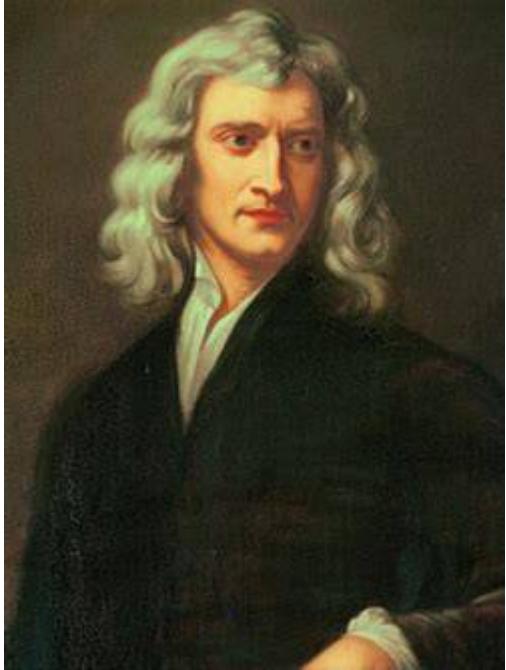


# **- Metabarcoding -**

## **- The ambition to scan Life -**

Arthur Cousson  
Ecos&Sols – IRD Montpellier  
[arthur.cousson@ird.fr](mailto:arthur.cousson@ird.fr)

**Let's make fun of physicists first**



# Let's make fun of physicists first

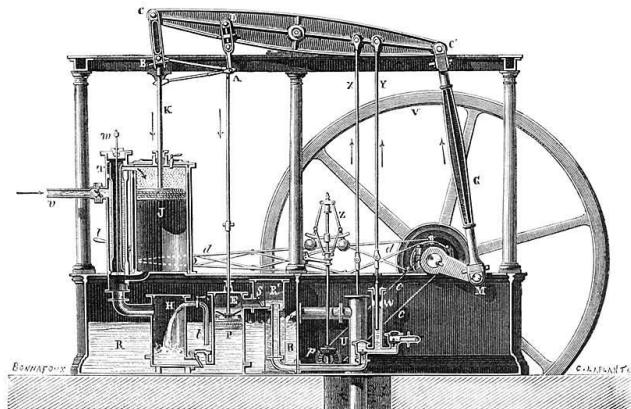
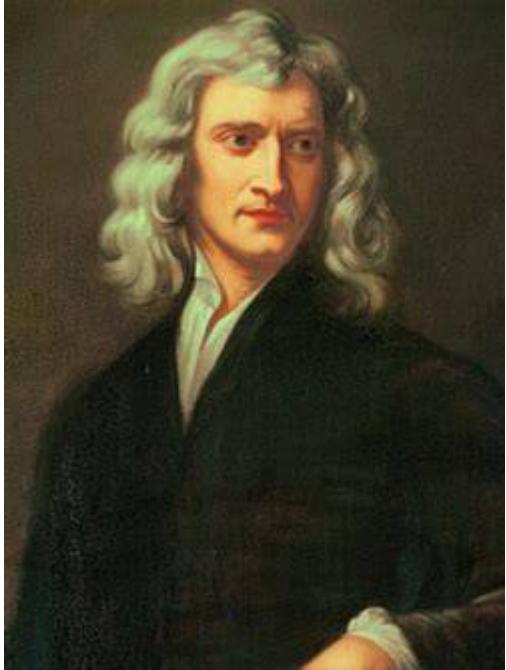
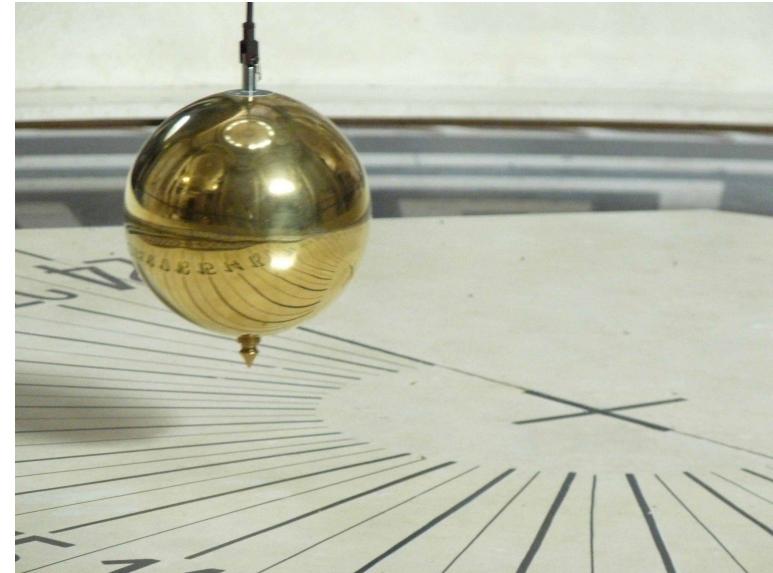
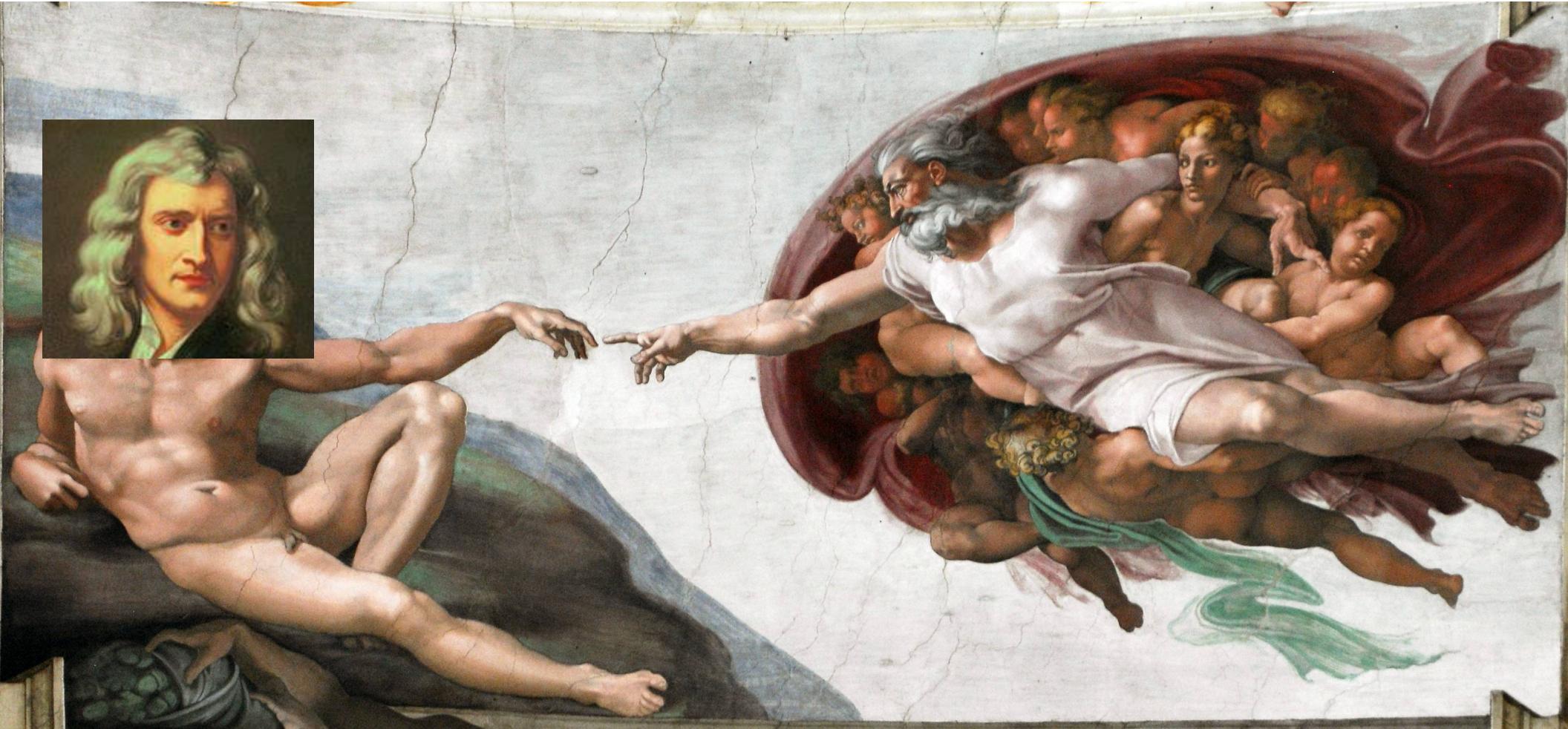


Fig. 59. — Machine à balancier de Watt.

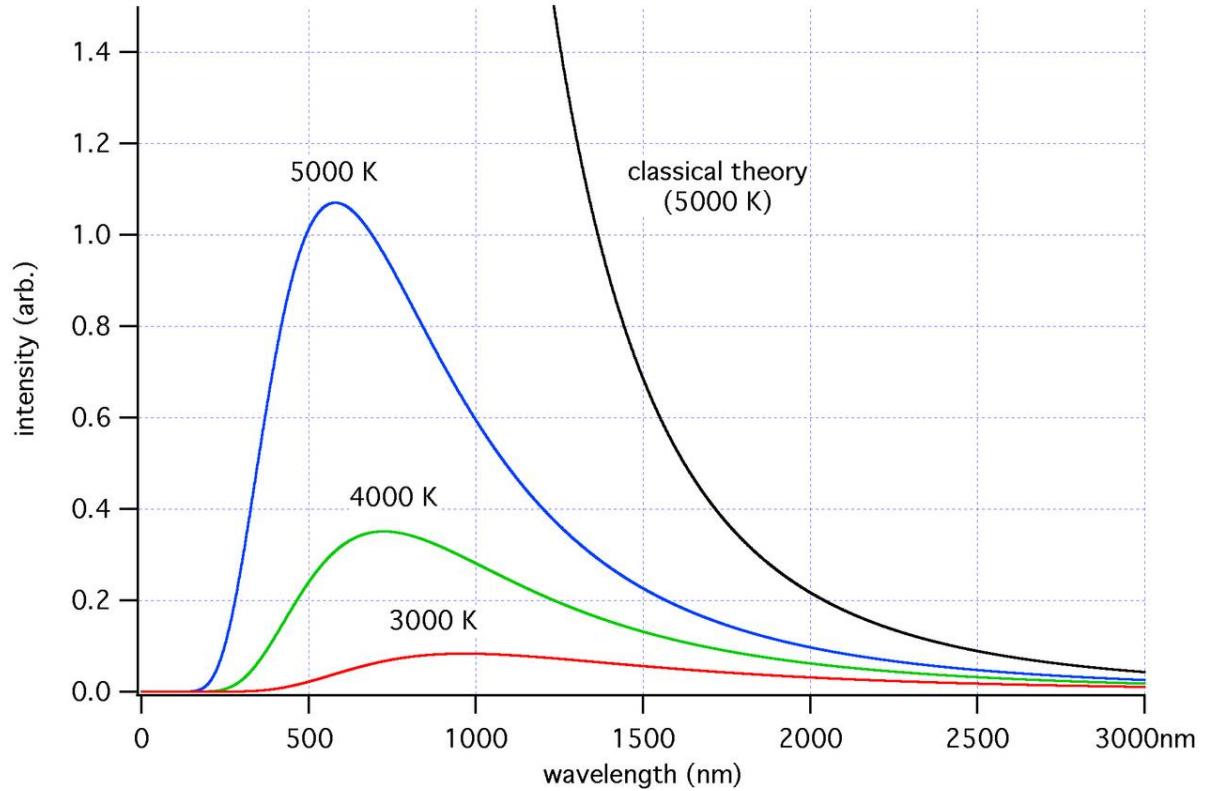
v. Tuyau de prise de vapeur; T, tiroir; J, cylindres; H, condenseur; PE pompe d'épuisement; WY pompe alimentaire de la chaudière; UX pompe d'alimentation de la bâche R; P, ZR régulateur; dd excentrique; ABCD parallélogramme; GM bielle et manivelle; V, volant.



