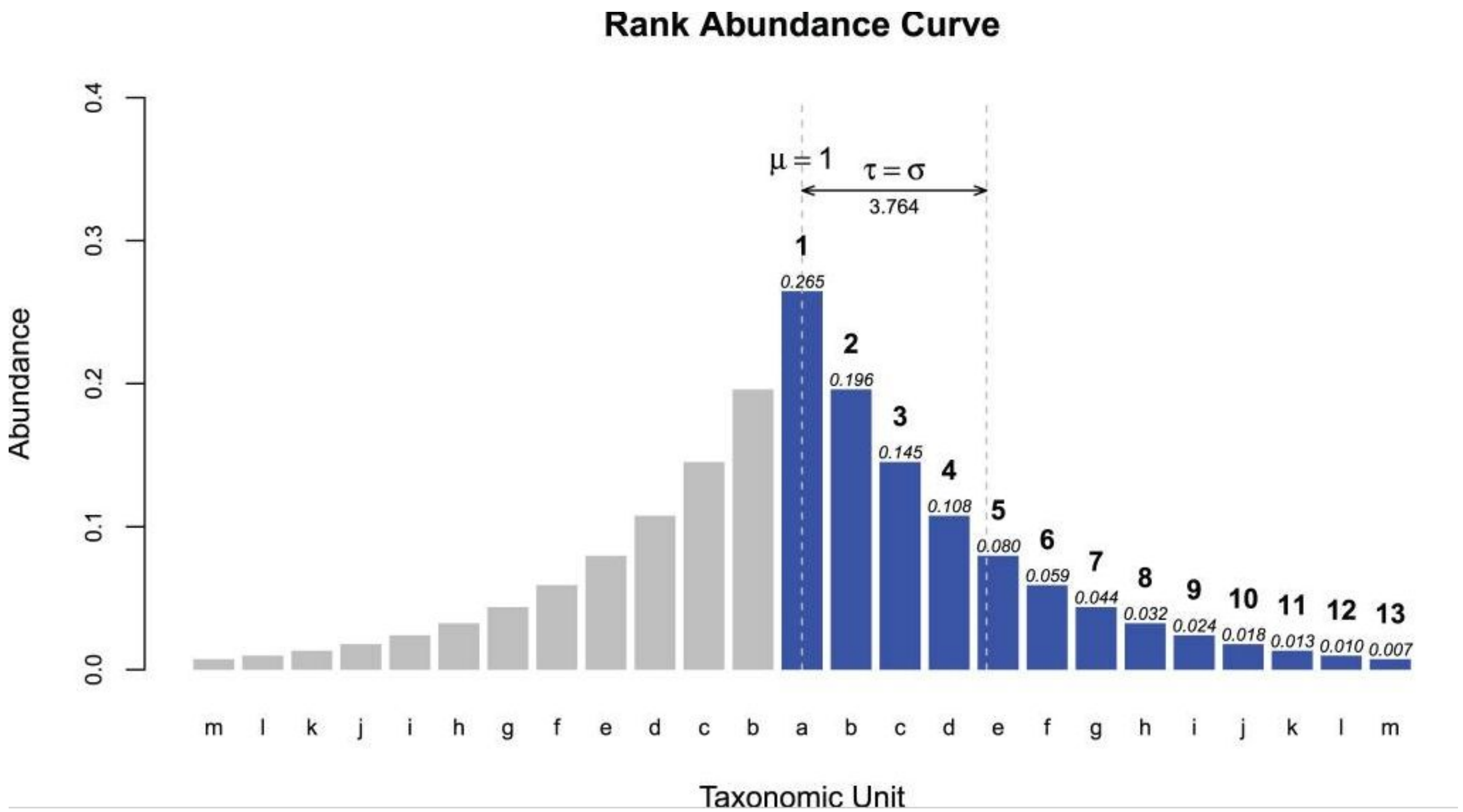
PLoS One. 2012; 7(6): e32118. PMCID: PMC3374608

Published online 2012 Jun 13. doi: 10.1371/journal.pone.0032118

Analyses of the Microbial Diversity across the Human Microbiome

Kelvin Li, Monika Bihan, Shibu Yooseph, and Barbara A. Methé*

- Tail statistic



doi: 10.1371/journal.pone.0032118

Richness versus Evenness

- Pop. 1: 20 ants and 1 centipede
- Pop. 2: 10 ants and 10 centipedes
- Both have 20 organisms
- Both have 2 species
- How to represent the difference?
- Need something that scales with complexity.

