



# LECTURE I: INTRODUCTION

ALEXANDER EILER  
DEPARTMENT OF BIOSCIENCES - AQUA

[alexander.eiler@ibv.uio.no](mailto:alexander.eiler@ibv.uio.no)

<https://www.mn.uio.no/ibv/english/people/aca/alexaei/>

# BIODIVERSITY, WHAT IS THAT?

**What statements are correct?**

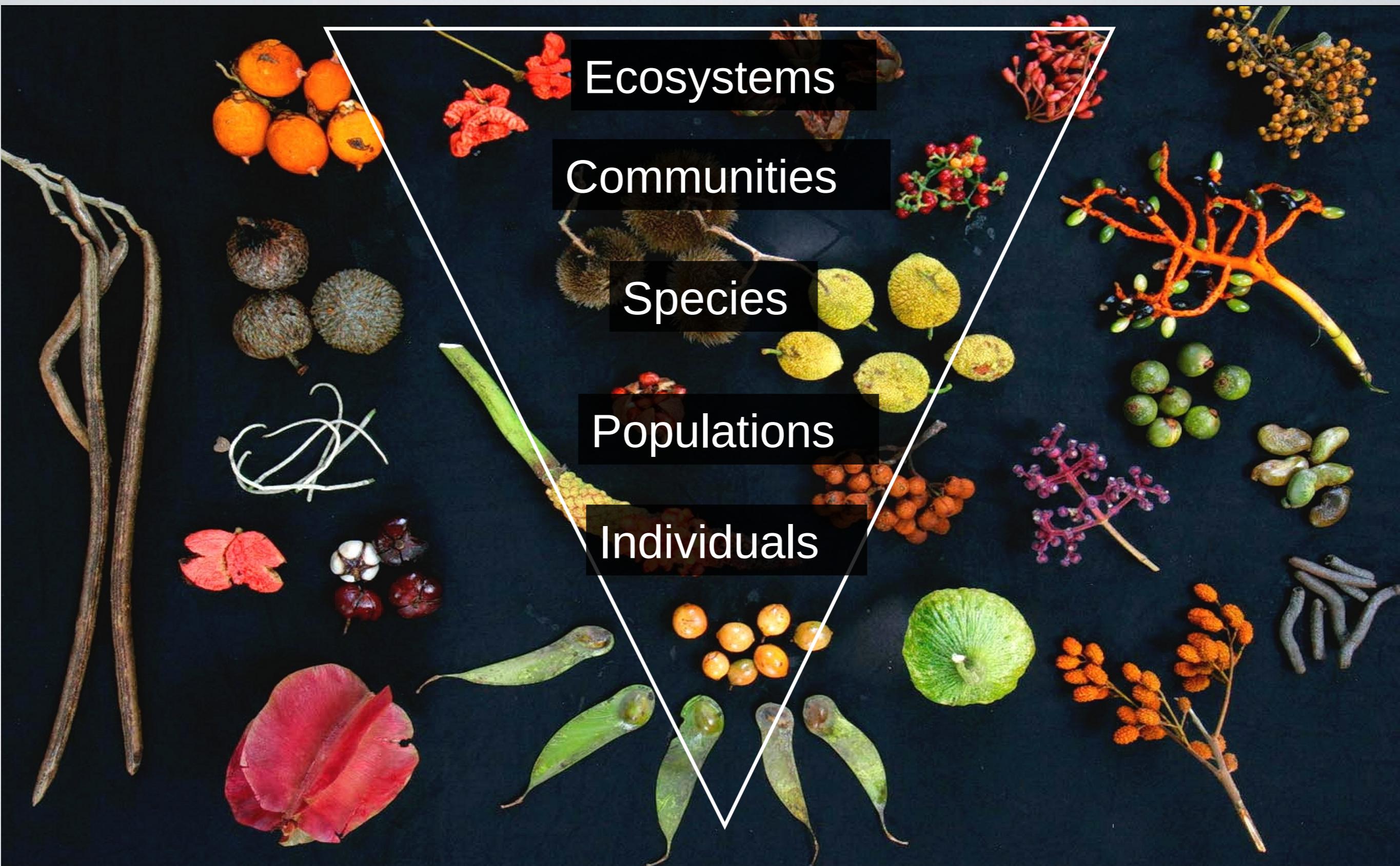
- There are multiple levels of biodiversity - it is not just about species diversity.
- Diversity includes measures for richness and evenness of the community.
- Biodiversity forms the foundation of the vast array of ecosystem services that critically contribute to human well-being.
- Ecological indicators are scientific constructs that use quantitative data to measure aspects of biodiversity, ecosystem condition, services, or drivers of change, but no single ecological indicator captures all the dimensions of biodiversity.

# WHAT IS (BIO)DIVERSITY?



"the variability among living organisms from all sources including terrestrial marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and ecosystems"  
(United Nations Environment Programme, 1995)

# BIODIVERSITY



# FORMS OF DIVERSITY

- functional diversity
- genetic diversity
- morphological diversity
- community diversity
- species diversity
- strain diversity

# DEFINITIONS OF DIVERSITY

- alpha diversity
  - richness
  - evenness

# DEFINITIONS OF DIVERSITY

- alpha diversity

- richness
- evenness

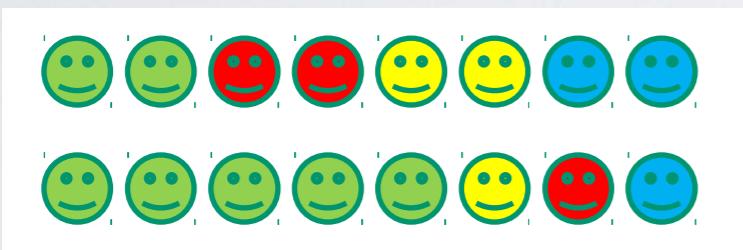


high evenness  
low evenness

# DEFINITIONS OF DIVERSITY

- **alpha diversity**

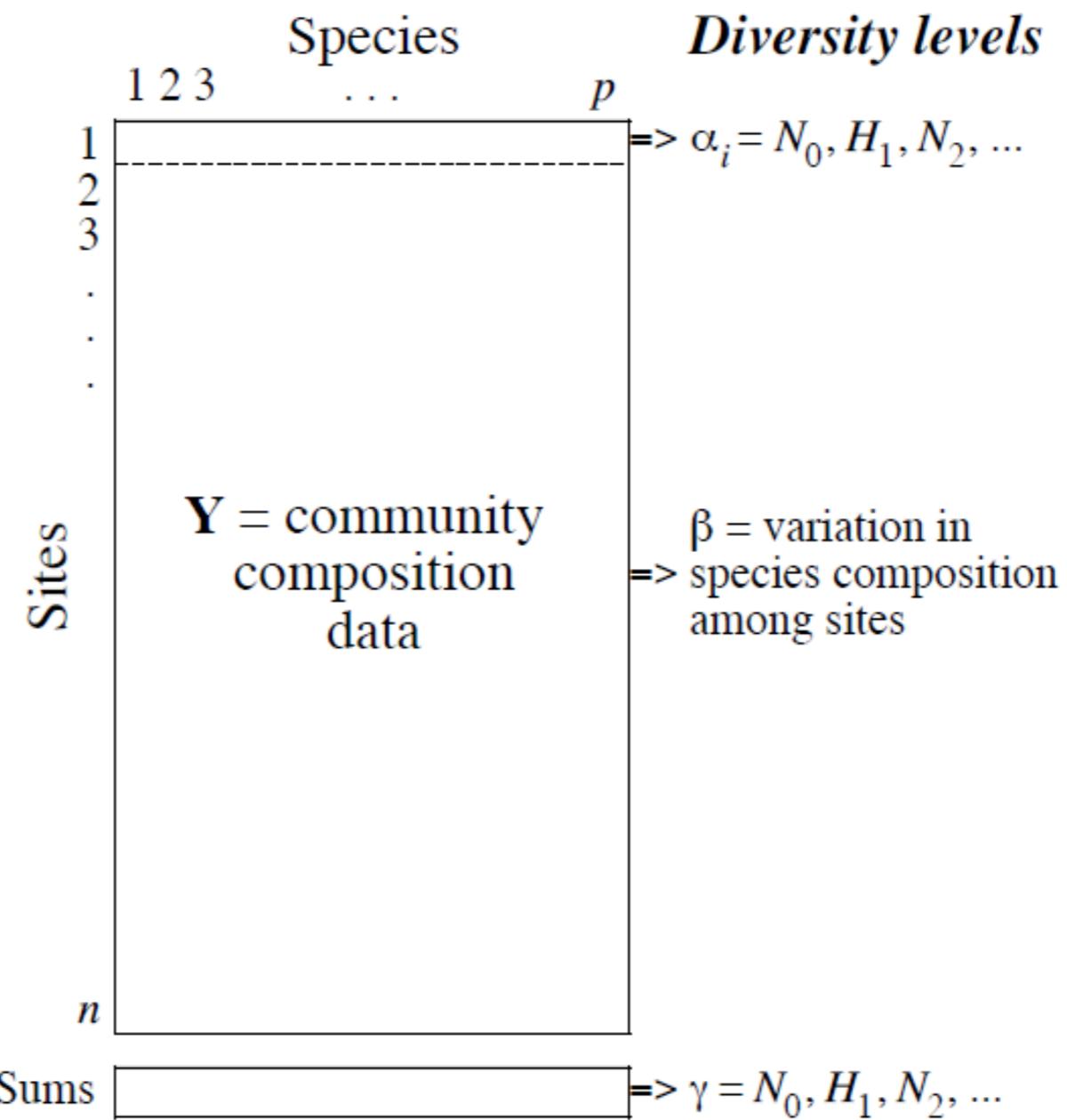
- richness
- evenness



high evenness  
low evenness

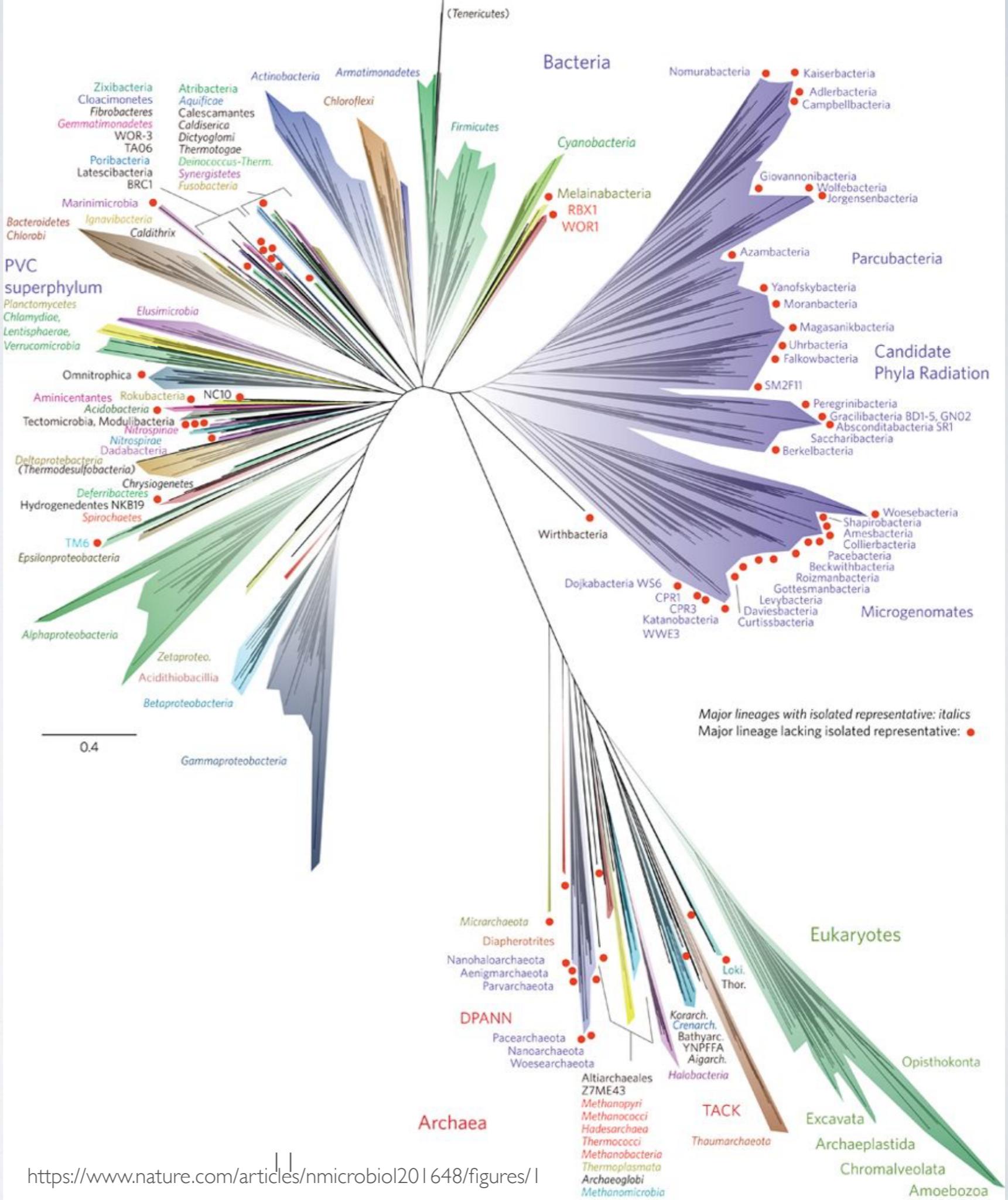
- **beta diversity**

- composition (a list of kinds/species and their distribution)
- phylogenetic diversity

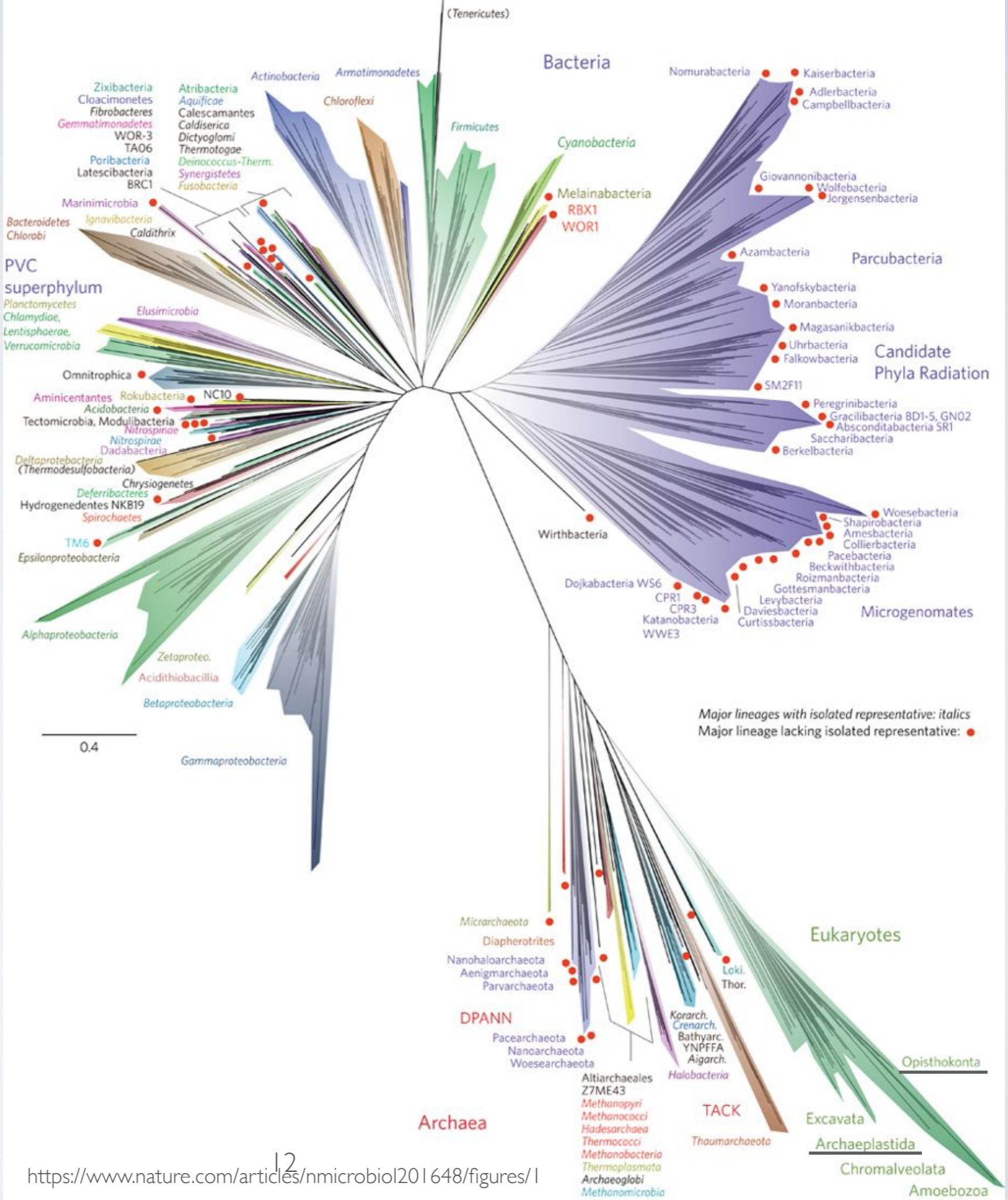


**Figure 6.3** Species diversity indices are computed from the community composition data (matrix  $\mathbf{Y}$ ). Alpha ( $\alpha$ ) diversity indices are computed for individual sites (rows)  $i$ . Gamma diversity ( $\gamma$ ) is computed from the vector of column sums of the data matrix using the same indices as for alpha diversity. Beta ( $\beta$ ) diversity is of a different nature: it is the variation in community composition among sites. It cannot be computed with the usual entropy of diversity number indices.