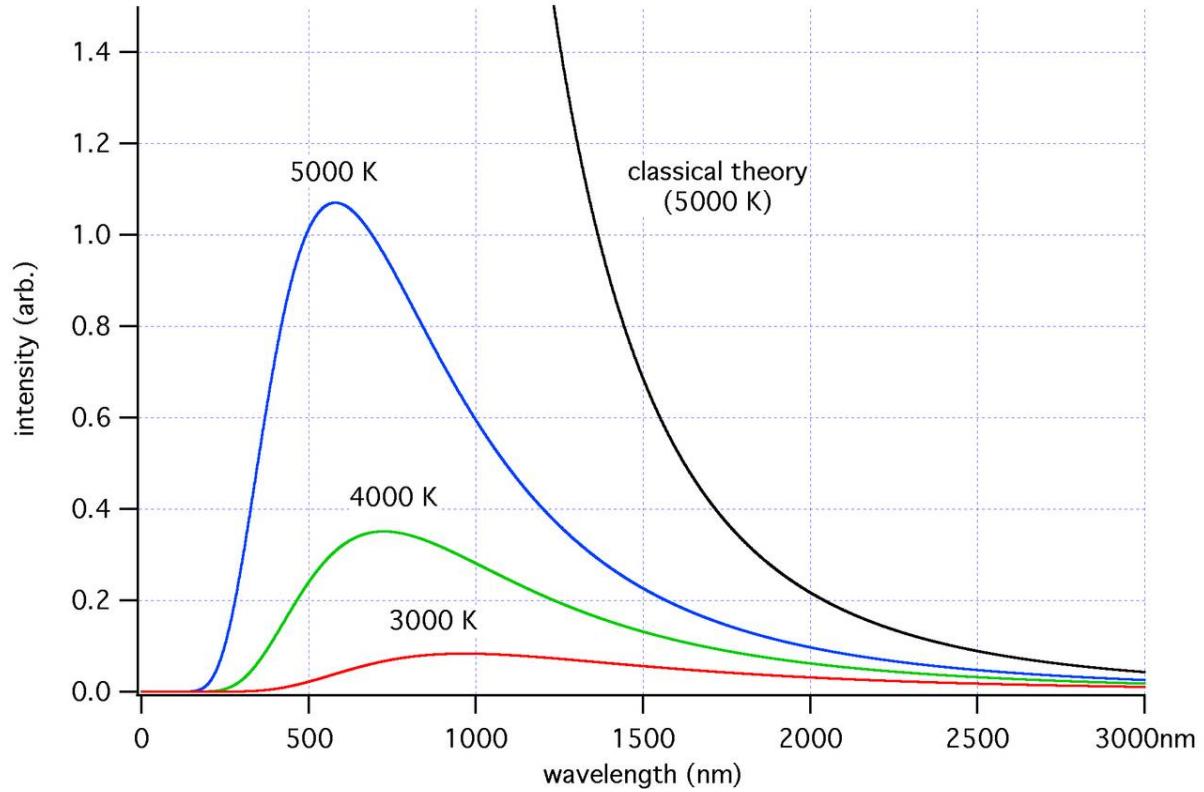
**Let's make fun of physicists first**



Let's make fun of physicists first



# Let's make fun of physicists first

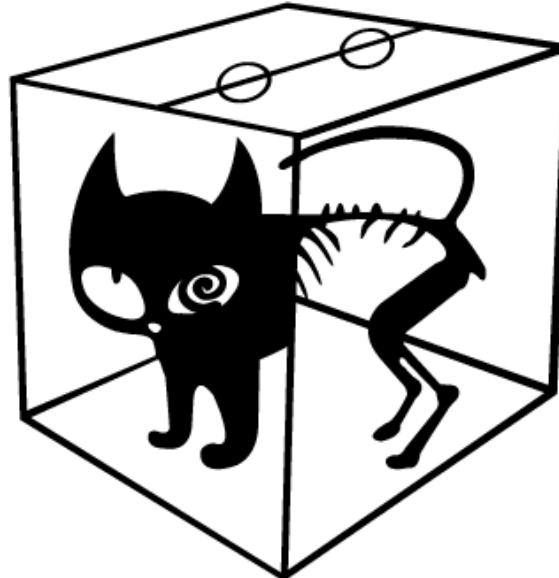


Let's make fun of physicists first



Let's make fun of physicists first

SCHRÖDINGER'S CAT IS  
ALIVE

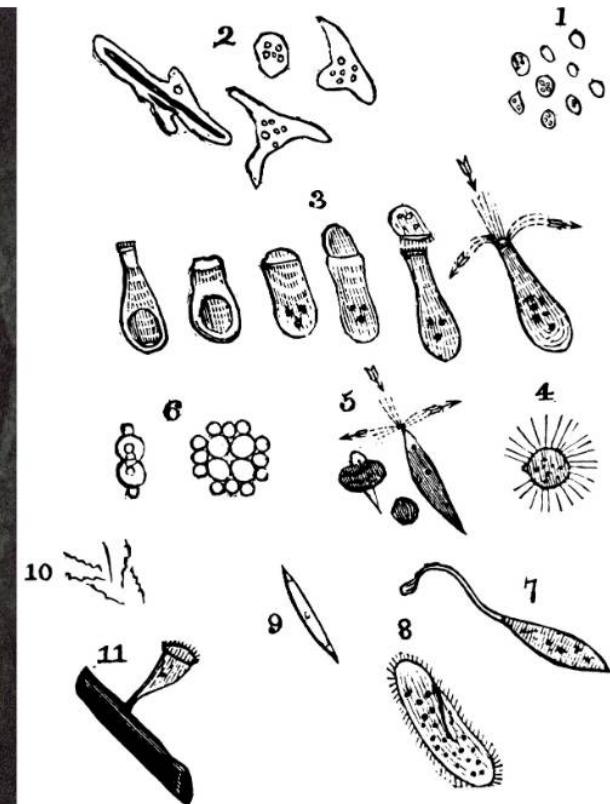




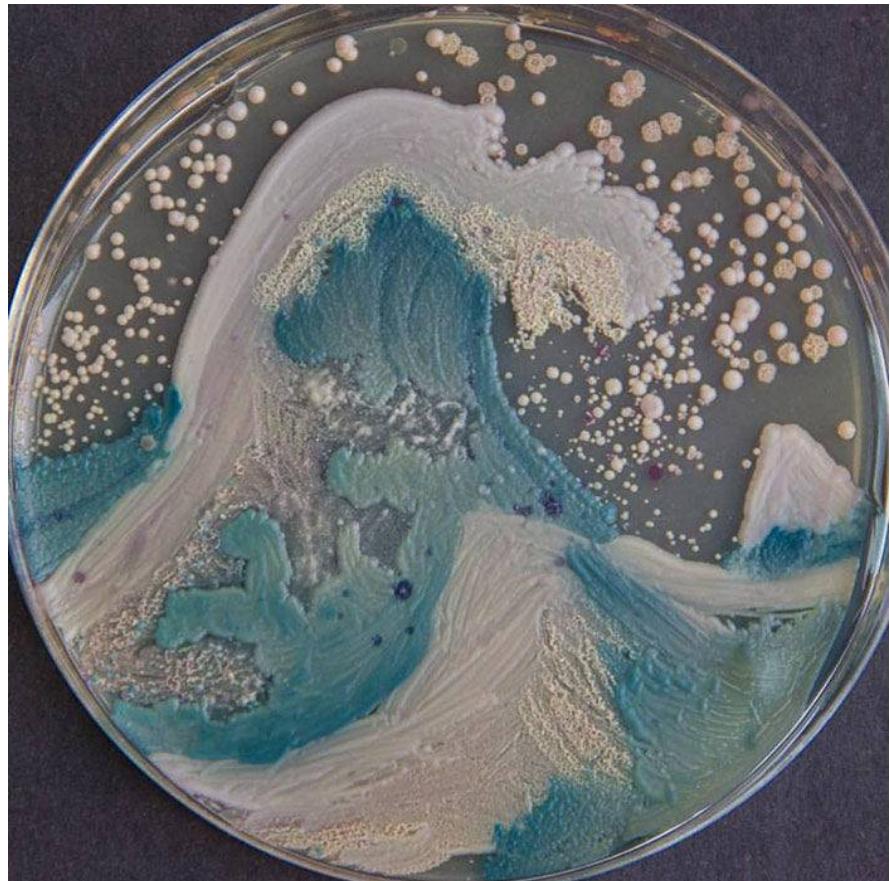
**Now, let's be humble**



## Now, let's be humble – Antonie Van Leeuwenhoek

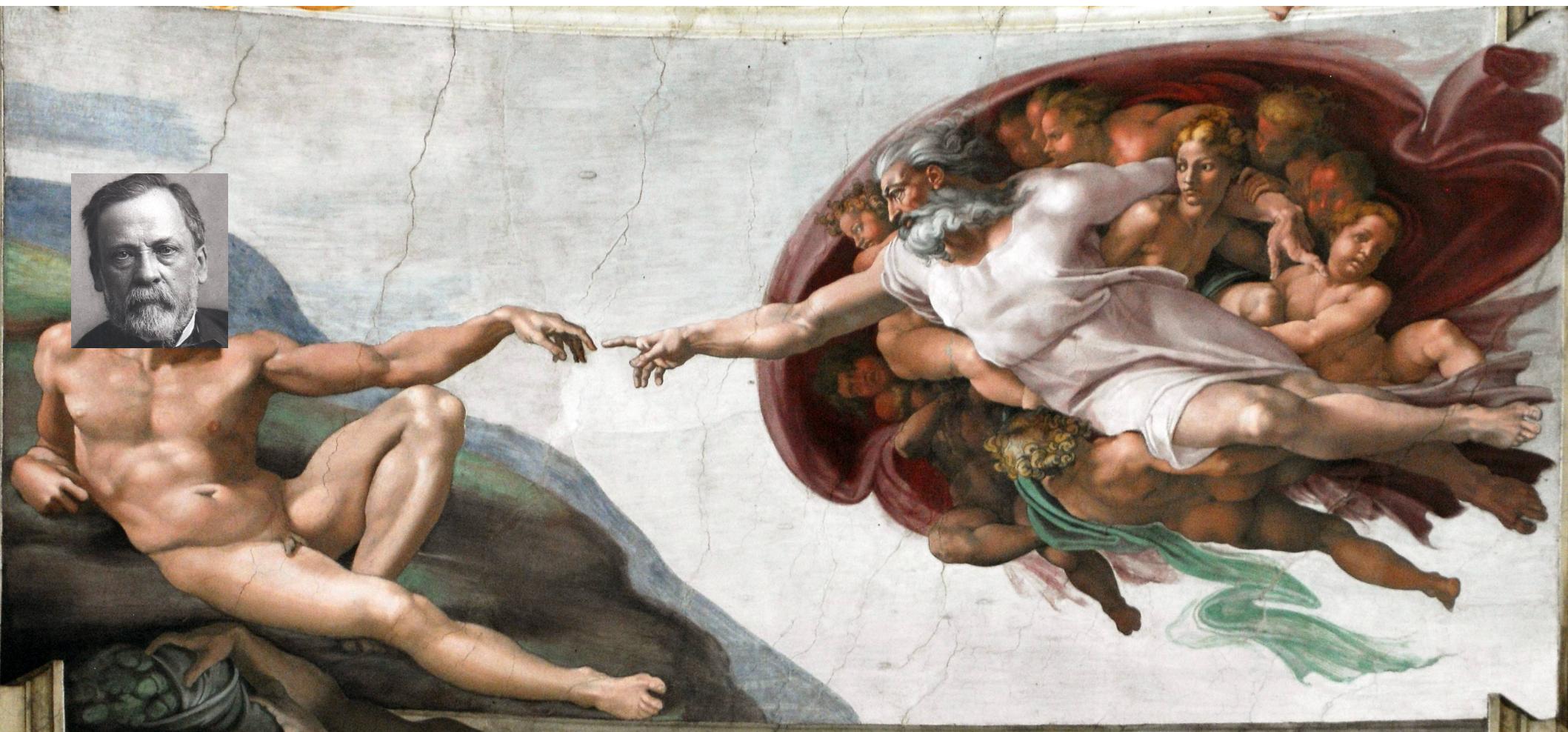


## **Birth of current Microbiology – End of 70's**



**Around 30 bacterial phyla  
described and cultivated**

## Birth of current Microbiology – End of 70's



# Birth of current Microbiology

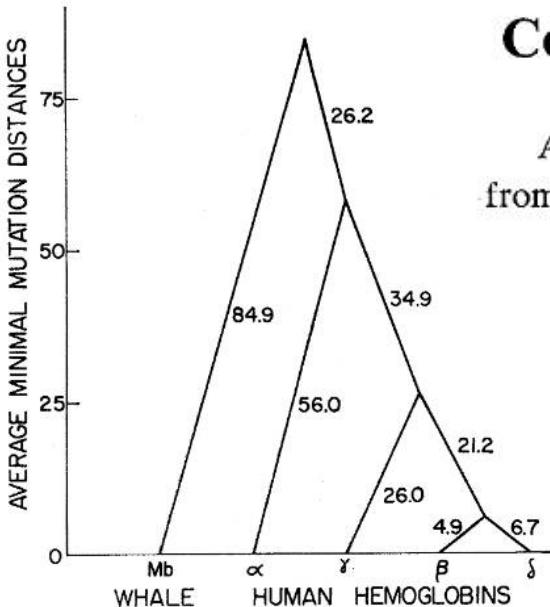
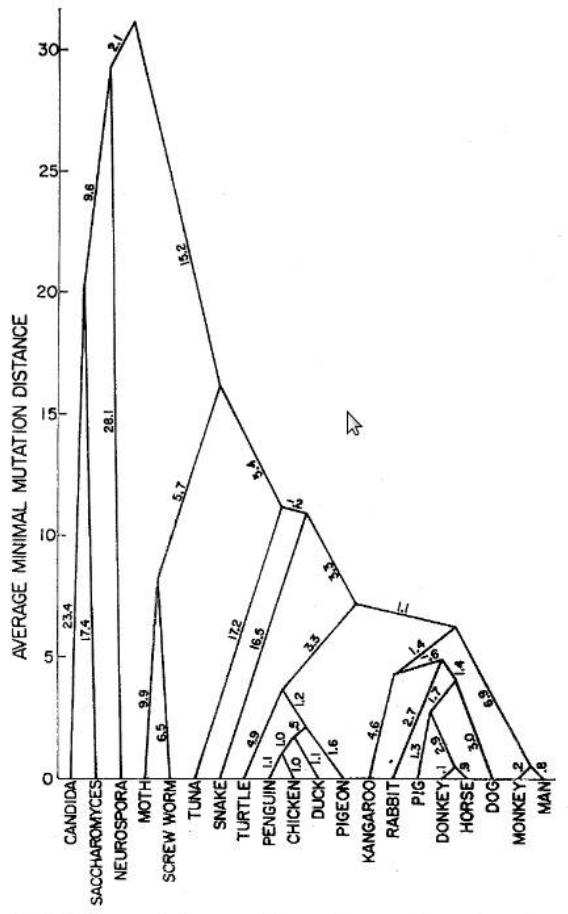


Fig. 2 (left). Phylogeny as reconstructed from observable mutations in the cytochrome *c* gene. Each number on the figure is the corrected mutation distance (see text) along the line of descent as determined from the best computer fit so far found. Each apex is placed at an ordinate value representing the average of the sums of all mutations in the lines of descent from that apex.

Fig. 3 (right above). A gene phylogeny as reconstructed from observable mutations in several heme-containing globins. See Fig. 2 for details. The percent "standard deviation" (7) for this tree is 1.33.

## Construction of Phylogenetic Trees

A method based on mutation distances as estimated from cytochrome *c* sequences is of general applicability.

Walter M. Fitch and Emanuel Margoliash

1965

# Birth of current Microbiology – End of 70's

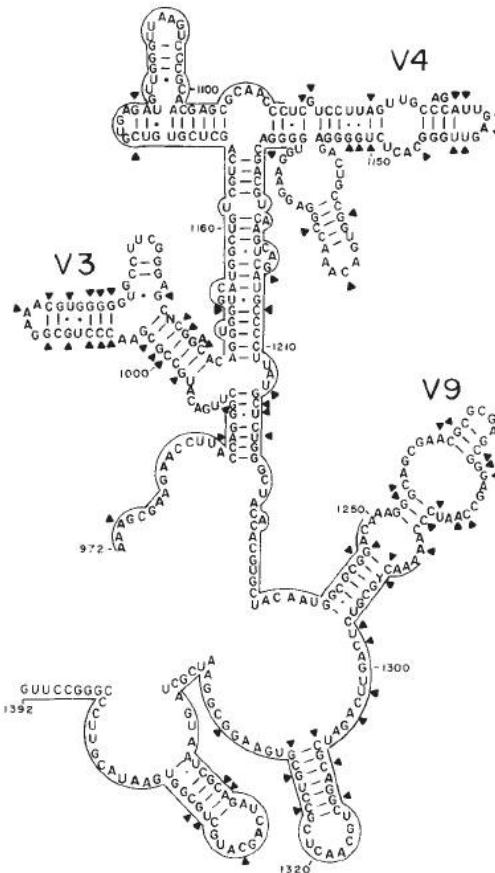
## 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community

David M. Ward, Roland Weller & Mary M. Bateson

Department of Microbiology, Montana State University, Bozeman, Montana 59717, USA

MICROBIOLOGISTS have been constrained in their efforts to describe the compositions of natural microbial communities using traditional methods. Few microorganisms have sufficiently distinctive morphology to be recognized by microscopy. Culture-dependent methods are biased, as a microorganism can be cultivated only after its physiological niche is perceived and duplicated experimentally. It is therefore widely believed that fewer than 20% of the extant microorganisms have been discovered<sup>1,2</sup>, and that culture methods are inadequate for studying microbial community composition<sup>3–7</sup>. In view of the physiological and phylogenetic diversity among microorganisms<sup>8</sup>, speculation that 80% or more of microbes remain undiscovered raises the question of how well we know the Earth's biota and its biochemical potential. We have performed a culture-independent analysis of the composition of a well-studied hot spring microbial community, using a common but distinctive cellular component, 16S ribosomal RNA. Our results confirm speculations about the diversity of uncultured microorganisms it contains.

We developed a cloning method for retrieving naturally occurring 16S rRNA sequences<sup>9</sup> that is more selective than methods previously proposed for recovering 16S rRNA genes<sup>10,11</sup>. The selectivity is based on the synthesis of com-



# Birth of current Microbiology – End of 70's

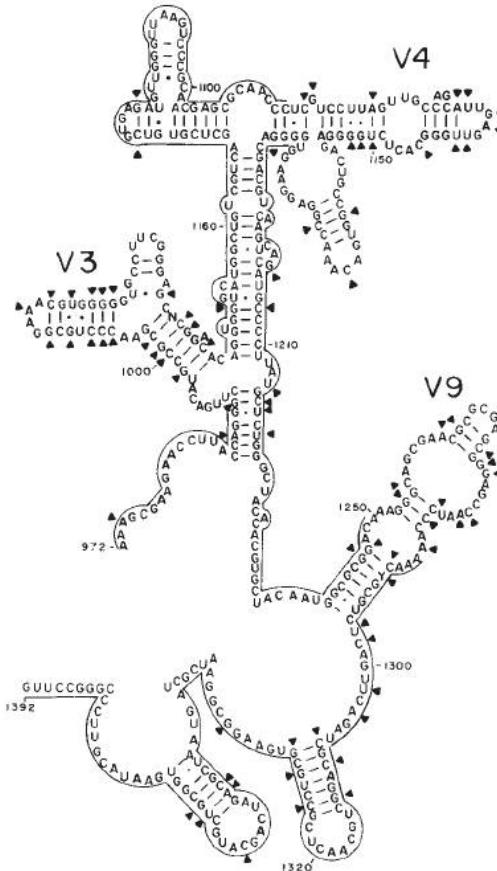
## 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community

David M. Ward, Roland Weller & Mary M. Bateson

Department of Microbiology, Montana State University, Bozeman, Montana 59717, USA

MICROBIOLOGISTS describe the complexity of microbial communities using traditional methods such as distinctive morphological features or distinctive metabolic pathways. However, only about 10% of the species in a natural community have been identified. This is because genetic diversity is often low, and less than 10% or more of microorganisms can be cultured. We know a great deal about the potential of how well we know a natural community, using a common but well-studied marker, 16S ribosomal RNA. We have performed a detailed analysis of the composition of a natural community, using a common but distinctive cloning method for retrieving naturally occurring 16S rRNA sequences<sup>9</sup> that is more selective than methods previously proposed for recovering 16S rRNA genes<sup>10,11</sup>. The selectivity is based on the synthesis of com-

Hello  
my name is  
Molecular  
Microbiology



# Plan

## Introduction

### I – Metabarcoding definition

- Principle
- Targeted genes
- Technical aspects

### II – Metabarcoding applications

- Human Microbiome project
- Communities investigation
- Environmental biomonitoring

### III – Perspectives for Metabarcoding

- Problems related to chosen barcodes
- Alternatives
- For the future

## Conclusion



Let's discuss

# Metabarcoding

Let's discuss

# Metabarcoding

Met(a)-

gr. *Meta* -

Indicating succession  
(after - beyond)

Let's discuss

# Metabarcoding

Met(a)-

gr. *Meta* -  
Indicating succession  
(after - beyond)

Barcode

(Taxonomy)  
Taxonomic method using short  
DNA sequences to identify an  
organism to a particular species

Let's discuss

# Metabarcoding

Met(a)-

gr. *Meta* -  
Indicating succession  
(after - beyond)

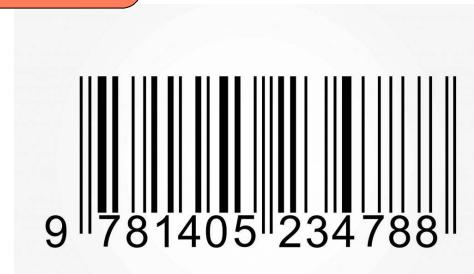
Barcode

(Taxonomy)  
Taxonomic method using short  
DNA sequences to identify an  
organism to a particular species

