An introduction to molecular approaches in microbial ecology (a.k.a. EcoGenomics)

Loïs Maignien
Associate professor
University of Western Brittany - Brest

Lab. Microbiologie des Environnements Extrêmes – IUEM

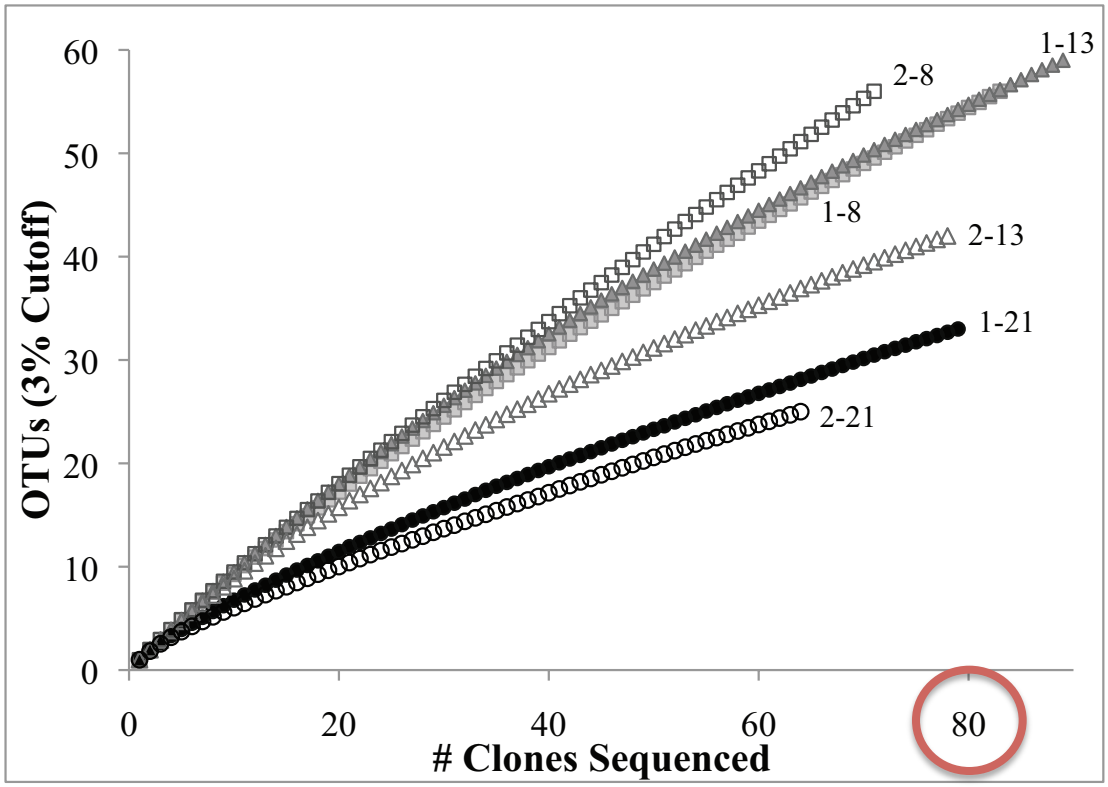
Lois_maignien@univ-brest.fr
http://pagesperso.univ-brest.fr/~maignien/

1.

Using NGS in microbial ecology

1. Using NGS in microbial ecology

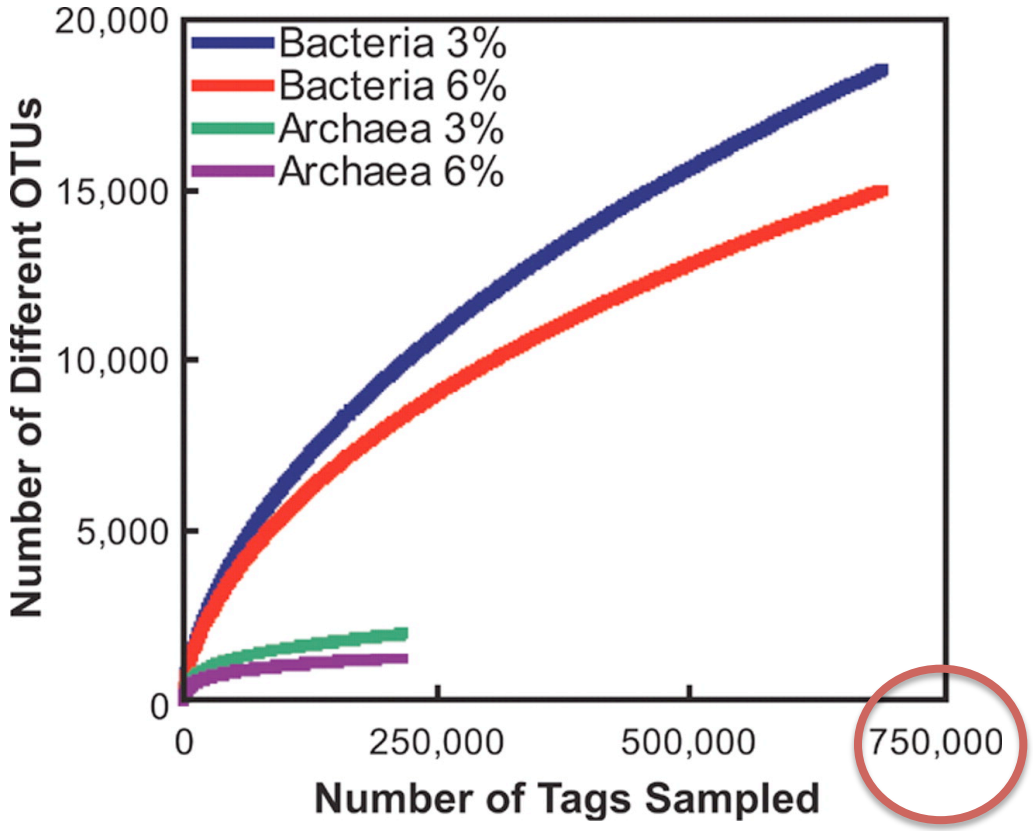
- preNGS: *clone libraries*. Sylvan *et al.*, 2012
- **« Time-series analysis of two hydrothermal plumes at 9°50'N East Pacific Rise reveals distinct, heterogeneous bacterial populations »**



1. Using NGS in microbial ecology

- NGS: 454 Huber *et al.,* Science 2007

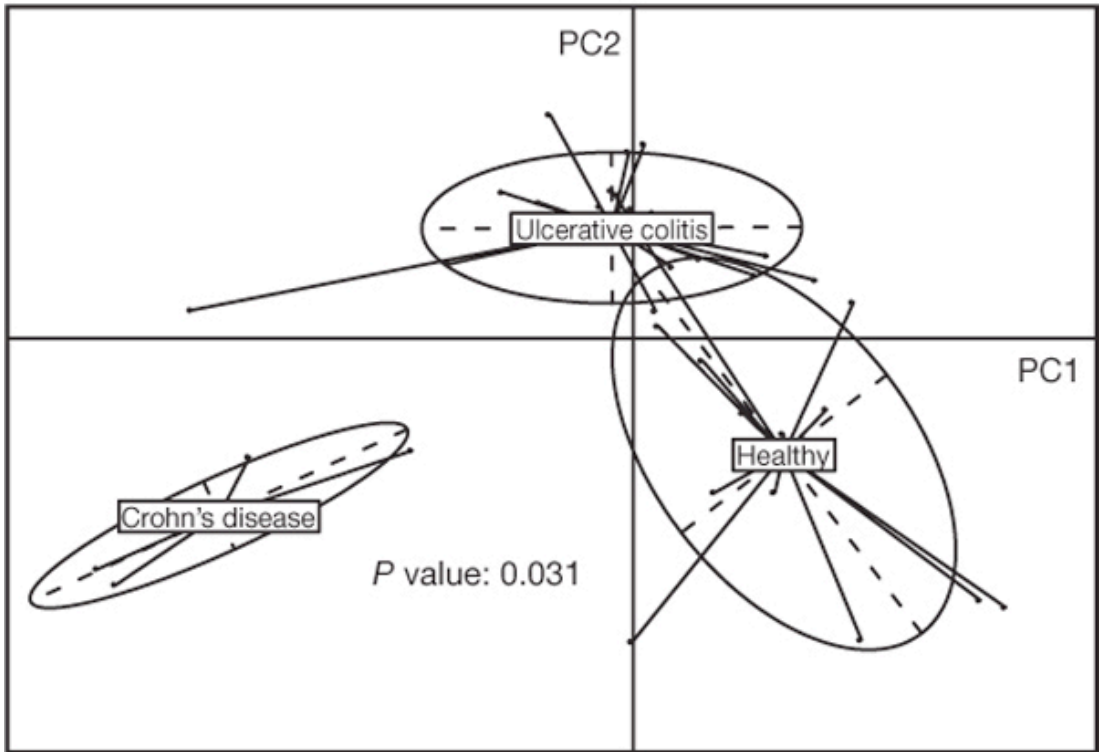
« Microbial Population Structures in the Deep Marine Biosphere »



1. Using NGS in microbial ecology

- NGS: Ilumina Quin et al. Nature 2010

« A human gut microbial gene catalogue established by metagenomic sequencing »

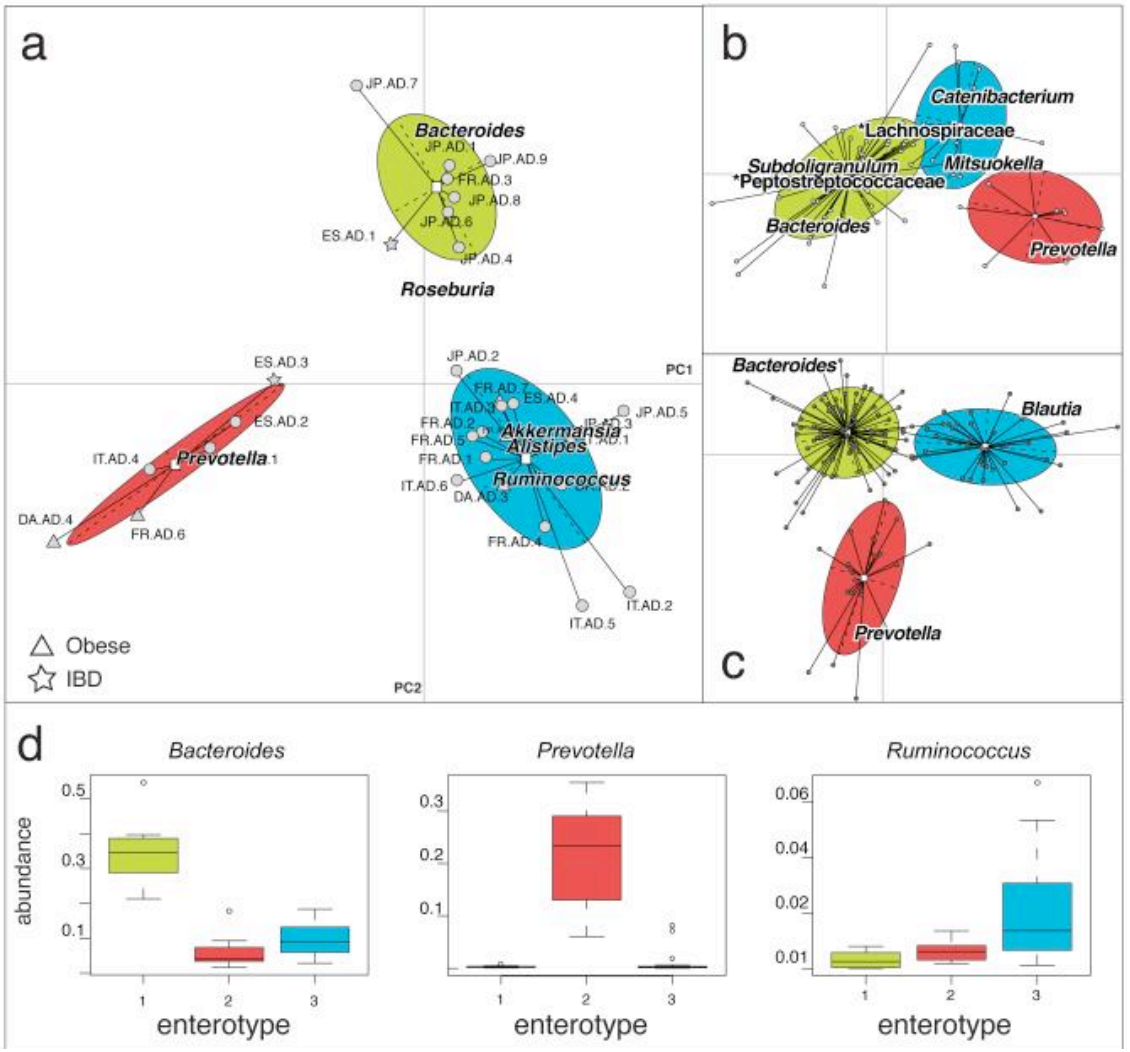


1. Using NGS in microbial ecology

- NGS: Sanger + 454

Arumugam *et al.* Nature 2011

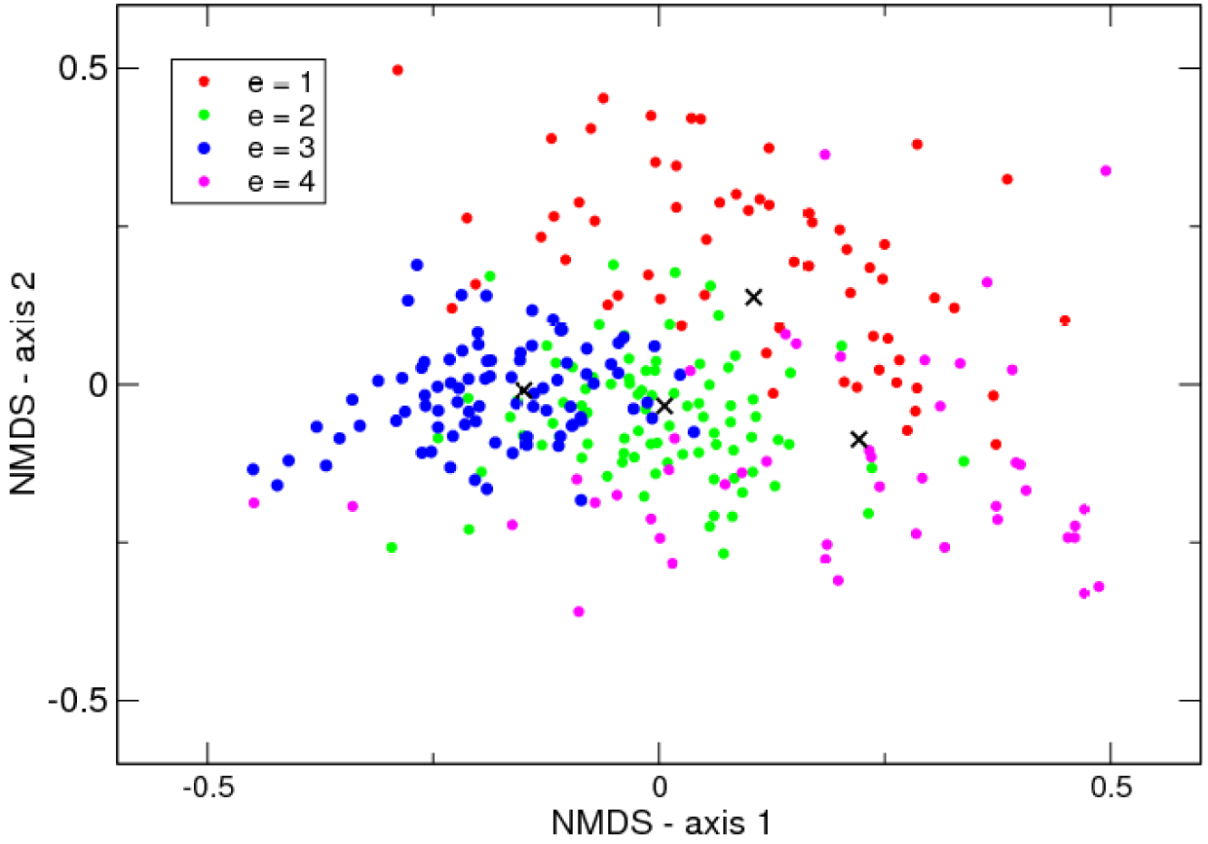
« **Enterotypes of the human gut microbiome**»



1. Using NGS in microbial ecology

- FYI: Enterotype hypothesis reanalysis with more robust statistics and a larger dataset. Enterotype is a trend but no discrete cluster

Holmes et al. plos one 2012 0.1371/journal.pone.0030126



2.