# TREE OF LIFE



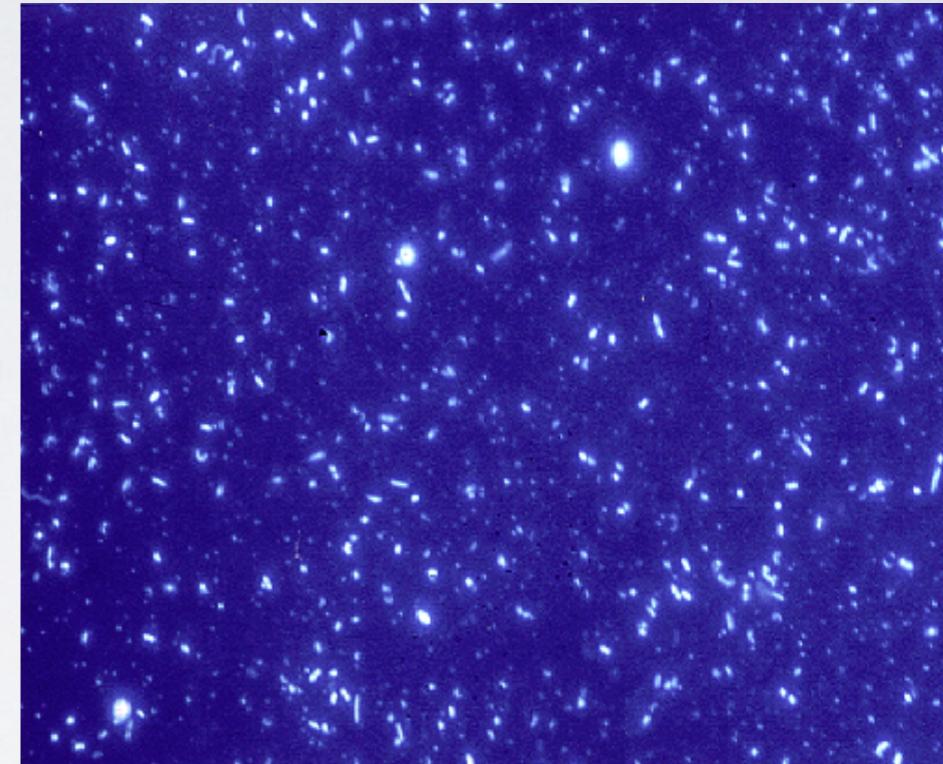
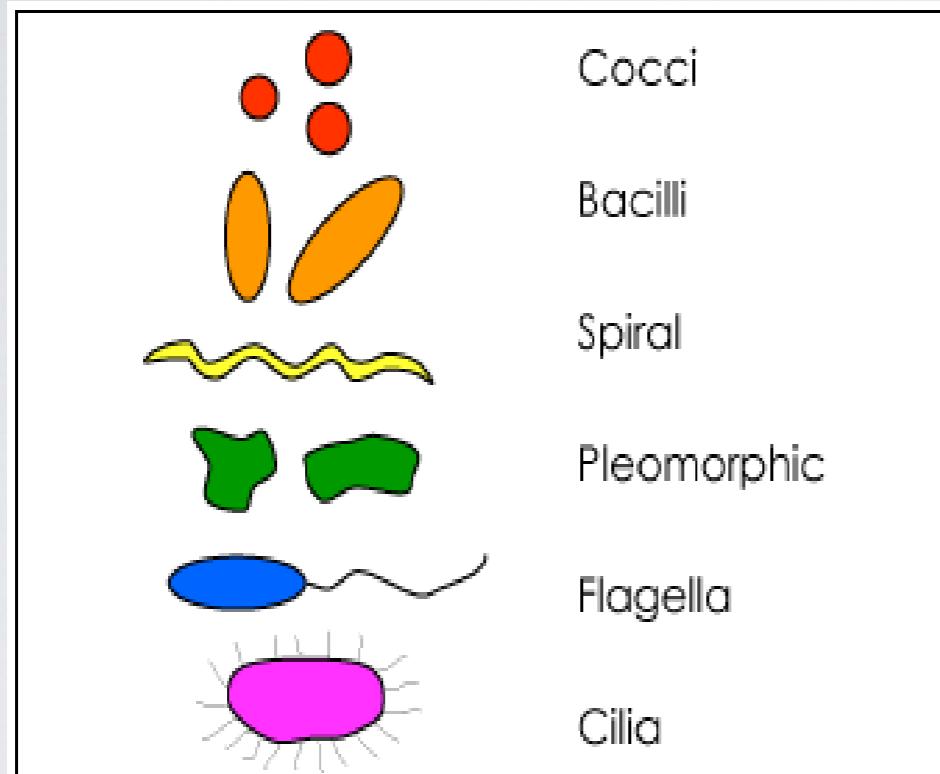
# TREE OF LIFE



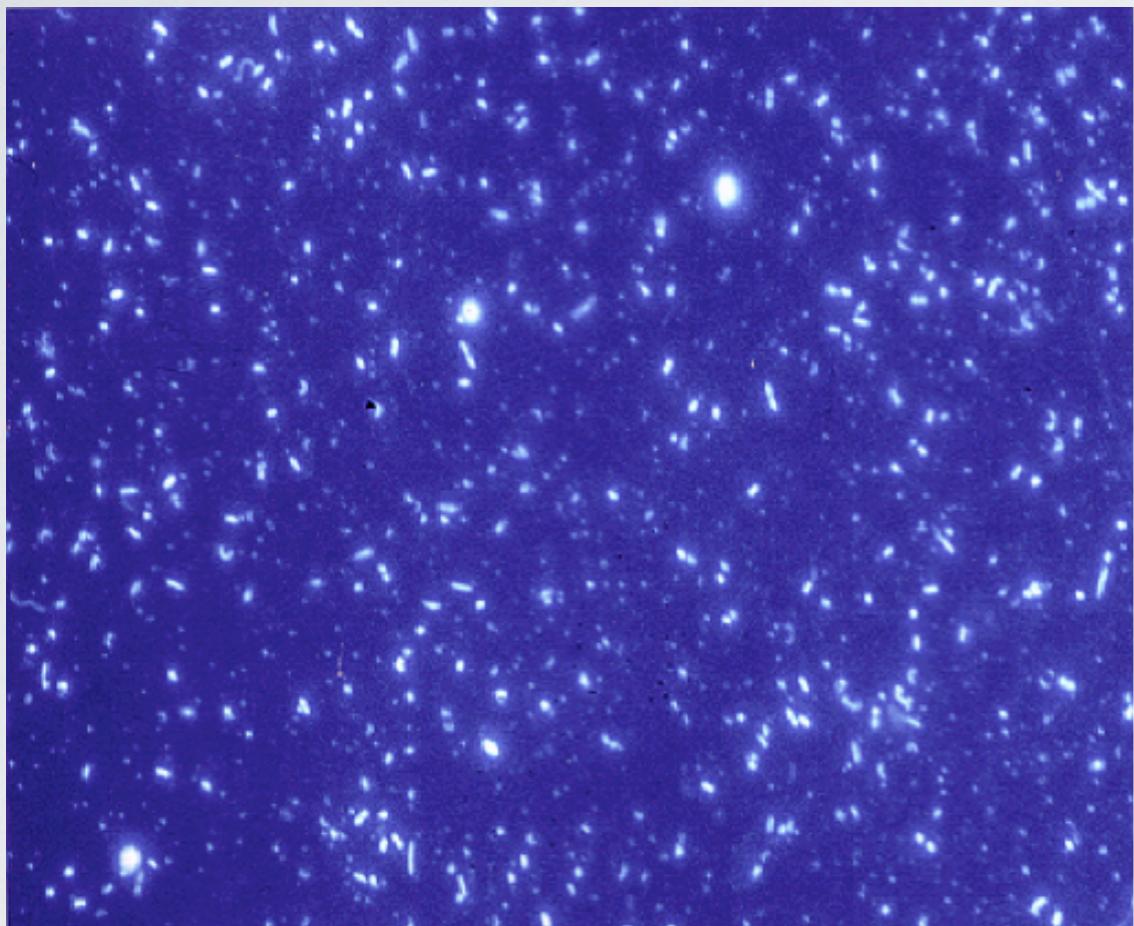
# METHODS



# MORPHOLOGICAL DIVERSITY



# THE GREAT PLATE ANOMALY



$10^6$  cells ml<sup>-1</sup>

< 0.1 %



$10^3$  cells ml<sup>-1</sup>

# CULTURE INDEPENDENT METHODS

- FISH - Fluorescence In Situ Hybridization
- FISH combined with MicroAutoRadiography
- Bulk sequencing of DNA, RNA or protein (meta-'omics')
- Sequencing of single cells
- Etc.

amplicon sequencing

DNA metabarcoding

metatranscriptomics

rDNA phlyotyping

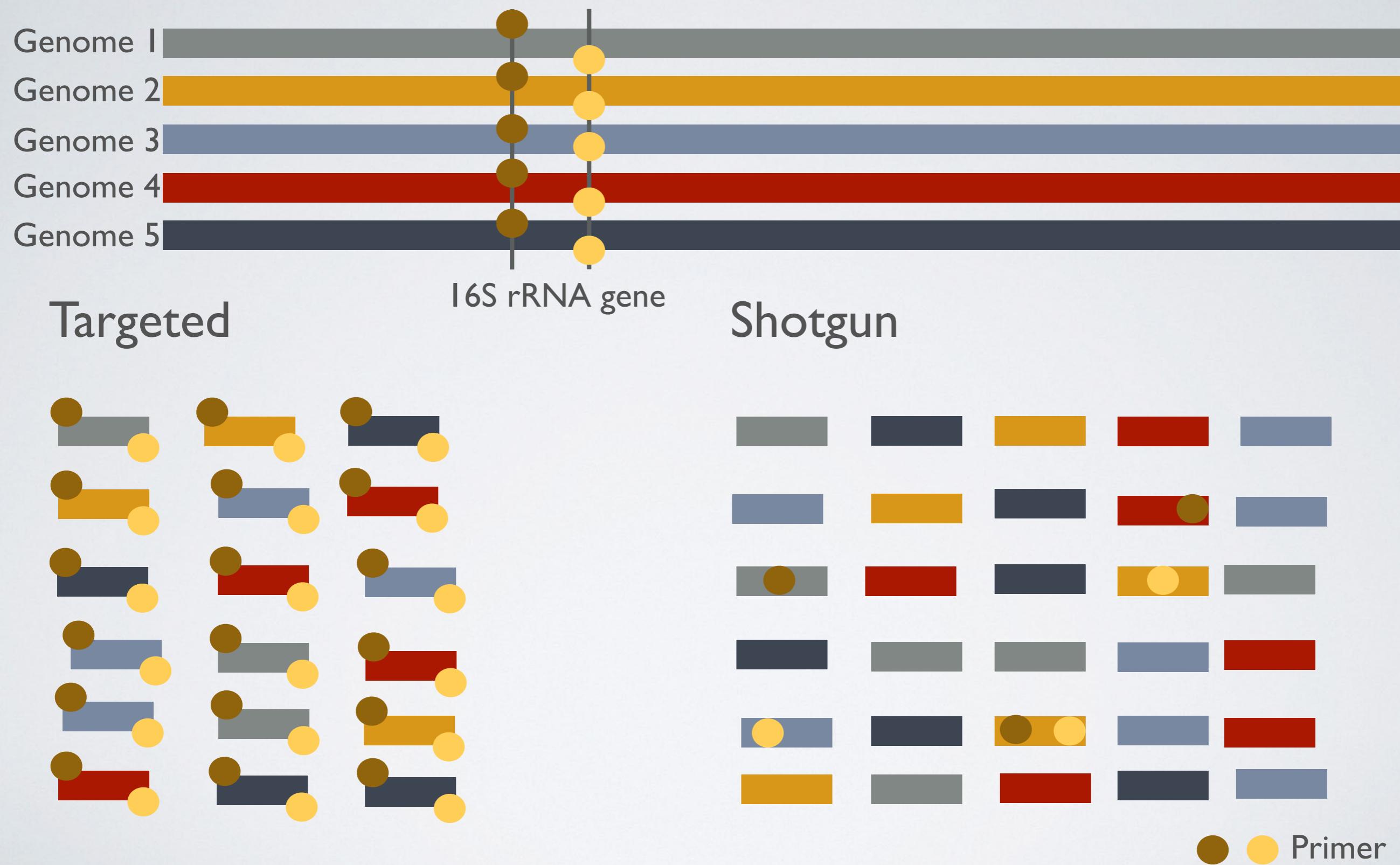
metagenetics

metagenomics

environmental sequencing

Community profiling

# METAGENOMICS



- **Metabarcoding:** Sequence variation in a single locus (e.g. 16S)
- **Metagenomics:** Genome wide sequence variation
- **Metatranscriptomics:** cDNA sequence variation

# TARGETED METAGENOMICS - METABARCODING

## Research questions:

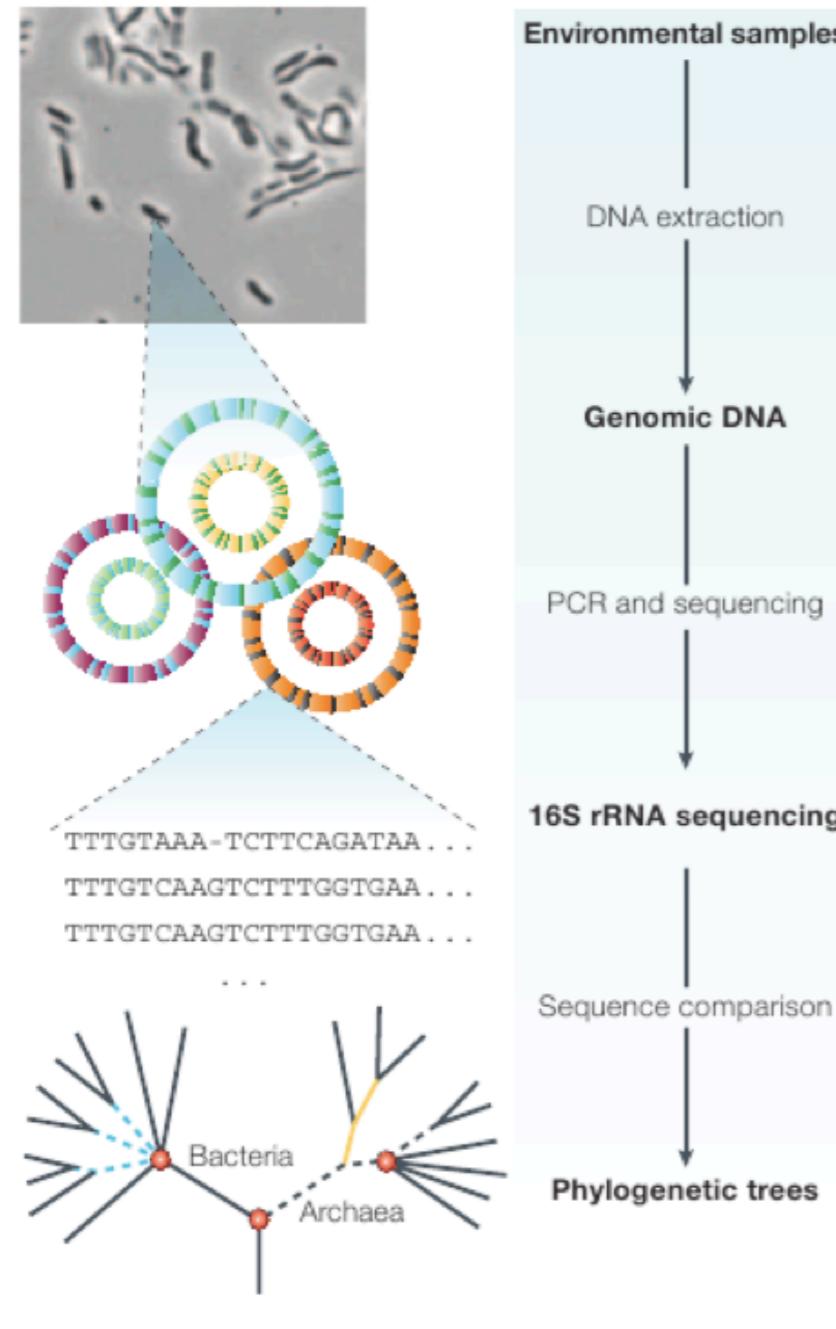
- Who are there?
- How many taxa / operational taxonomic units (alpha/gamma diversity)?
- Compositional differences (beta diversity)?
- Who are common – who are rare?
- Co-occurrence patterns
- Which processes and drivers are shaping the communities?
- Ecosystem processes: What are they doing..?