Metabarcoding is the approach going beyond  
traditional barcoding

# Barcode

**Barcode highly shared sequence**



~~Abundances~~

PCR

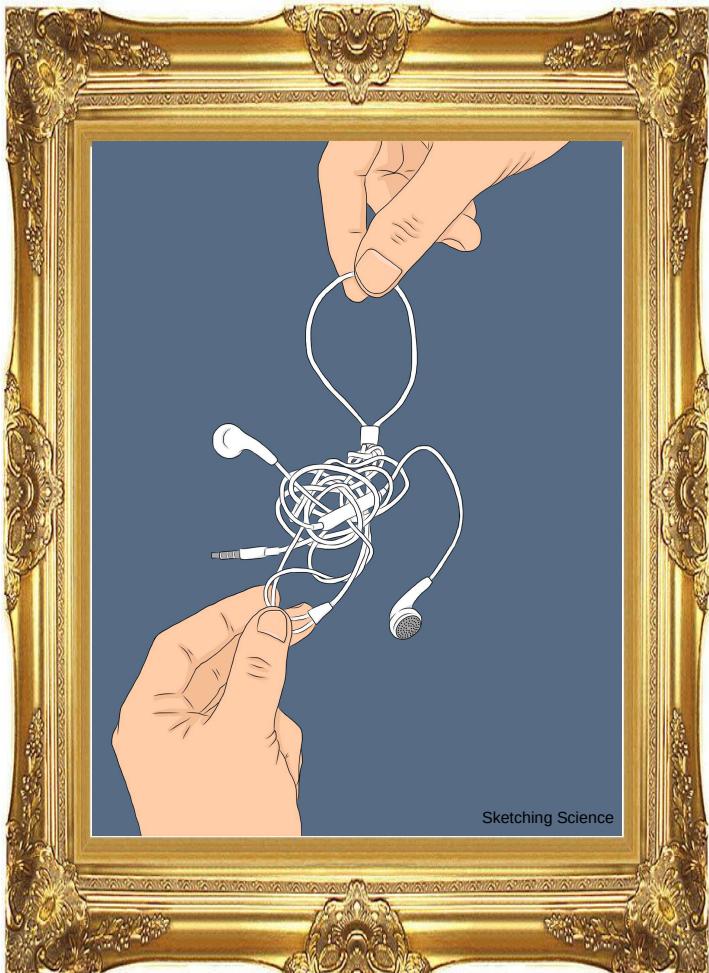
**Universal primer**

**Slightly variable sequence**

**Variable lengths**

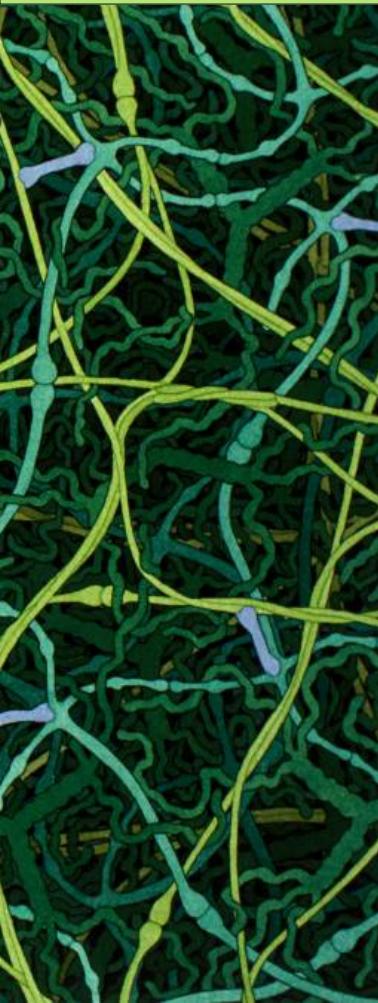
**Universal primer**

TD



**Your turn to think  
the experiment**

## Targeted Genes



Animals :

Cytochrome C oxydase subunit I gene (**COI**)

Plants :

Ribulose-bisphosphate carboxylase  
and Maturase K genes  
**rbcL** and **matK**

Fungi :

Internal Transcribed Spacer (**ITS-1 -2**)  
**18S** (AMF)

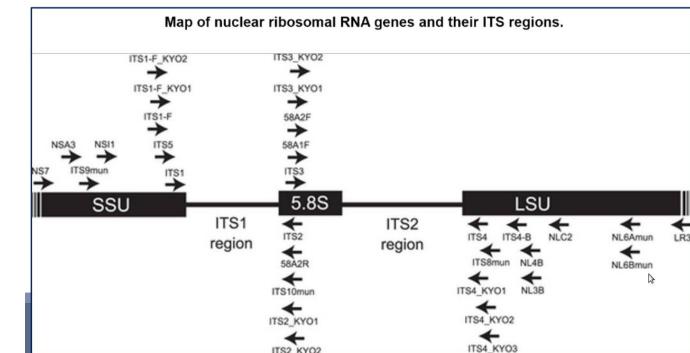
Bacteria :

**16S** Ribosomal subsunit

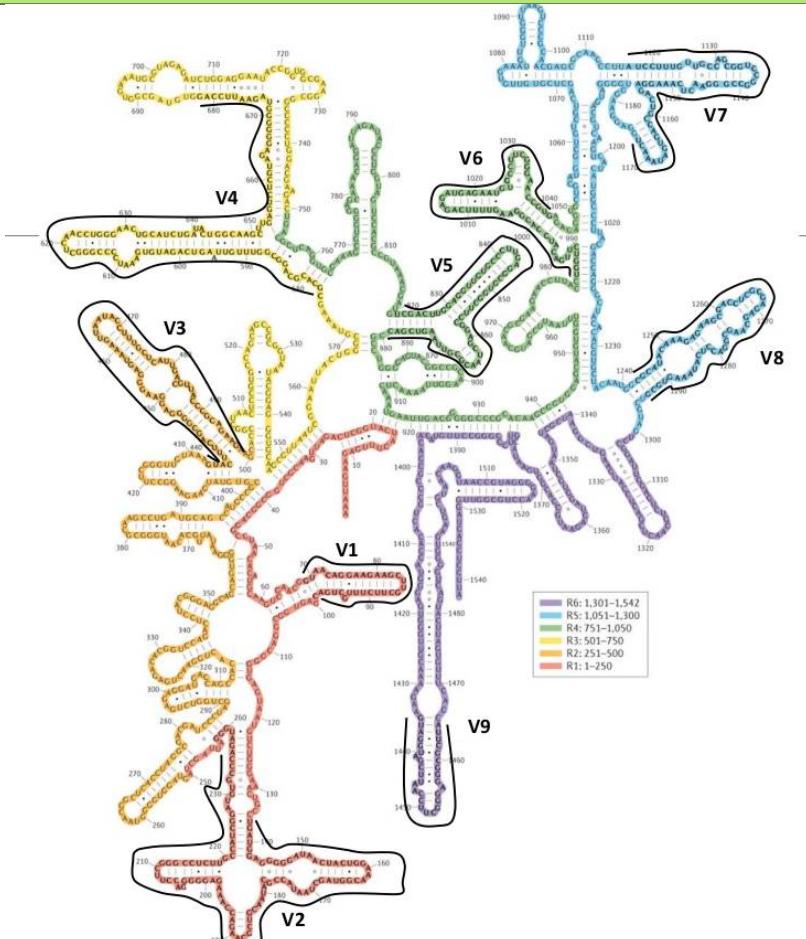
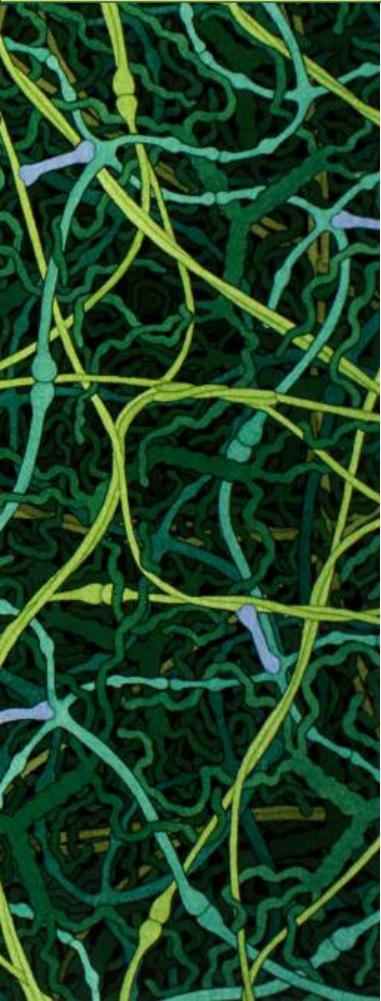


**CONSERVED REGIONS:** unspecific applications

**VARIABLE REGIONS:** group or species-specific applications

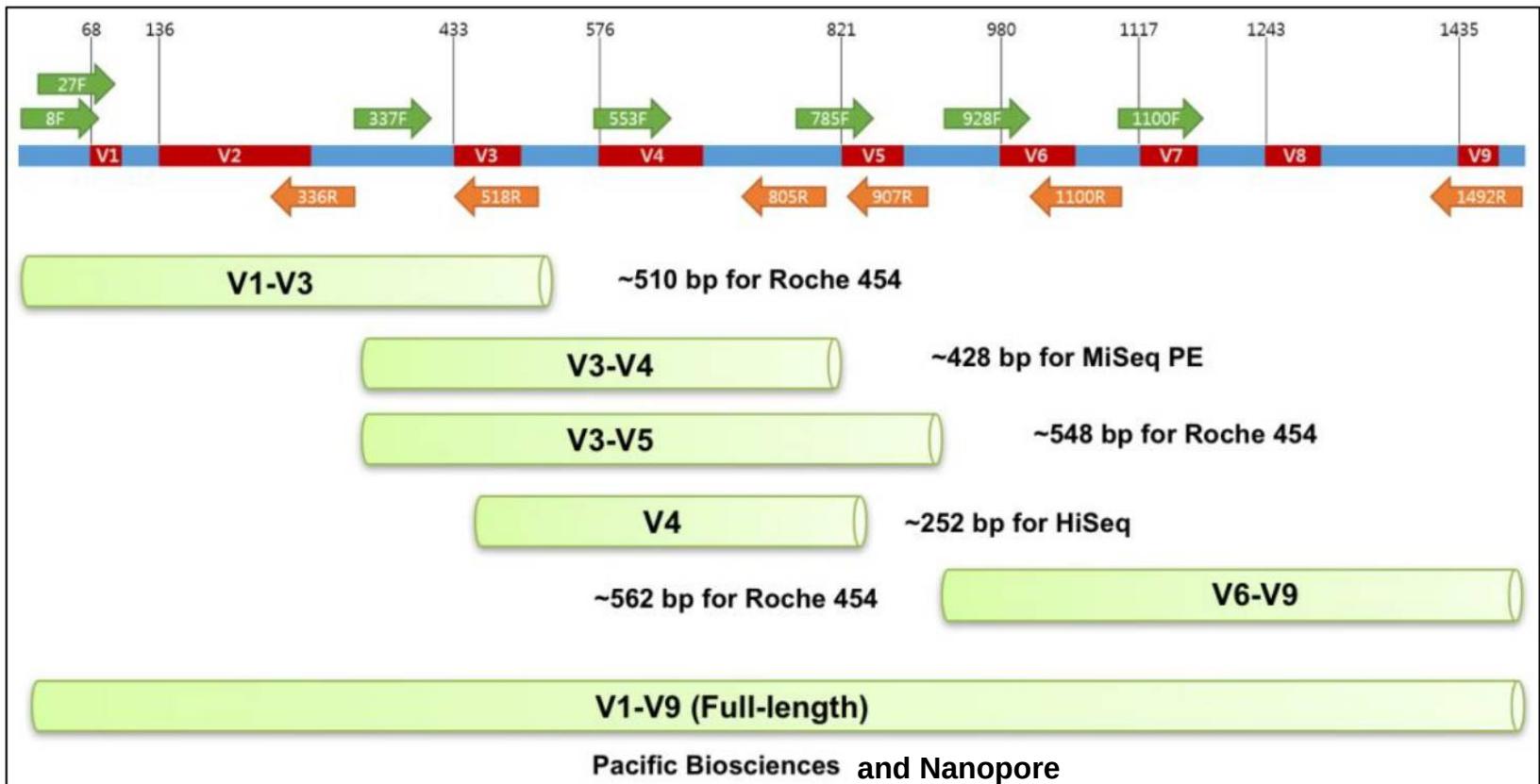
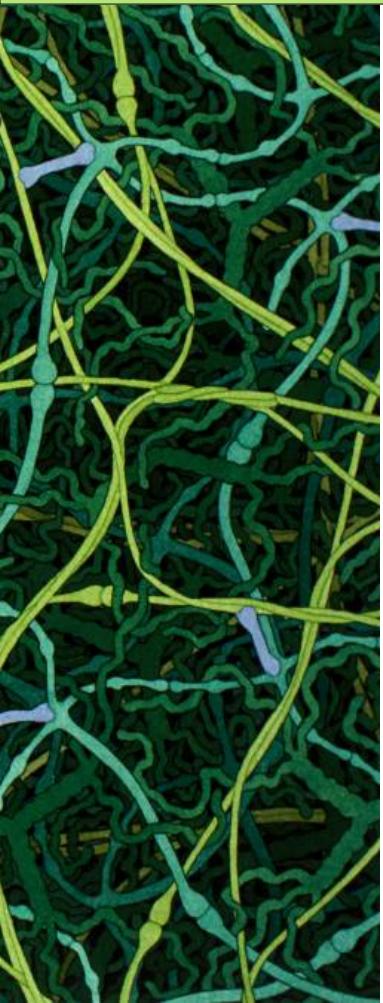


## Targeted Genes – Bacterial focus

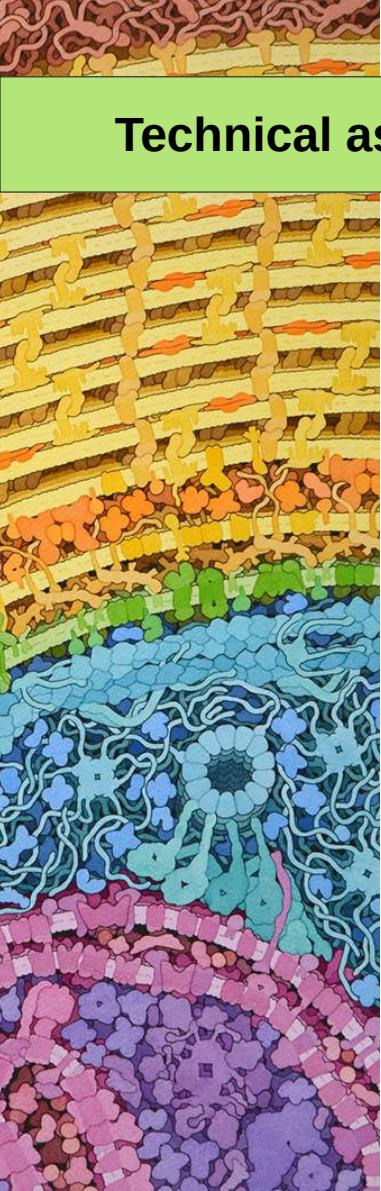


16S small ribosomal subunit

## Targeted Genes – Bacterial focus



Pacific Biosciences and Nanopore



## Technical aspects

### Wet-lab

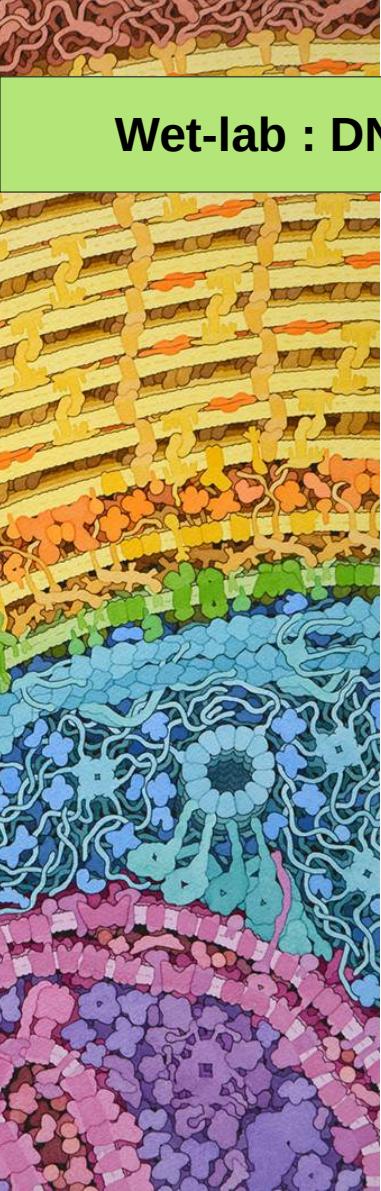
- DNA extraction
- Blanks
- Amplification
- Sequencing
- Illumina - Nanopore

### Dry-lab

- Quality check
- Sequences merging
- Clustering
- Chimera Removal
- Filtering
- Affiliations