Shannon Diversity Index

$$D = -\sum_{i=1}^{s} p_i \ln p_i$$

- Where *p* is the proportion of species *i* in the community
- More even = greater value
- Quantifies uncertainty in predicting identity of a species chosen at random

Simpson Index

$$D = \sum_{i=1}^{s} p_i^2$$

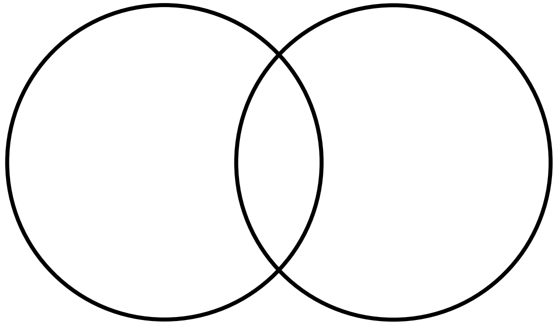
- Where *p* is the proportion of species *i* in the community
- Less even = greater value
- Probability of two random members of the population being the same type

Beta-diversity

- Beta-diversity measures community structure differences (taxon composition and relative abundance) between two or more samples
- For example, beta-diversity indices can compare similarities and differences in microbial communities in healthy and diseases states
- Many qualitative (presence/absence taxa) and quantitative(taxon abundance) measures of community distance are available using several tools
- LIBHUFF, TreeClimber, DPCoA, UniFrac (QIIME)