# TARGETED METAGENOMICS - METABARCODING

## Research field:

- A largely descriptive field of science
- Often functions as a first step looking into largely uncharacterized habitats and study systems
- Often generates hypotheses that can be addressed more in-depth with more taxon specific approaches

# SINGLE GENE INVESTIGATIONS



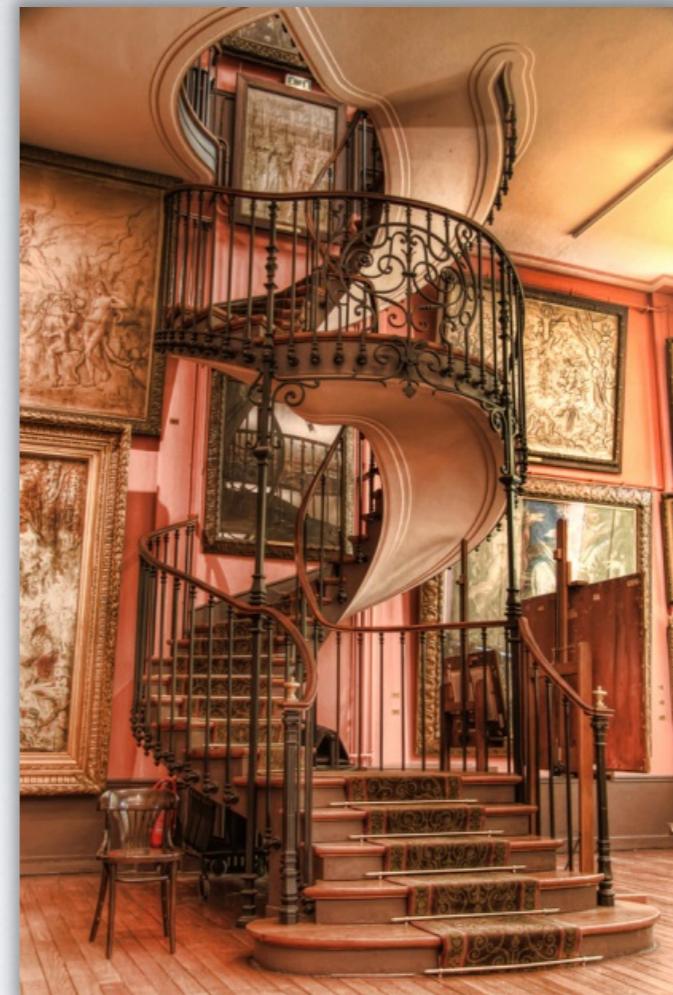
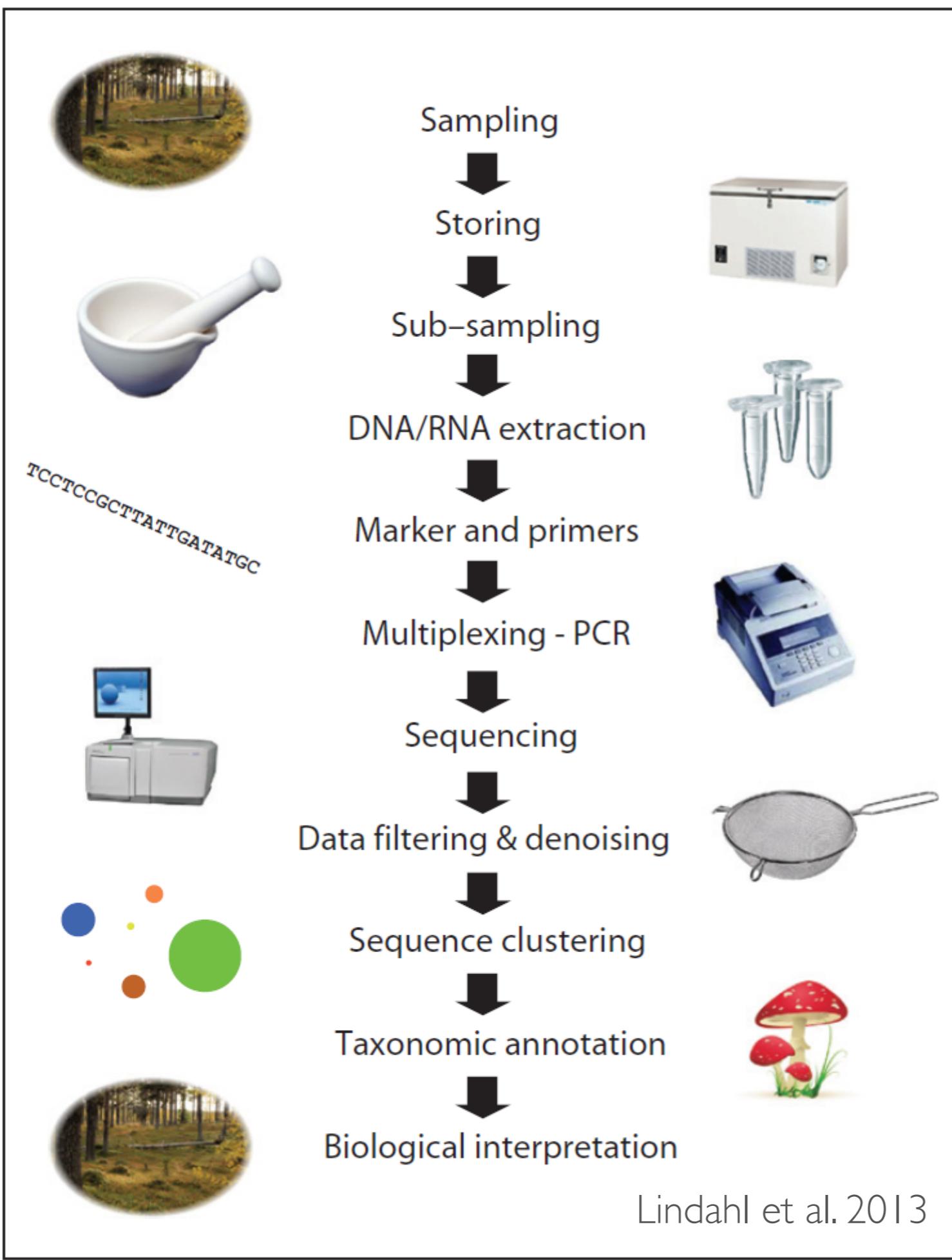
Direct extraction of total DNA from environmental samples

PCR amplification with specific primers

DNA sequencing

Homologies:

- Taxonomic affiliation
- Evolutionary relationships



Many steps



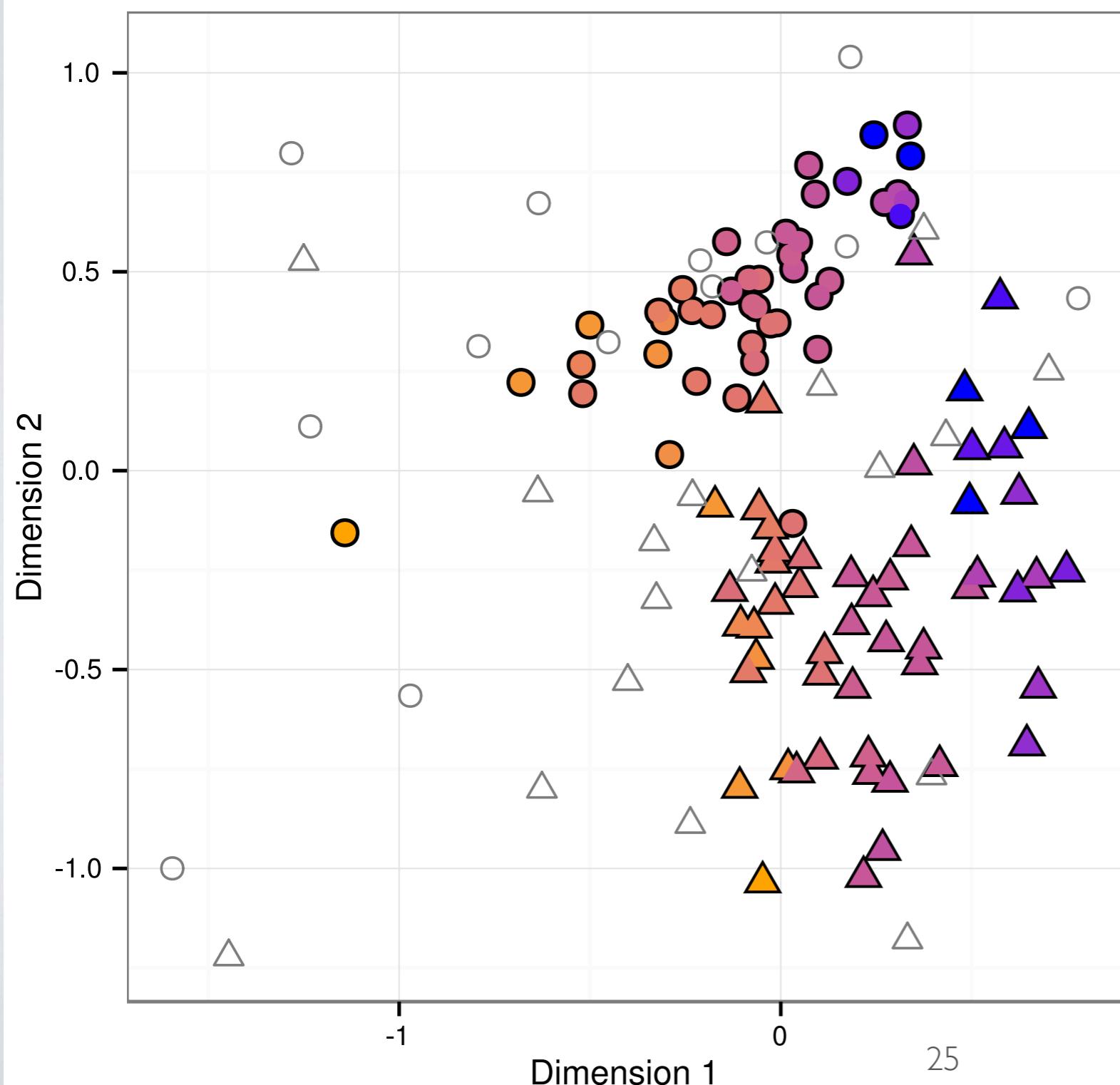
That can go wrong

# SAMPLING/SURVEY DESIGN

## Biological vs. Technical replicates

- Which samples should be analysed multiple times?
- Research question needs to be aligned with survey design and statistical analyses
- Variability (experimental error) of the analysis technique sets limits for replication and what is meaningful / significant data

# Bacterial community composition (Bray-Curtis dissimilarity)



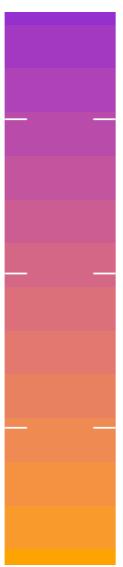
## Size fraction

- free-living: 0.2-3.0  $\mu\text{m}$
- ▲ particle-ass.: >3.0  $\mu\text{m}$

## Tributaries

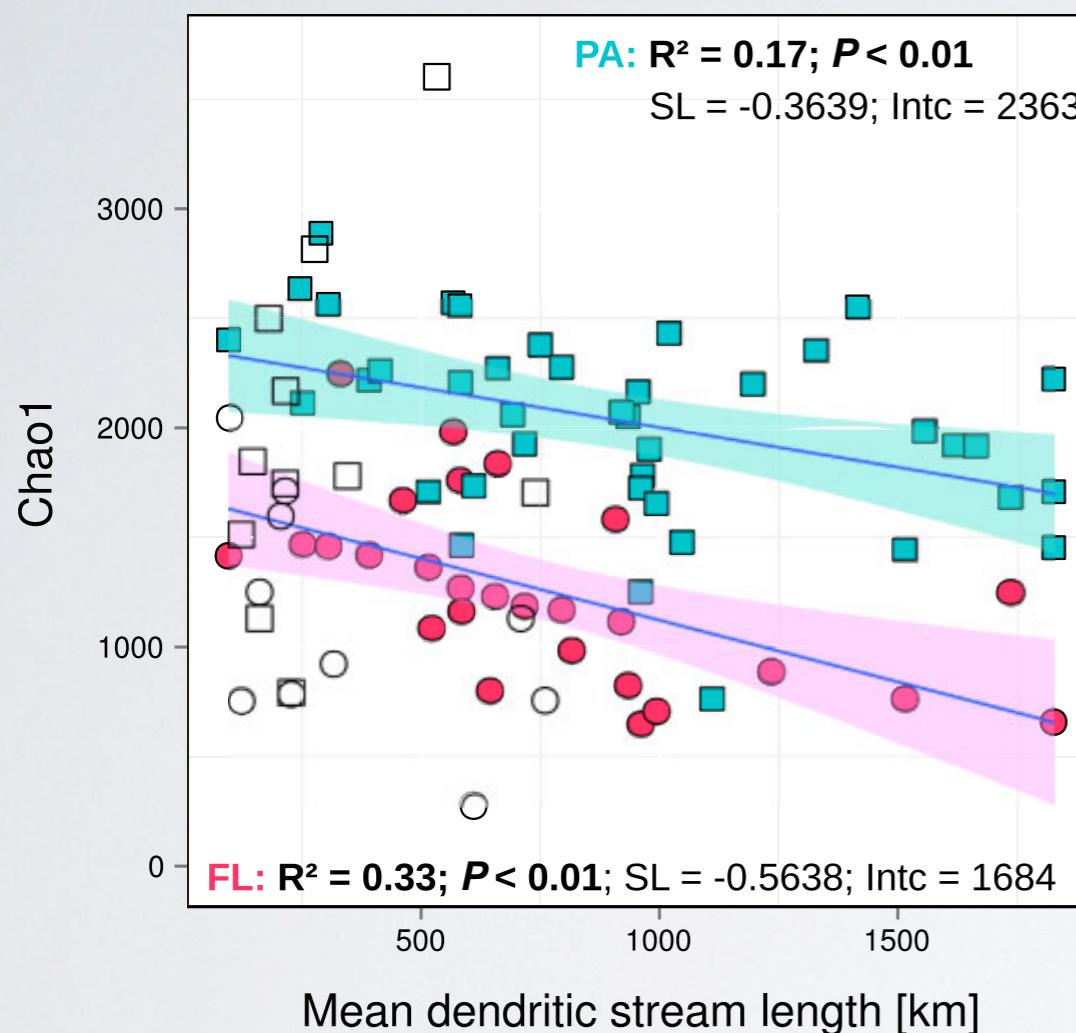
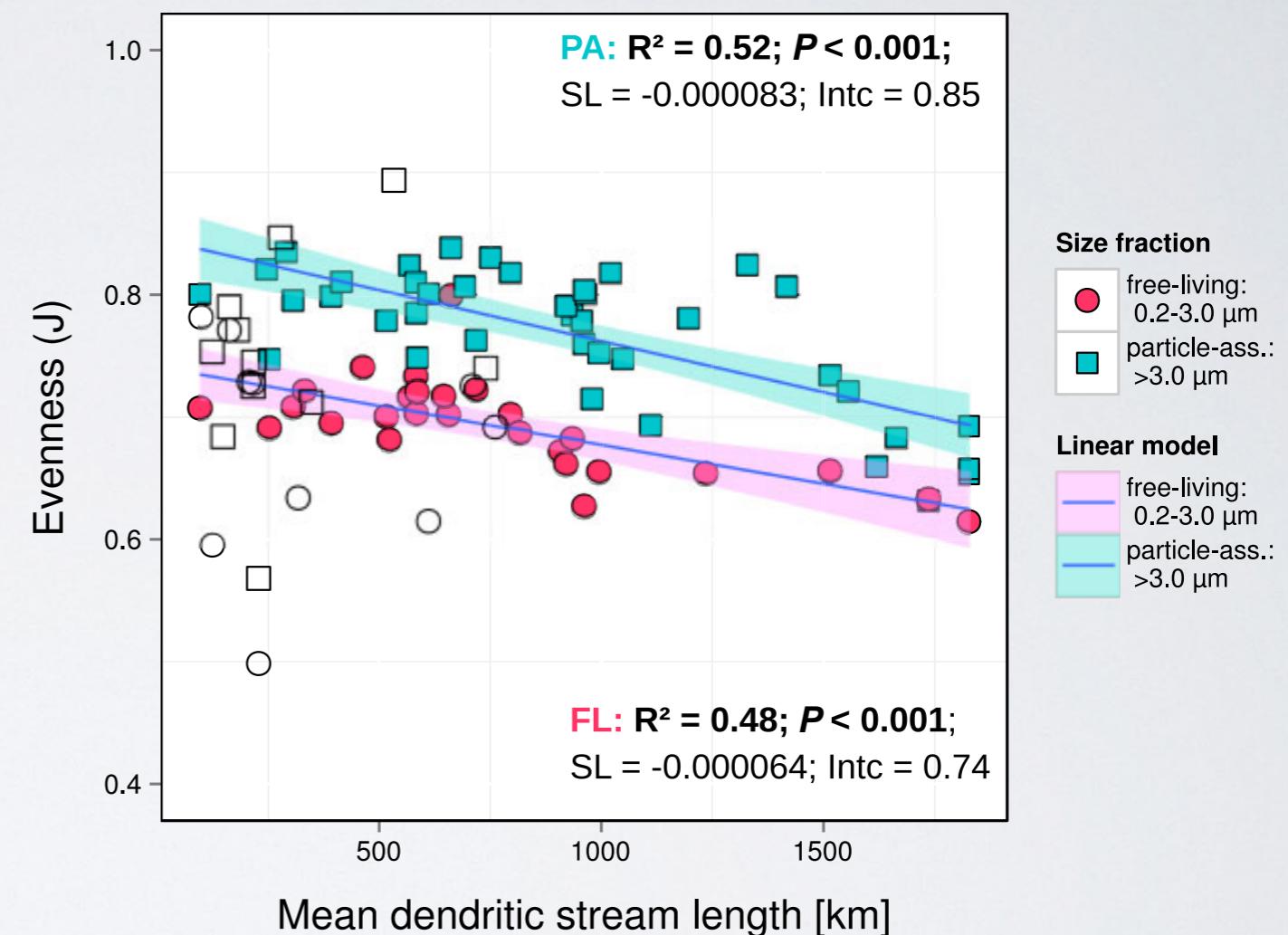