Central problem: describing microbial communities

2. Central problem: describing microbial communities

- **Contribution to ecosystem functioning through observation**

2. Central problem: describing microbial communities

- **Contribution to ecosystem functioning through observation**

M. Roberts ocean.si.edu

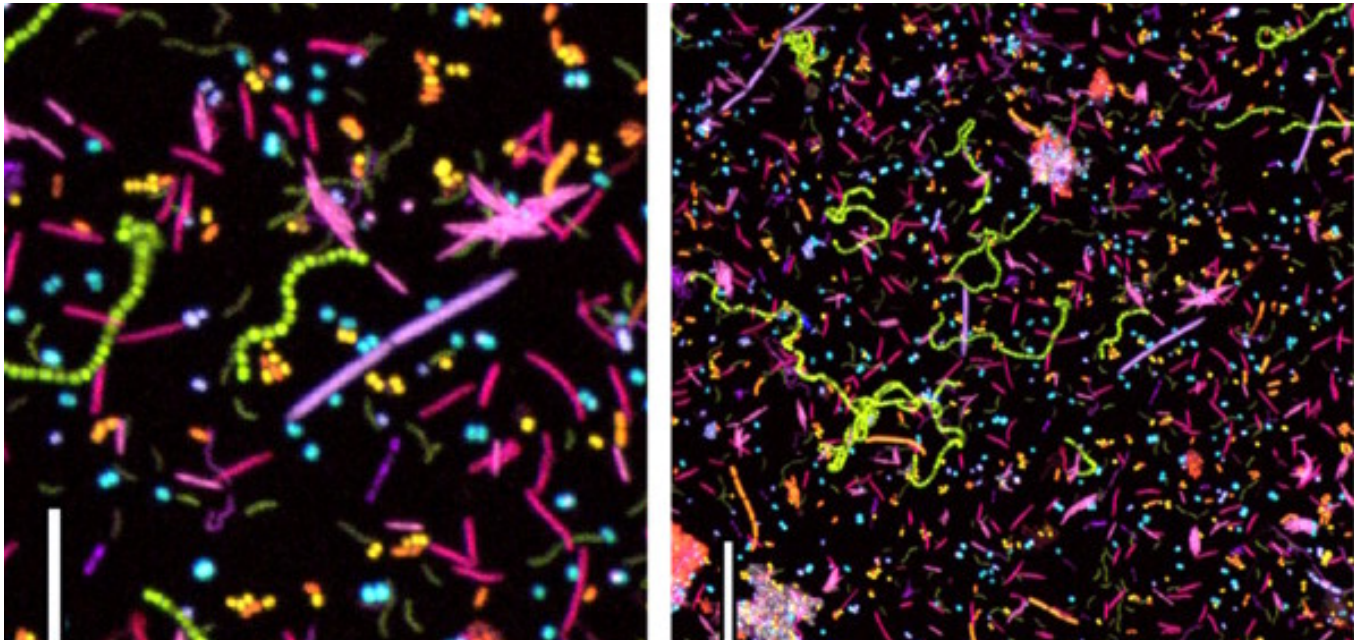
Lophelia pertusa (deep-sea coral)
Feeds on particulate OM….

S. Ross et al., UNCW, NOAA/USGS DISCOVRE Cruise

… and create 3D frames for other organisms

2. Central problem: describing microbial communities

- **Here is the problem with microbes…**
  - Almost useless phenotype
  - Very low cultivability (<1% of microorganisms)
  - Huge diversity
  - Huge number

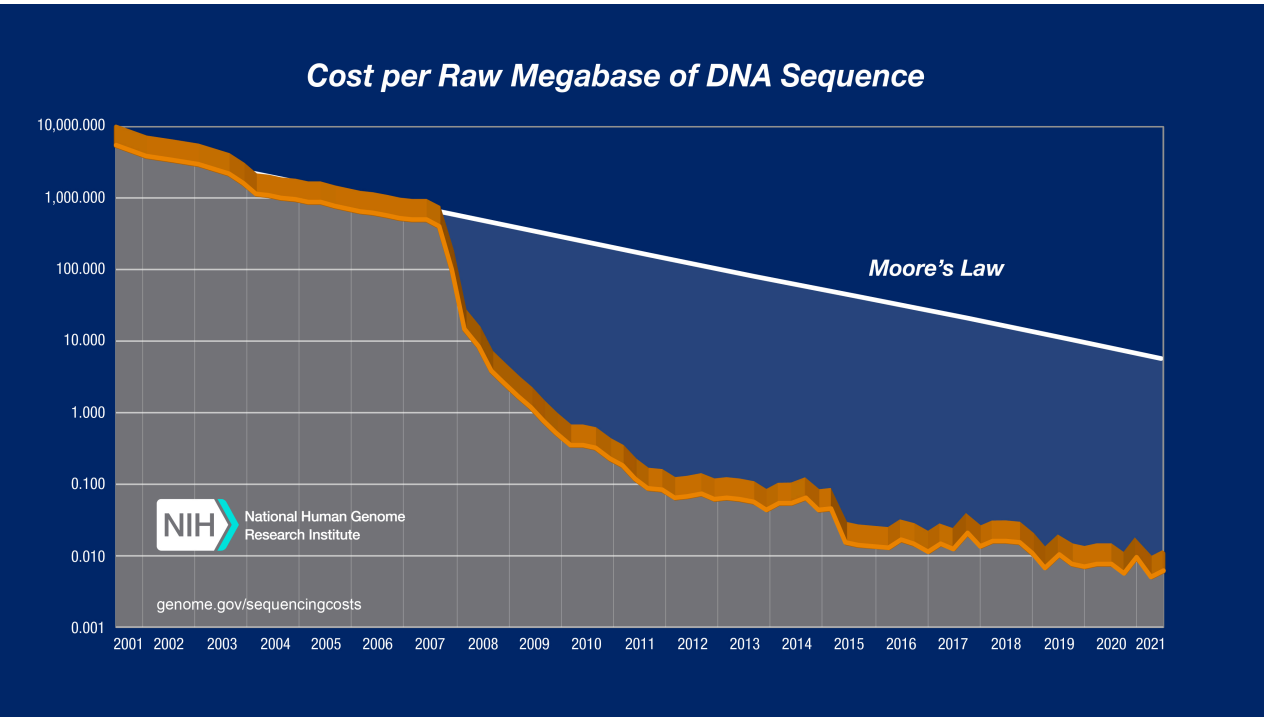


CLASI-FISH Valm et al. 2011 PNAS

2. Central problem: describing microbial communities

- **DNA libraries** can provide taxonomic diversity
- When gene function is known, DNA libraries may indicate microorganisms function in the environment
- Several approaches:
  - Sequencing of PCR products => **Amplicon**
  - Sequencing of genomic / transcript fragment from
    - an isolated strain => **Genomics / Transcriptomics**
    - an environmental sample => **Metagenomics / Metatranscriptomics**

2. Central problem: describing microbial communities



Census of the 10^5 microbes in 1 mL of seawater
2005: 50.000 euros (Sanger)
2022: 0.5 euros (Illumina)

A brand new world!

Large scale comparison of

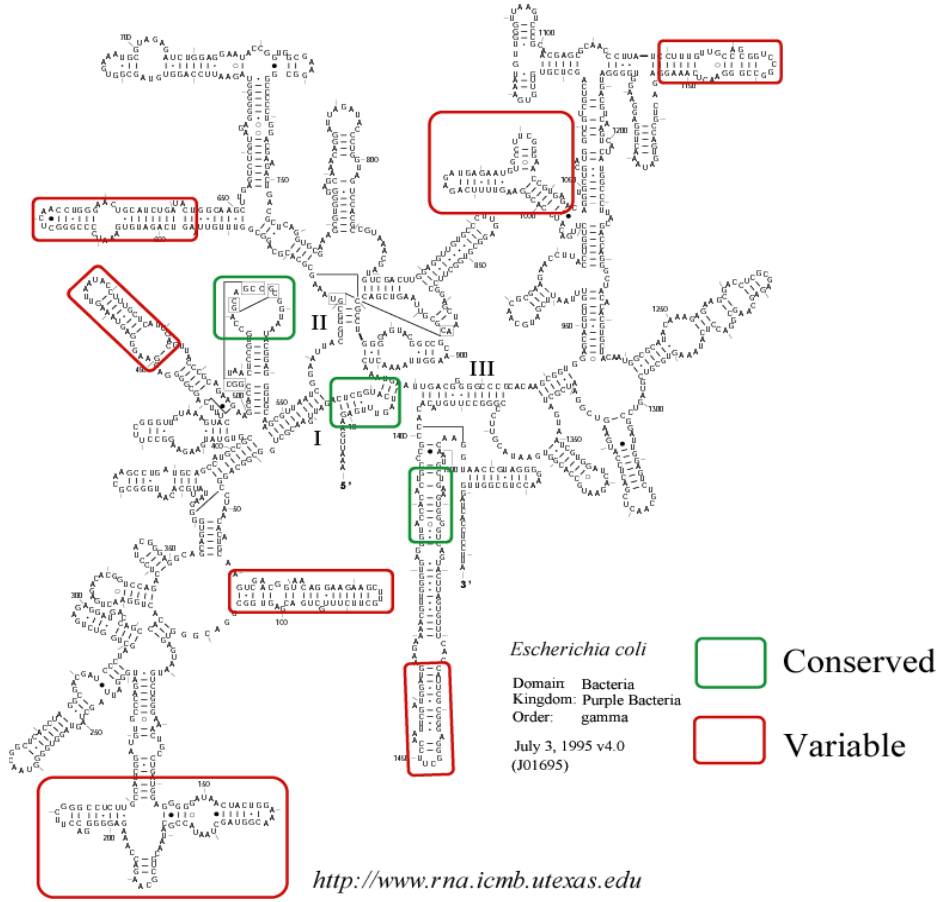
- Genes
- Genomes
- Transcriptomes
- Populations
- Communities

3.

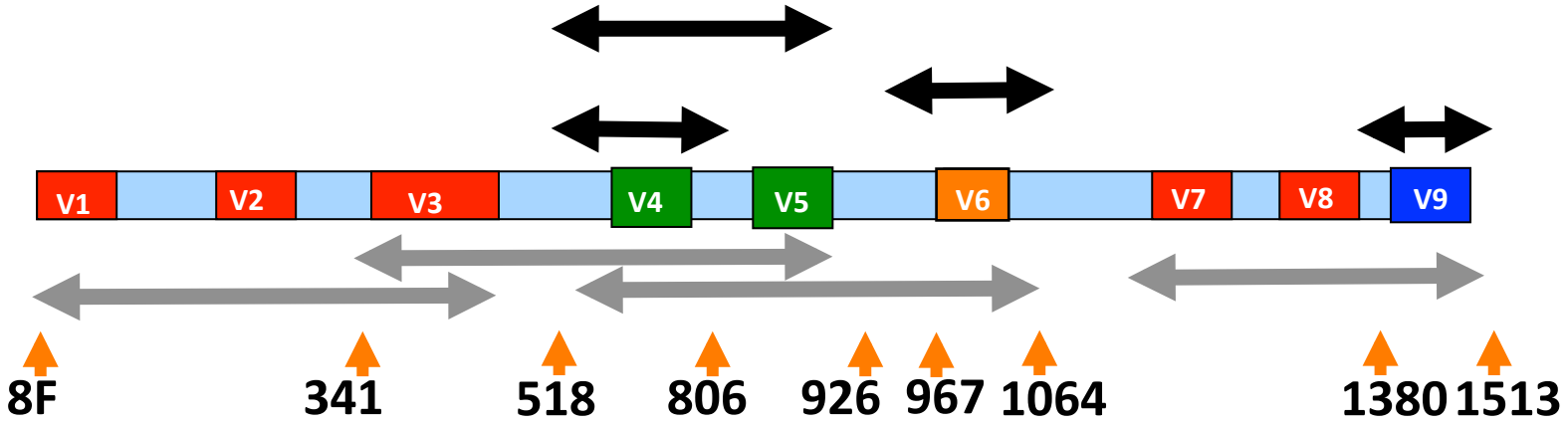
Using 16S ribosomal RNA gene as a taxonomic marker for microbes

3. Using 16S ribosomal RNA gene as a taxonomic marker for microbes

Contains both conserved and hypervariable regions



Using 16S ribosomal RNA gene as a taxonomic marker for microbes



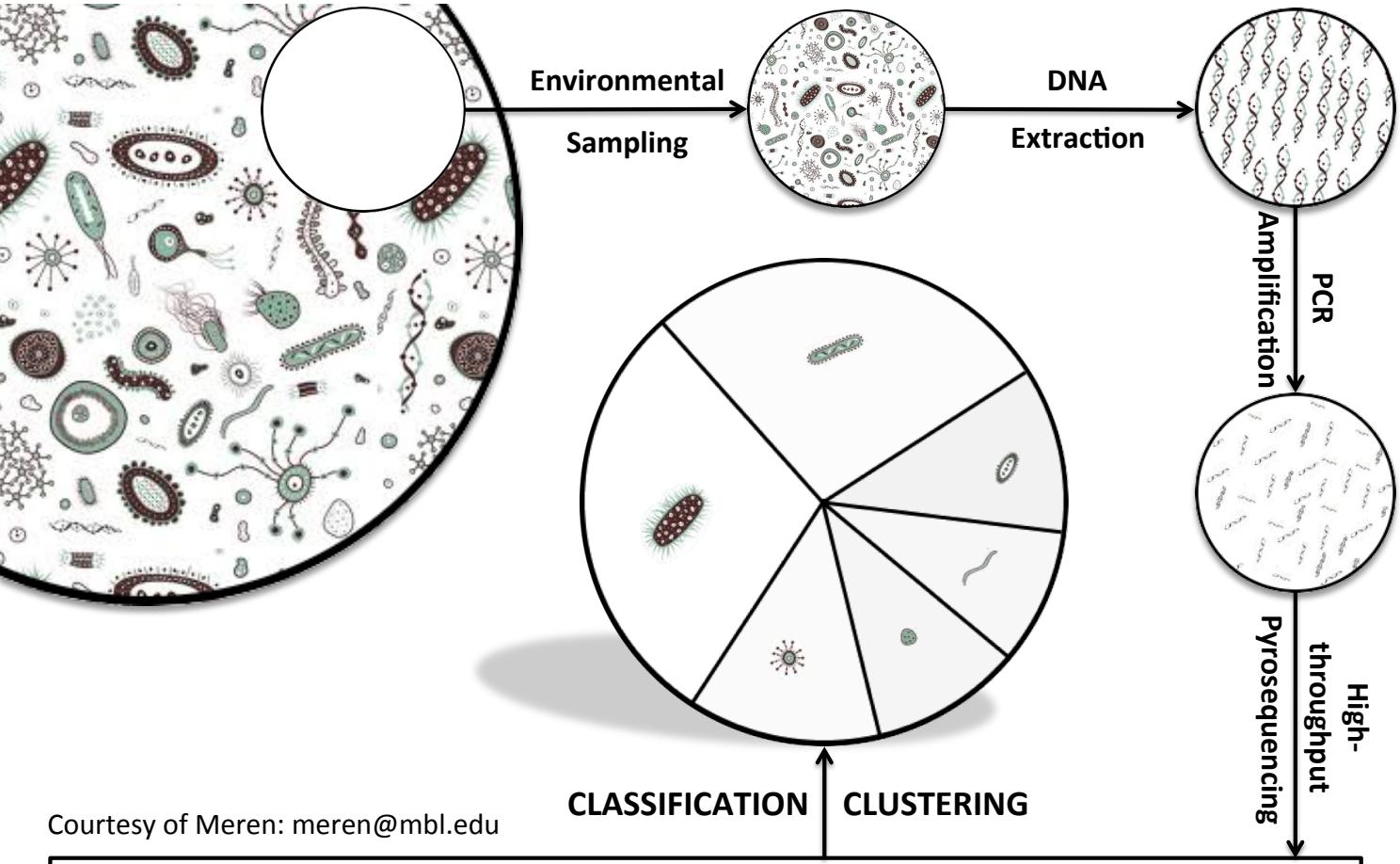
V6 (967F-1046R) 60 bp
V4 (518F-806R) 288 bp
V4V5 (518F-1064R) 550 bp
V9 (1380/1389-1513) for eukaryotes

Courtesy H.Morison, www.jbpc.mbl.edu

4

Molecular approaches for microbial taxonomic diversity

4. Molecular approaches for microbial taxonomic diversity



Courtesy of Meren: meren@mbl.edu

4. Molecular approaches for microbial diversity

Defining ecological units from sequences

(NON TRiVIAL!):

1. **Reference based**: taxonomic classification

Using specific reference databases, assign a taxonomy to each sequence.