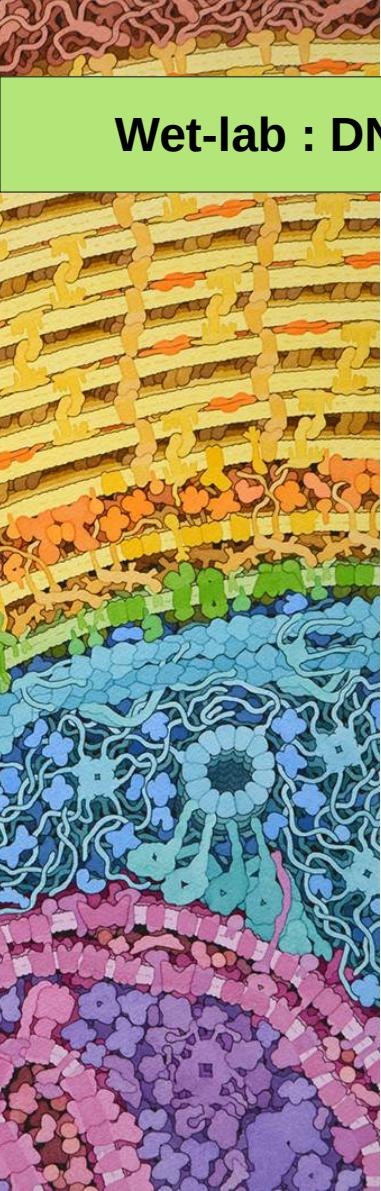
### Behind the scene

- All of your Science



## Wet-lab : DNA extraction

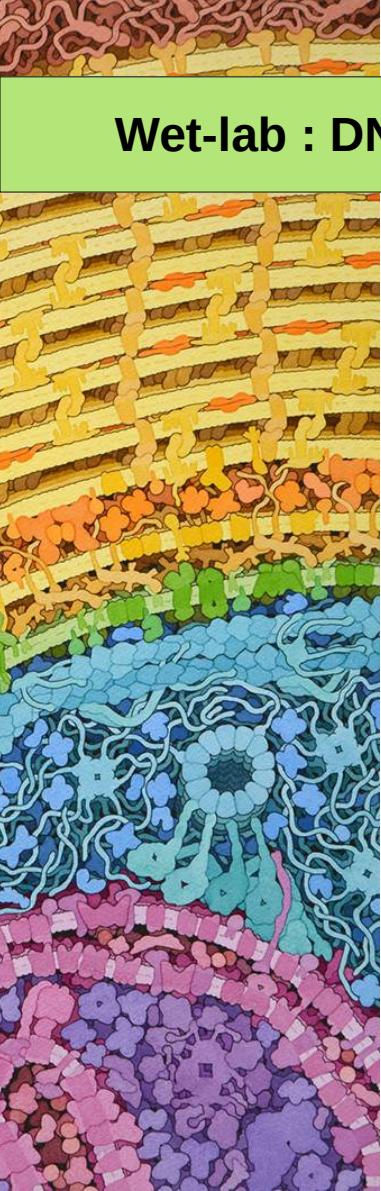
**Are you sure to sample  
your community appropriately ?**



## Wet-lab : DNA extraction

**Are you sure to sample  
your community appropriately ?**

What are the organisms you are looking for ?

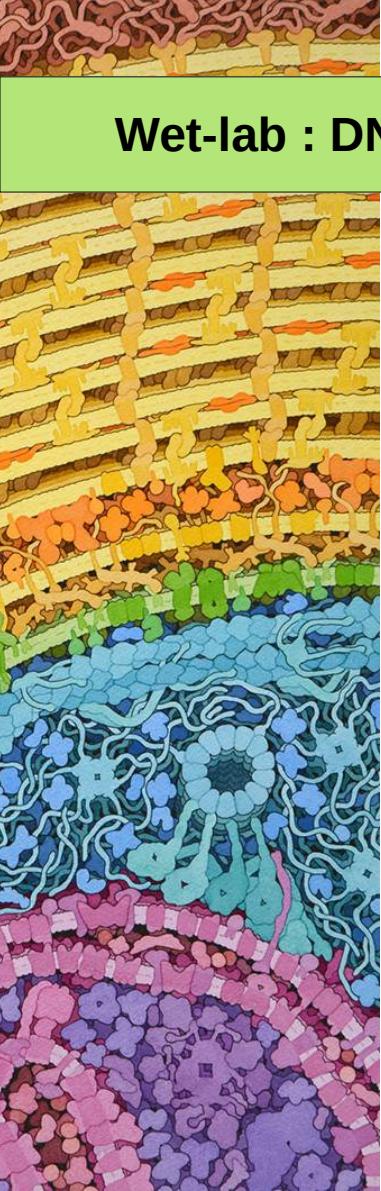


## Wet-lab : DNA extraction

**Are you sure to sample  
your community appropriately ?**

What are the organisms you are looking for ?

Do you choose your sample accurately ?



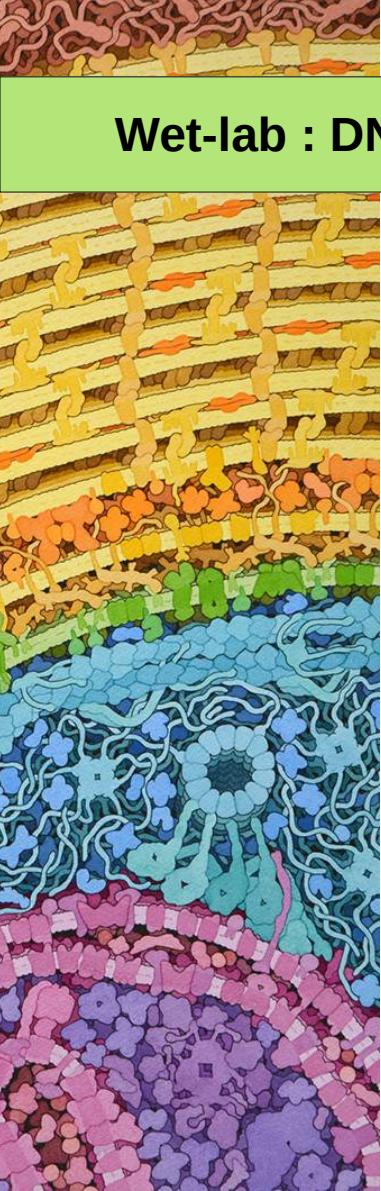
## Wet-lab : DNA extraction

**Are you sure to sample  
your community appropriately ?**

What are the organisms you are looking for ?

Do you choose your sample accurately ?

Do you extract all of your organisms of interest ?



## Wet-lab : DNA extraction

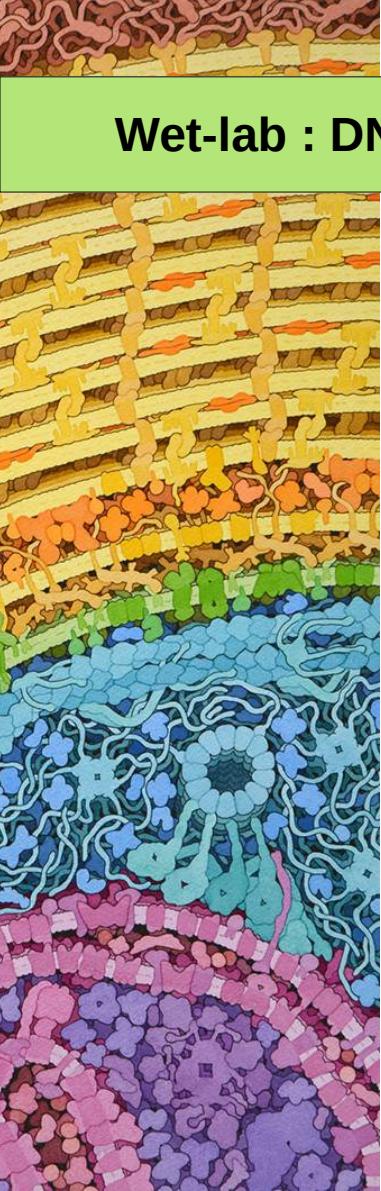
**Are you sure to sample  
your community appropriately ?**

What are the organisms you are looking for ?

Do you choose your sample accurately ?

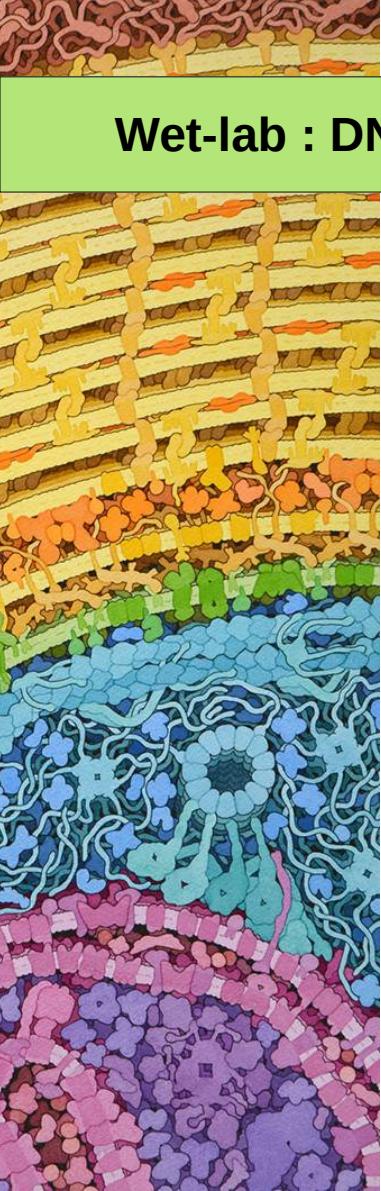
Do you extract all of your organisms of interest ?

Are you sure you did not cross-contaminate your samples ?



## Wet-lab : DNA extraction

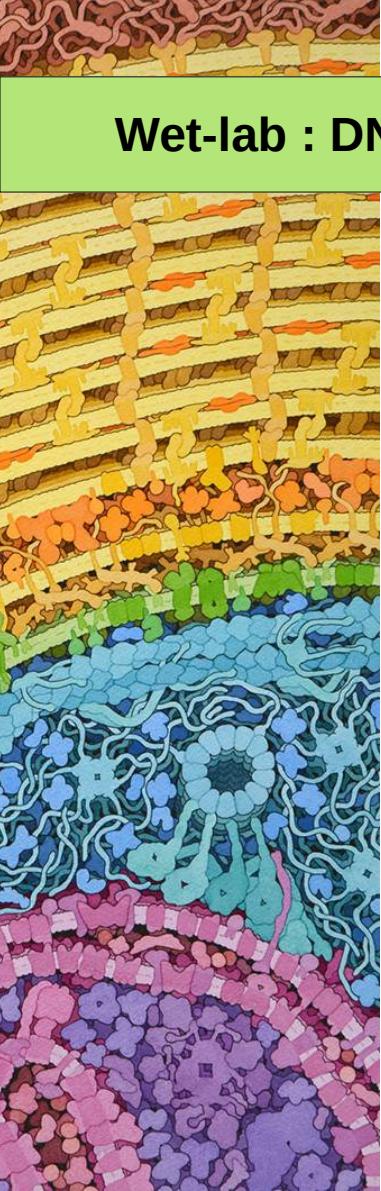
### Cross-contamination tips



## Wet-lab : DNA extraction

### Cross-contamination tips

**Perfectly clean your workplace and tools before starting anything**

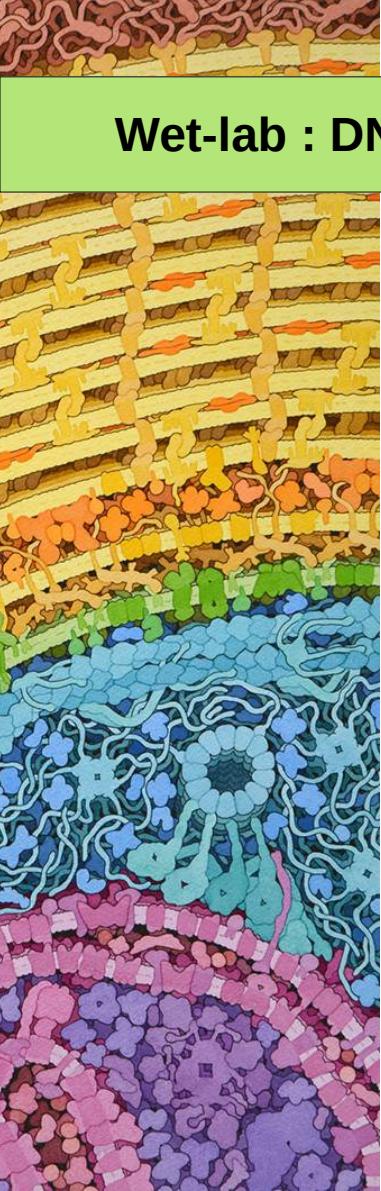


## Wet-lab : DNA extraction

### Cross-contamination tips

**Perfectly clean your workplace and tools before starting anything**

- Change your labcoat for a clean one

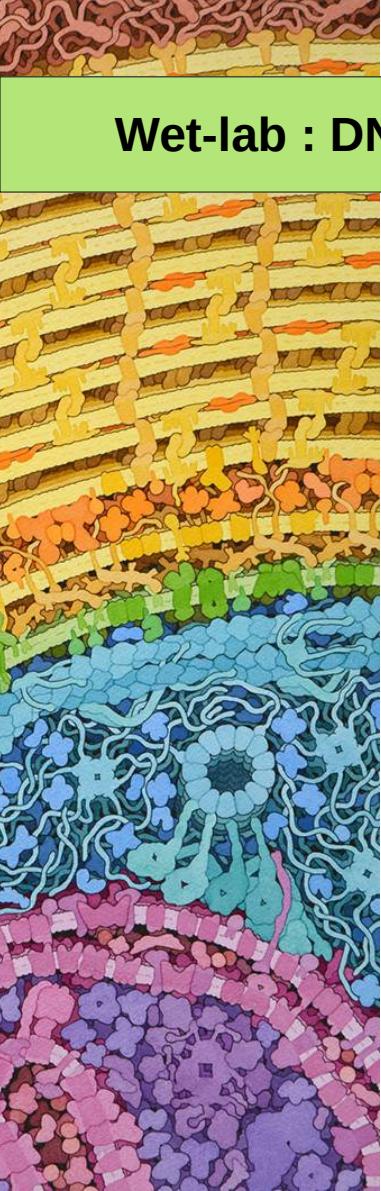


## Wet-lab : DNA extraction

### Cross-contamination tips

**Perfectly clean your workplace and tools before starting anything**

- Change your labcoat for a clean one
- Wear a mask



## Wet-lab : DNA extraction

### Cross-contamination tips

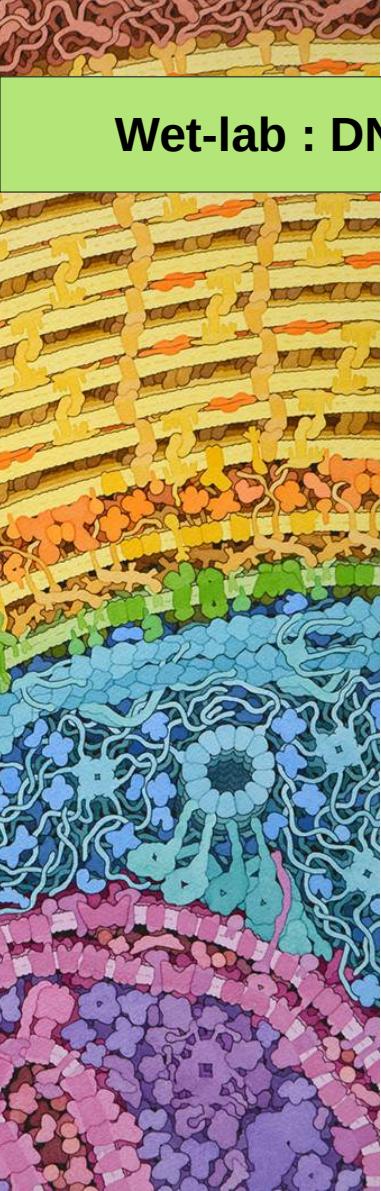
**Perfectly clean your workplace and tools before starting anything**

- Change your labcoat for a clean one
- Wear a mask
- Open new materials for the occasion

### Cross-contamination tips

**Perfectly clean your workplace and tools before starting anything**

- Change your labcoat for a clean one
- Wear a mask
- Open new materials for the occasion
- Do not assume reagents history

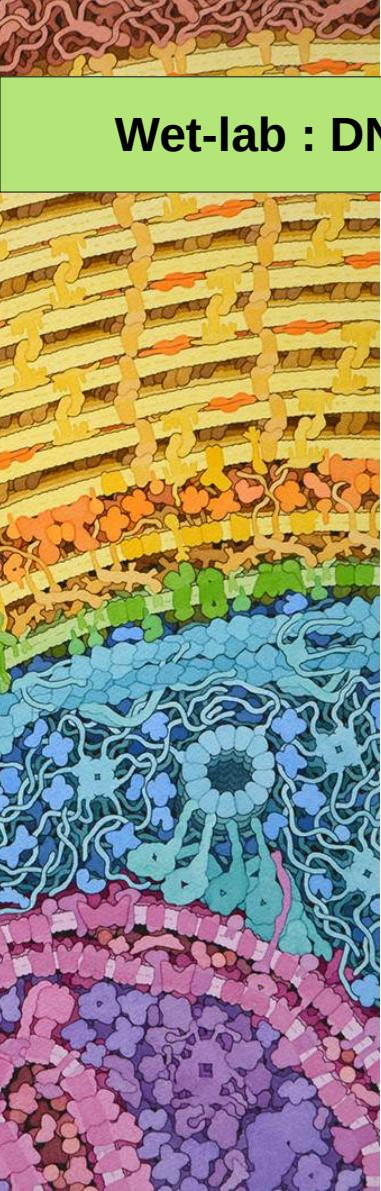


## Wet-lab : DNA extraction

### Cross-contamination tips

**Perfectly clean your workplace and tools before starting anything**

- Change your labcoat for a clean one
- Wear a mask
- Open new materials for the occasion
- Do not assume reagents history
- Do not open two different tubes at the same time