## Mean dendritic stream length [km]

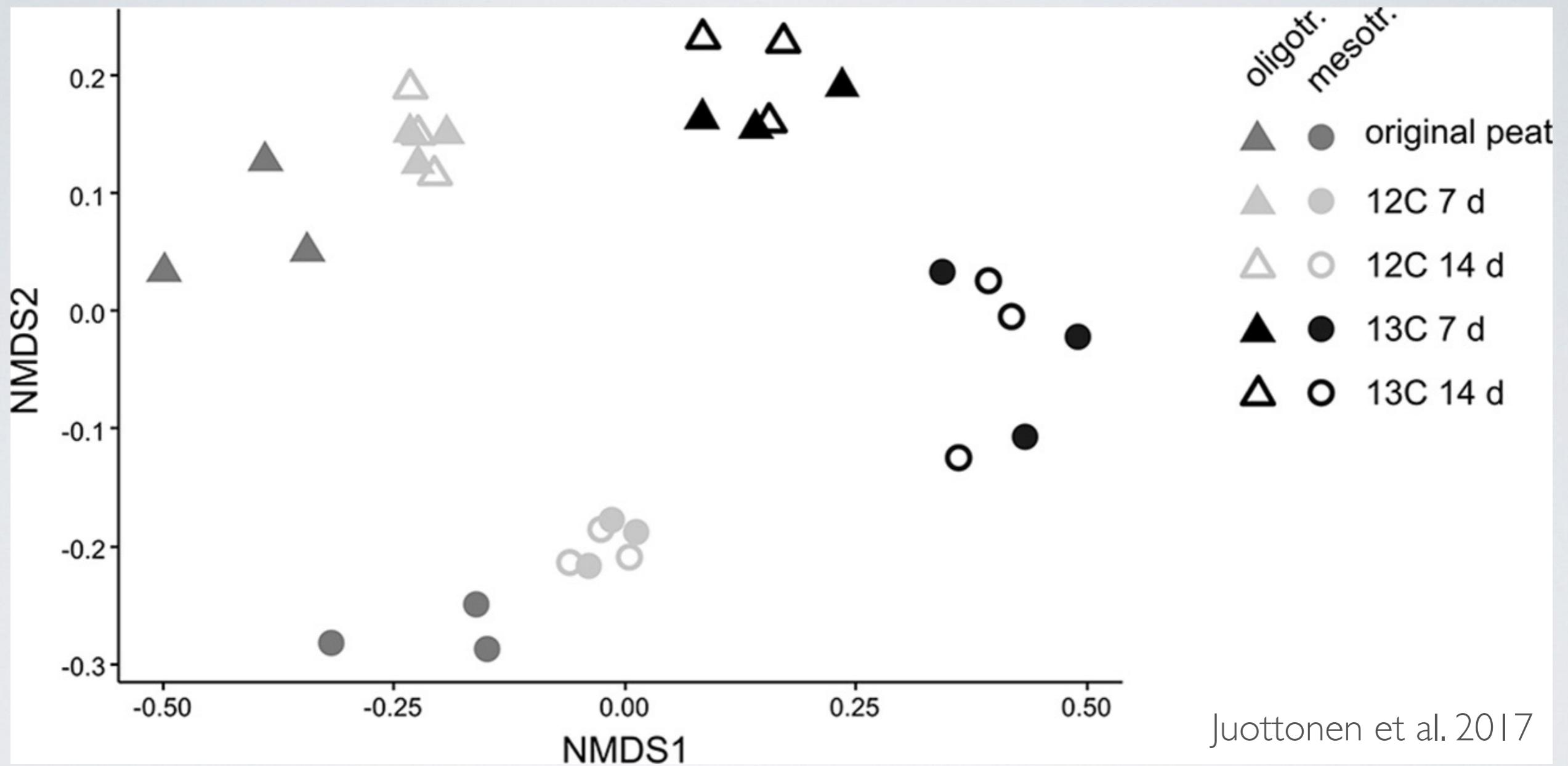


Savio et al. 2015

Lecture I: AeN metabarcoding course

**A****Chao1 richness****B****Pielou's Evenness (J)**

Savio et al. 2015



# DNA EXTRACTION

## An evaluation of commercial DNA extraction kits for the isolation of bacterial spore DNA from soil

S.M. Dineen<sup>1,2</sup>, R. Aranda IV<sup>1,2</sup>, D.L. Anders<sup>3</sup> and J.M. Robertson<sup>2</sup>

<sup>1</sup> Visiting Scientist, Federal Bureau of Investigation Laboratory, Quantico, VA, USA

<sup>2</sup> Counterterrorism and Forensic Science Research Unit, Federal Bureau of Investigation Laboratory, Quantico, VA, USA

<sup>3</sup> Hazardous Materials Science Response Unit, Federal Bureau of Investigation Laboratory, Quantico, VA, USA

Molecular biology, genetics and biotechnology

## Effect of DNA extraction and sample preservation method on rumen bacterial population

Katerina Fliegerova<sup>a,\*</sup>, Ilma Tapiö<sup>b</sup>, Aurelie Bonin<sup>e</sup>, Jakub Mrazek<sup>a</sup>, M. Paolo Bani<sup>c</sup>, Alireza Bayat<sup>d</sup>, Johanna Vilkki<sup>b</sup>, Jan Kopečný<sup>a</sup>, Kevin J. Frederic Boyer<sup>e</sup>, Eric Coissac<sup>e</sup>, Pierre Taberlet<sup>e</sup>, R. John Wallace<sup>f</sup>

OPEN ACCESS Freely available online

## Effect of DNA Extraction Methods and Sampling Techniques on the Apparent Structure of Cow and Sheep Rumen Microbial Communities

Gemma Henderson<sup>1</sup>, Faith Cox<sup>1</sup>, Sandra Kittelmann<sup>1</sup>, Vahideh Heidarian Miri<sup>1</sup>, Mi Samantha J. Noel<sup>1,2</sup>, Garry C. Waghorn<sup>3</sup>, Peter H. Janssen<sup>1</sup>

NOTE / NOTE

## Influence of DNA extraction and PCR amplification on studies of soil fungal communities based on amplicon sequencing

Lihui Xu, Sabine Davnskov, John Larsen, and Mogens Nicolaisen

## The Impact of Different DNA Extraction Kits and Laboratories upon the Assessment of Human Gut Microbiota Composition by 16S rRNA Gene Sequencing

Nicholas A. Kennedy<sup>1</sup>, Alan W. Walker<sup>2</sup>, Susan H. Berry<sup>3</sup>, Sylvia H. Duncan<sup>4</sup>, Freda M. Farquharson<sup>4</sup>, Petra Louis<sup>4</sup>, John M. Thomson<sup>5</sup>, UK IBD Genetics Consortium<sup>1</sup>, Jack Satsangi<sup>1</sup>, Harry J. Flint<sup>4</sup>, Julian Parkhill<sup>2</sup>, Charlie W. Lees<sup>1,6</sup>, Georgina L. Hold<sup>3,\*</sup>

PLOS ONE

RESEARCH LETTER – Environmental Microbiology

## Comparison of four DNA extraction methods for comprehensive assessment of 16S rRNA bacterial diversity in marine biofilms using high-throughput sequencing

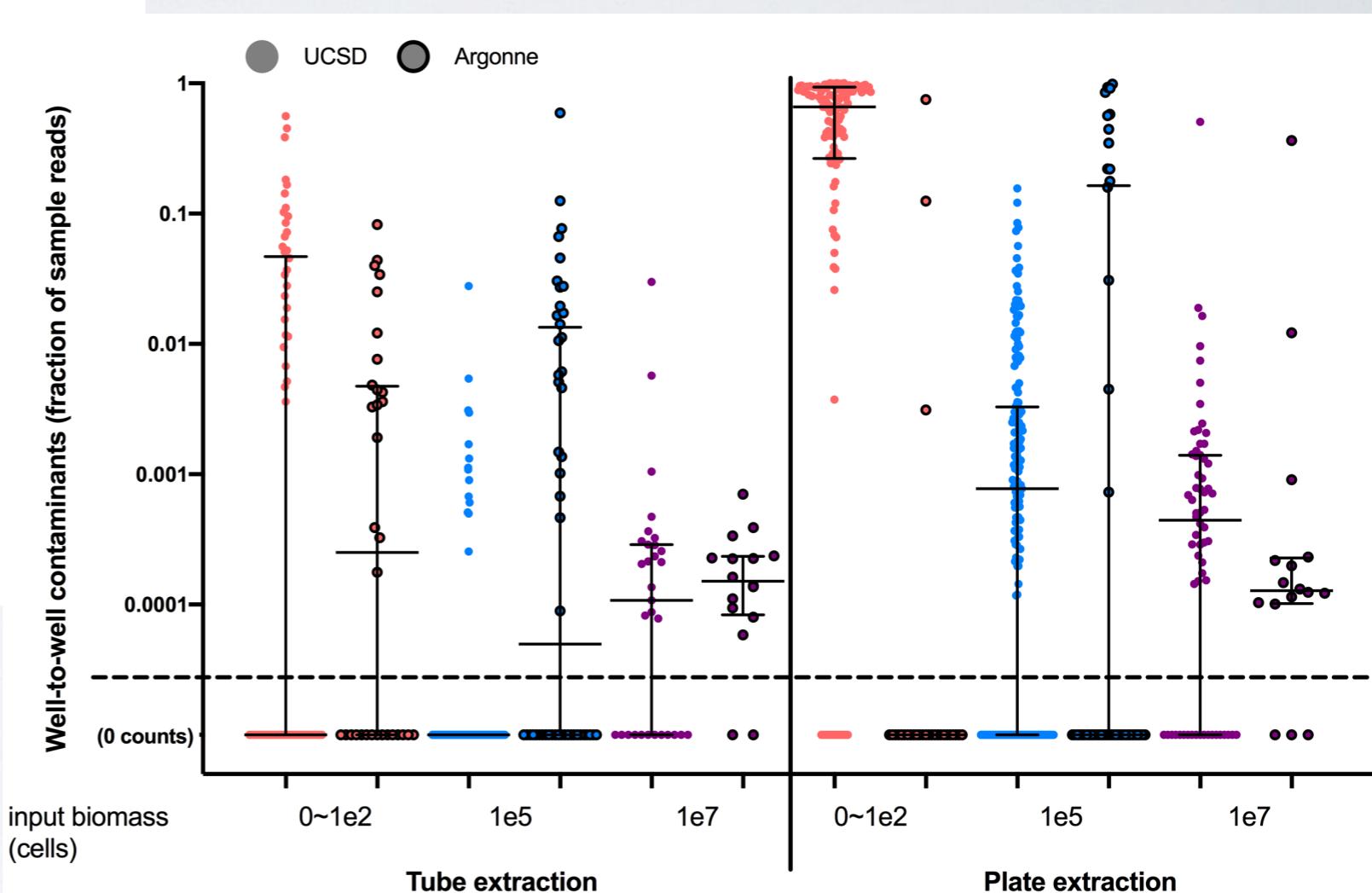
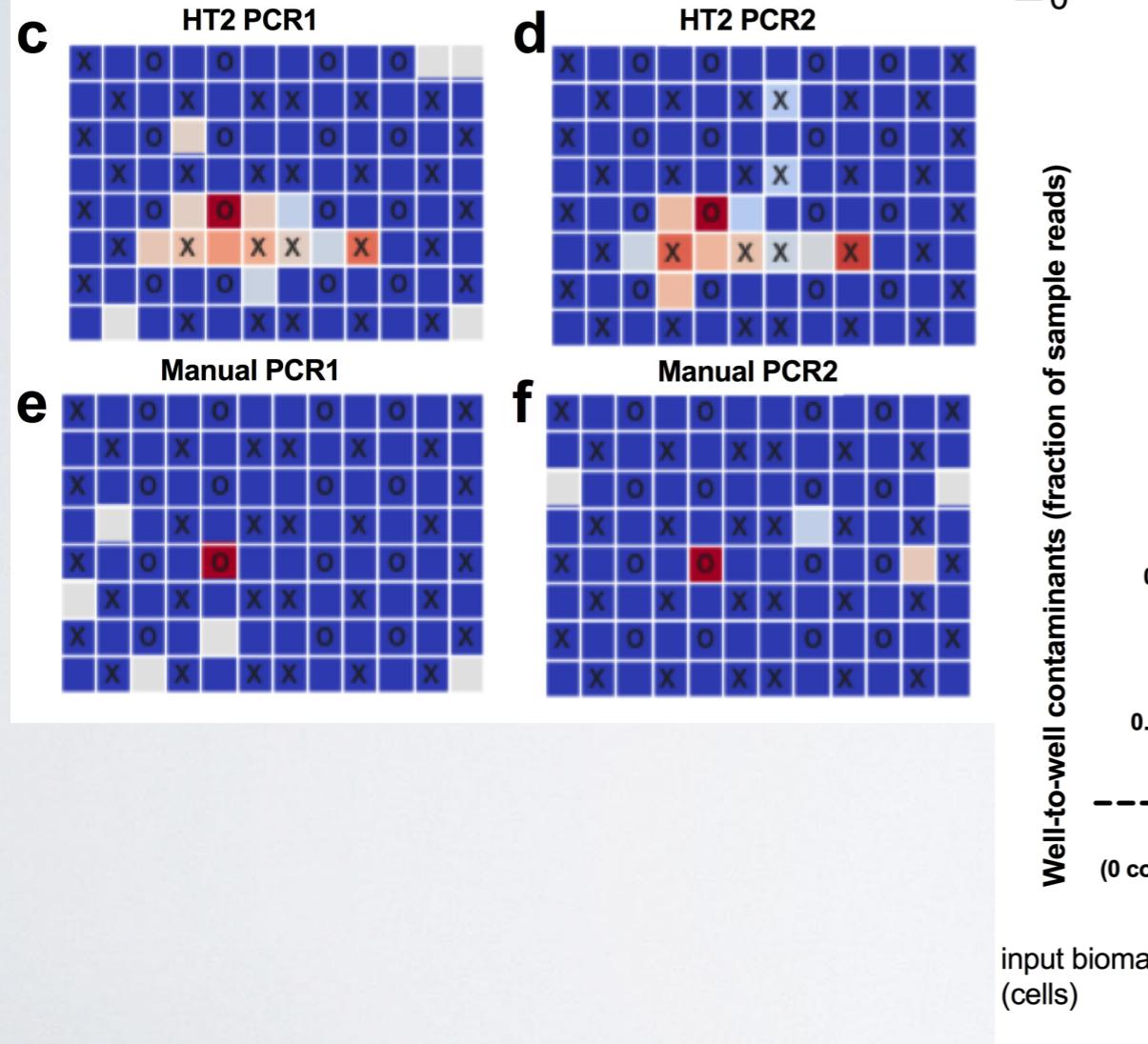
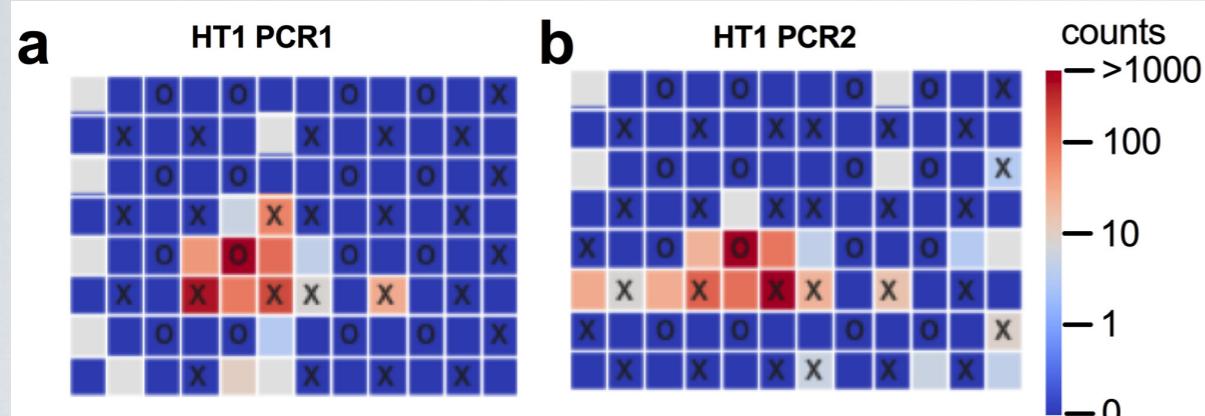
Natàlia Corcoll<sup>1,\*†‡</sup>, Tobias Österlund<sup>2,†</sup>, Lucas Sinclair<sup>3</sup>, Alexander Eiler<sup>3</sup>, Erik Kristiansson<sup>2</sup>, Thomas Backhaus<sup>1</sup> and K. Martin Eriksson<sup>4</sup>

# DNA EXTRACTION

- Extraction protocols should yield high and uniform amounts of DNA, while the concentration of PCR inhibitors is minimised
- Same protocol should be used for all samples!
- If no proper literature or experimental evidence are available on your study system / substrate – conduct a pilot?!
- Extraction negatives are a must.