OR

2. **De novo**: clustering

Cluster similar sequence together

4.1 Classification

1/ Submit all sequences to a « **classifier** » (web-based or locally) that uses reference database (presented later…). BLAST is not an option here!

2/ Gather all taxonomic annotations:

Bacteria;Proteobacteria;Betaproteobacteria;Nitrosomonadales;Nitrosomonadaceae;Nitrosomonas;

3/ Create groups of sequences having the same taxonomy

4.1 Classification

Results: an observation matrix

	Echantillon1	Echantillon2	Echantillon3	Echantillon4
Taxon1	200	480	30	0
Taxon2	100	220	100	0
Taxon3	50	100	200	200
Taxon4	10	50	200	50
Taxon5	2	1	2	50
	362	851	532	300

4.2 clustering

Clustering reads that present sequence similarity into OTU (operational taxonomic Units)

Allows to account for intraspecific variability

(1 OTU is supposed to represent 1 « ecological unit » = species?)

Reduces the impact of sequencing errors

PCR: ~10^{-4} - 10^{-6} (nt) / Sequencing: 0.1% des nt with Illumina

Major inconvenient: requires the use of hard similarity cutoff (97% or 98%) that in fact **do not correspond to any biological reality! do not correspond to any biological reality! do not correspond to any biological reality! do not correspond to any biological reality! do not correspond to any biological reality!...**

Emerging approaches in amplicon sequencing for microbial community analysis

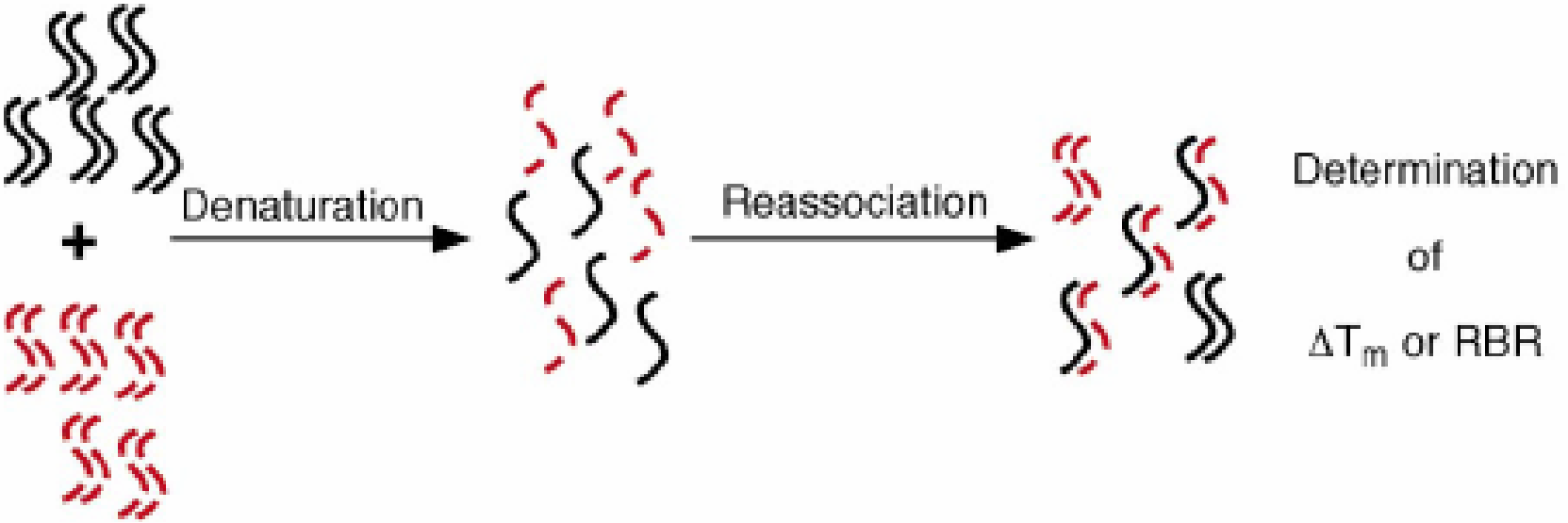
Defining coherent units based on DNA-DNA hybridation (DDH)

*"The **phylogenetic definition** of species generally would include strains with approximately 70% or greater DNA-DNA relatedness and with 5°C or less ΔT. "*

Wayne et al., Report of the **Ad** Hoc Committee on Reconciliation of Approaches to Bacterial Systematics JOURNAOLF SYSTEMATIC BACTERIOLOGY. 1987

Emerging approaches in amplicon sequencing for microbial community analysis

Defining coherent units based on DNA



Time consuming, requires isolates, difficult to standardize => fails to establish a general system for microorganisms classification

Courtesy of TOM COENYE, UGENT. http://www.cbs.dtu.dk/courses/brasil09/lectures/poltax+Burk.pdf

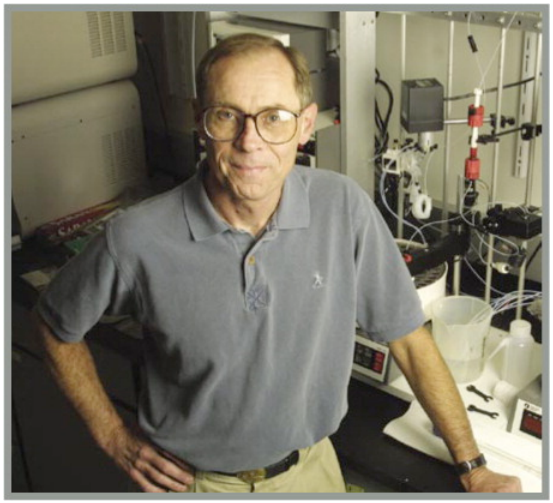
Emerging approaches in amplicon sequencing for microbial community analysis

Meanwhile…

Emerging approaches in amplicon sequencing for microbial community analysis

Defining coherent units based on rDNA genes

Mitch Sogin



Norman Pace

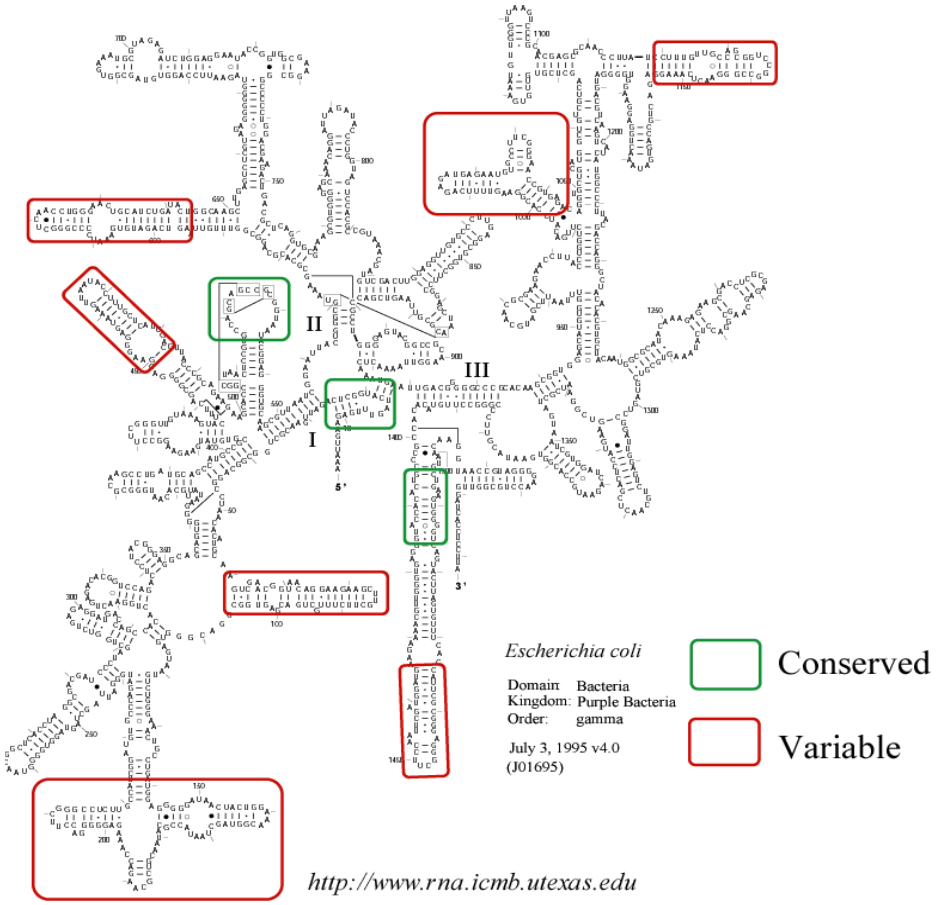
Carl Woese

And others…

Emerging approaches in amplicon sequencing for microbial community analysis

Defining coherent units based on rDNA genes

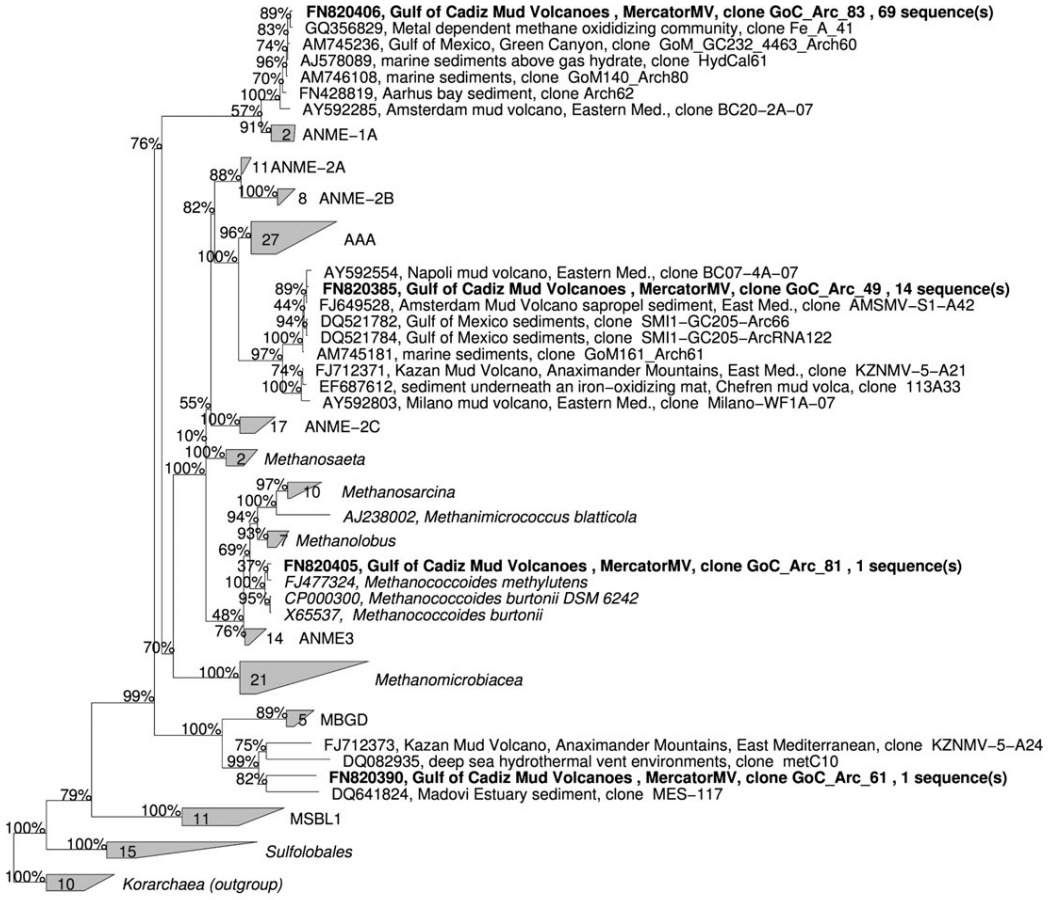
- Universal molecule present in all cells
- Strongly conserved domain can be used for PCR primer design
- Hypervariable region under lower selective pressure can be used to discriminate different cells



H.Morison, www.jbpc.mbl.edu

Emerging approaches in amplicon sequencing for microbial community analysis

Defining coherent units based on rDNA genes



Comparative 16S rRNA gene sequence analysis provides phylogenetic relationships between microorganisms, directly from enviropnemental samples.

Weeee!

From phylogeny to coherent units?

Emerging approaches in amplicon sequencing for microbial community analysis

- From molecular phylogeny to coherent units
  - Taxonomic **Classification**:

    Classify sequences based on their relatedness with already described species.
    - Very good for well studied taxa with many isolates (E. *Coli*, *Lactobacilus*, etc…)
    - Weak in tree branches with poor taxonomic resolution due to lack of isolates
  - Classification is just as good as the 16S reference database

Emerging approaches in amplicon sequencing for microbial community analysis

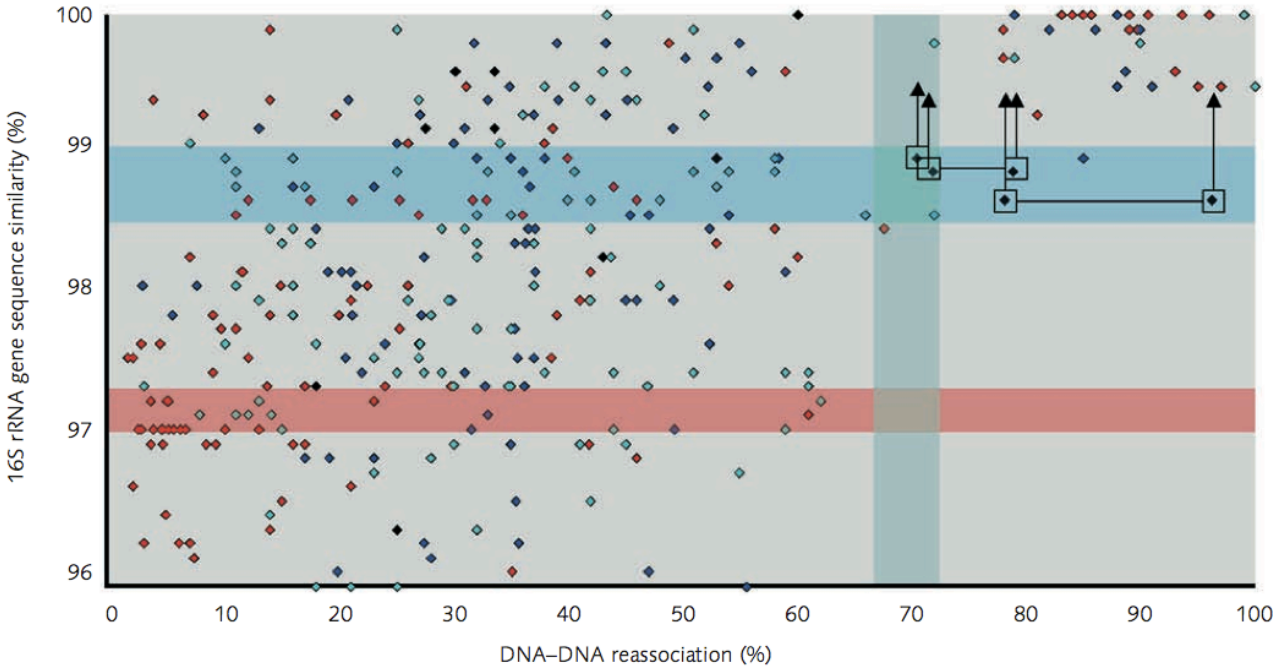
- From molecular phylogeny to coherent units
  - Clustering

    Group similar sequences together to account for **intra-specific variability.**

    How similar should they be?

Emerging approaches in amplicon sequencing for microbial community analysis

- From molecular phylogeny to coherent units
  - A correpondance between 16s rDNA similarity and DNA-DNA Hybridation



Stackebrandt and Ebers, 2006

Emerging approaches in amplicon sequencing for microbial community analysis

- From molecular phylogeny to coherent units
  - The Great Original Mistake: Clustering 16s rDNA at 97% similarity to define « Operational Taxonomic Units » **has no scientific basis** and may **masks useful diversity** among microorganisms.