## Wet-lab : DNA extraction

### Cross-contamination tips

**Perfectly clean your workplace and tools before starting anything**

- Change your labcoat for a clean one
- Wear a mask
- Open new materials for the occasion
- Do not assume reagents history
- Do not open two different tubes at the same time
- Always change your cones between samples

### Cross-contamination tips

**Perfectly clean your workplace and tools before starting anything**

- Change your labcoat for a clean one
- Wear a mask
- Open new materials for the occasion
- Do not assume reagents history
- Do not open two different tubes at the same time
- Always change your cones between samples
- Immediately change gloves if contaminated

### Cross-contamination tips

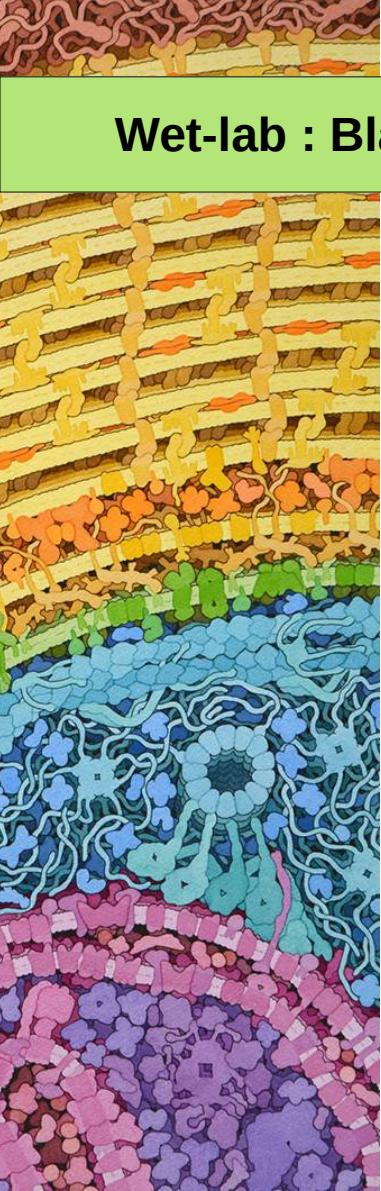
**Perfectly clean your workplace and tools before starting anything**

- Change your labcoat for a clean one
- Wear a mask
- Open new materials for the occasion
- Do not assume reagents history
- Do not open two different tubes at the same time
- Always change your cones between samples
- Immediately change gloves if contaminated
- Regularly change gloves (ex: every 10 different opened tubes etc.)

### Cross-contamination tips

**Perfectly clean your workplace and tools before starting anything**

- Change your labcoat for a clean one
- Wear a mask
- Open new materials for the occasion
- Do not assume reagents history
- Do not open two different tubes at the same time
- Always change your cones between samples
- Immediately change gloves if contaminated
- Regularly change gloves (ex: every 10 different opened tubes etc.)
- Prepare all needed tubes in advance. Identify them all



## Wet-lab : Blanks



CRITICAL

- Extraction blank
- H<sub>2</sub>O blank
- PCR blank
- Mock community (ex: Logarithmic DNA)

Blanks have to be MEANINGFUL ! If several extractions, several extraction blanks: avoid pooling them if you can. If several PCR : H<sub>2</sub>O, PCR and Mock each time.

Consider blanks before experimenting. Always count them as samples.



CRITICAL

**Blanks are not just to publish**

**Without them, your experiment is  
WORTHLESS.**

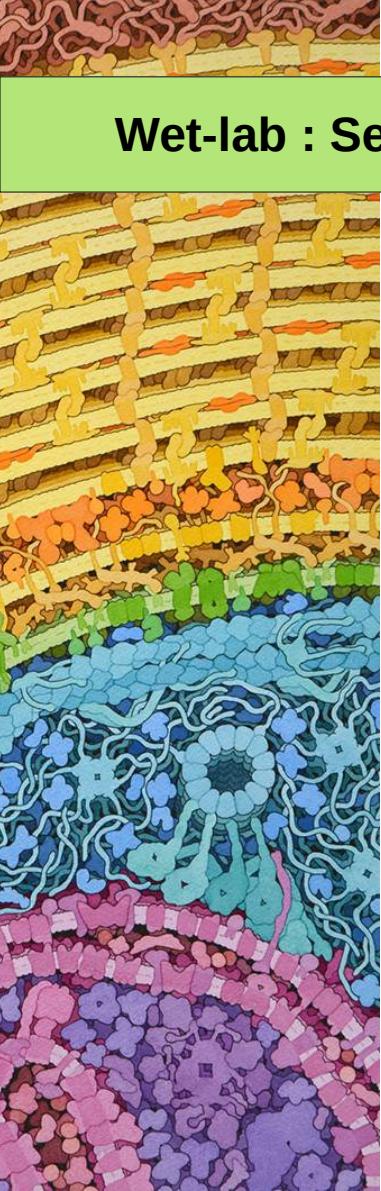
**If a contamination occur, you can retrace it,  
and maybe keep some data, avoiding loosing  
all at once**

## Wet-lab : Amplification

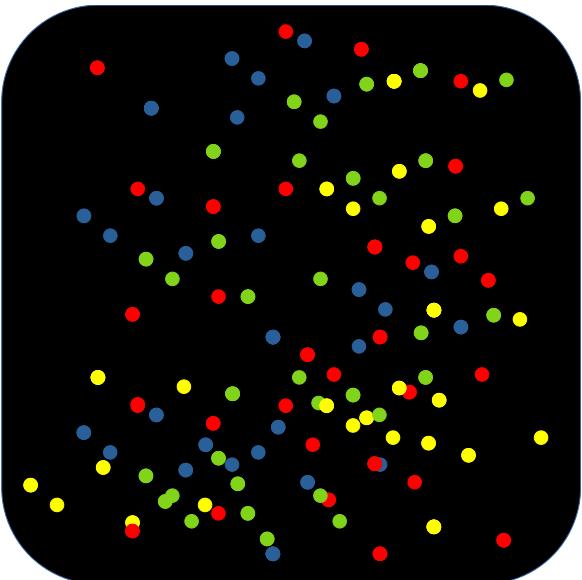
# Amplification



Mix and retrieve sample origin through multiplexing

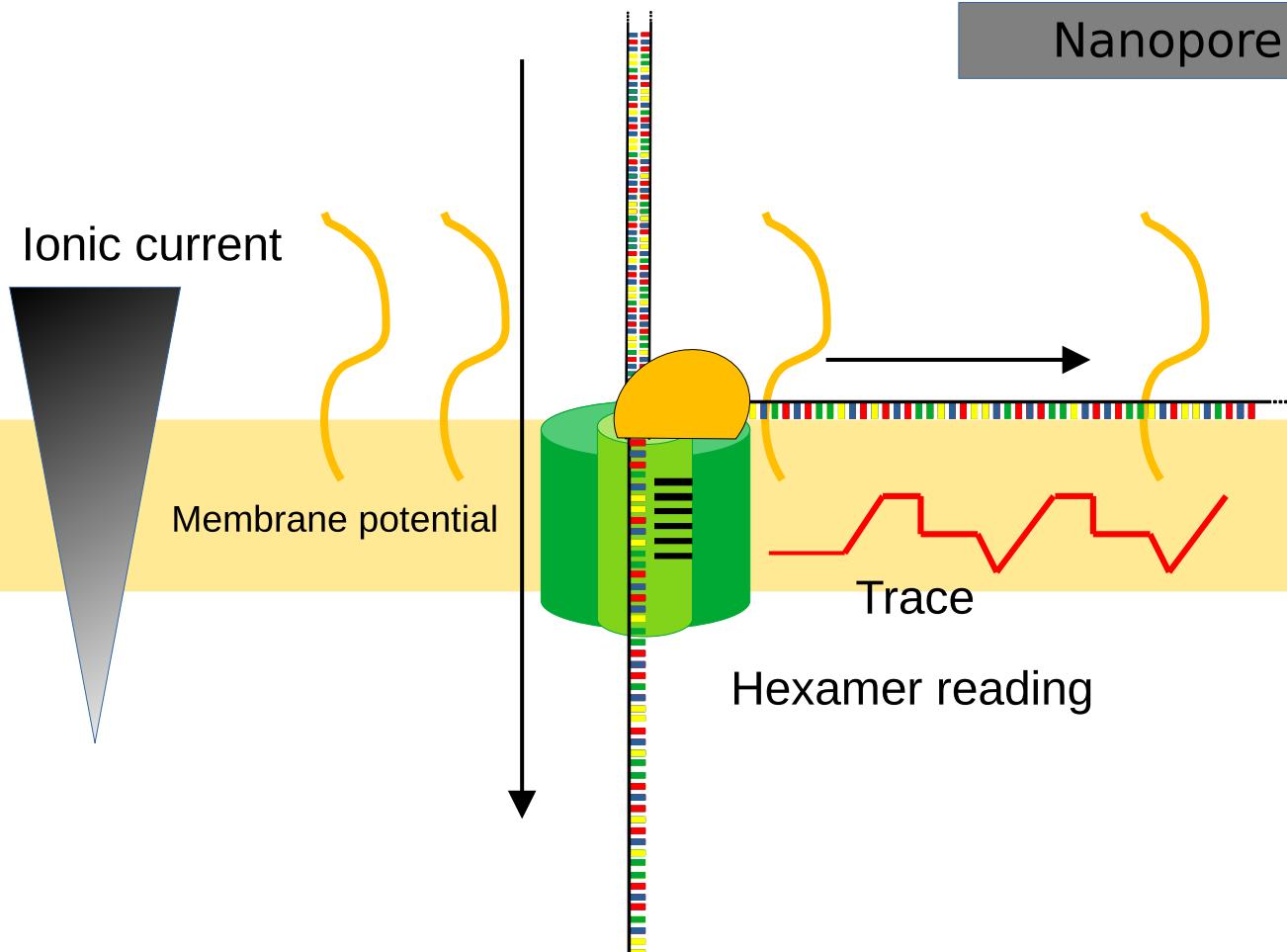
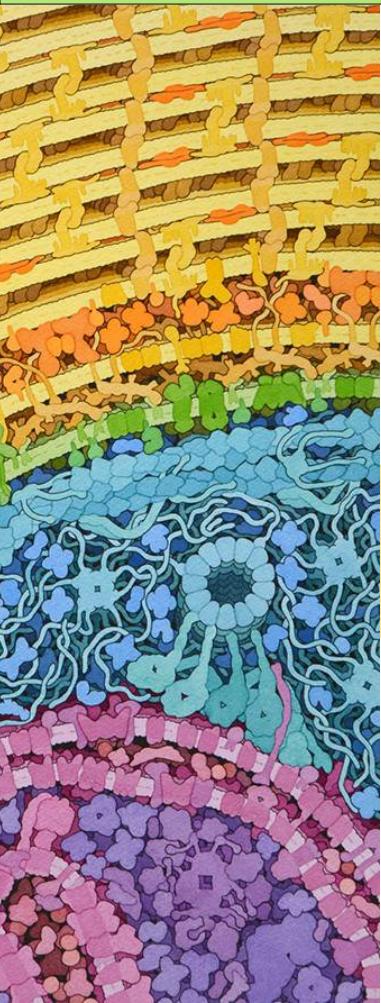


## Wet-lab : Sequencing



**Illumina library  
preparation**

## Wet-lab : Sequencing



## Dry-lab : Quality Check



- Primer free sequences
- Below quality threshold sequences
- Inappropriate CG%
- ...

## Dry-lab : Software Cheat Sheet

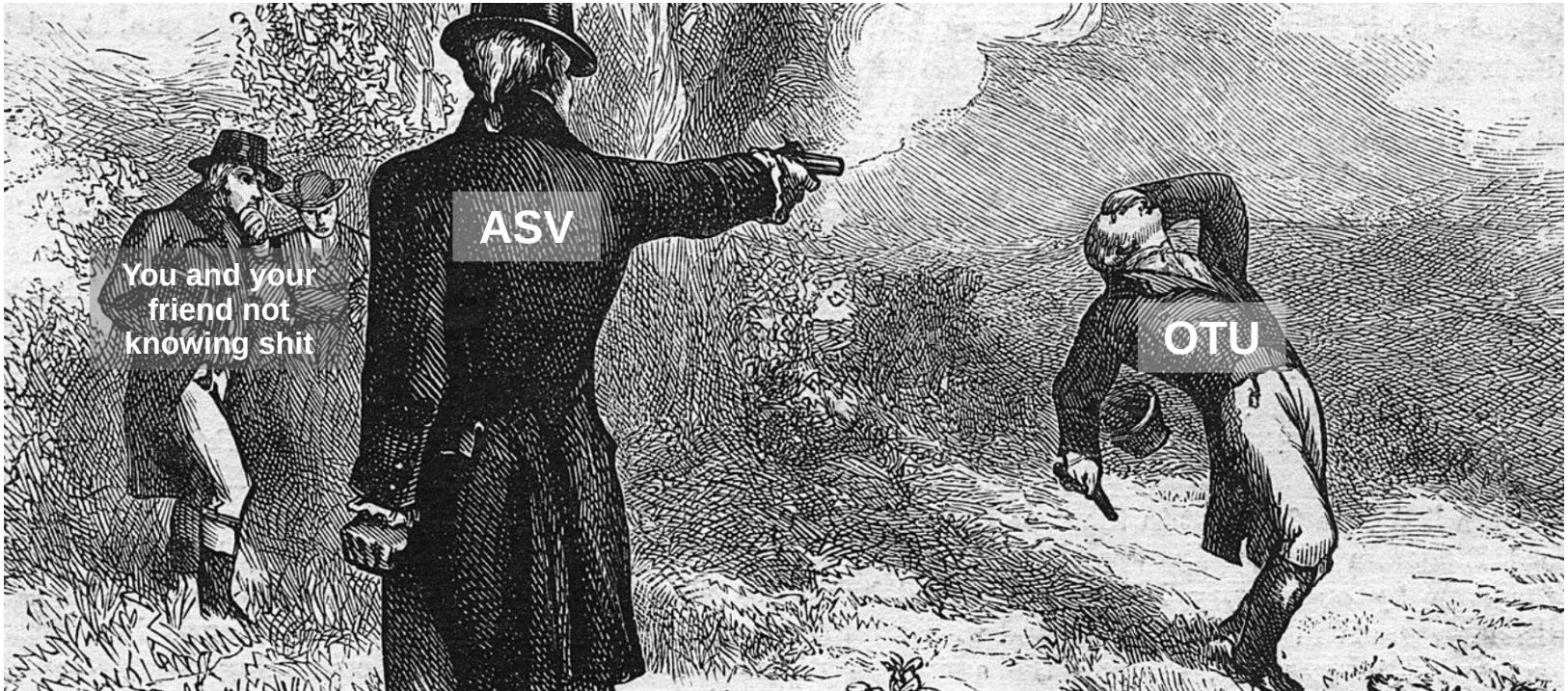
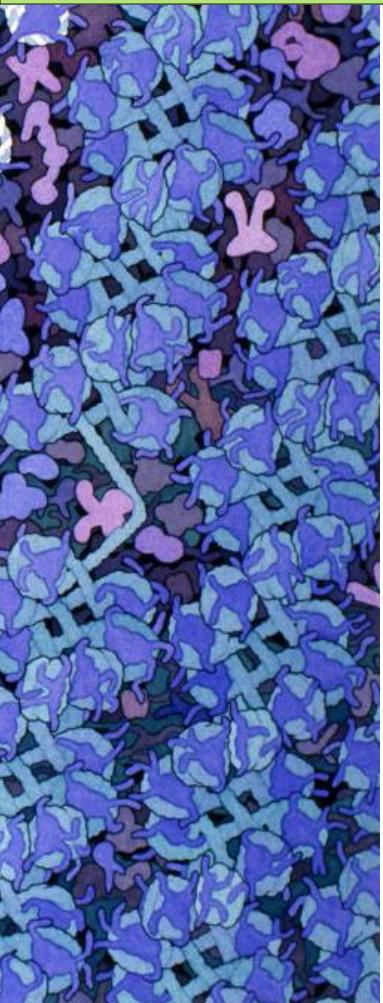
**Quality check :**  
**FastQC – MultiQC**