- DNeasy PowerWater Sterivex Kit
- DNeasy PowerSoil kit
- DNA purification kits

# DNA EXTRACTION



<https://www.biorxiv.org/content/10.1101/577718v1.full.pdf>

# TARGETS AND PRIMERS

- Use primer to target a specific region of the genome
- The “ 16 Svedberg small subunit of prokaryotic ribosome” is a popular marker gene

# RRNA

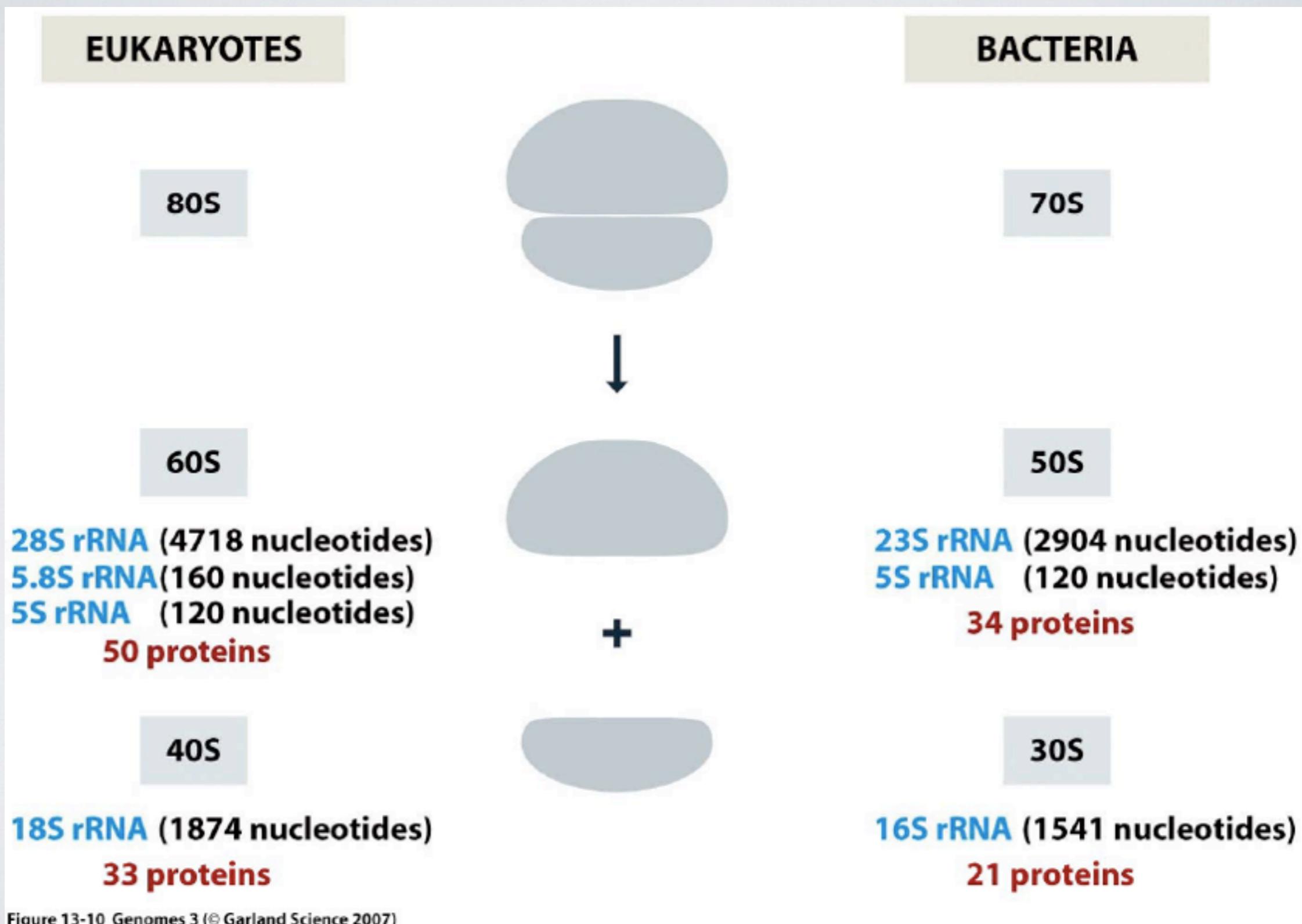


Figure 13-10 Genomes 3 (© Garland Science 2007)

# MARKERS IN METABARCODING

Standard markers:

- 18S: Eukaryotes
- 16S: Bacteria/archaea
- ITS: Fungi
- COI: Metazoa
- RbcL: Plants
- trnL: Plants
- ITS: Plants

# MARKERS IN METABARCODING

## The ideal marker:

- Has primer sites that are shared by all target organisms and easy to amplify
- Is of appropriate length for efficient amplification and sequencing
- Amplifies fragments of similar length across targets
- Lacks intragenomic variation (i.e. no paralogs)
- Facilitates alignment (i.e. variable and conserved sites)
- Has high interspecific variation
- Has low intraspecific variation

**No known markers meet all these requirements!**

# VARIABLE/CONSERVED SITES

RESET

Primers Info

Primers Forward

- 8F
- CC[F]
- 357F
- 16S.100.F16
- 1237F

Primers Reverse

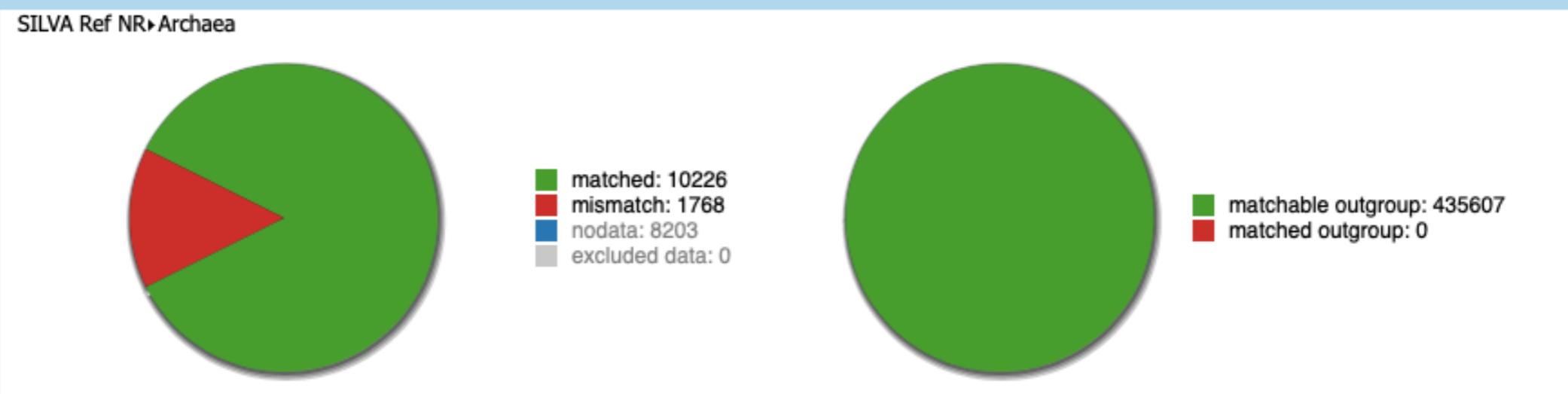
- 519R
- CD[R]
- 907R
- 1100R
- 1391R

# SPECIFICITY OF PRIMERS

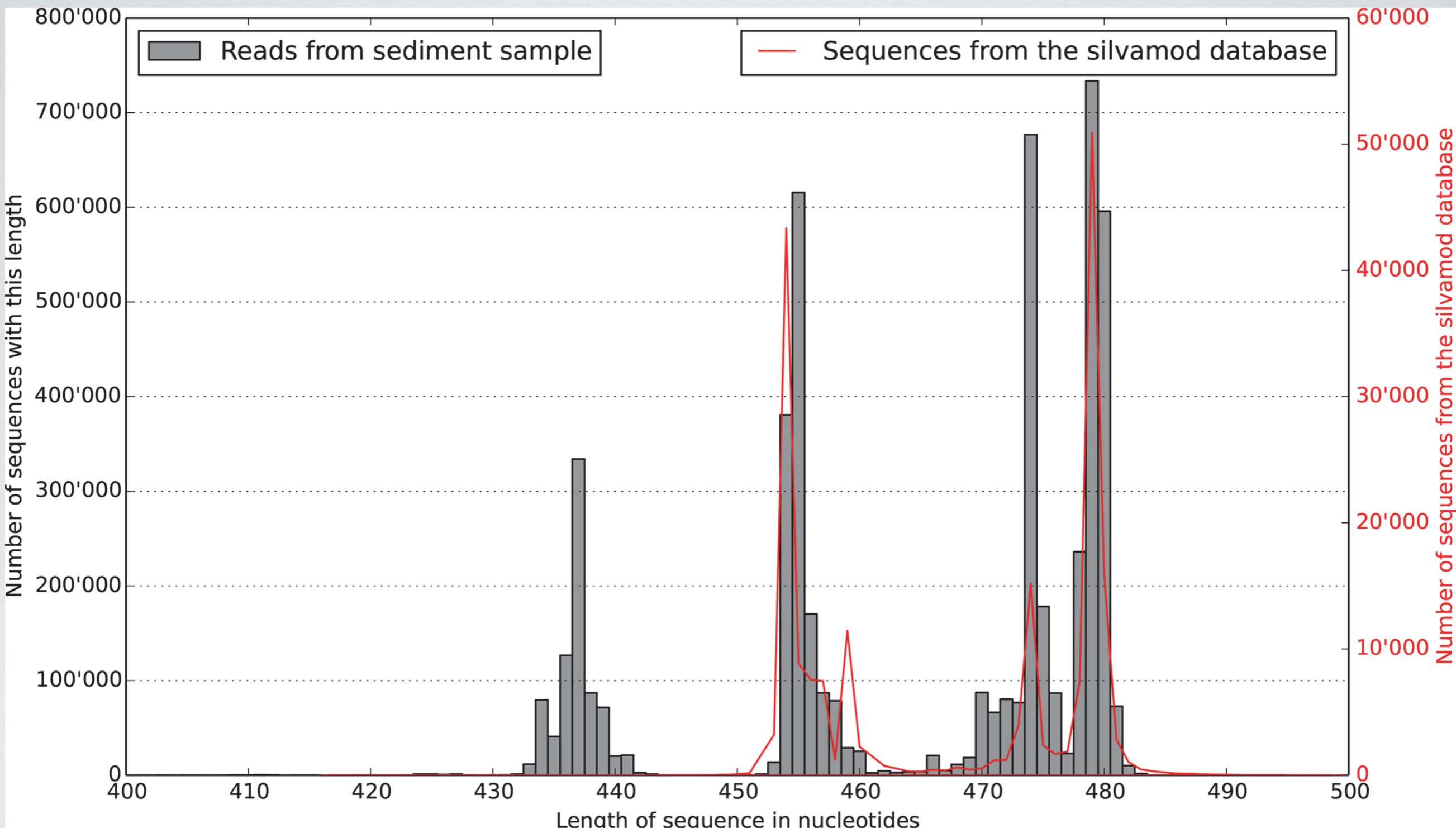
SILVA Ref NR (0.11%) (2.3%)	Archaea (51%) (85.3%)
▼ (4)	▼ (15) ● (1)
<b>Archaea</b> (51%) (85.3%)	<b>Aenigmarchaeota</b> (19%) (40.5%)
Bacteria	<b>Altarchaeota</b> (24%) (40.0%)
Eukaryota	<b>Asgardarchaeota</b> (4.5%) (7.1%)
Unclassified	<b>Crenarchaeota</b> (37%) (89.2%)
	<b>Euryarchaeota</b> (77%) (88.5%)
	<b>Hadarchaeota</b> (60%) (91.9%)
	<b>Halobacterota</b> (69%) (90.6%)
	<b>Hydrothermarchaeota</b> (43%) (97.9%)
	<b>Iainarchaeota</b> (30%) (46.7%)
	<b>Korarchaeota</b> (71%) (78.6%)
	<b>Micrarchaeota</b> (24%) (55.3%)
	<b>Nanoarchaeota</b> (45%) (63.4%)
	<b>Nanohaloarchaeota</b> (82%) (84.8%)
	<b>Thermoplasmatota</b> (50%) (91.2%)
	<b>uncultured</b> (63%) (100.0%)

<https://www.arb-silva.de/search/testprime/>

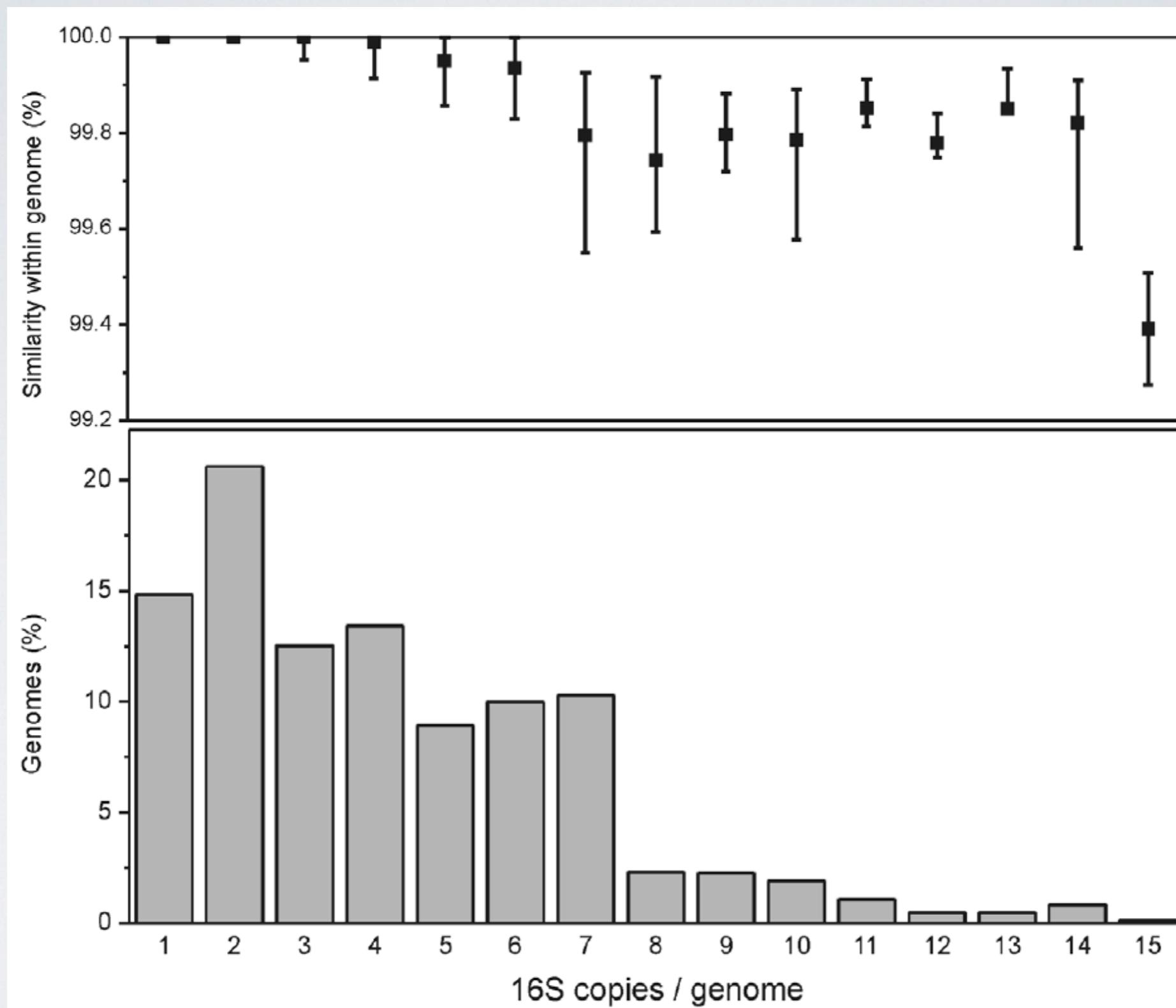
probeForward	probeReverse	db	databasetype	matchMismatches	mismatchRegion
CCCTAYGGGGYGCASCAG	GGCCATGCACYWCYTCTC	ssu-138	nr	1	0