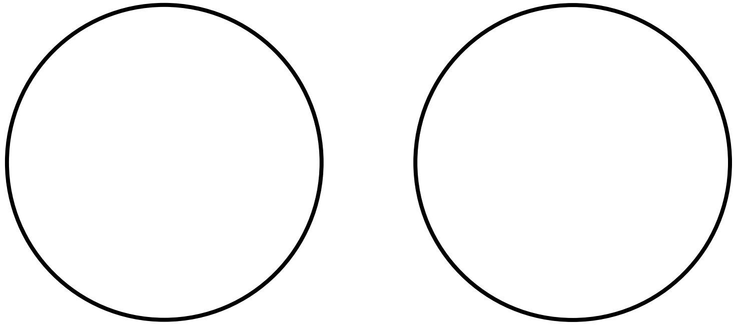
Jaccard Index

Two sets of OTU A,B



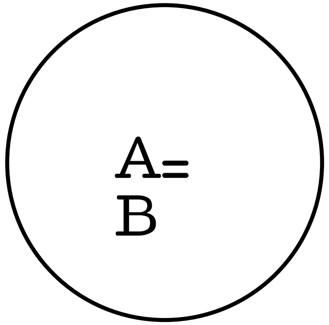
$$\frac{A \cap B}{A \cup B}$$

Jaccard Index



$$\frac{A \cap B}{A \cup B} = 0$$

Jaccard Index

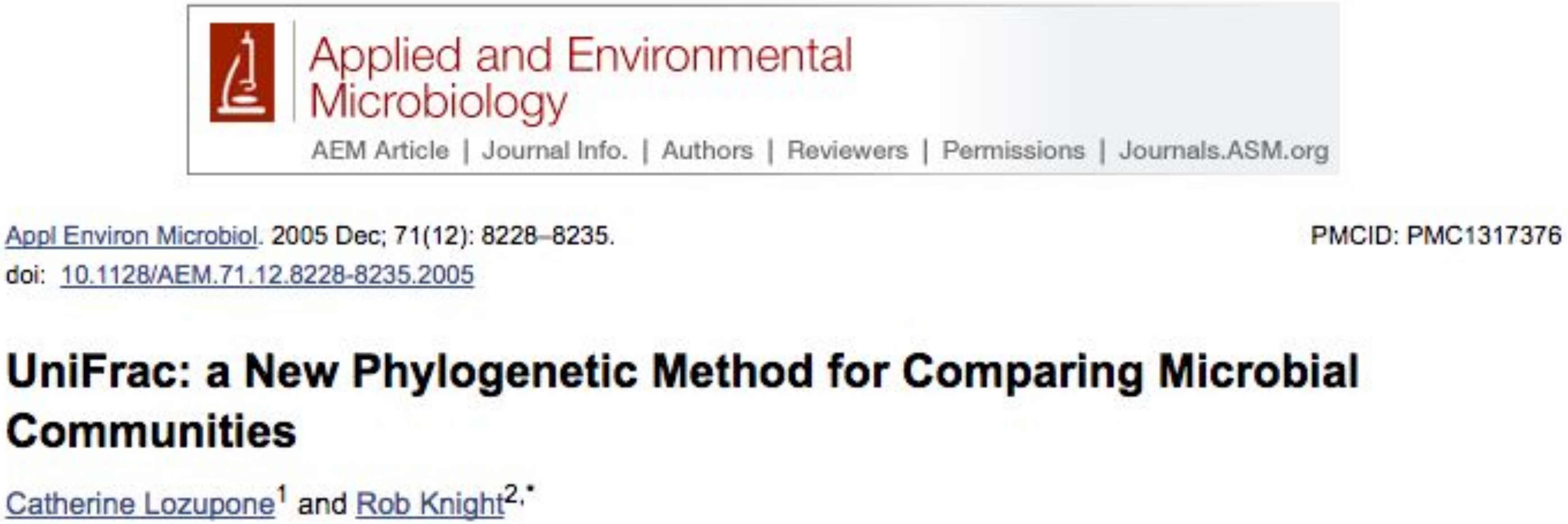


$$\frac{A \cap B}{A \cup B} = 1$$

Jaccard Index

- Problem: Relatedness/phylogeny is ignored
- As long as the union and intersection OTU numbers are the same, two highly related communities will have the same index as two distantly related communities

Unifrac



Applied and Environmental Microbiology

AEM Article | Journal Info. | Authors | Reviewers | Permissions | Journals.ASM.org

Appl Environ Microbiol. 2005 Dec; 71(12): 8228–8235. PMCID: PMC1317376
doi: 10.1128/AEM.71.12.8228-8235.2005

UniFrac: a New Phylogenetic Method for Comparing Microbial Communities

Catherine Lozupone[1] and Rob Knight[2,*]