**Data processing :**  
**End-to-end solutions :**  
**FROGS**  
**nf-core Ampliseq**  
**DADA2**

**Other software :**  
**QIIME**  
**Vsearch**  
**Blast**  
...

**Check IFB (Institut Français de Bioinformatique) and  
Genotoul Plateform for more insights**

# Dry-lab : OTU or ASV ?



## Dry-lab : OTU or ASV ?

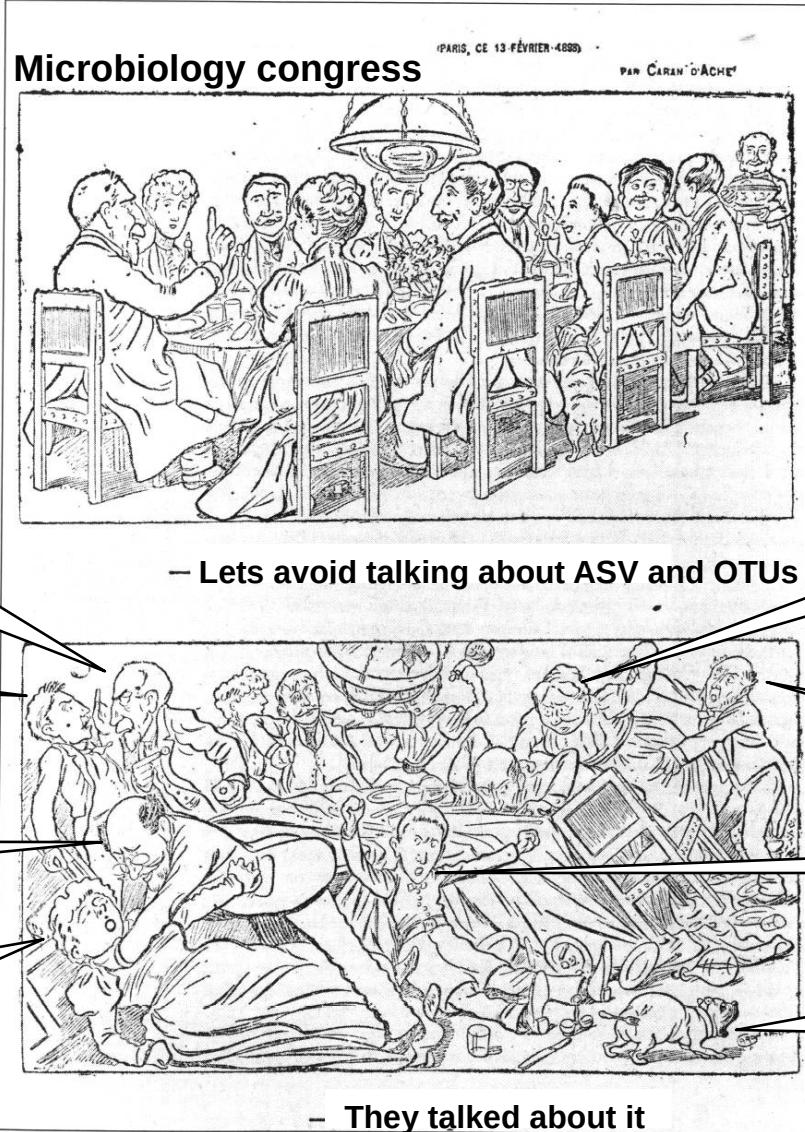
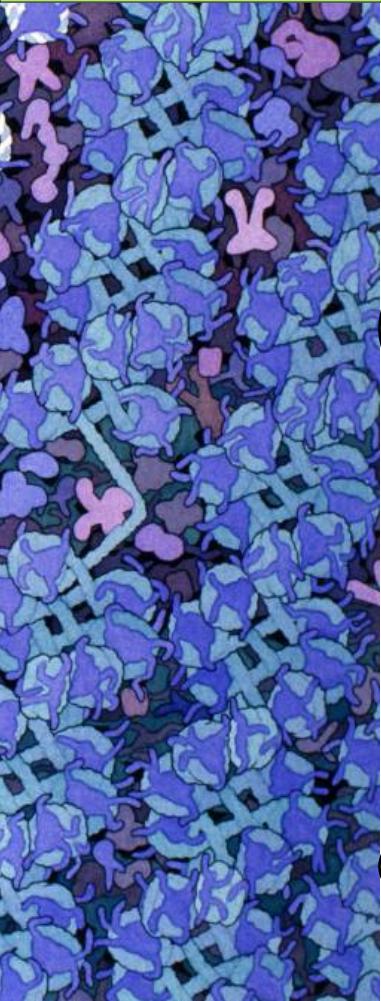
### **ASV : Amplicon Sequence Variant**

An ASV represents an exact sequence as acquired by the sequencer

### **OTU : Operational Taxonomic Unit**

An OTU is a consensus sequence built from highly close sequences (97-99% identity)  
Technically speaking, an OTU is a grouping of closely related ASV...

# Dry-lab : OTU or ASV ?



Back in my days  
there were  
No millenial ASV  
bullshit

ASV are biological  
entities

OTUs are  
unreproducible

ASVs explode my  
Diversity indices

OTUs are  
Outdated !

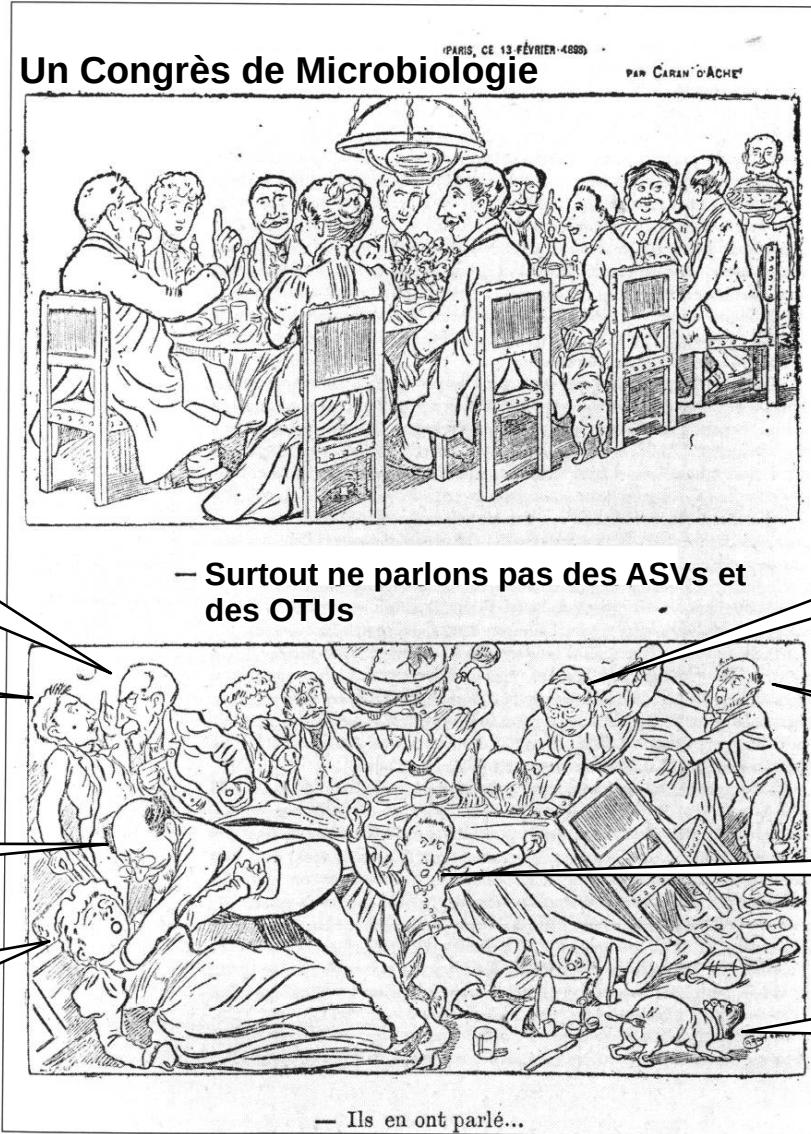
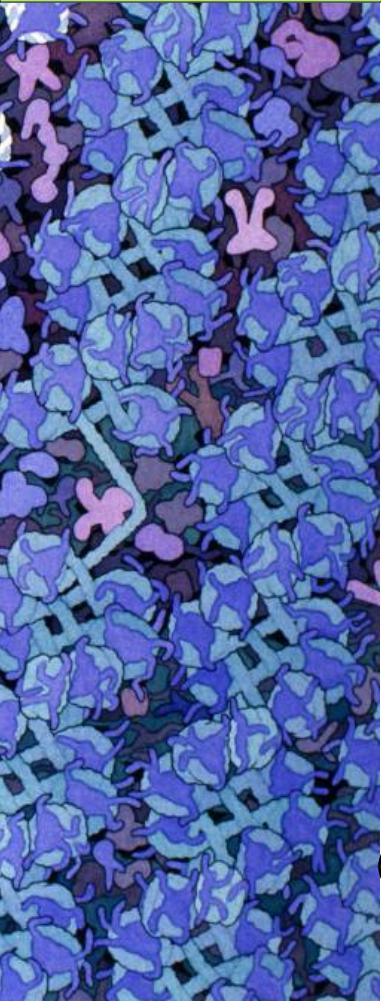
But so easy  
on Galaxy

We cluster ASVs  
In my lab !

PCR and sequencing are too  
Unreliable to directly use  
Amplicons !

– They talked about it

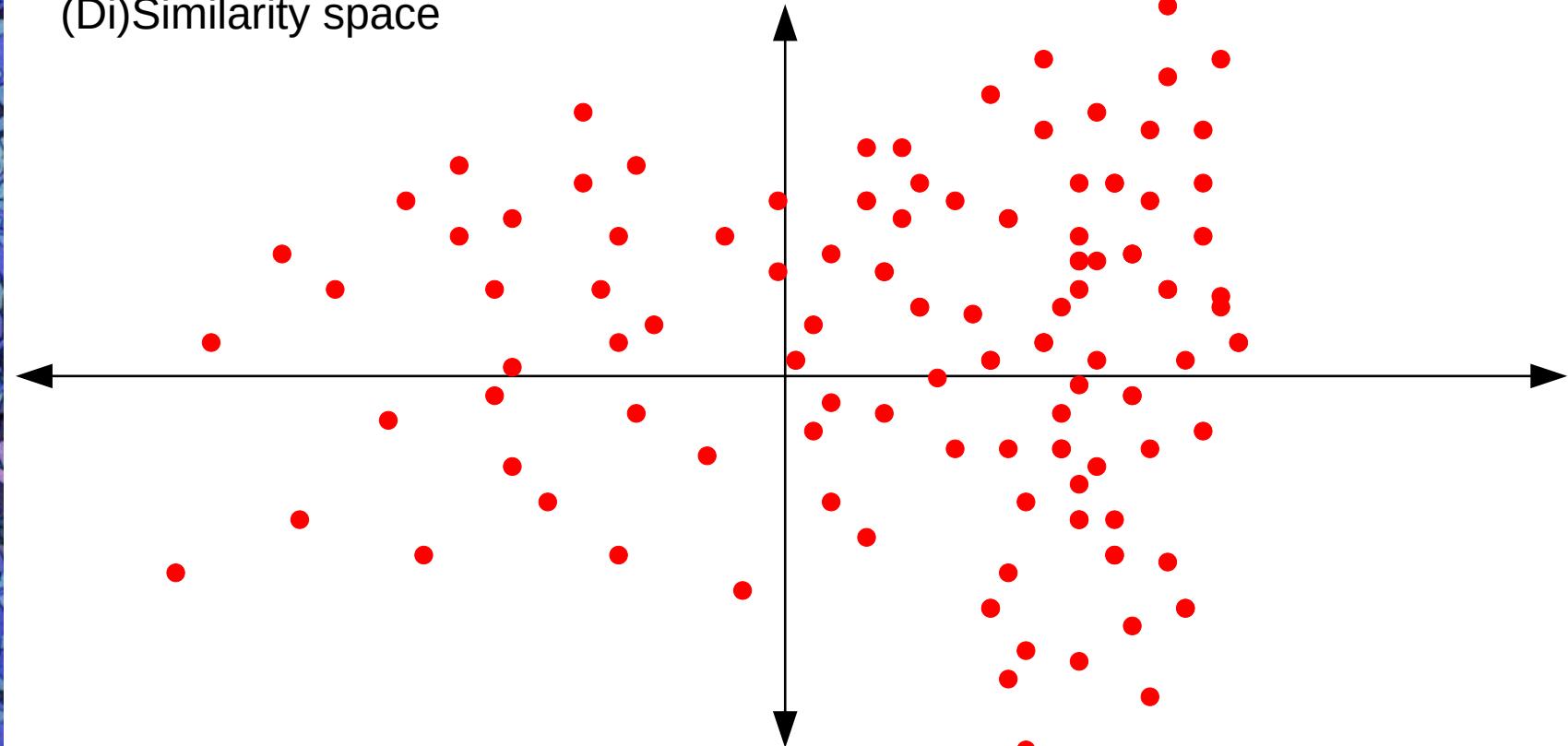
# Dry-lab : OTU or ASV ?