# LENGTH VARIATION OF AMPLICONS

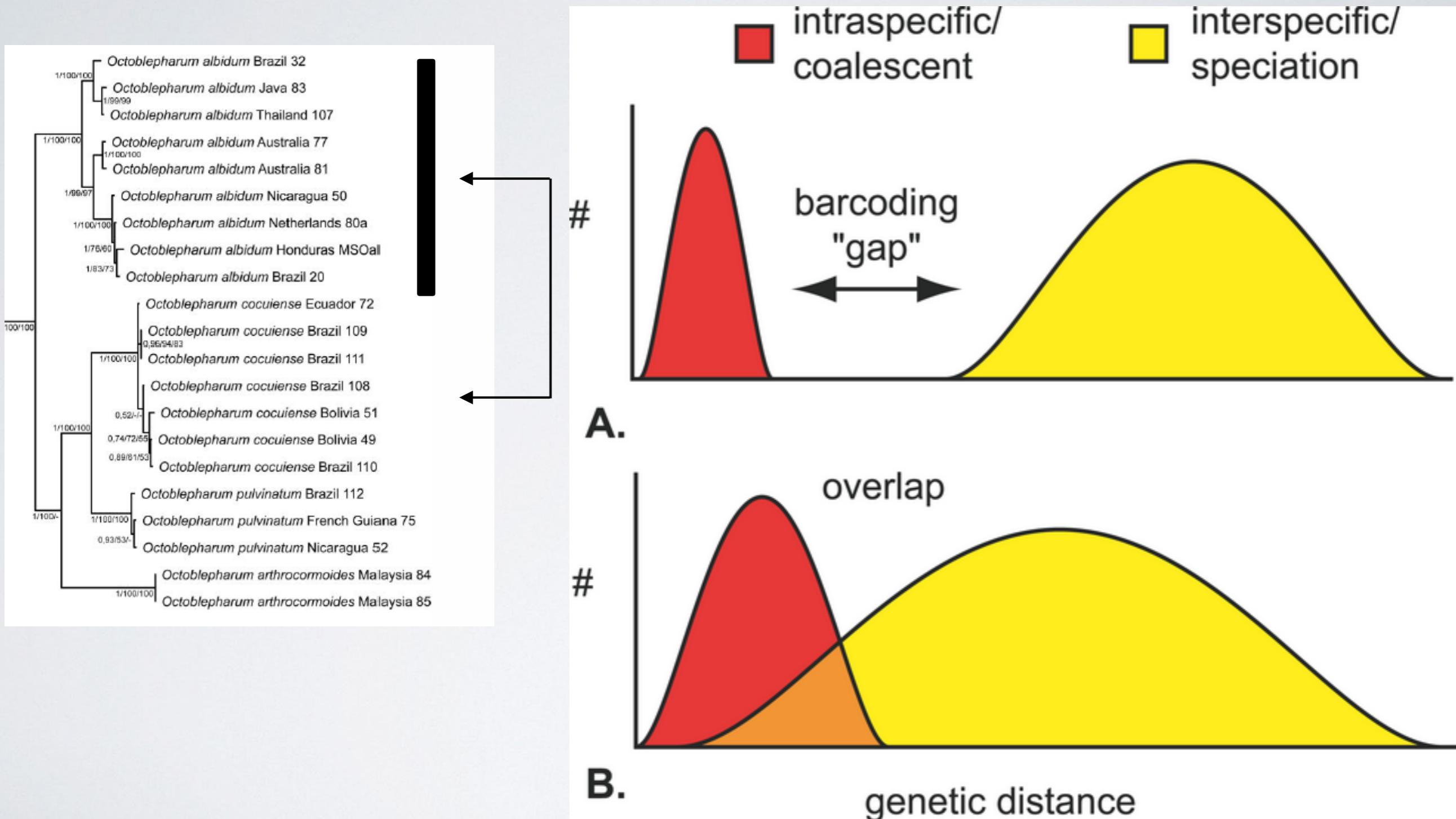


# (INTRA)GENOMIC VARIABILITY IN 16S



Vetrovsky and Baldrian 2013

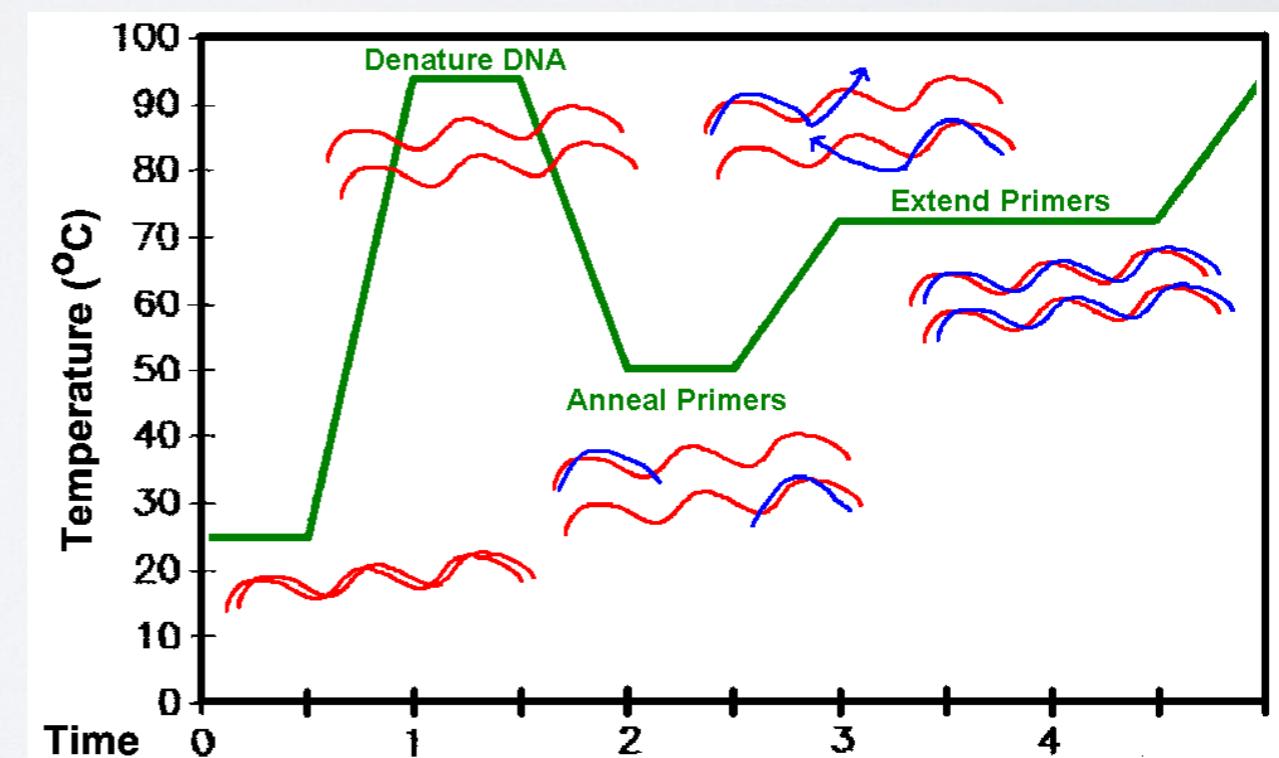
# THE BARCODING GAP



# POLYMERASE CHAIN REACTION - PCR

Different relevant factors during PCR:

- Which polymerase enzyme (proofreading, hot-start)?
- Which RAMP speed?
- How many cycles?
- Which annealing temperature?
- How many PCR replicates followed by pooling?
- PCR negatives!

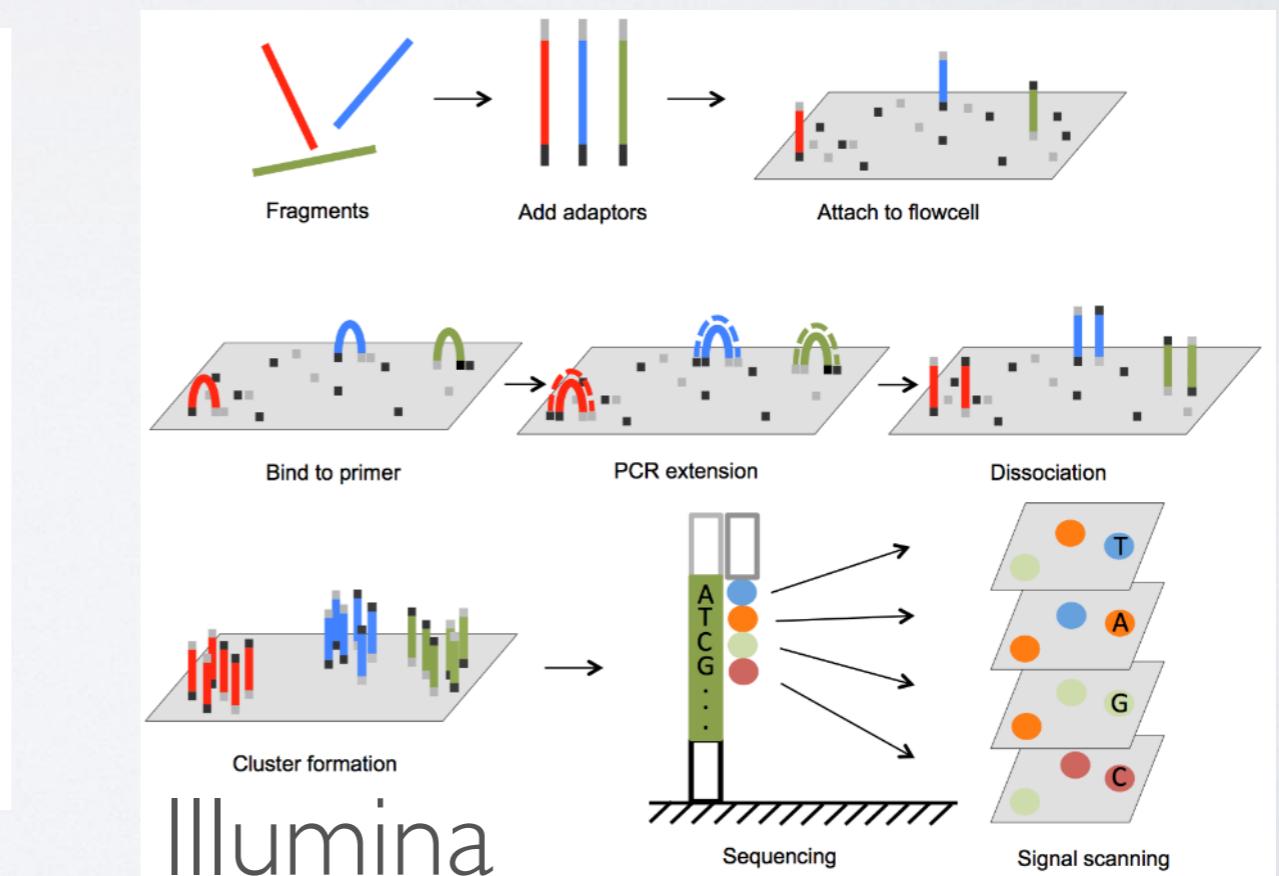
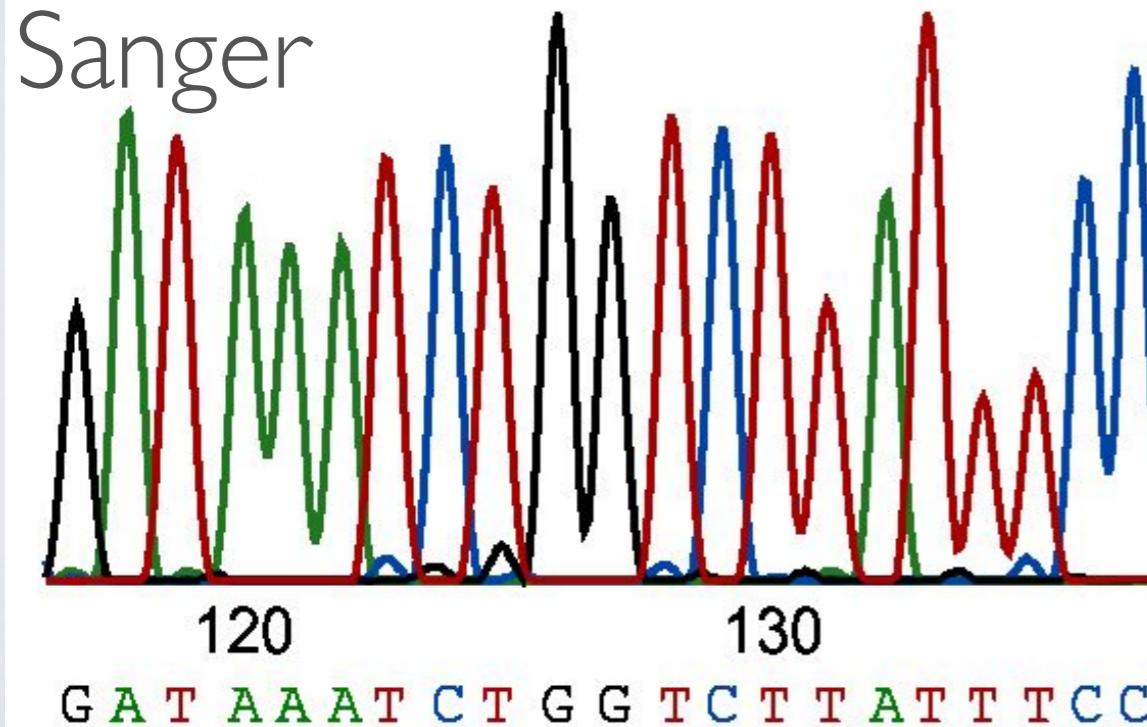