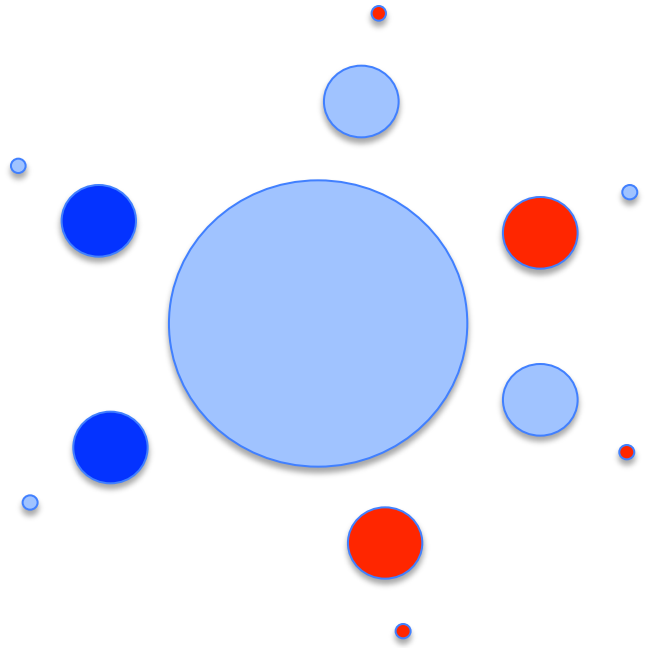
Emerging approaches in amplicon sequencing for microbial community analysis

- From molecular phylogeny to coherent units
  - The Great Original Mistake: Clustering 16s rDNA at 97% similarity to define « Operational Taxonomic Units » **has no scientific basis** and may **masks useful diversity** among microorganisms.
  - So why has everybody (including me) been doing this for years to define ecological units?

Emerging approaches in amplicon sequencing for microbial community analysis

- Because of sequencing errors!

(note that we slightly departed from a species definition)

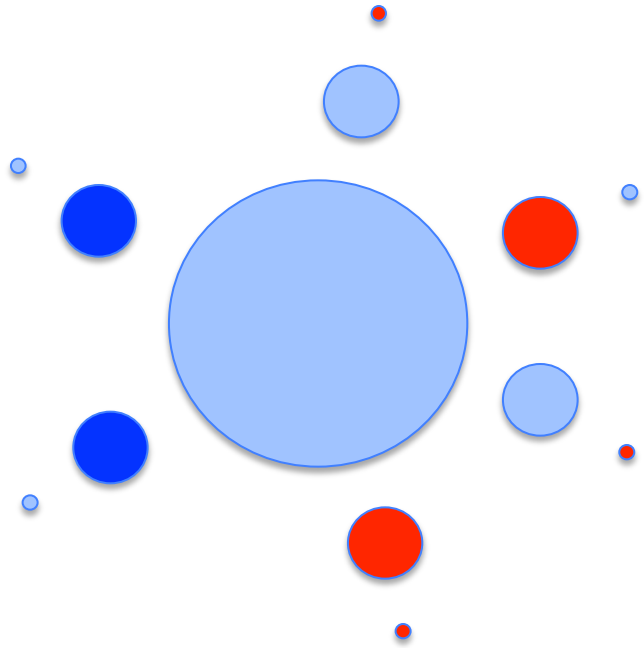


related sequences
with true variants and 5 errors…

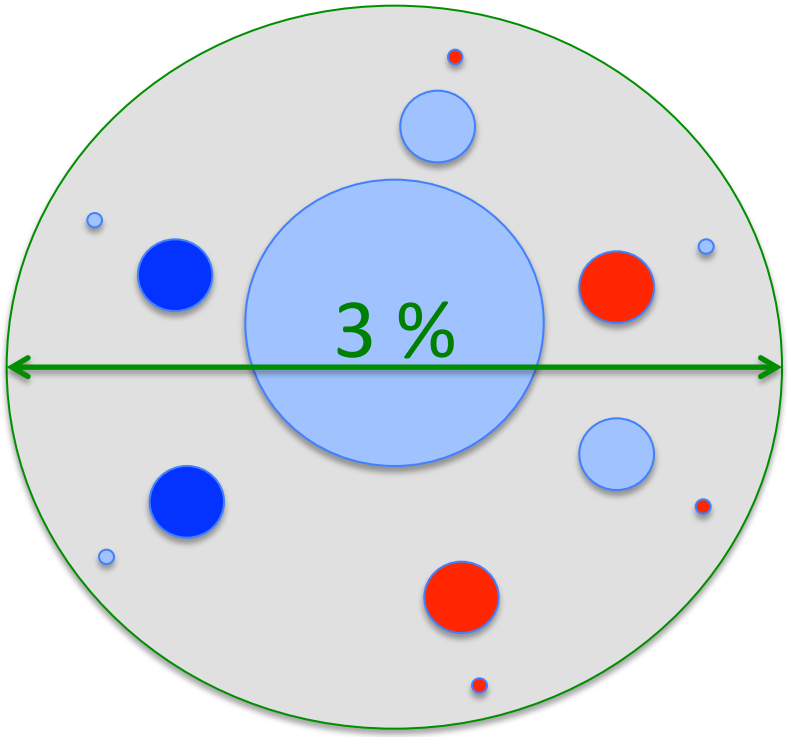
Emerging approaches in amplicon sequencing for microbial community analysis

- Because of sequencing errors

(note that we slightly departed from a definition of microbial species)

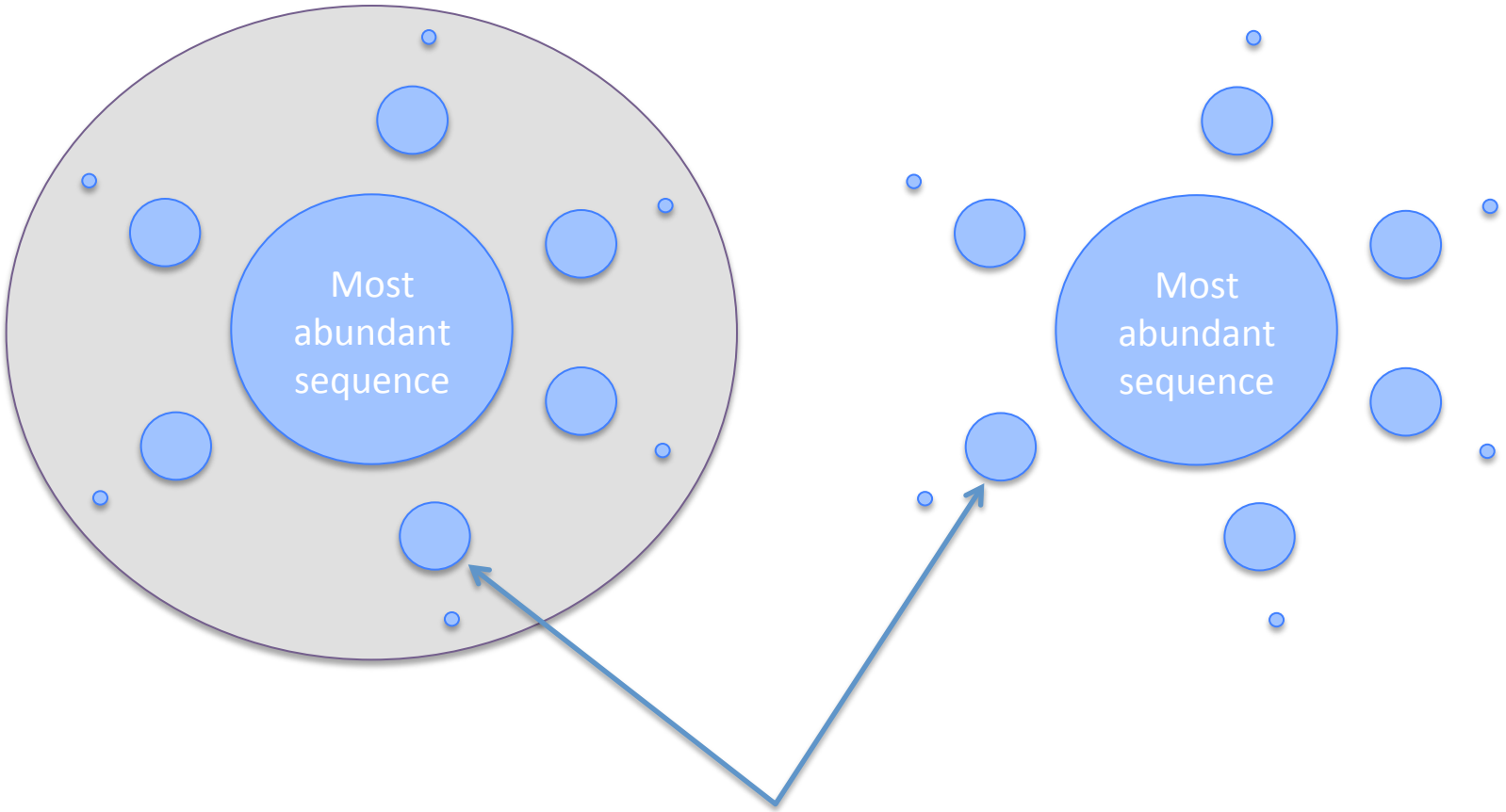


related sequences
with true variants and 5 errors…



…Grouped into 1 single OTU

Effect of sequencing errors on diversity measure



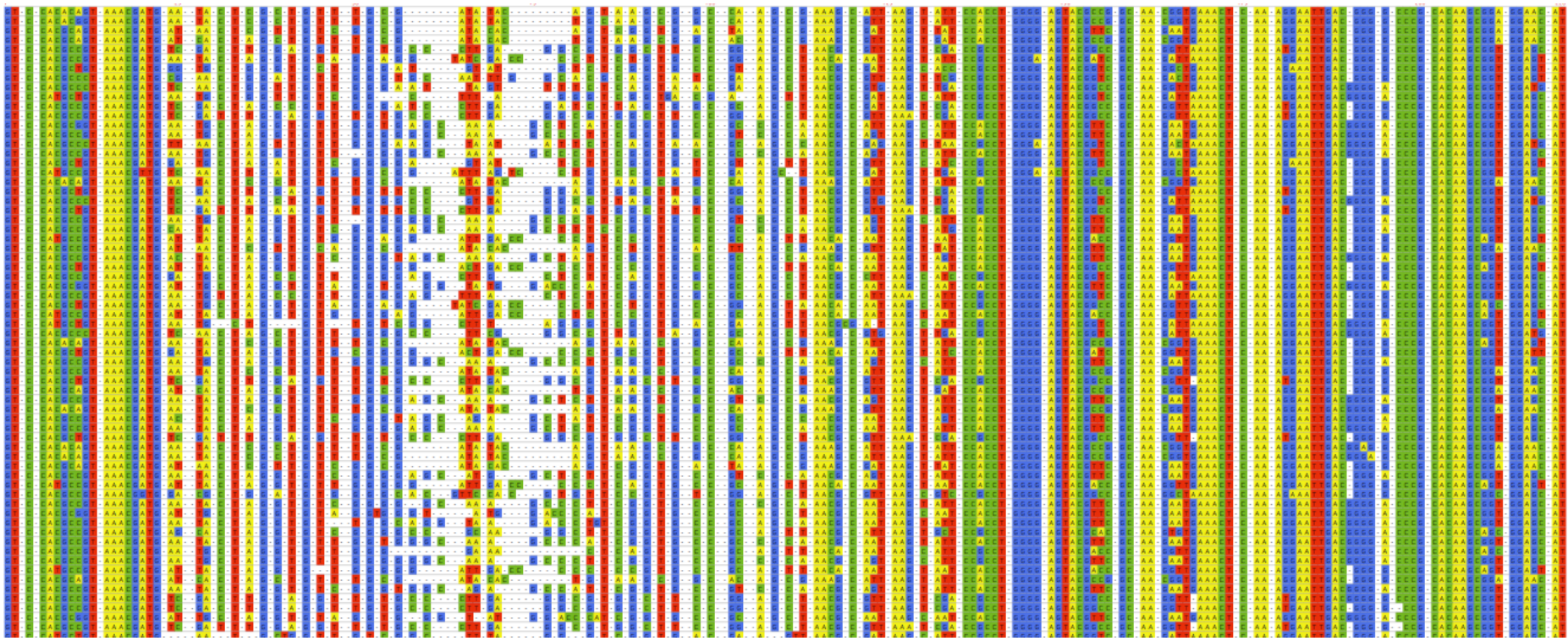
Sequencing errors or closely related variants?

1 ecological unit (97% OTU)

13 ecological units (100% OTUs)

4.2 Clustering

1/ Align sequences, possibly using ref. DB as a template (silva-greengenes-RDP)



4.2 Clustering

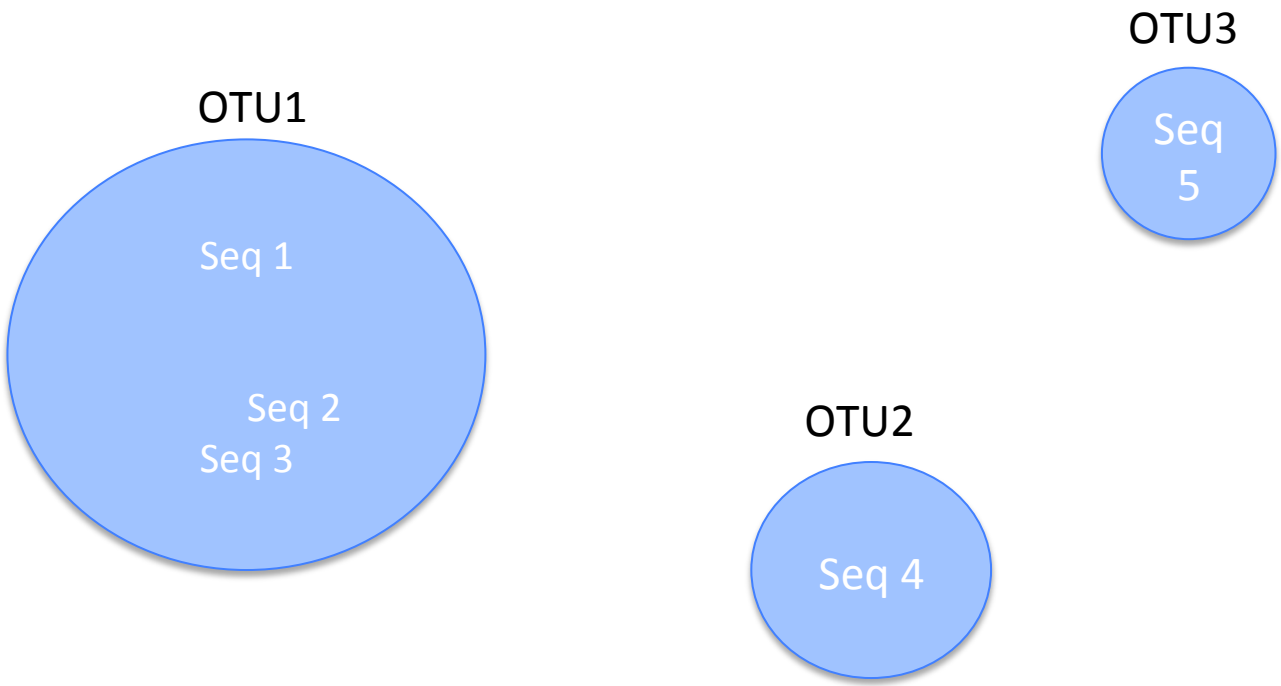
2/ Compute genetic distances => **sequence disimilarity matrix**

	seq1	seq2	seq3	seq4	seq5
seq1	0				
seq2	0.01	0			
seq3	0.02	0.03	0		
seq4	0.01	0.15	0.3	0	
seq5	0.05	0.08	0.09	0.5	0

4.2 Clustering

3/ OTU formation by clustering algorythms

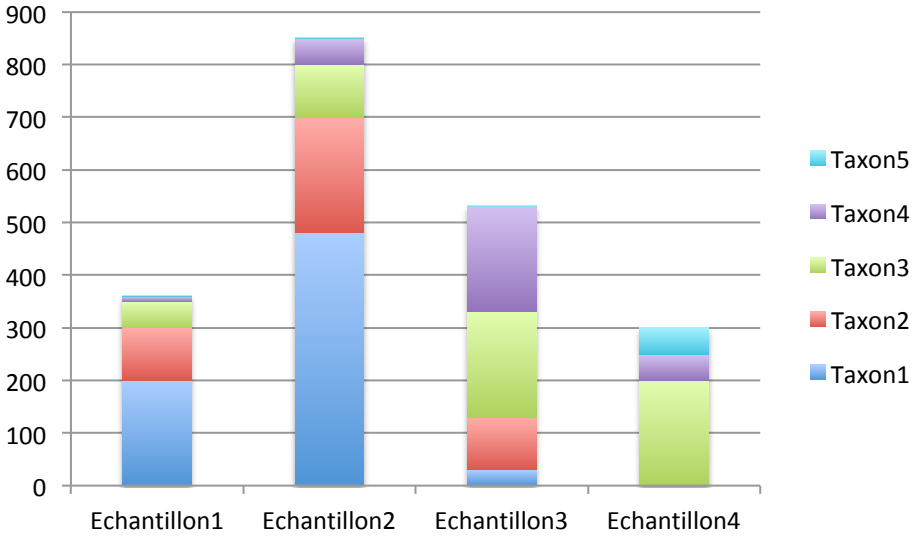
(nearest / furthest / average neighbor clustering)



4.3 Observation Matrix

An observation matrix is **the central object in ecology**.
At this point microbial ecology can use generic statistical methods in ecology (see « **Numerical Ecolology** », Legendre & Legendre)

	Echantillon 1	Echantillon 2	Echantillon 3	Echantillon 4
OTU1	200	480	30	0
OTU2	100	220	100	0
OTU3	50	100	200	200
OTU4	10	50	200	50
OTU5	2	1	2	50
	362	851	532	300



(a.k.a. Matrice de communauté, table OTU, …)

4.4 Emerging methods

- Both Clustering and classification approaches have their own caveats
- New methods:
  - Oligotyping and Minimum Entropy Decomposition

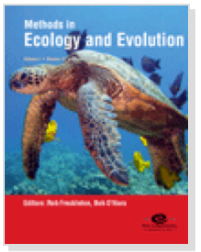
Methods in Ecology and Evolution
British Ecological Society
Explore this journal >

Open Access Creative Commons

Research Article

Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data

A. Murat Eren ✉, Loïs Maignien, Woo Jun Sul, Leslie G. Murphy, Sharon L. Grim, Hilary G. Morrison, Mitchell L. Sogin



View issue TOC
Volume 4, Issue 12
December 2013
Pages 1111–1119

A human-guided computational method that allows researchers to investigate the diversity of closely related but distinct bacterial organisms in final operational taxonomic units identified by canonical approaches.