## Dry-lab : Clustering (OTU)



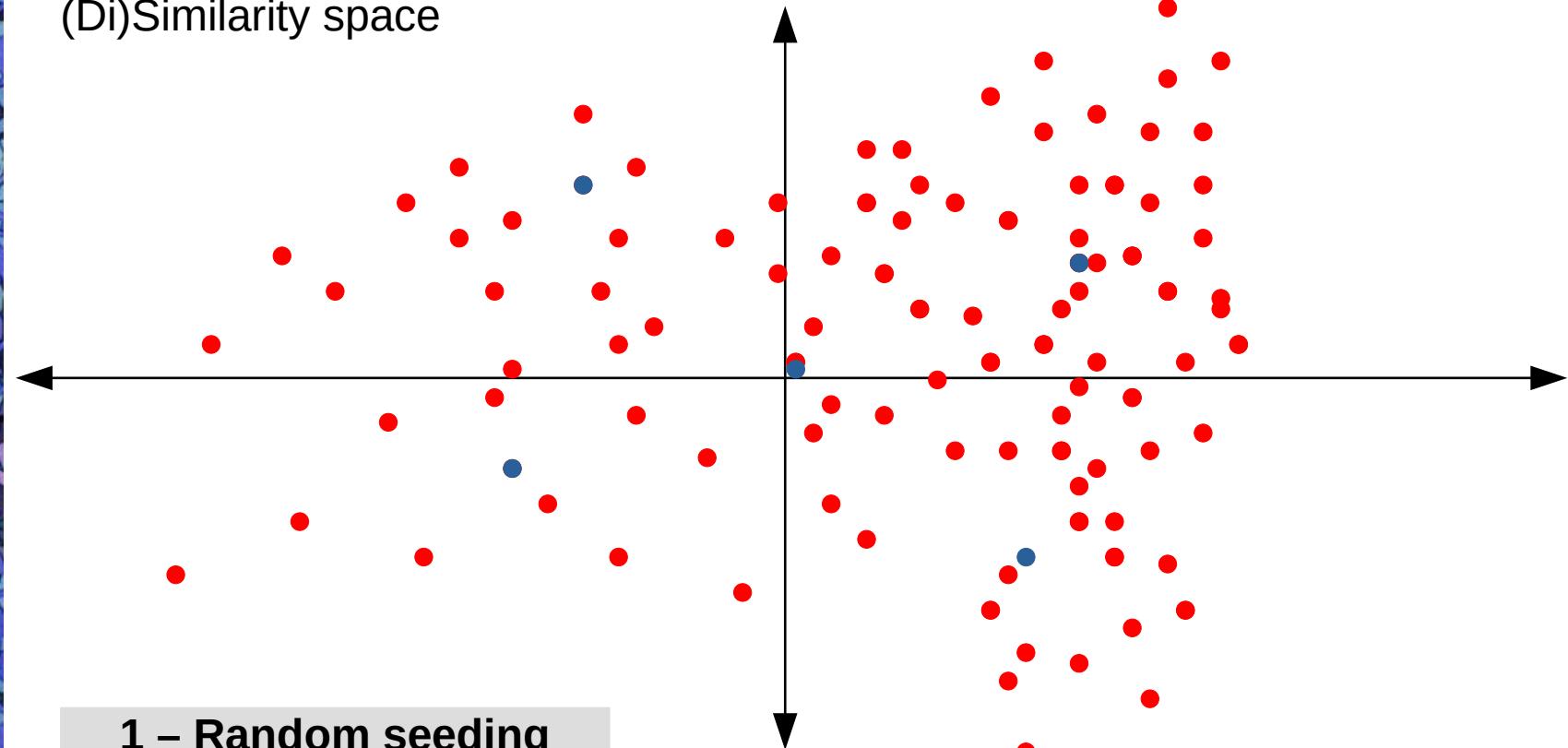
(Di)Similarity space



## Dry-lab : Clustering (OTU)



(Di)Similarity space

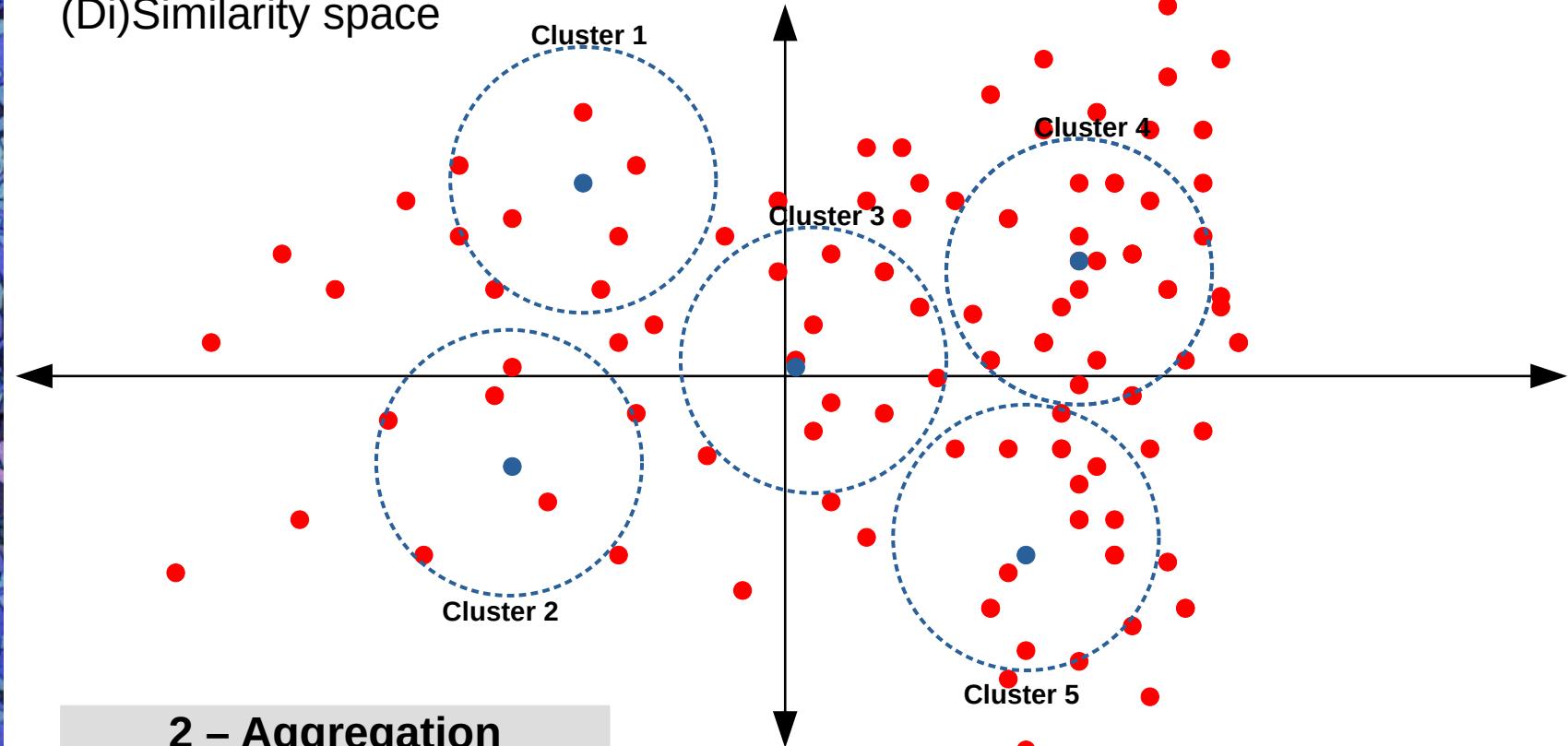


1 – Random seeding

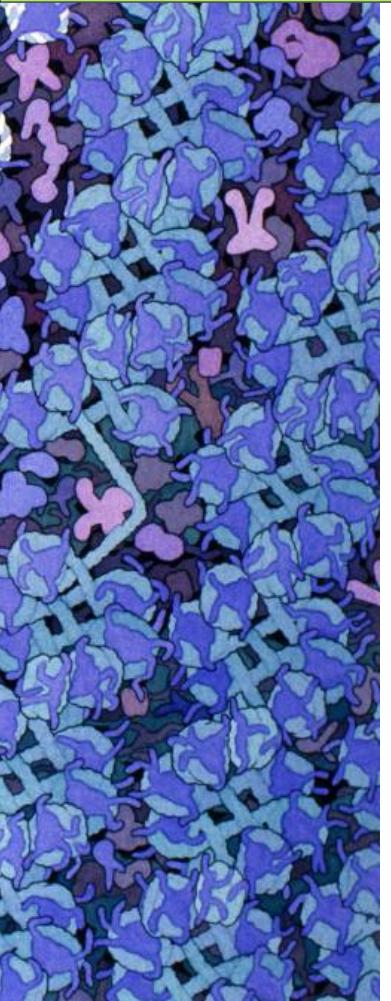
## Dry-lab : Clustering (OTU)



(Di)Similarity space



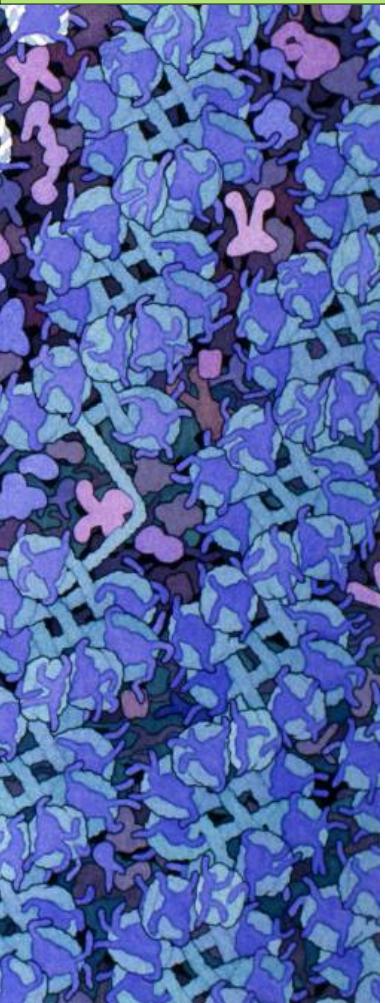
## Dry-lab : Clustering (OTU)



```
>Cluster_1  
AAATTCTCGGCTACGGCTAACGT CGCGCTACGTTAGCGCCTGACTGCG  
CTCGACTCCCGCGTCATCGTAGCTAGCTCGTAGCTG  
>Cluster_2  
AAATTCTCGGCTACGGCTTTCGT CGCGCTACGTTAGCGCCTGACTGCG  
CTCGACTCCCGCGTCATCGTAGCTAGCTCGTAGCTG  
>Cluster_3  
AAATTCTCGATCGGCTAACGT CGCGCTACGTTAGCGCCTGACTGCGCT  
CGACTCCCGCGTCATCGTAGCTAGCTCGTAGCTG  
>Cluster_4  
AAATTCTCGGCTACGGCTAACGT CGCGCTACGTTAGCGCCTGACTGCG  
CTCGACTCTTCGTCACTCGTAGCTAGCTCGTAGCTG
```

### 3 – Consensus sequences

## Dry-lab : Chimera Removal



**Cluster 1**

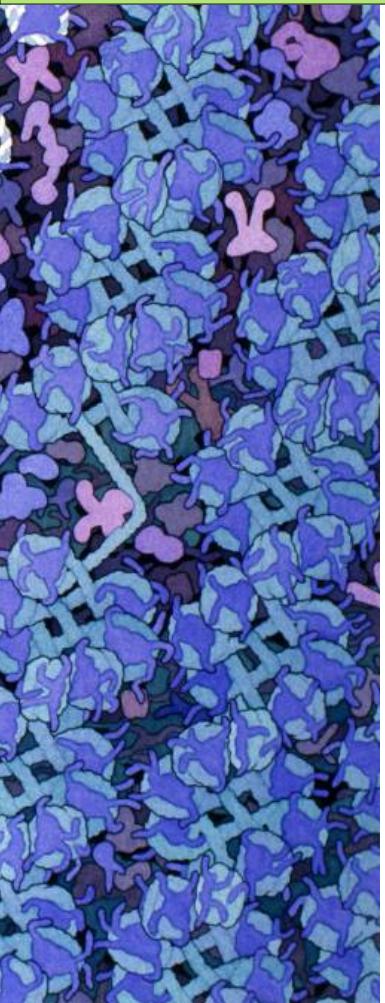
**Cluster 2**

**Cluster 3**

**Cluster 4**

**Cluster 5**

## Dry-lab : Chimera Removal



**Cluster 1**



**Cluster 2**



**Cluster 3**



**Cluster 4**

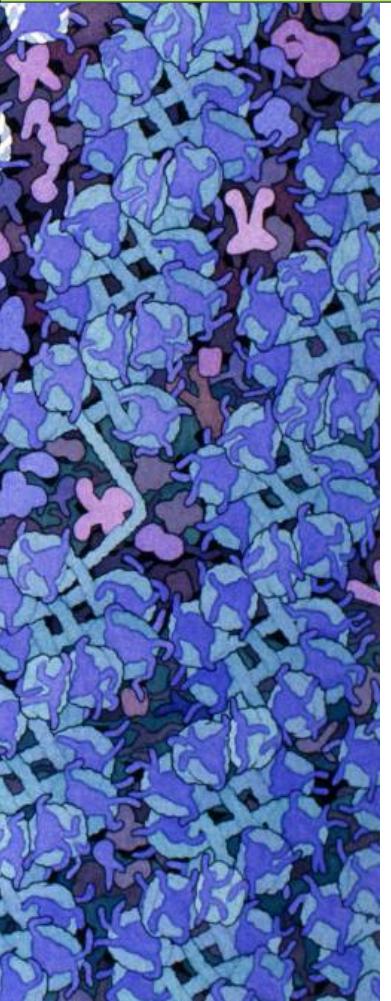


**Cluster 5**



Chimeric read = PCR artifact (?)

## Dry-lab : OTU filtering

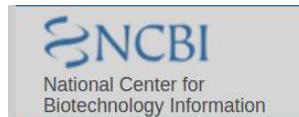


Use your favorite parameters to discriminate what you will keep from what you'll discard

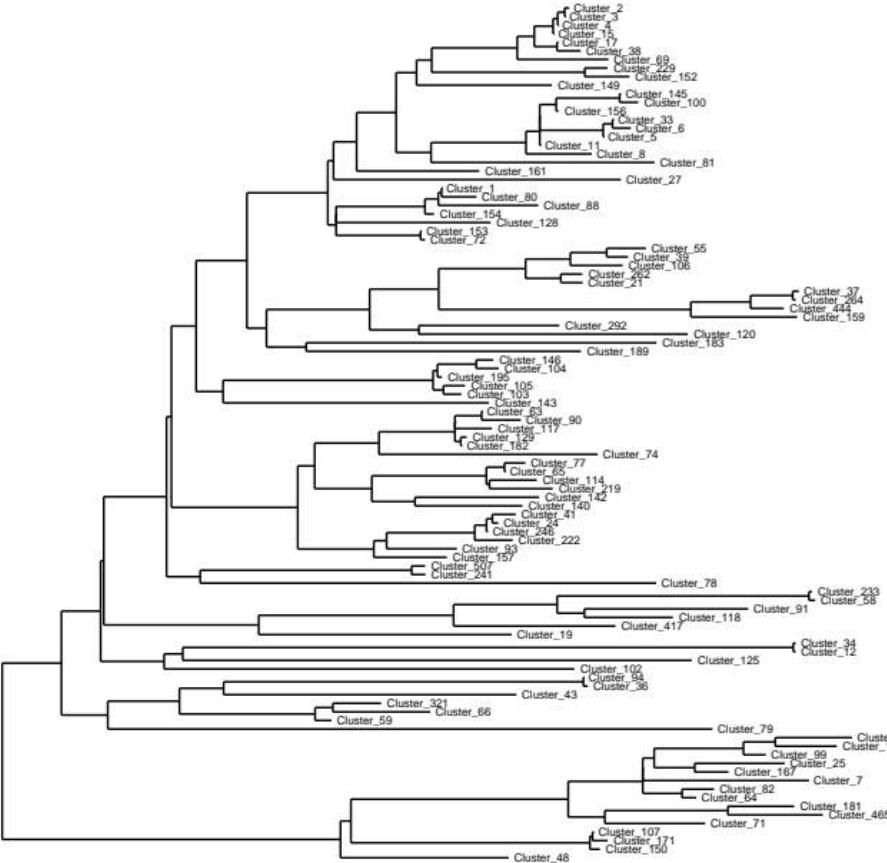
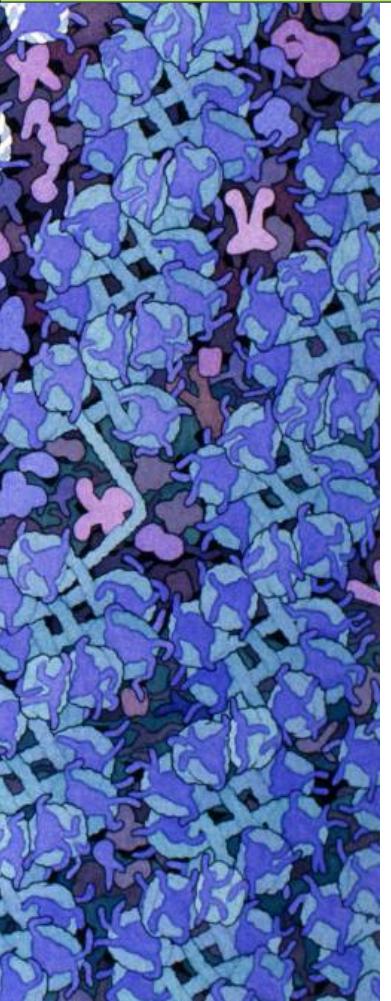
Check your blanks and control, discard contaminants found in these

## Dry-lab : Taxonomic Affiliation

Blast all your consensus sequences against a cured database  
(ex : Sylva, GreenGenes, 16S\_nr\_NCBI...)



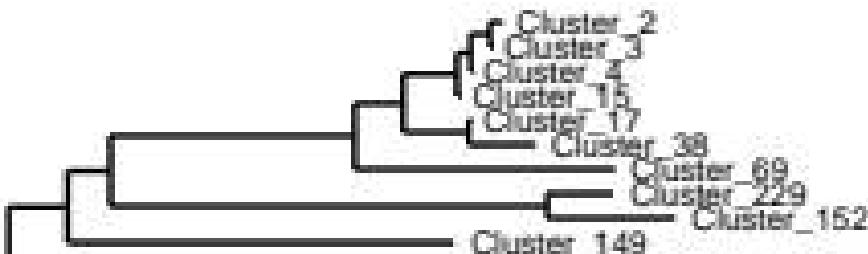
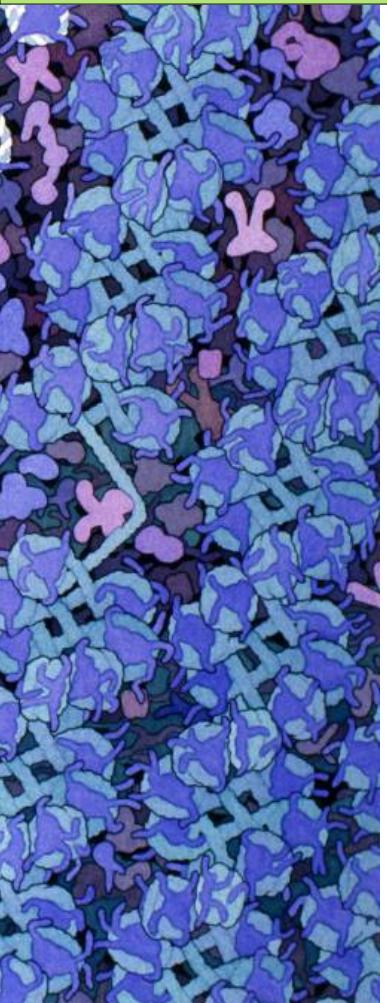
## Dry-lab : Build the tree



Here the tree is made  
with only ~500bp highly  
conserved 16S rRNA

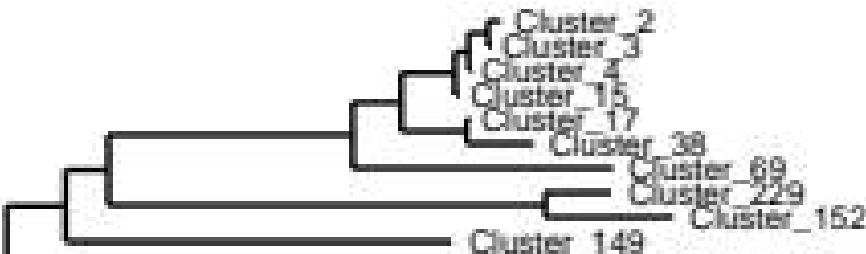
Multi-copy 16S rRNA  
operon ?

## Dry-lab : Understand the tree



Here, I think (no real proof) that 2-3-4 are the same Bacteria with multi-copy genomes or 16S rRNA operon.