From: H Kauserud

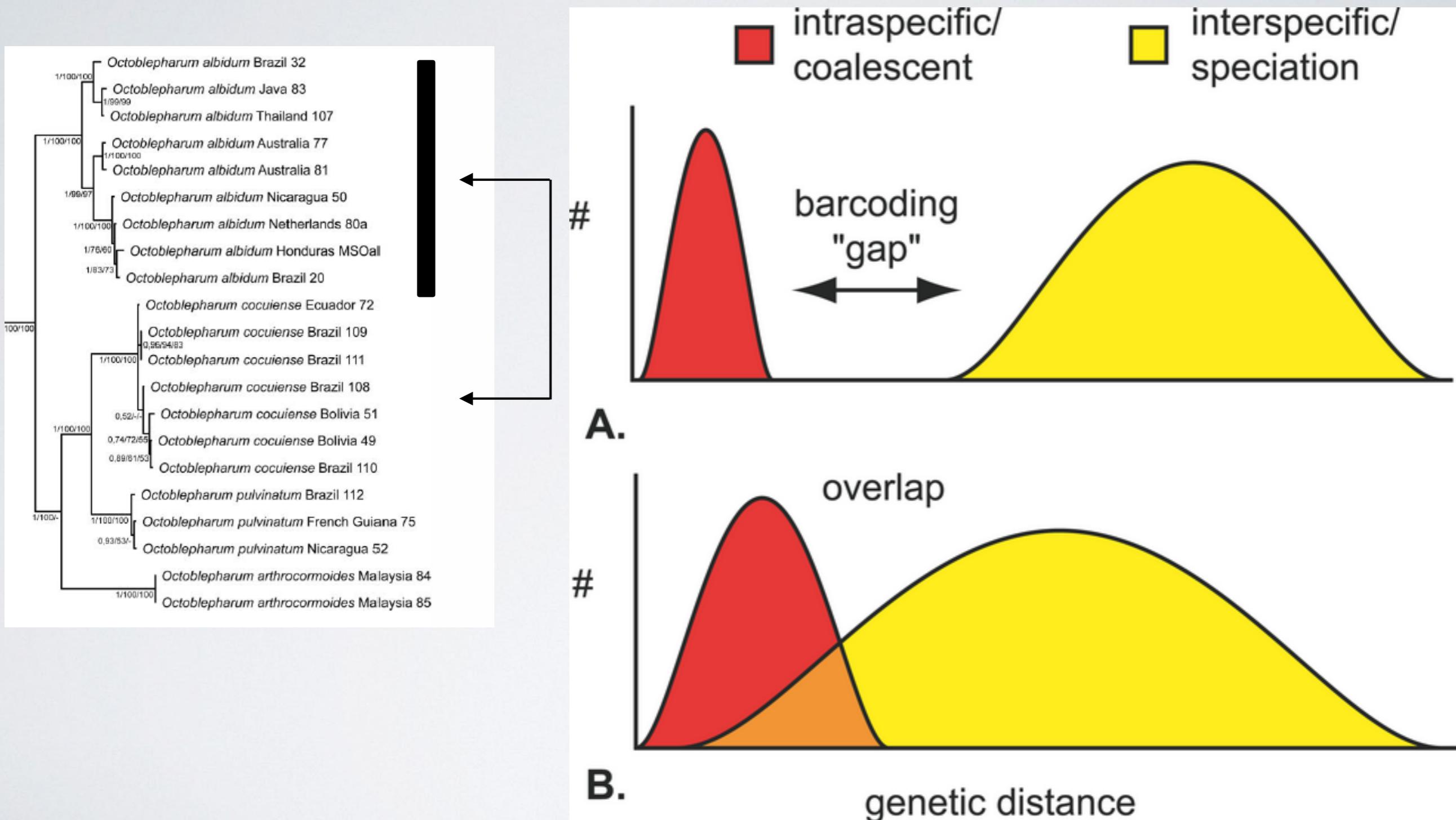
# PCR INDUCED ERRORS

PCR mutations: polymerase enzyme introduces an erroneous nucleotide

- Dependent on the technology whether these becomes «visible»
- In classic (direct) Sanger sequencing such errors become «diluted»
- In methods where your final sequence is derived from a single PCR-amplified DNA template they become visible and must be corrected for!



# THE BARCODING GAP



# PCR INDUCED ERRORS

## Chimeric sequences:

- The level of chimeric sequences depends on how variable the marker is!
- The level of chimeric sequences can be reduced with certain PCR settings

## Tips:

- The duration of the extension step should be substantially increased over the standard recommended/used times
- Low initial amount of DNA template should be used (approximately  $2 \times 10^6$  molecules)
- Two-round amplification and low number of PCR cycles (20 + 12) should be used
- Bioinformatic tools can remove chimeric sequences

<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12864-019-5847-2>

# TAG SWITCHING (JUMPING, BLEEDING, LEAKING, ETC.)

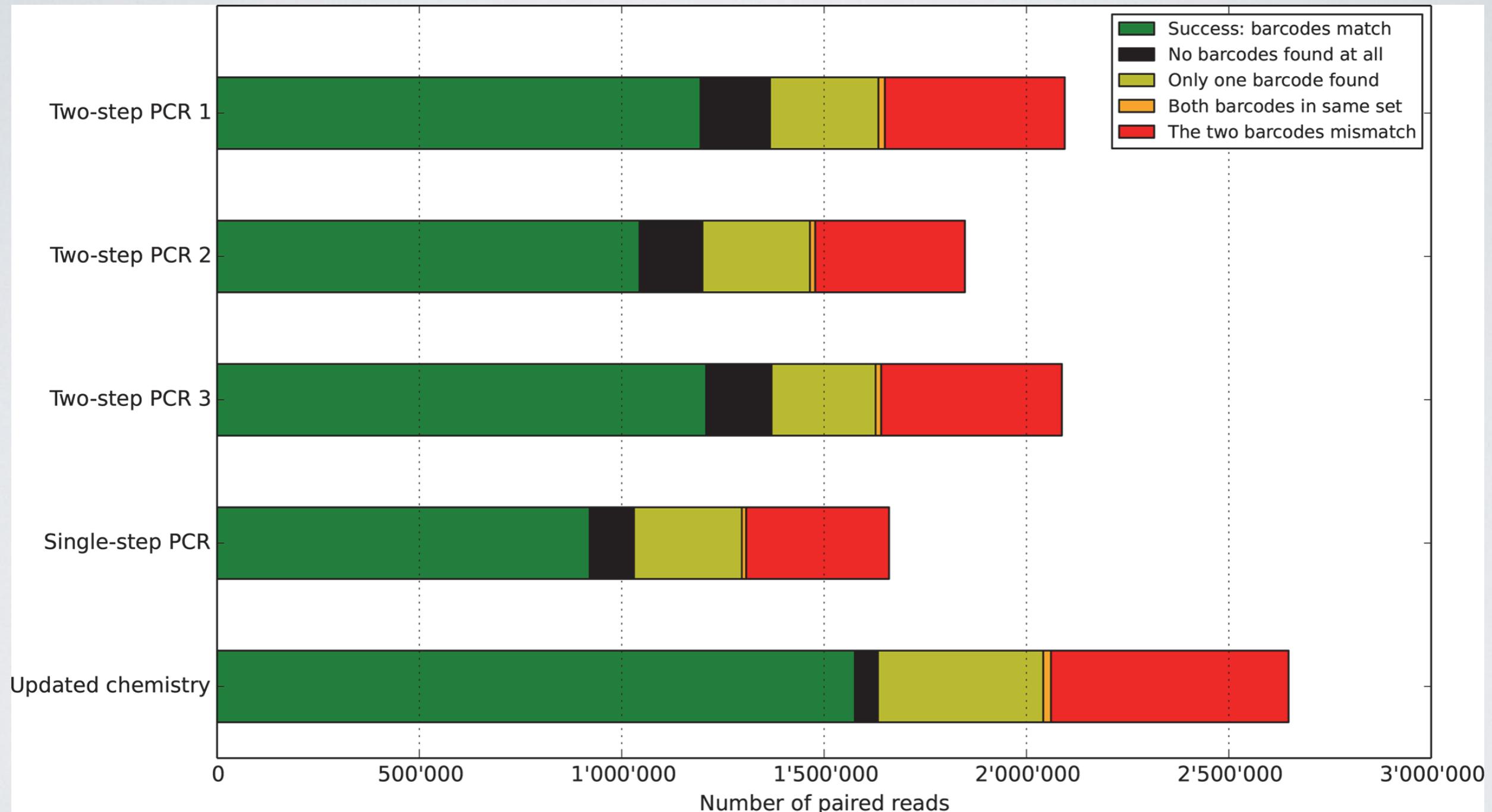


	Samples						
	1	2	3	4	5	6	7
OTU1	0	0	0	0	0	0	0
OTU2	2	0	10000	0	0	5	0
OTU3	0	0	0	0	0	0	0
OTU4	0	0	0	0	0	0	0
OTU5	0	0	0	0	0	0	0
OTU6	0	500	0	0	0	4	0
OTU7	0	0	0	0	0	0	0
OTU8	0	0	0	0	0	0	0
OTU9	0	0	23	0	0	30000	0
OTU10	0	0	0	0	0	0	0

Can lead to  
numerous false  
positives!



# TAG SWITCHING



Sinclair et al. 2015

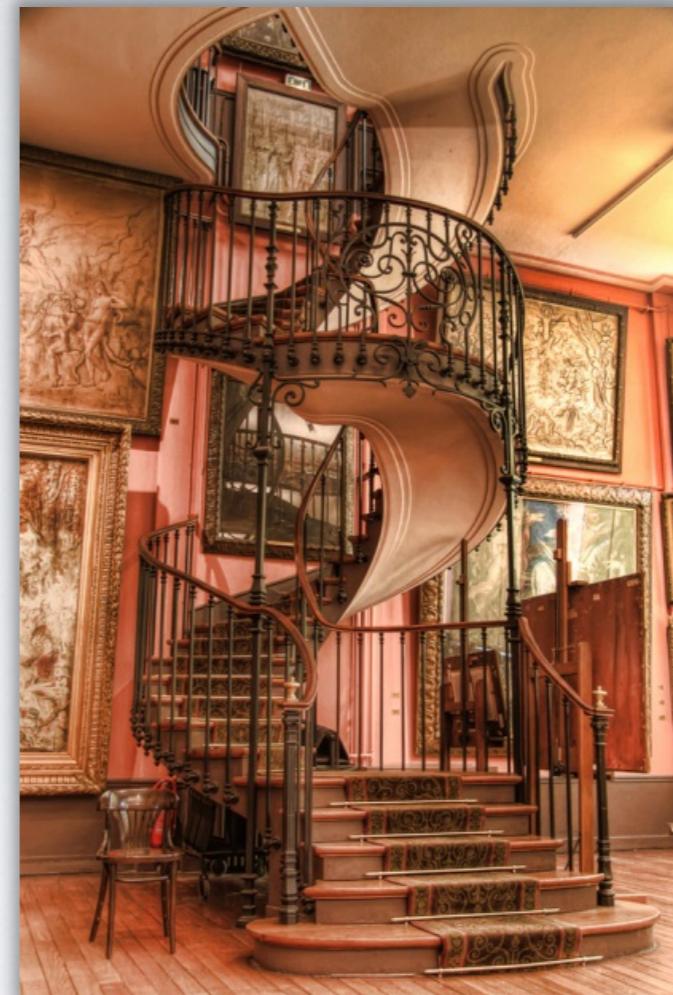
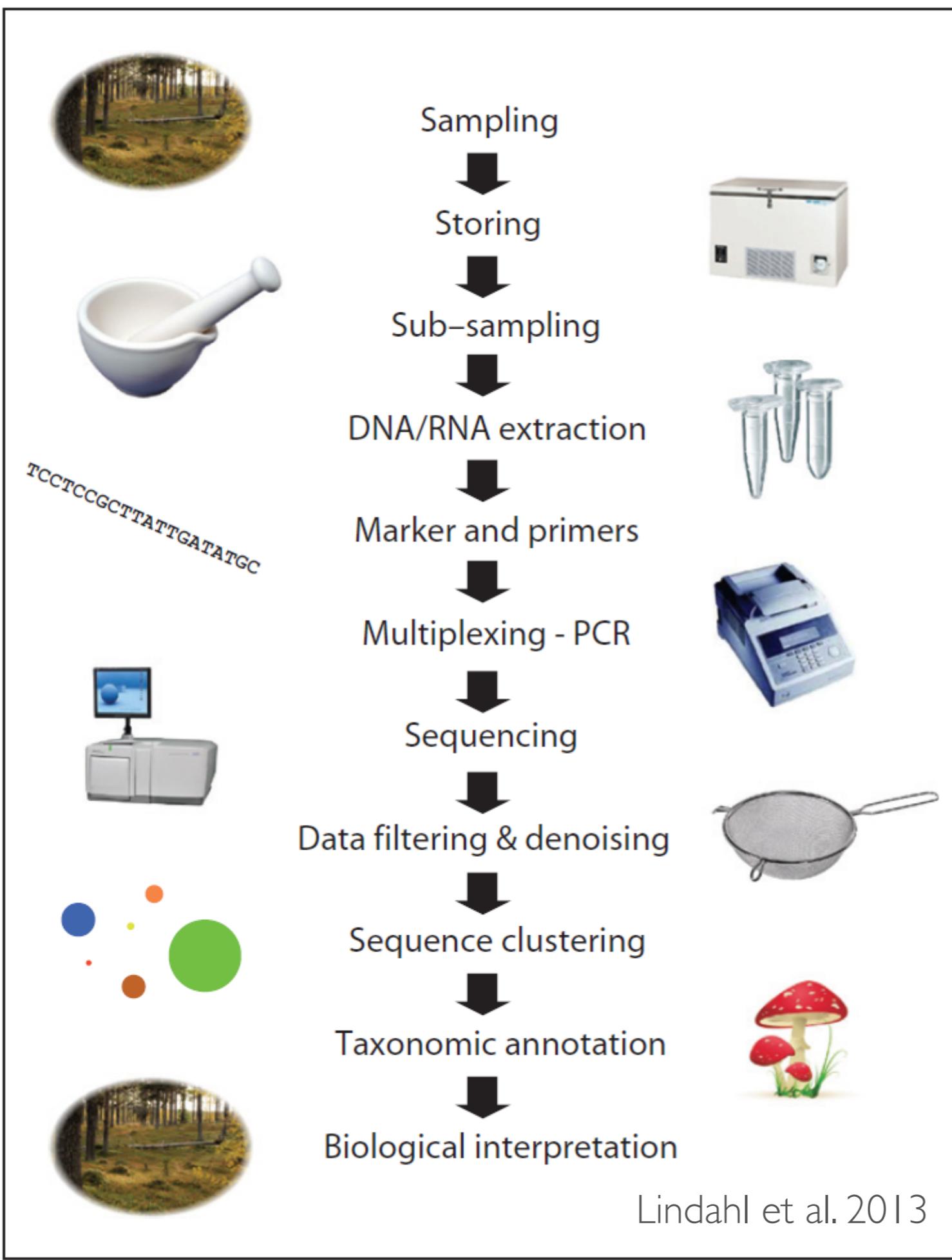
# TAG SWITCHING

The problem can be reduced or controlled for by:

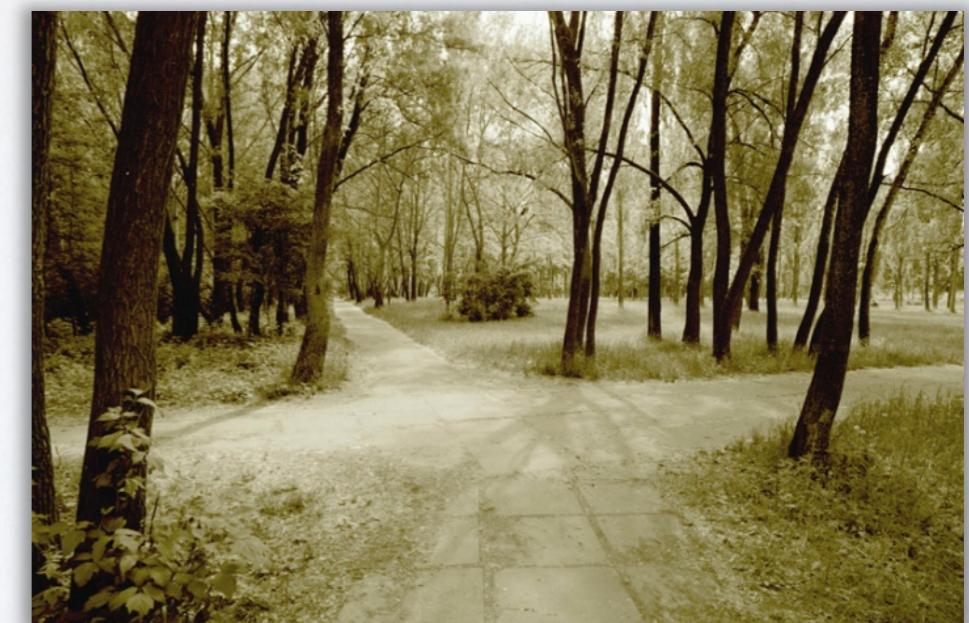
- Tagging in both ends with unique tag combinations
- Rinse the PCR amplicons thoroughly
- Avoid PCR steps during the final library preparations steps before sequencing (i.e. when adaptors are introduced)
- Include positive and negative controls to identify the level of switching/leakage
- (Drastic) ad hoc solution: Process your OTU/sample matrix by removing low frequency occurrences

# LIBRARY PREPARATION

- Tagging of amplicons (modified primers with illumina adapters)
- Normalisation of samples (equal molarity)



Many steps



That can go wrong

# LUNCH BREAK

# LECTURE II: LAB METHODS

ALEXANDER EILER  
DEPARTMENT OF BIOSCIENCES - AQUA

[alexander.eiler@ibv.uio.no](mailto:alexander.eiler@ibv.uio.no)