- UNIFRAC measures the phylogenetic relatedness of communities

- Beta Diversity - UniFrac

  - Measures how different two samples' component sequences are

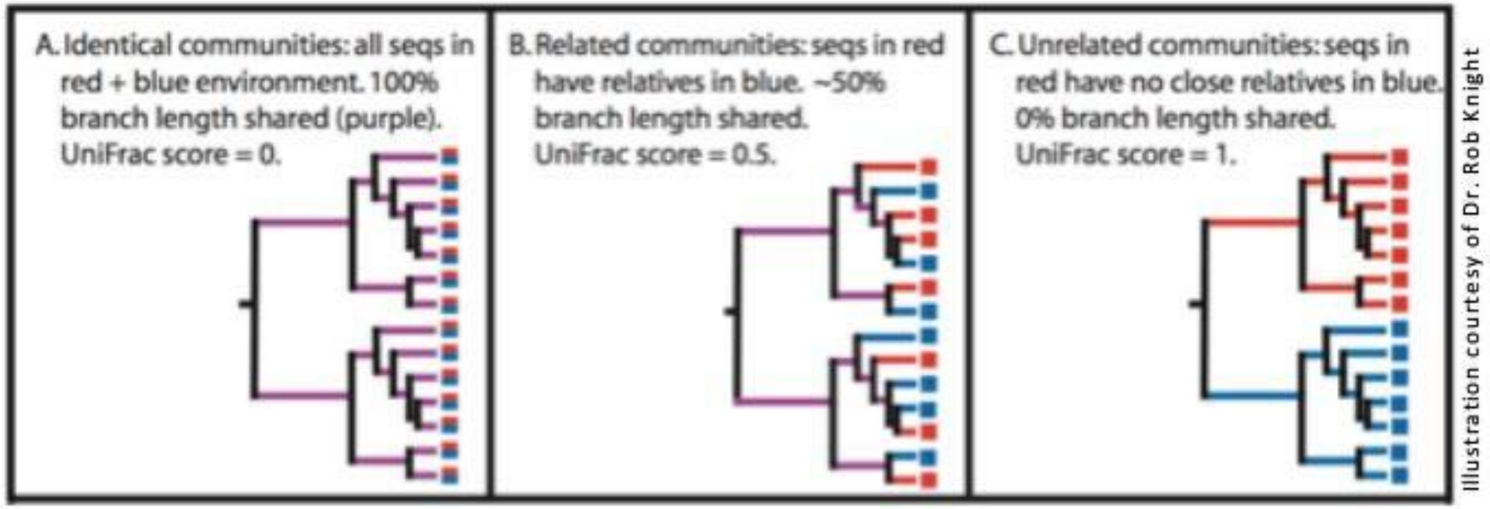
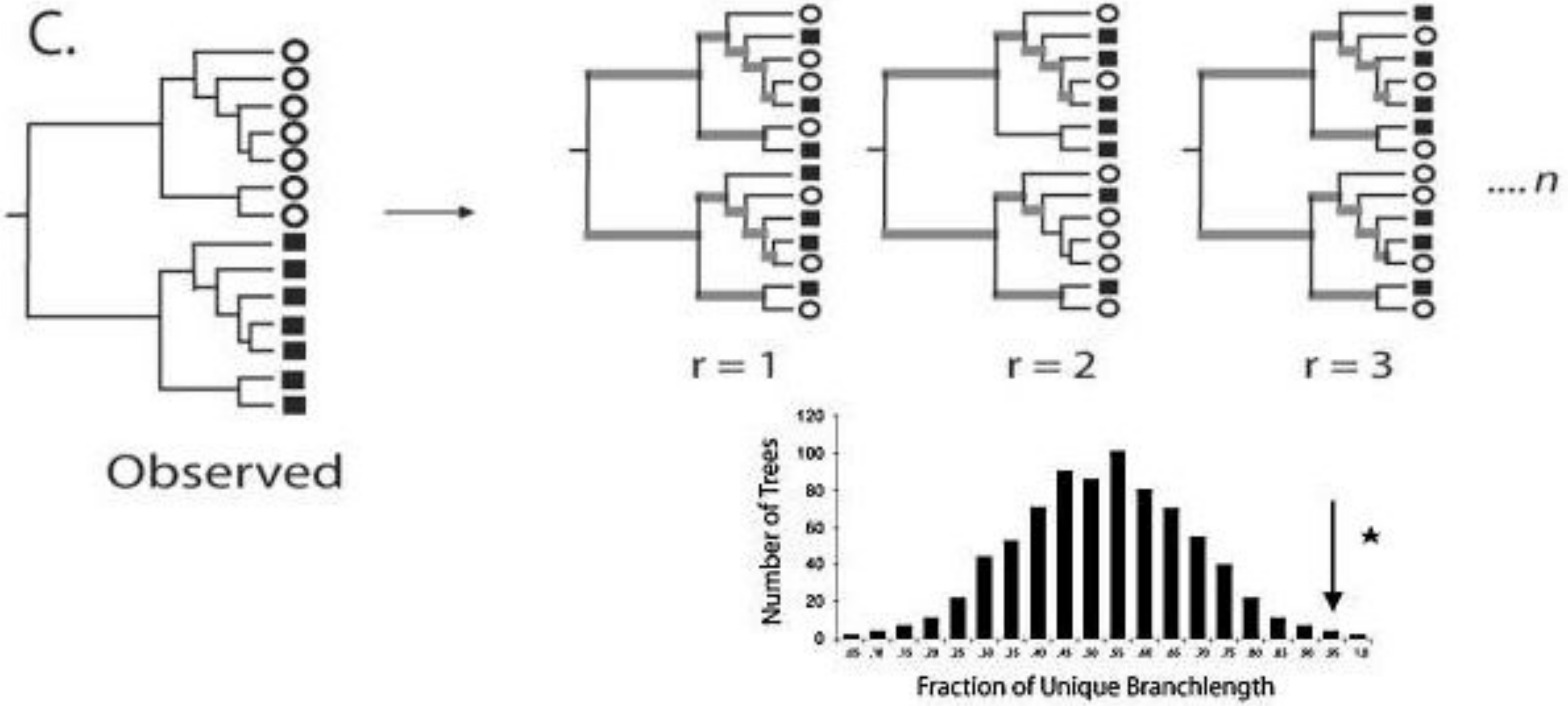

A. Identical communities: all seqs in red + blue environment. 100% branch length shared (purple). UniFrac score = 0.
B. Related communities: seqs in red have relatives in blue. ~50% branch length shared. UniFrac score = 0.5.
C. Unrelated communities: seqs in red have no close relatives in blue. 0% branch length shared. UniFrac score = 1.

Illustration courtesy of Dr. Rob Knight

  - Weighted Unifrac: takes abundance of each sequence into account

Computing significance



The bioinformatics workshop is supported by New Mexico INBRE through an Institutional Development Award (IDeA) from National Institute of General Medical Sciences of the NIH grant number P20GM103451

New Mexico INBRE
IDeA Networks of Biomedical Research Excellence

NCGR
National Center for Genome Resources