GTTGTAAACCGCTTTTGATTGGGAGCAAGCTTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCCTTCGGGTGAGTGTACCTTTCGAATAAGCG

Reads from Environment 1

```
GTTGTAAACCGCTTTTGATTGGGAGCAAGCTTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCTTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCTTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCTTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCTTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCTTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCTTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCTTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCTTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCTTTCGGGTGAGTGTACCTTTCGAATAAGCG
```

Reads from Environment 2

```
GTTGTAAACCGCTTTTGATTGGGAGCAAGCCTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCCTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCCTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCCTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCCTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCCTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCCTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCCTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCCTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCCTTCGGGTGAGTGTACCTTTCGAATAAGCG
```

(in theory)

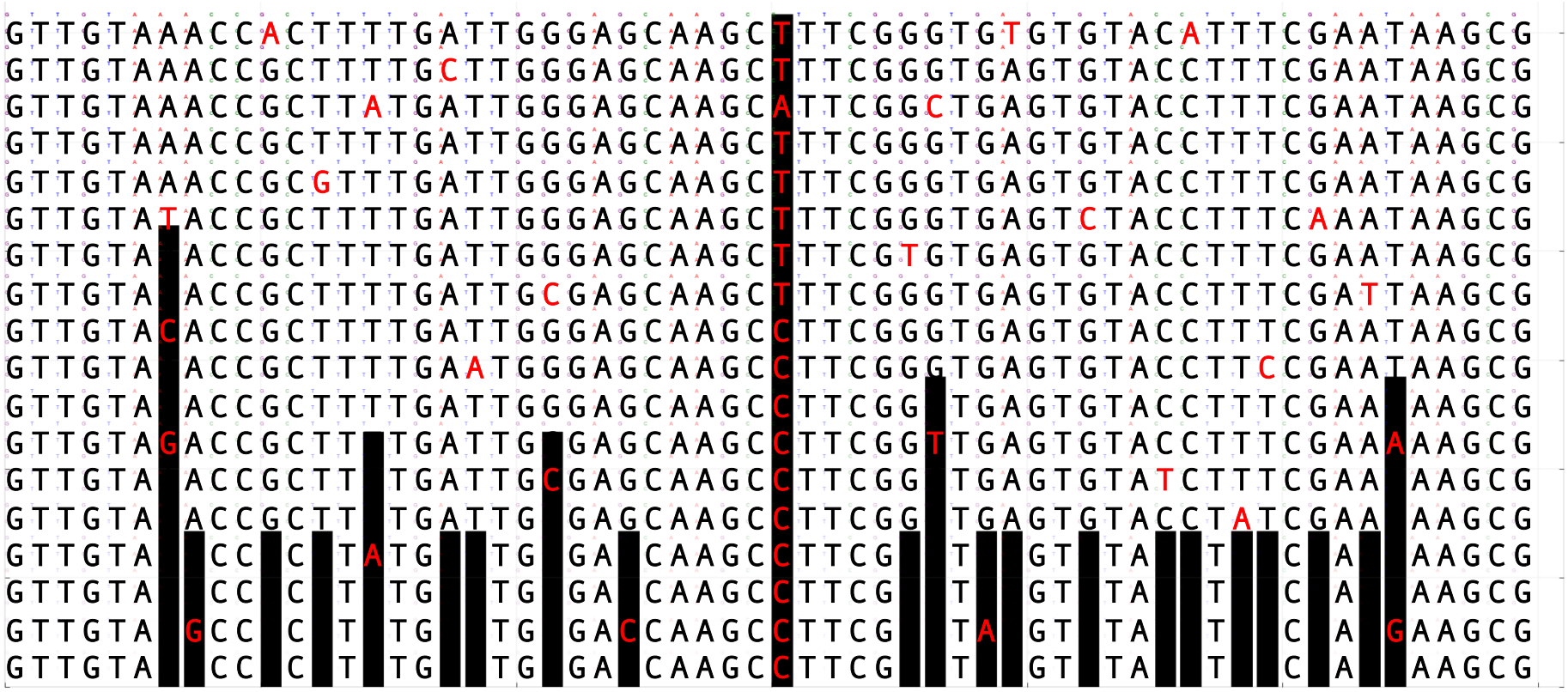
Reads from Environment 1

GTTGTAAACCGCTTTTGATTGGGAGCAAGCTTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCTTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCACTTTTGATTGGGAGCAAGCTTTCGGGTGTGTGTACATTTCGAATAAGCG
GTTGTAAACCGCTTTTGCTTGGGAGCAAGCTTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTATGATTGGGAGCAAGCATTCGGCTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCTTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCGTTTGATTGGGAGCAAGCTTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTATACCGCTTTTGATTGGGAGCAAGCTTTCGGGTGAGTCTACCTTTCAAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCTTTCGTGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGCGAGCAAGCTTTCGGGTGAGTGTACCTTTCGATTAAGCG

Reads from Environment 2

GTTGTACACCGCTTTTGATTGGGAGCAAGCCTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGAATGGGAGCAAGCCTTCGGGTGAGTGTACCTTCCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCCTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAGACCGCTTTTGATTGGGAGCAAGCCTTCGGTTGAGTGTACCTTTCGAAAAAGCG
GTTGTAAACCGCTTTTGATTGCGAGCAAGCCTTCGGGTGAGTGTATCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCCTTCGGGTGAGTGTACCTATCGAATAAGCG
GTTGTAAACCGCTTATGATTGGGAGCAAGCCTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCCTTCGGGTGAGTGTACCTTTCGAATAAGCG
GTTGTAAGCCGCTTTTGATTGGGACCAAGCCTTCGGGTAAGTGTACCTTTCGAAGAAGCG
GTTGTAAACCGCTTTTGATTGGGAGCAAGCCTTCGGGTGAGTGTACCTTTCGAATAAGCG

(in practice)



For a random variable X with n outcomes $\{n_i : i = 1, ..., n\}$, the Shannon entropy, is defined as:

$$H(X) = -\sum_{i=0}^{n} p(x_i) log_b(p(x_i))$$

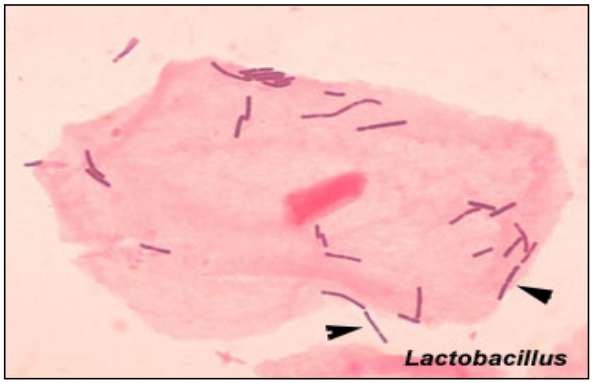
Entropy analysis of DNA sequences: A basic example

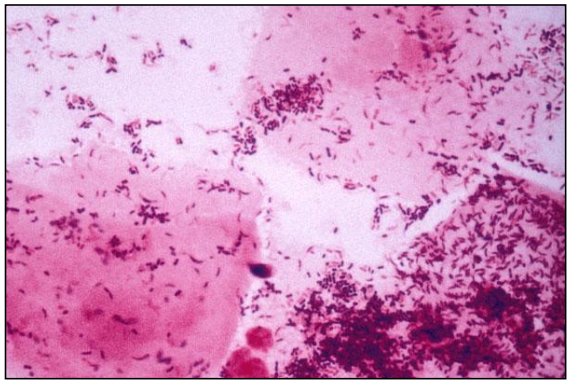
(slides by the author A. M. Eren)

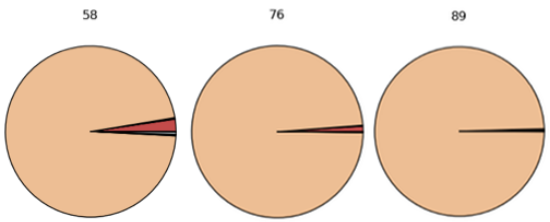
A. Murat Eren (Meren)
http://meren.org

1. Exemples d'utilisation des NGS

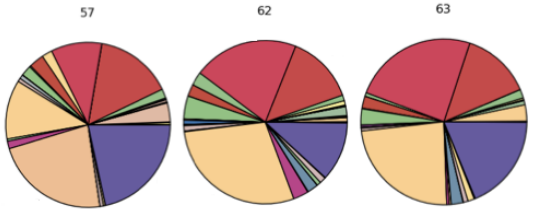
- Microbiome vaginal et Bacteriose vaginal (B.V.): infection à *Gardnerela vaginalis*







(genus level bacterial community composition of three healthy women)

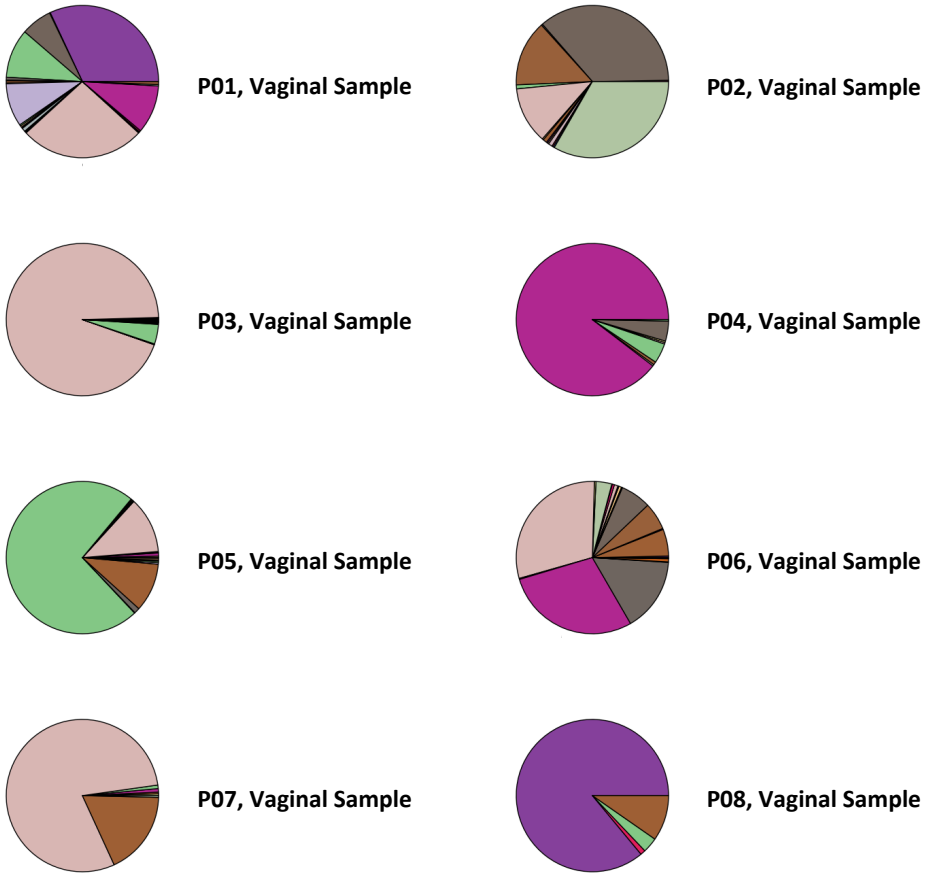


(genus level bacterial community composition of three women with BV)

Photo credit: Janice Carr, © 1982.

1. Exemples d'utilisation des NGS

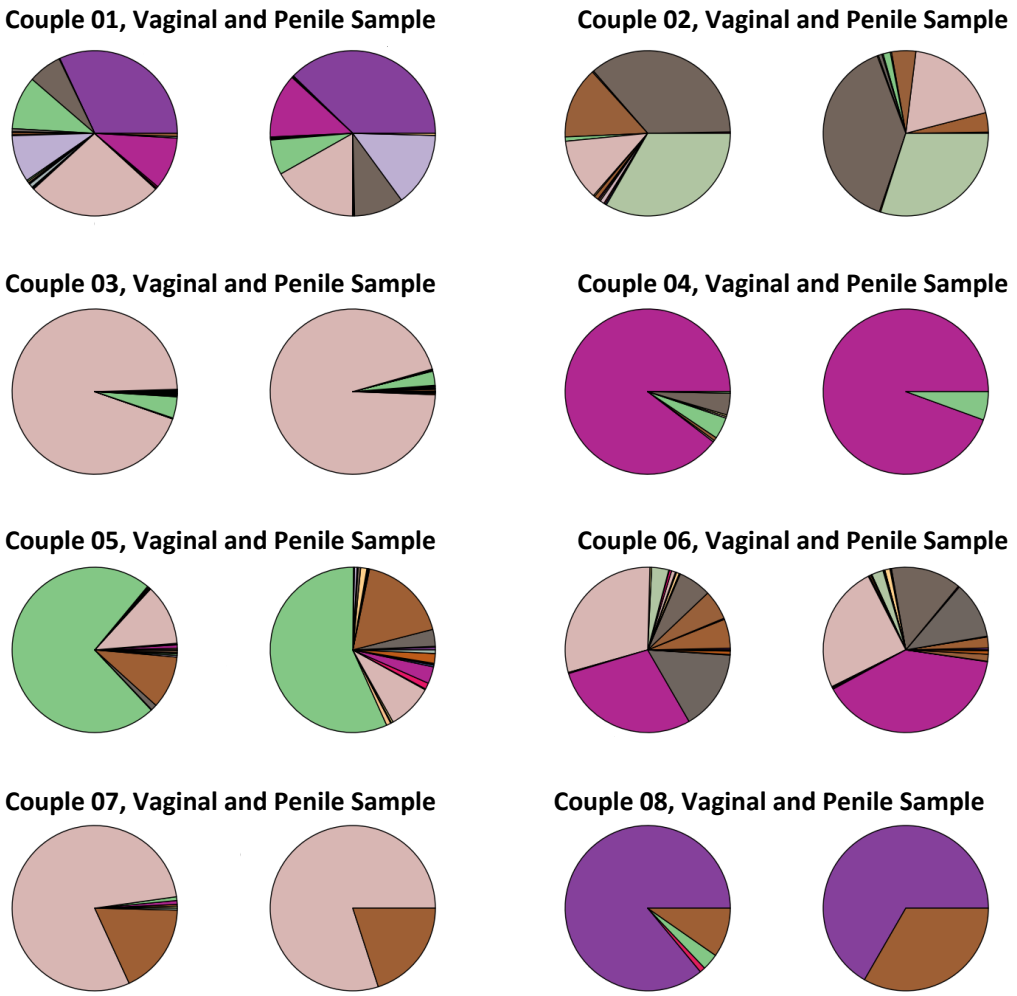
- Microbiome vaginal et Bacteriose vaginal (B.V.)



Classification par gène marker permet rarement d'identifier les souches pathogènes

1. Exemples d'utilisation des NGS

- Microbiome vaginal et Bacteriose vaginal (B.V.)