<https://www.mn.uio.no/ibv/english/people/aca/alexaei/>

# GLASS OF WATER TO FASTA FILE



- filter
- extract
- amplify
- sequence

1. CAUC  
CGGGUAA

2. AACCGUA  
UGCGAG

3. GCAAAGU

# LIBRARY CONSTRUCTION

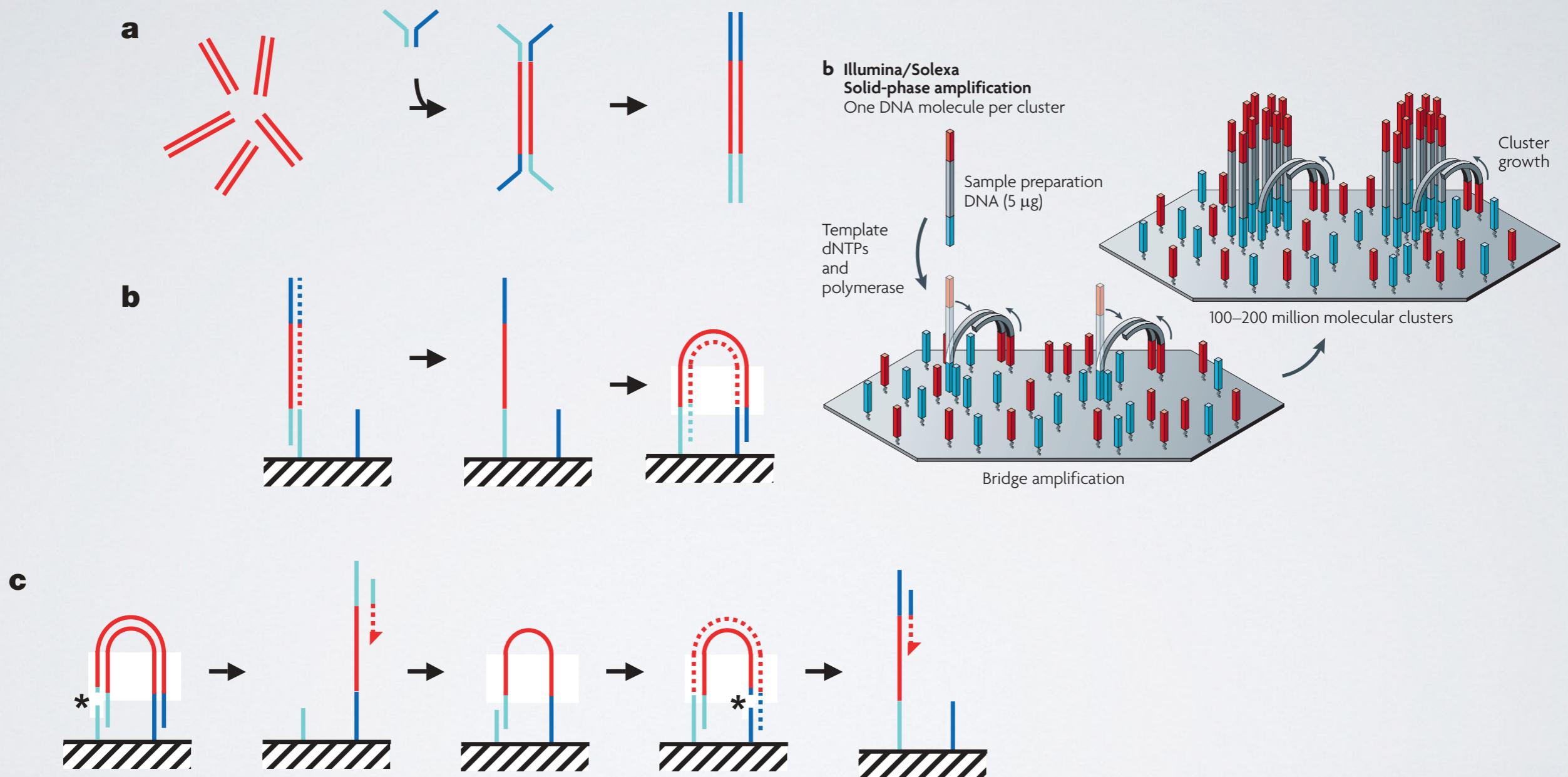
# SEQUENCING

- All cellular organisms have genomes
- All cellular organisms transcribe DNA into RNA and then translate RNA into protein
- Sequencing involves reading a string of letters (bases or amino acids) forming DNA, RNA or protein
- Sequencing is getting cheaper and cheaper
- Sequencing is very useful for studying microbial diversity

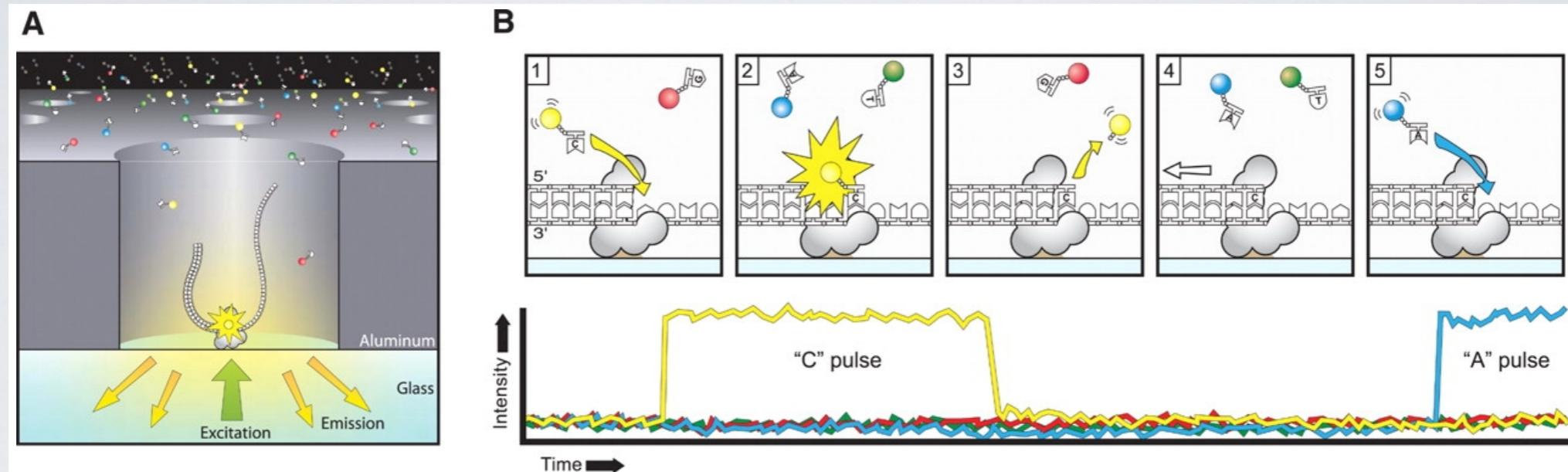
# ILLUMINA

<https://www.youtube.com/watch?v=fCd6B5HRaZ8>

# ILLUMINA



# PACBIO



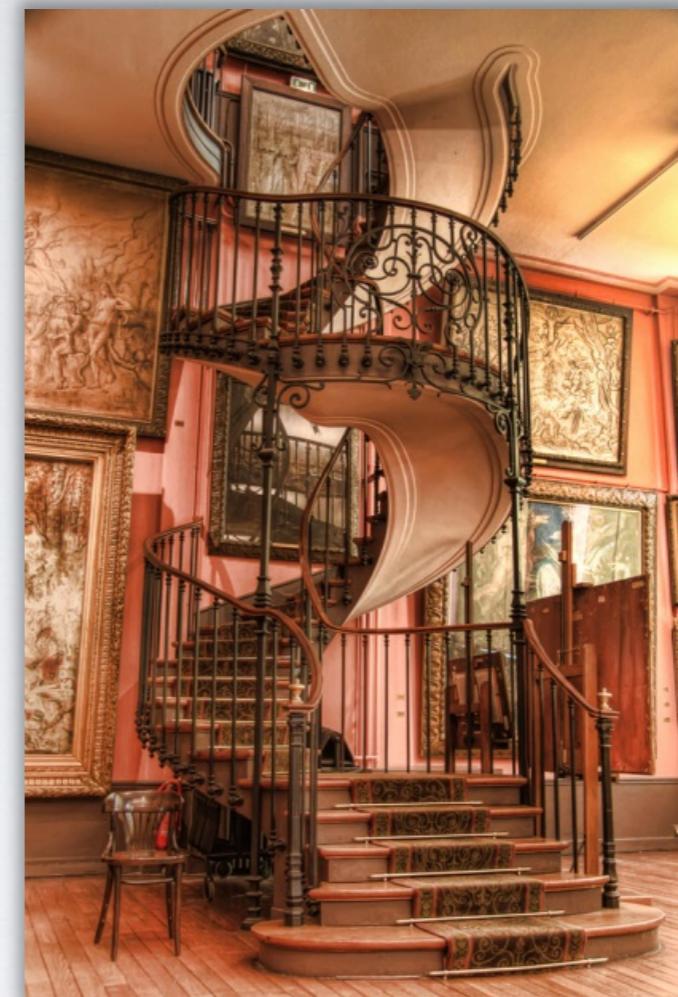
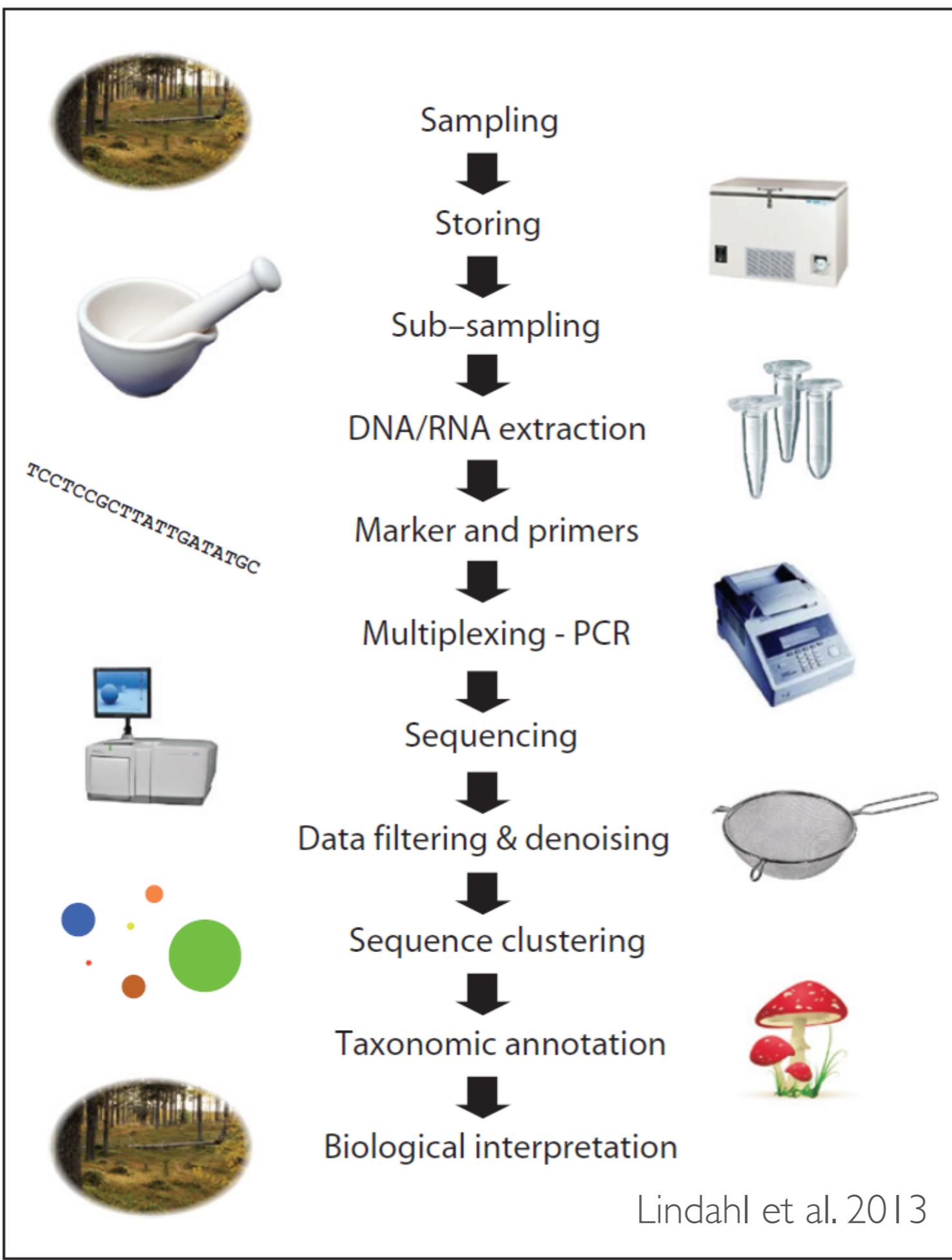
<https://www.pacb.com/videos/video-introduction-to-smrt-sequencing/>

# FASTQ FILE

/My Folders/PHD/Server/Uppmax\_home/proj35/INBOX/131126\_M00485\_0087\_00000000-A6BWBGS55RED

pool1\_fwd.fastq \*

```
1 @HWI-ST344:246:D1W9VACXX:7:1101:1489:2153 1:N:0:TTAGGC
2 NTGAGATGAAAGGTTCCATAGATAACGTCGCTCACCTCGAGCGTTACCCAGCTTGTGATGAGTACCACT
TGATACGATGTAGGCGTAAAGATCAAC
3 +
4 #1:BDDFFHHHHHFHJJJIIJJFIJJJIJHIIJJJGIJJJIJJIIJJHHHFFFFEEEEEEDBBECDCDDDD
DDCDEDBBDBACCCD@DDDDCCCCC
5 @HWI-ST344:246:D1W9VACXX:7:1101:1372:2188 1:N:0:TTAGGC
6 NAAAGGGGCCACCACGGCAAAGGGAAAAGATGACAGAGCCAACACCAACCAACCGACACCCCTGGGC
CGGGTTCATCTGCATTCCAAGGGCCG
7 +
8 #1:BDA11CBFDHIGDIGGIG<FGH6F@FFE@EFGBFEEFG;CCGHIDDEE5=D;6?CCCBBC'8=?A?B>;
>99>@BBCD:>@DCCC>AD######
9 @HWI-ST344:246:D1W9VACXX:7:1101:1736:2146 1:N:0:TTAGGC
10 NTTATCTAACATGTCAATCGTCAATCCAACGATGGGTGAAACATTAGTGTACATATCTTAAGGCACTGACGCT
CAATCCGTCTTATGTTGGTGCAATT
11 +
12 #1=DDFFFHHHHHJJJIIGIJJJJJJJJJFJJJJJJGGIIJIIJJJJGIDIJIJGHGGHHH
HFFFFFDDCDDDDC@DDDBBDDCCDD
13 @HWI-ST344:246:D1W9VACXX:7:1101:1735:2174 1:N:0:TTAGGC
14 NAGCTGTGAAGCCAGAAAAGGTCACTTTGCGCTGCTGGGGCCGATGCCGGAGTGGCGGACTGTCCGGCAC
GCGGATCTATTCTGACGTGGCGCCGCA
15 +
16 #1:BDDDEDFFAFHGGHIIIIIGGHGG@HGIGIGFHIIIFIIG8BHHFDADB?=??=B;-799:@@B@305
-95>&)5ACED(:A@CB<<7&555<B<
17 @HWI-ST344:246:D1W9VACXX:7:1101:1505:2193 1:N:0:TTAGGC
18 NCACGGCTACCCCTAACGGTCAAGGTTGCGAGGTGAACGCATTAGGATTGTATGGCGTACCATCAGGTACCC
AGGTCGAAATGACCATCTTCTTCAAAG
19 +
20 #1=DFFFFHHHHHJJJJJJHIIJIIJHIIJJJJJJ?GHJIIJJJJJJHIIHIHFFFFDEEEEDDCDDDD
DDBCDDDDDDDDDDDDDDDDDEDDDDCDD
21 @HWI-ST344:246:D1W9VACXX:7:1101:1586:2216 1:N:0:TTAGGC
22 TATCAGATGTTTTTACAGGTACAGAAGGAAGAGAGCCGTATCTGTCCGATCGTGAAGTAATCACGAAAGGTT
GCATAGTTGCCACATTCCGCCAGTGAC
```



Many steps



That can go wrong

# BIOGEOGRAPHY

- the distribution of biodiversity over space and time
- what are the underlying mechanisms: Who is where, at which abundance and why?

# HISTORICAL VS. ECOLOGICAL

- **Historical biogeography:** patterns in  $\beta$  diversity are due to speciation, dispersal and extinction of species
- **Ecological biogeography:** patterns in  $\beta$  diversity are caused by contemporary interactions, among organisms as well as with their physical and biotic environment

# ECOLOGICAL BIOGEOGRAPHY

What regulates biodiversity?

- Composition ( $\beta$  diversity)
- Richness and Evenness ( $\alpha$  diversity)

# CO-OCCURRENCE

**Species co-occurrences:**

- Statistical inferences
- Co-occurrence networks

# NATURAL HISTORY

Realised niches:

- Statistical inferences
- Environmental boundaries of species