Eren *et al.* (2011) PLOS ONE

Minimum Entropy Decomposition

- Unsupervised version of oligotyping
  - The entire dataset is split according of the position of highest entropy => 4 nodes (A, T, C or G)
  - Each of these nodes is in turn split according to its highest entropy position and so on.
  - Decomposition stops when all nodes reach minimum entropies.

The ISME Journal (2015) **9,** 968–979; doi:10.1038/ismej.2014.195; published online 17 October 2014

Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences

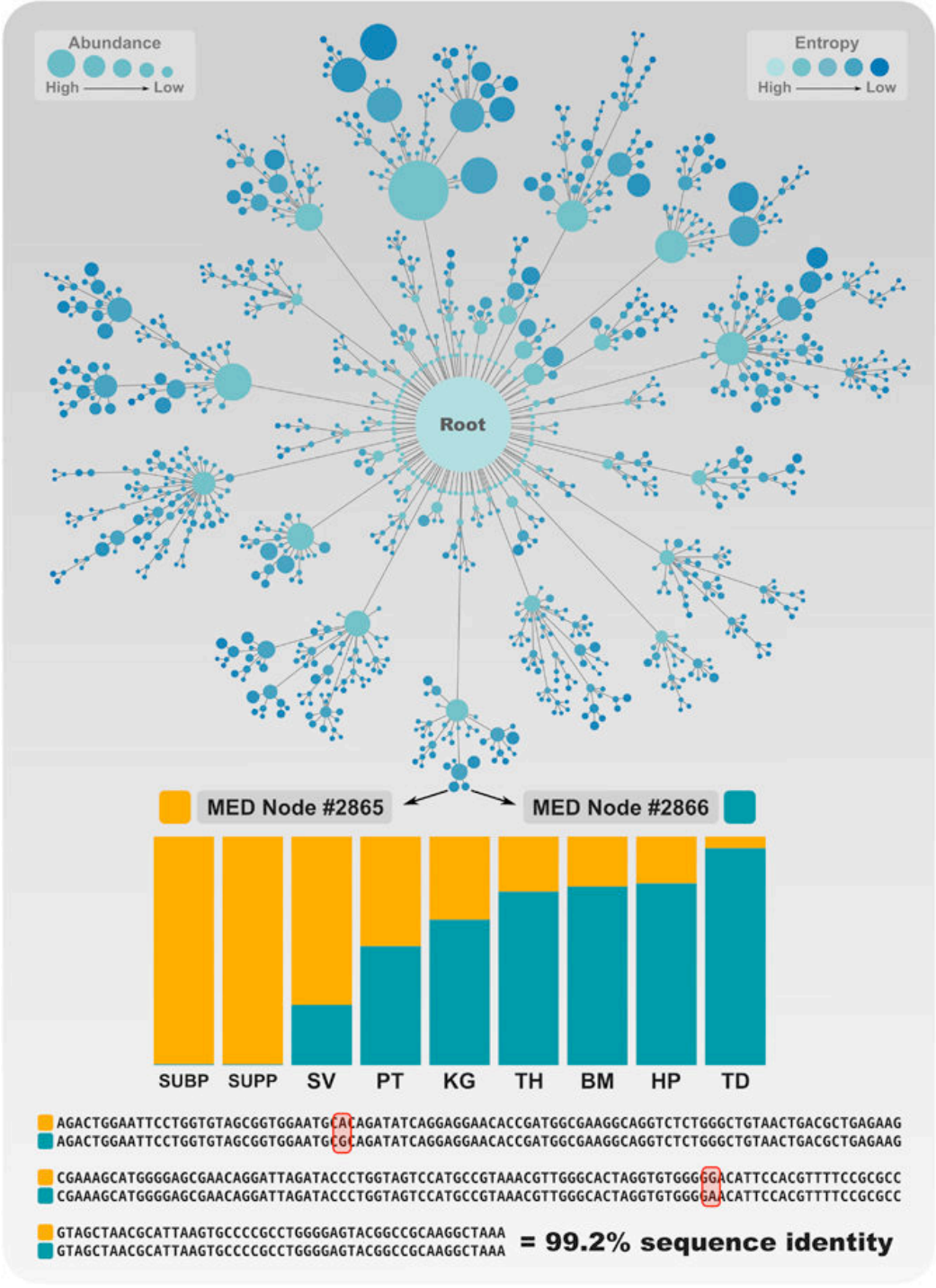
OPEN

A Murat Eren[1], Hilary G Morrison[1], Pamela J Lescault[1], Julie Reveillaud[1], Joseph H Vineis[1] and Mitchell L Sogin[1]

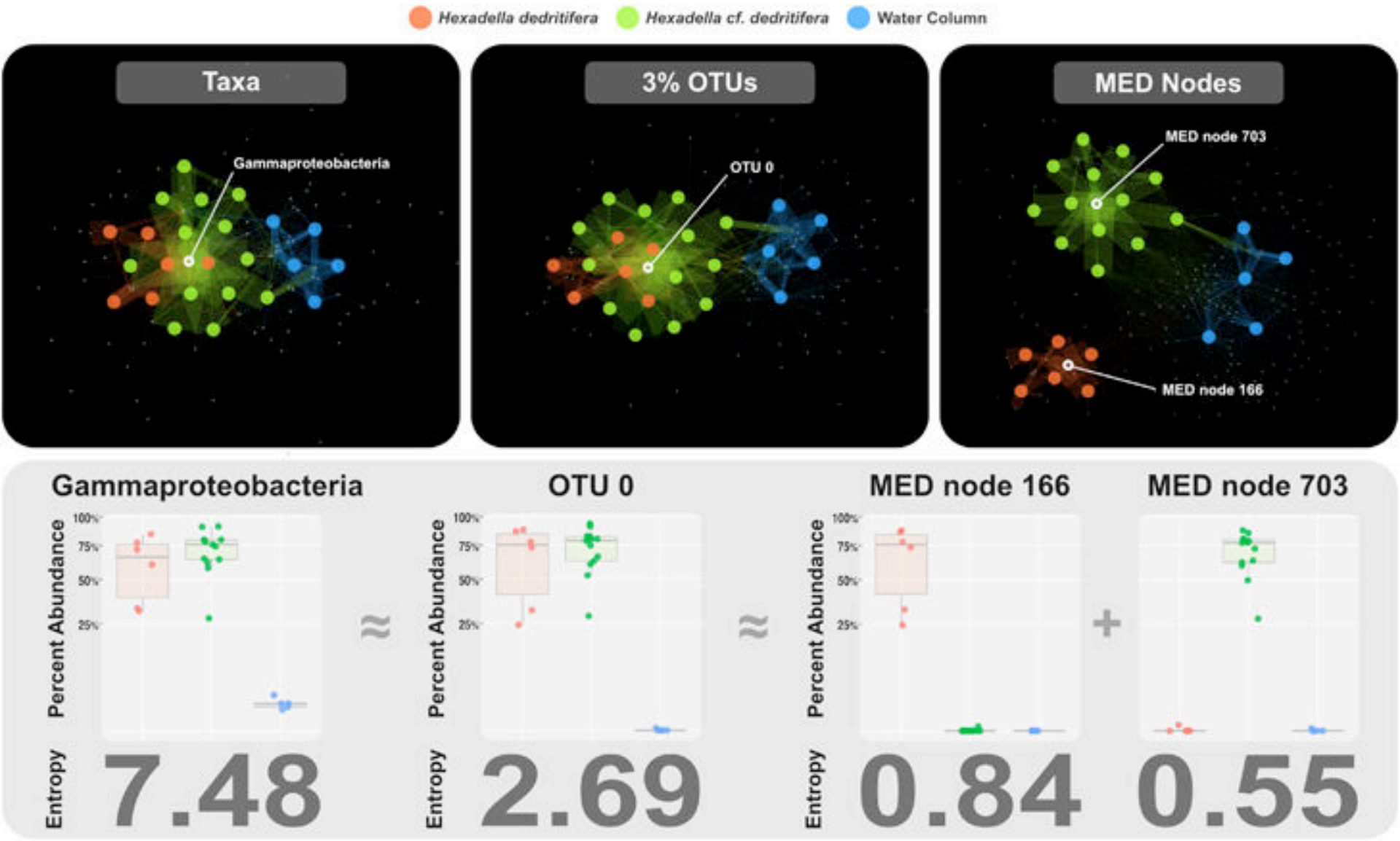
[1]Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA, USA

Minimum Entropy Decomposition

The topology of the MED process. The top panel shows the decomposition of the oral microbiome dataset composed of ~6 M reads into intermediate and final nodes by MED. Two final nodes in the topology are marked (nodes #2865 and #2866) and the bar-chart plot shows the their distribution across oral sites in human mouth; subgingival plaque (SUBP), supragingival plaque (SUPP), saliva (SV), palatine tonsils (PT), keratinized gingiva (KG), throat (TH), buccal mucosa (BM), hard palate (HP) and tongue dorsum (TD). The lower panel shows the alignment for the representative sequences of the two nodes, which are 99.2% identical



Minimum Entropy Decomposition



4.4 Emerging methods

- New methods:
  - Oligotyping and Minimum Entropy Decomposition
  - DADA2 (probabilistic approach for sequencing error detection and correction)

DADA2: High resolution sample inference from amplicon data

Benjamin J Callahan, Paul J McMurdie, Michael J Rosen, Andrew W Han, Amy J Johnson, Susan P Holmes

doi: http://dx.doi.org/10.1101/024034

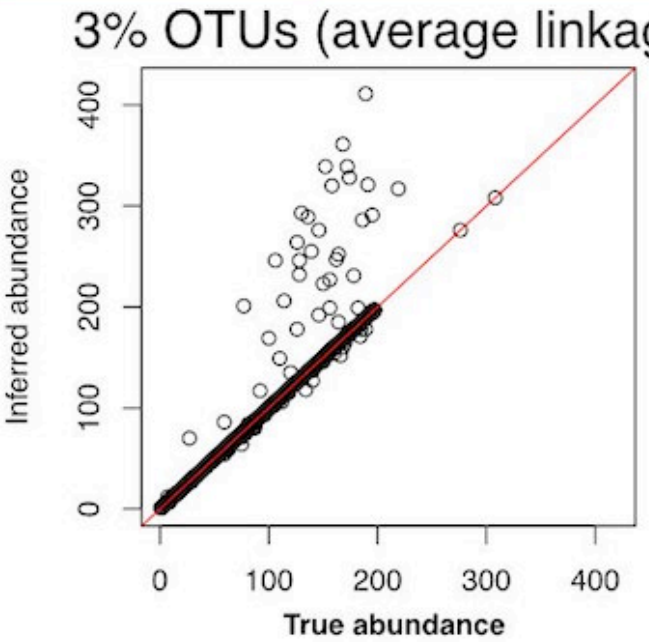
Now published in *Nature Methods* doi: 10.1038/nmeth.3869

DADA2

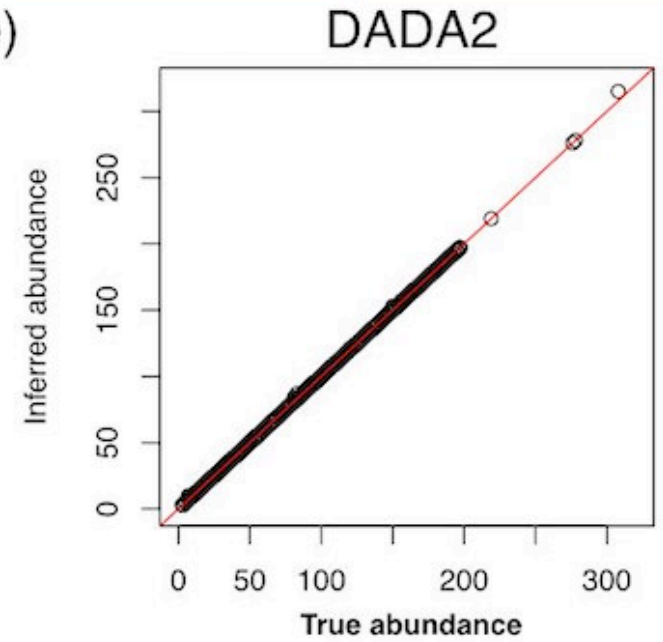
- DADA2 uses a probabilistic error model from the data itself
- Results in a dataset partitioning into Amplicon Sequence Variants (e.g. unique sequences) instead of clusters of similar sequences into OTUs

DADA2

Accuracy: Simulated data



TP: 978
FP: 272
FN: 77
cor: 0.935



TP: 1042
FP: 0
FN: 13
cor: 0.999

Data: Kopylova, et al. mSystems, 2016.

4.4 Emerging methods

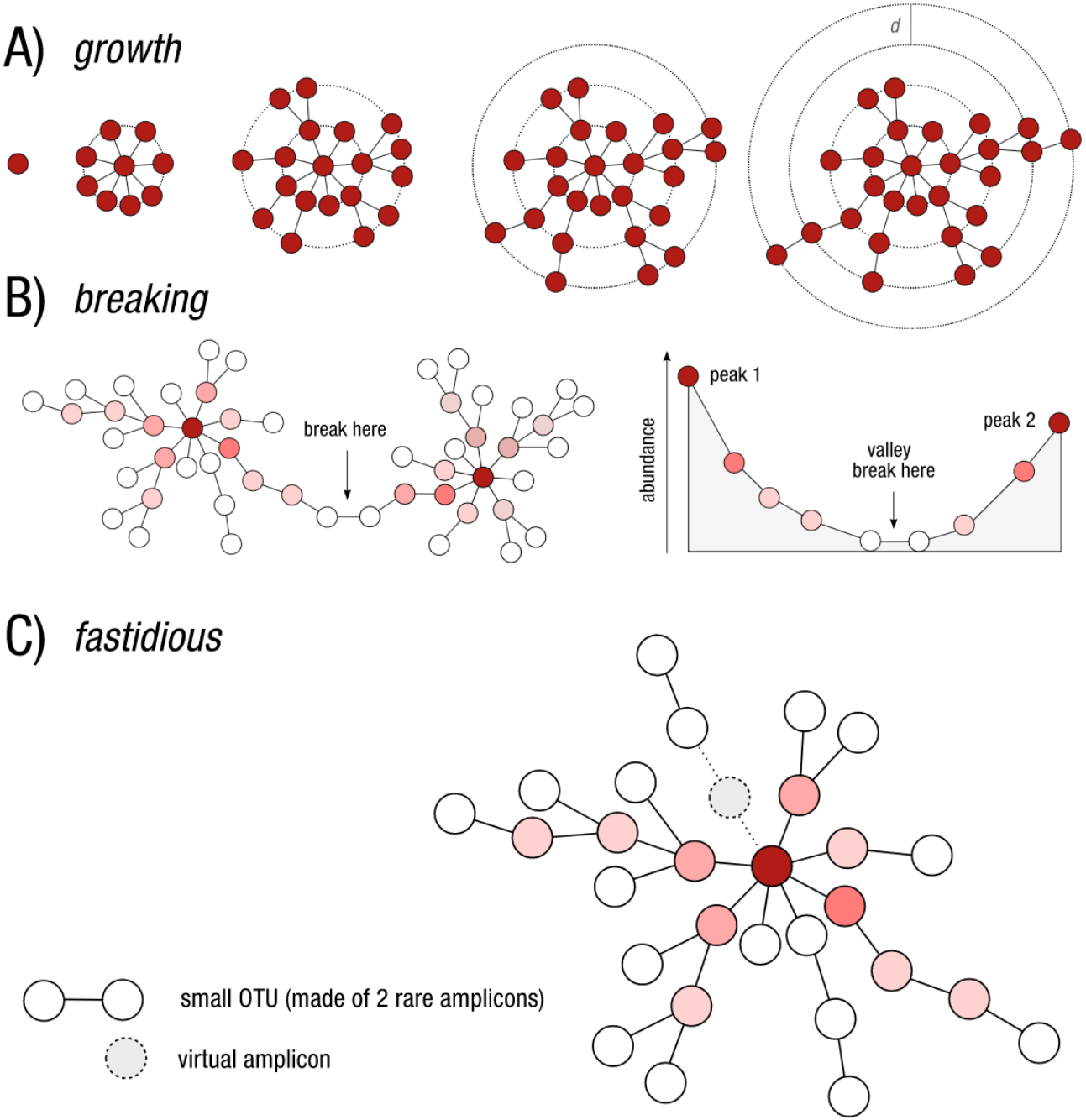
- New methods:
  - Oligotyping and Minimum entropy decomposition
  - DADA2
  - SWARM-V2: exploration of OTU natural boundaries around most abundant sequences

Swarm v2: highly-scalable and high-resolution amplicon clustering

Biodiversity | Bioinformatics | Environmental Sciences | Microbiology | Molecular Biology

Frédéric Mahé[1], Torbjørn Rognes[2,3], Christopher Quince[4], Colomban de Vargas[5,6], Micah Dunthorn[1]

SWARM2


A) growth
d
B) breaking
break here
peak 1
peak 2
valley
break here
abundance
C) fastidious
small OTU (made of 2 rare amplicons)
virtual amplicon

5.
Comparing microbial communities

5. Comparing microbial communities

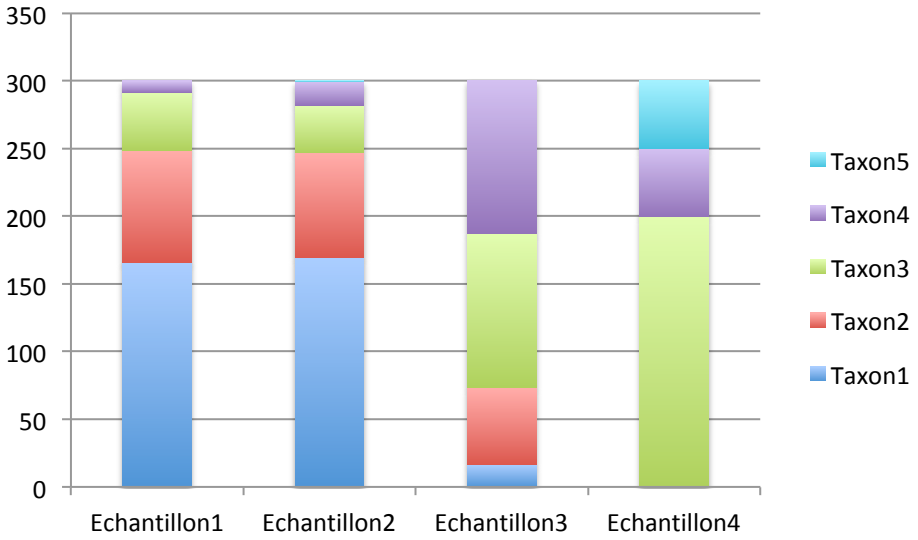
- Matrix normalization!

  Accounting for differences in sampling depth

5. Comparing microbial communities

- Normalisation by random subsampling:

	Echantillon 1	Echantillon 2	Echantillon 3	Echantillon 4
Taxon1	166	169	17	0
Taxon2	83	78	56	0
Taxon3	43	35	114	200
Taxon4	8	18	113	50
Taxon5	0	0	0	50
	300	300	300	300



All communties have the same number of individuals
Loss of information

5. Comparing microbial communities

PLOS | COMPUTATIONAL BIOLOGY A Peer-Reviewed, Open Access Journal

View this Article | Submit to PLOS | Get E-Mail Alerts | Contact Us

PLoS Comput Biol. 2014 Apr; 10(4): e1003531. PMCID: PMC3974642

Published online 2014 Apr 3.

doi: 10.1371/journal.pcbi.1003531

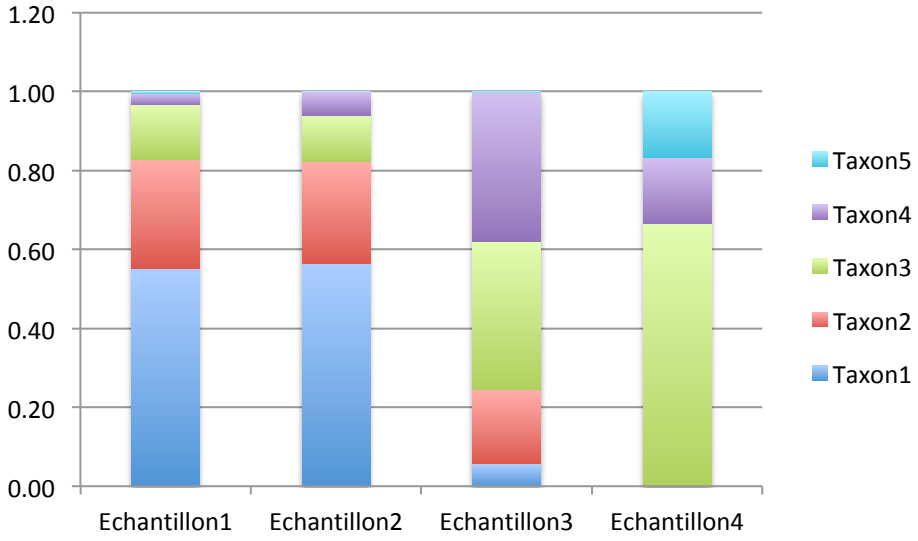
Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible

Paul J. McMurdie and Susan Holmes *

5. Comparing microbial communities

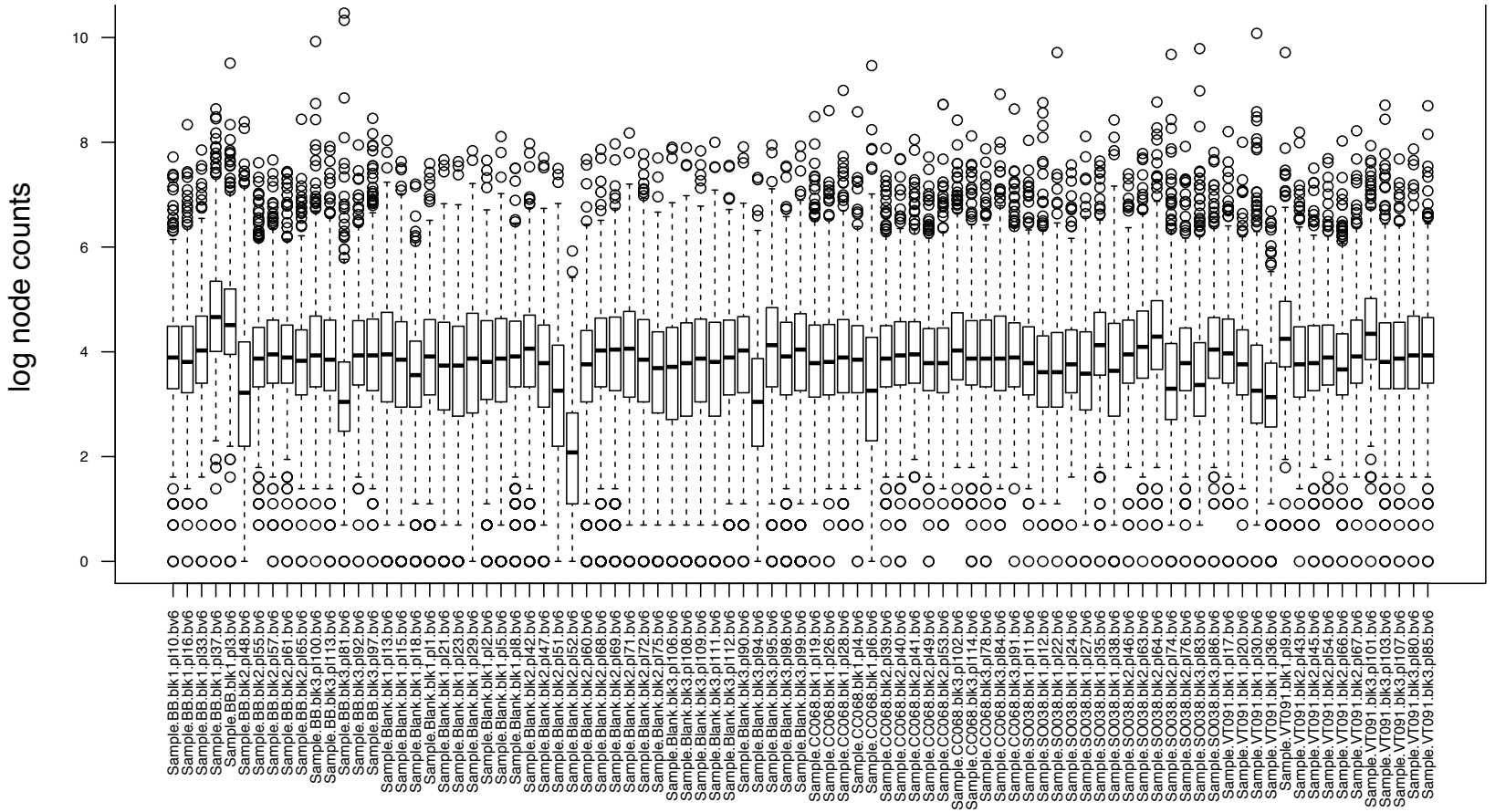
- Normalisation by fraction:

	Echantillon 1	Echantillon 2	Echantillon 3	Echantillon 4
Taxon1	0.55	0.56	0.06	0.00
Taxon2	0.28	0.26	0.19	0.00
Taxon3	0.14	0.12	0.38	0.67
Taxon4	0.03	0.06	0.38	0.17
Taxon5	0.01	0.00	0.00	0.17
	1.00	1.00	1.00	1.00



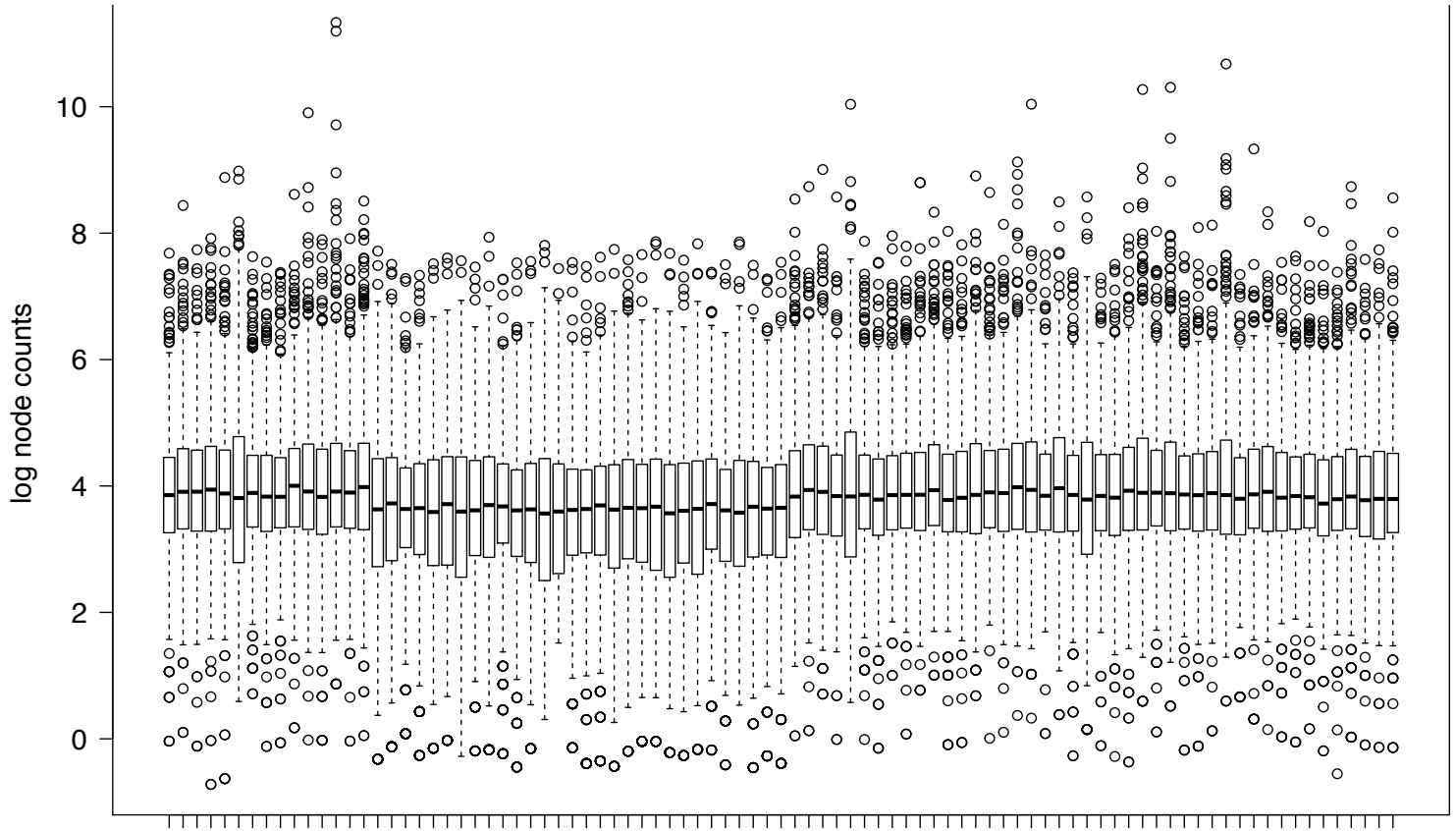
5. Comparing microbial communities

- Normalisation mean / variance stabilisation:



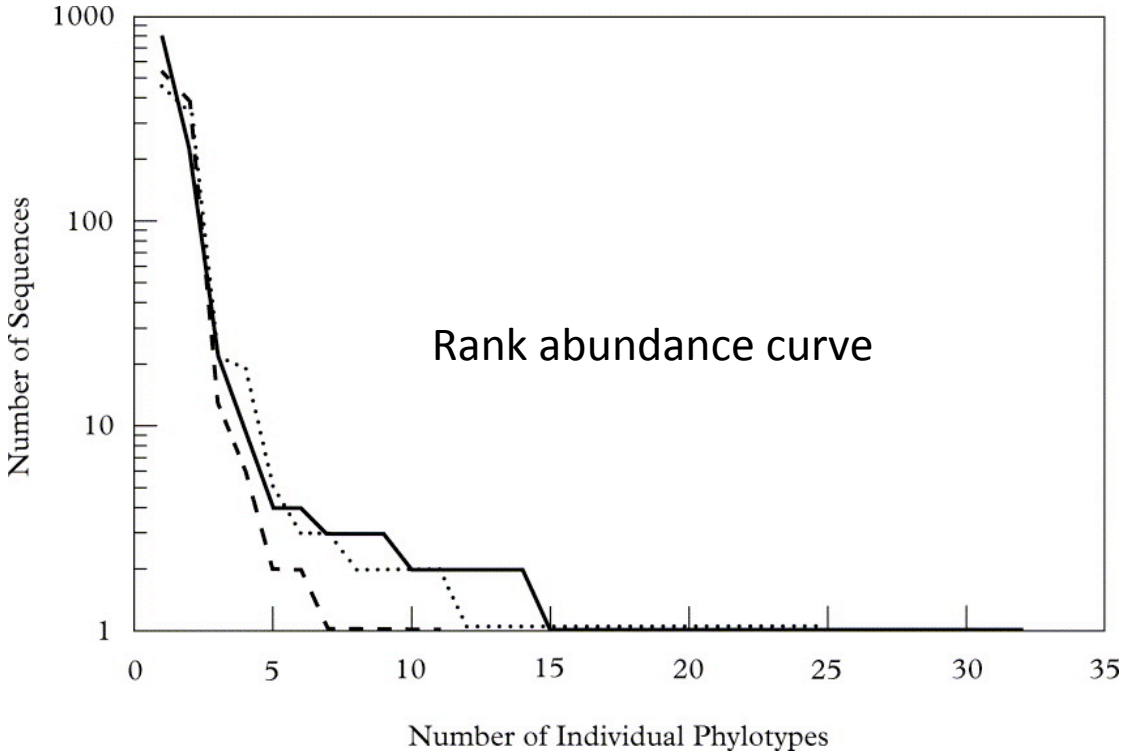
5. Comparing microbial communities

- Normalisation mean / variance stabilisation (DeSeq2 or MetagenomeSeq):



5. Comparing microbial communities

α-diversity: observed richness

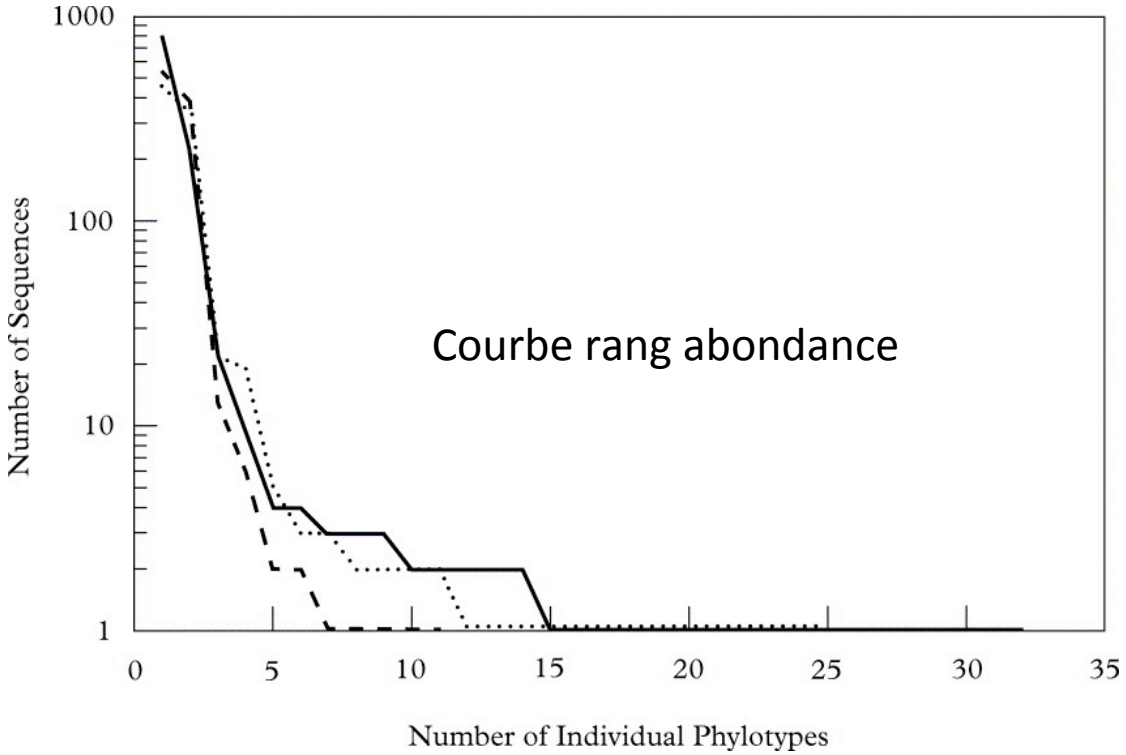


Sobs = 32

pino-bodas et al. The Lichenologist 2012

5. Comparing microbial communities

α-diversity: non-parametric richness estimator

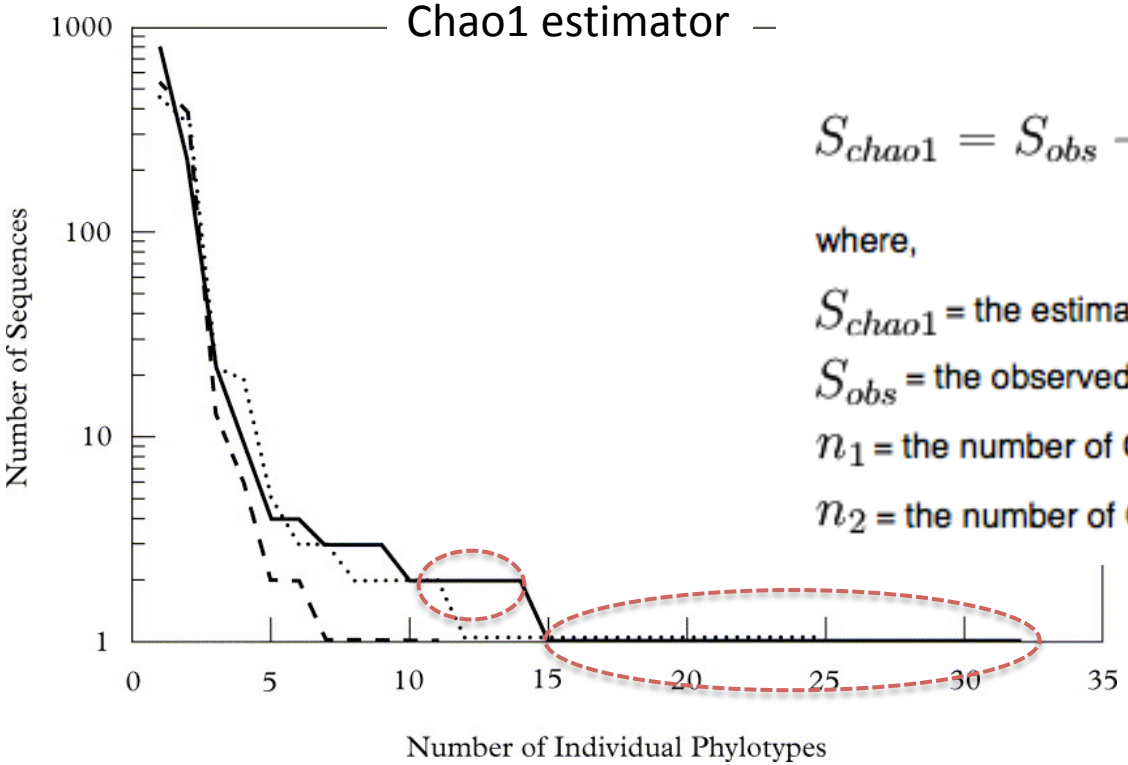


pino-bodas et al. The Lichenologist 2012

Given the richness in the sample, how to estimate the richness in the initial community?

5. Comparing microbial communities

α-diversity: non-parametric richness estimator



$$S_{chao1} = S_{obs} + \frac{n_1 (n_1 - 1)}{2 (n_2 + 1)}$$

where,

S_{chao1} = the estimated richness

S_{obs} = the observed number of species

n_1 = the number of OTUs with only one sequence (i.e. "singletons")

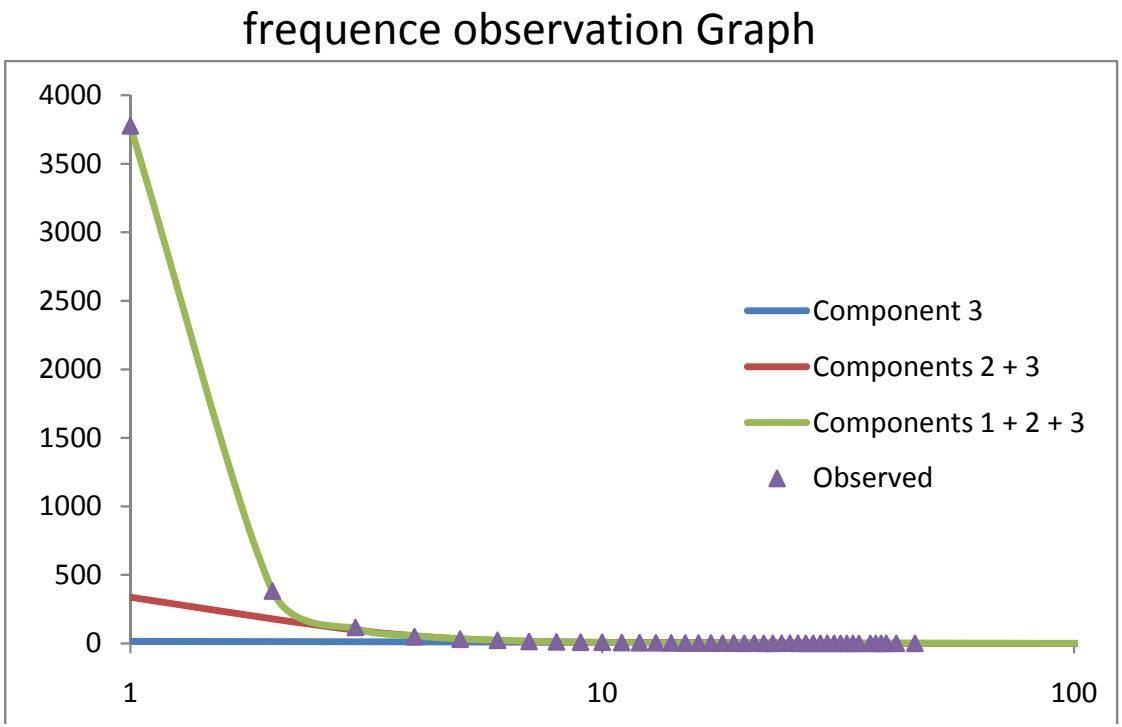
n_2 = the number of OTUs with only two sequences (i.e. "doubletons")

$S_{chao1}=59$

pino-bodas et al. The Lichenologist 2012

5. Comparing microbial communities

α-diversity: parametric richness estimator



Postulate that community structure is a (mixture of) canonical distributions (binomiale, poisson, lognormal…)
The model that fits best observed datapoint provides distribution parameters, hence richness
CatchAll (Bunge et al. 2013, Bioinformatics 28:1045—47)

5. Comparing microbial communities

β-diversity: Measuring relative distances between microbial communities

Using (di)similarity indices

- $0 < \text{Indice} < 1$
- $I_{A,B}=0$ => communities A et B are identical
- $I_{A,B}=1$ => communities A and B are “infinitely” different

5. Comparing microbial communities

β-diversity: Measuring relative distances between microbial communities

RELATIVE ABONDANCE : exemple Morisita-Horn index

	Echantillon 1	Echantillon 2	Echantillon 3	Echantillon 4
Taxon1	166	169	17	0
Taxon2	83	78	56	0
Taxon3	43	35	114	200
Taxon4	8	18	113	50
Taxon5	0	0	0	50
	300	300	300	300

$$D_{Morisita-Horn} = 1 - 2\frac{\sum \frac{S_{A,i}}{n}\frac{S_{B,i}}{m}}{\sum \left(\frac{S_{A,i}}{n}\right)^2 + \sum \left(\frac{S_{B,i}}{m}\right)^2}$$

where,

$S_{A,i}$ = the number of individuals from community A in the ith OTU

$S_{B,i}$ = the number of individuals from community B in the ith OTU

n = the number of individuals in community A

m = the number of individuals in community B

http://www.mothur.org/wiki/Morisitahorn

5. Comparing microbial communities

β-diversity: Measuring relative distances between microbial communities

	Echantillon1	Echantillon2	Echantillon3	Echantillon4
Echantillon1	0			
Echantillon2	0.000000	0		
Echantillon3	0.5958953	0.5934003	0	
Echantillon4	0.7826372	0.8036867	0.2327174	0

Sample distance matric based on M-H similarity distances

5. Comparing microbial communities

β-diversity: Measuring relative distances between microbial communities

INCIDENCE (OTUs Presence / Absence): Jaccard index

	Echantillon 1	Echantillon 2	Echantillon 3	Echantillon 4
Taxon1	166	169	17	0
Taxon2	83	78	56	0
Taxon3	43	35	114	200
Taxon4	8	18	113	50
Taxon5	0	0	0	50
	300	300	300	300

$$D_{Jaccard} = \frac{S_{AB}}{S_A + S_B - S_{AB}}$$

where,

S_{AB} = the number of shared OTUs between communities A and B

S_A = the number of OTUs in community A

S_B = the number of OTUs in community B

http://www.mothur.org/wiki/Jclass

5. Comparing microbial communities

β-diversity: Measuring relative distances between microbial communities

INCIDENCE (OTUs Presence / Absence):

	Echantillon1	Echantillon2	Echantillon3	Echantillon4
Taxon1	1	1	1	0
Taxon2	1	1	1	0
Taxon3	1	1	1	1
Taxon4	1	1	1	1
Taxon5	0	0	0	1

Equivalent to binary transformation of the distance matrix

http://www.mothur.org/wiki/Jclass

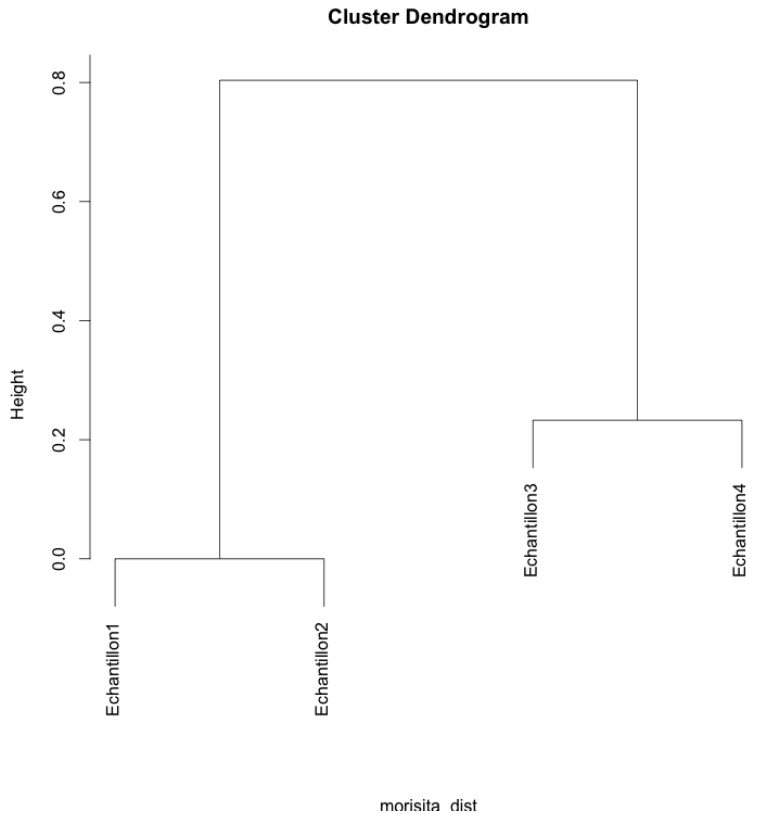
6.

Back to ecology…

6. Back to ecology… Visualization

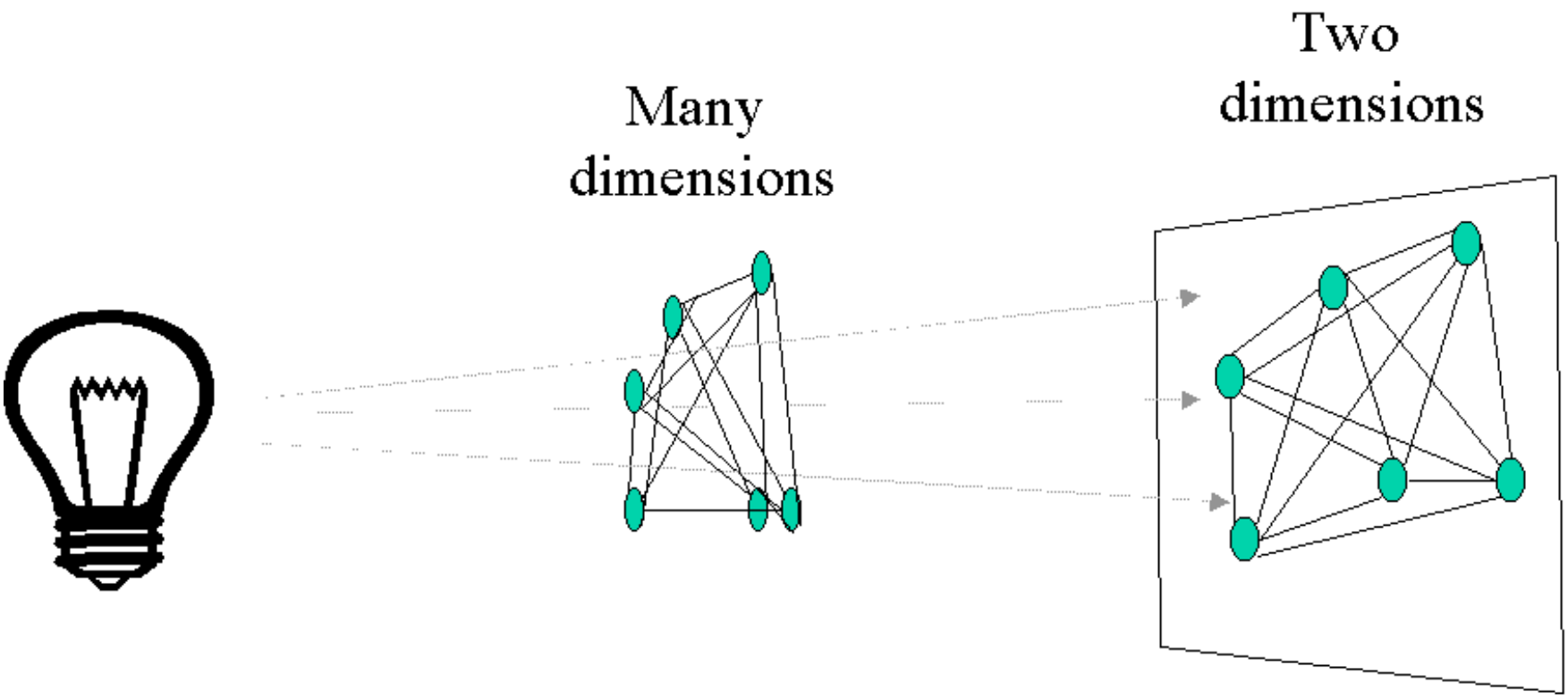
- Cladogram

(using classical clustering algorythm like UPGMA,…)



6. Back to ecology… Visualization

- PCoA (Principal coordinate analysis)



http://ordination.okstate.edu/overview.htm#Figure_5

6. Back to ecology… Visualization

- PCoA (Principal coordinate analysis)