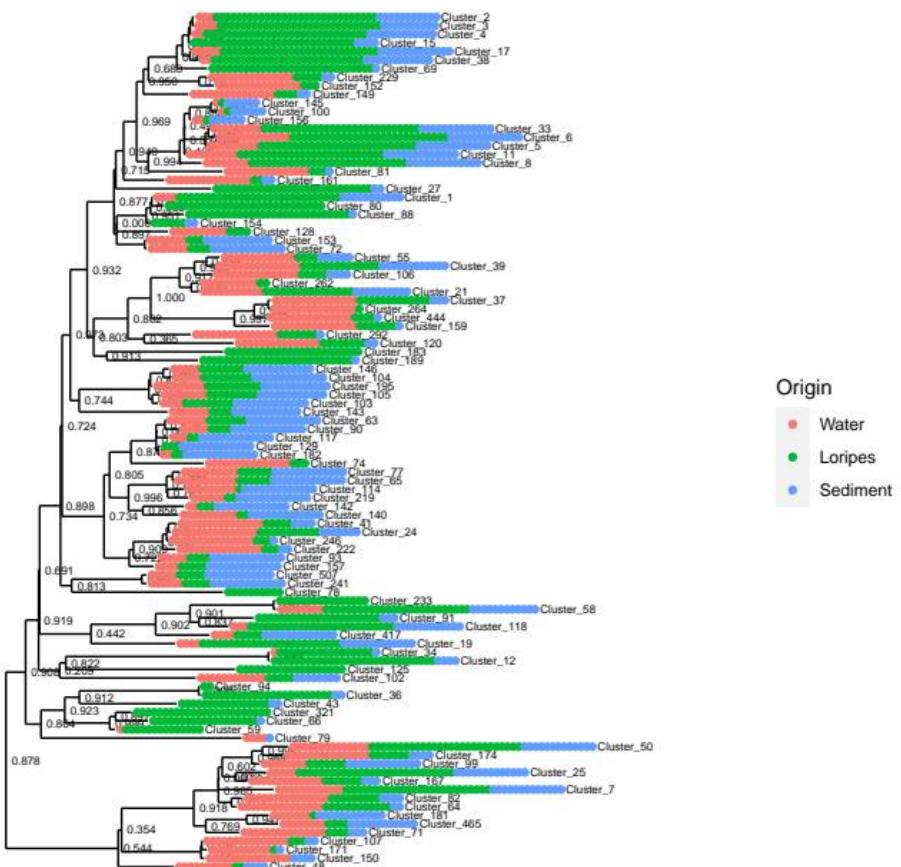
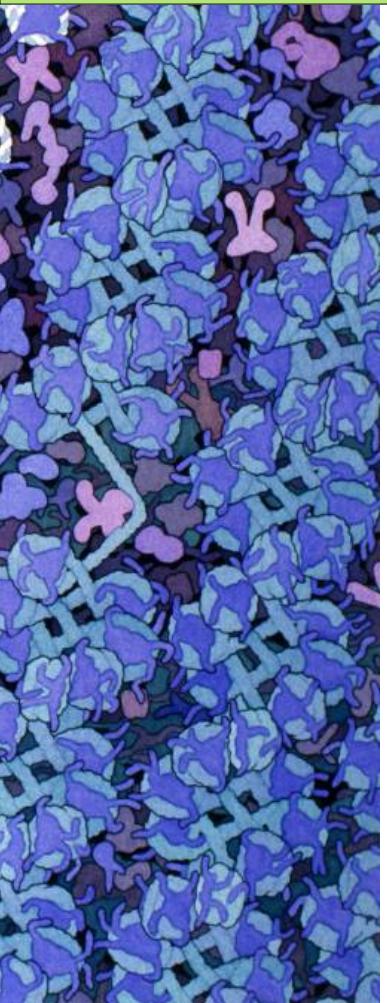
## Dry-lab : Understand the tree



Here, I think (no real proof) that 2-3-4 are the same Bacteria with multi-copy genomes or 16S rRNA operon.

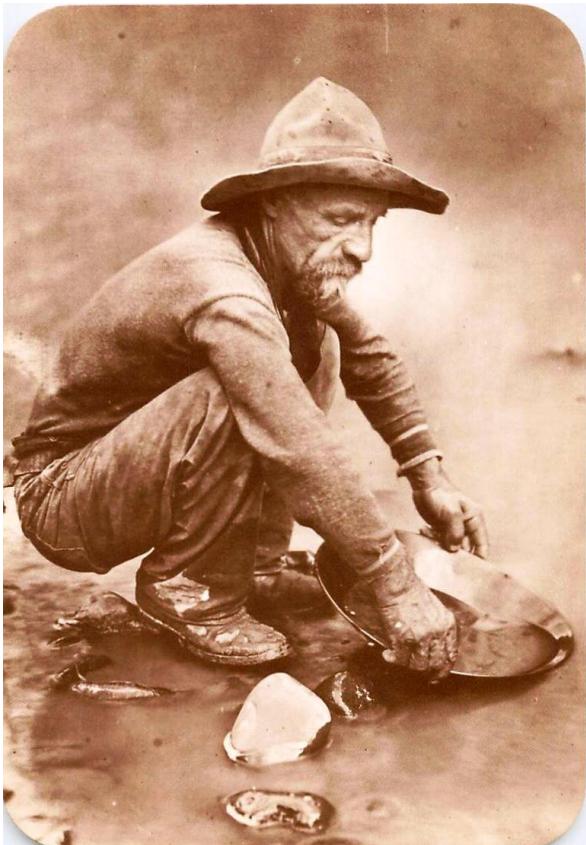
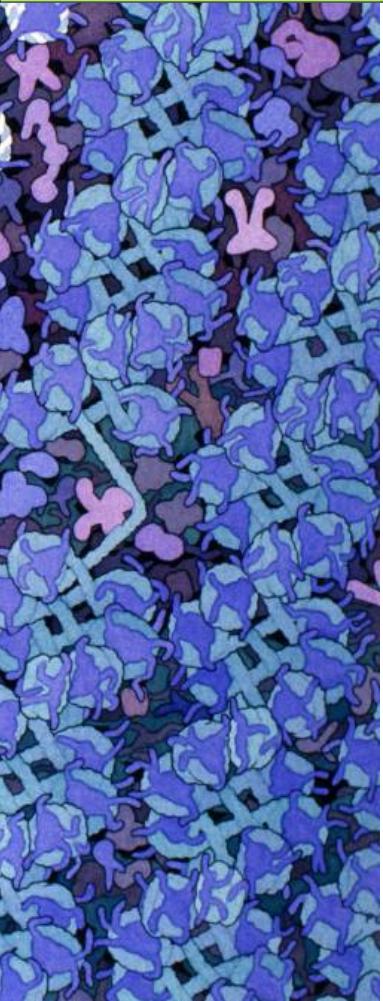
What should I do ?  
Do I have to acknowledge this for their relative abundance ?  
→ Metabarcoding strongest limitation

# Dry-lab : Understand the tree



Phylogenetic  
specialization ?

# Prospecting



If similarity threshold of 16S rRNA sequence with a cultured species is below XX, you discovered :



- 98.65% : A new species
- 95% : A new Genus
- 87% : A new Family
- 82% : A new Order
- 79% : A new Class
- 75% : A new phylum

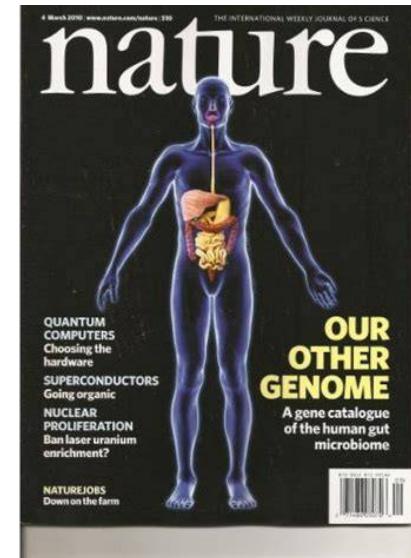
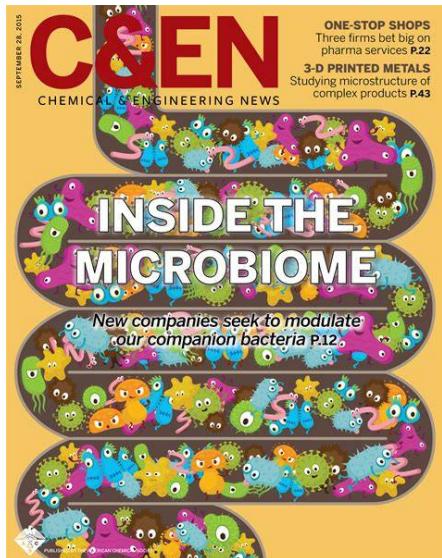
# Let's do Science !

## OTU table (Abundance table)

Let's discuss

# Applications

# The Human Microbiome Project



# The Human Microbiome Project

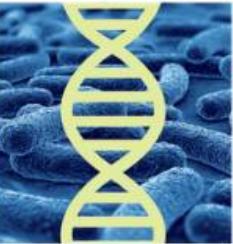


NIH Human Microbiome Project

A silhouette of a person standing with arms raised inside a circular wheel divided into colored segments.

Characterization of the microbiomes of healthy human subjects at five major body sites, using 16S and metagenomic shotgun sequencing.

Enter HMP1

A close-up image of a DNA double helix structure.

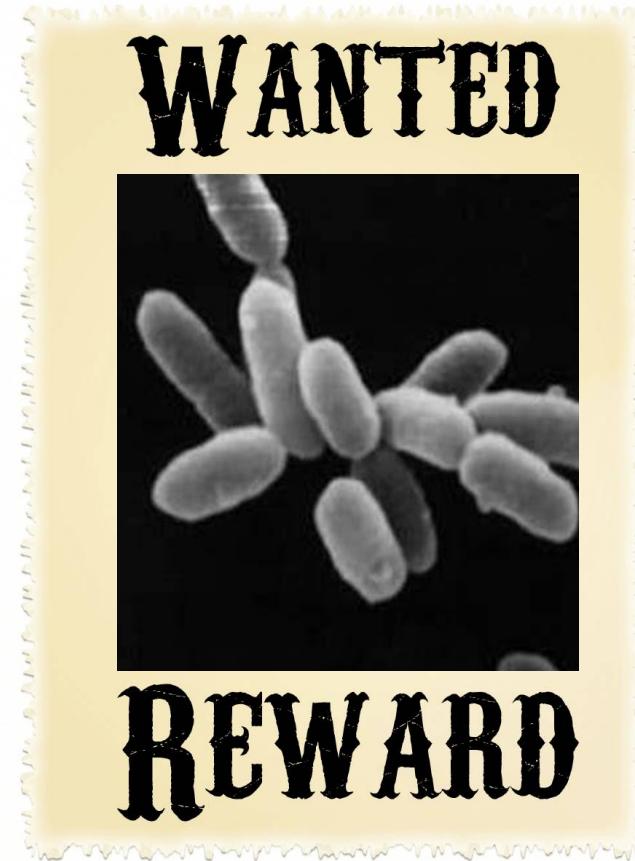
Characterization of microbiome and human host from three cohorts of microbiome-associated conditions, using multiple 'omics technologies.

Enter iHMP

Enter HMP Integrated Portal 



# The Human Microbiome Project





## Communities Investigation

Is it really Tuna ?

Is there Horse in my Beef ?

Which animals do live in this cave ?

Are those individuals from the same species ?



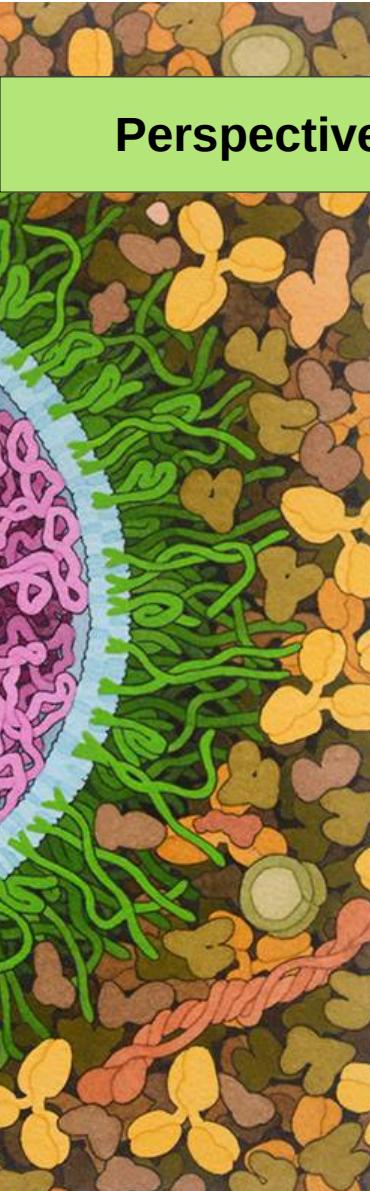
## Environmental biomonitoring

Assess spatio-temporal variation in your communities.

Predict shifts, inform authorities.

Reveal pollution in natural environment.

Investigate Soil health



## Perspectives for Metabarcoding – Biases linked to *barcodes*

**Multi-copy 16S operon**

**Don't blindly trust databases,  
especially for your favorite organisms**

**Was it really a chimera ?**

## Alternatives

**RpoB gene**

**House keeping genes**

**Problem of horizontal transfers**

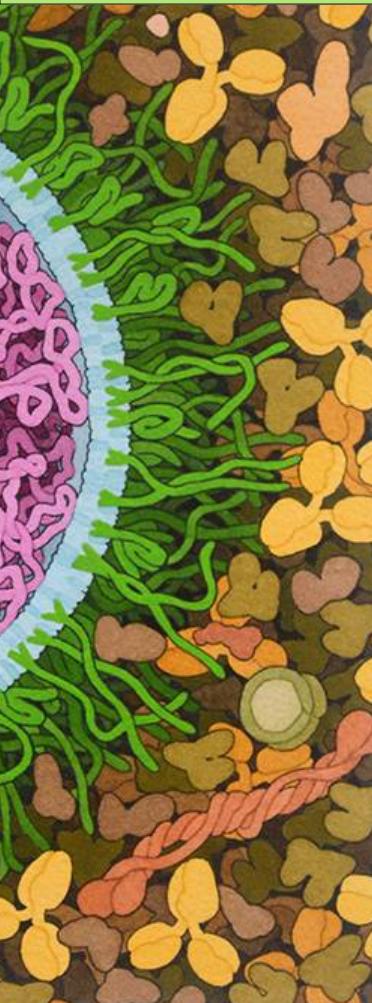
**Highly disturbed environment ?**

# Biases in Databases

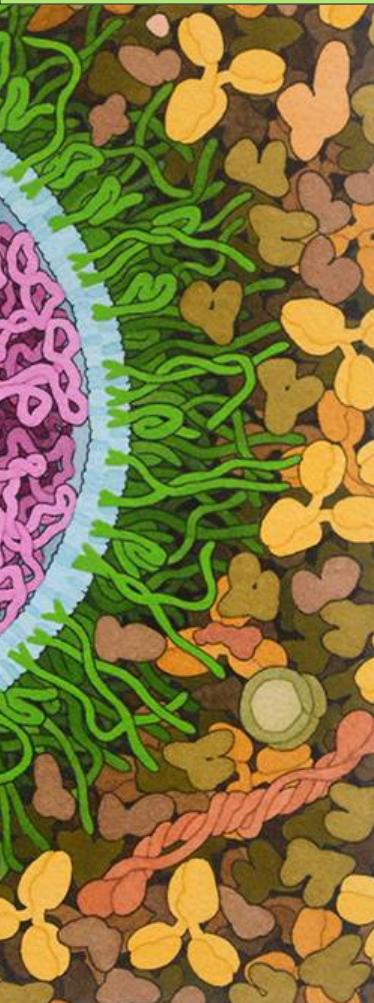
This is an important OTU in my dataset  
Do I trust it ?

blast_subject	blast_perc_identity	blast_perc_query_coverage	blast_evalue	blast_align_length	seed_id	observation_name	observation_sum
GQ853555.1.1569	99.766	100	0	427	M04485:39:000000000-K2GWT:1:1101:28336:131	Cluster_3	173324

# Biases in Databases



# Biases in Databases



The matrix, 1999

# Biases in Databases

First : Get the reference sequence's accession number

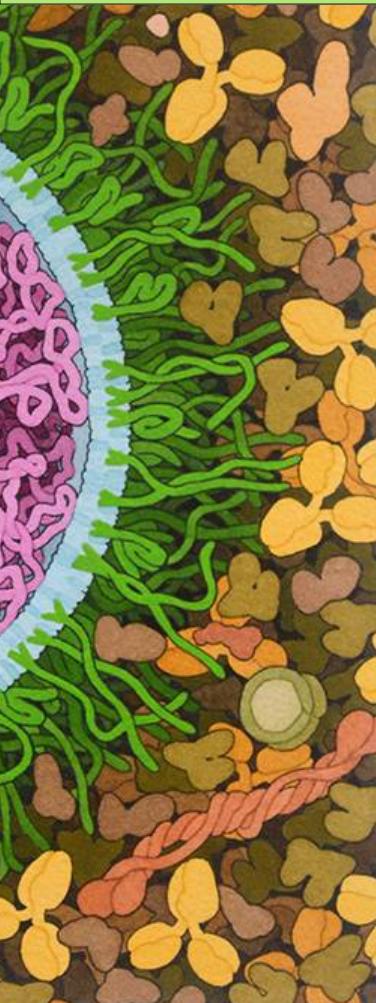
blast_subject	blast_perc_identity	blast_perc_query_coverage	blast_evalue	blast_align_length	seed_id	observation_name	observation_sum
GQ853555.1.1569	99.766	100	0	427	M04485:39:000000000-K2GWT:1:1101:28336:131	Cluster_3	173324

# Biases in Databases

Go back to the database : Who are you ?

```
File Edit View Search Terminal Help
>G0853555.1.1569 Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Endozoicomonadaceae;Endozoicomonas;Loripes lacteus gill symbiont
GAGAGUUUAGAUUAUGGCUCAGAUUGAACGCUUGCGCAGGCCUAACACAUGCAAGUCGAGCGGUAAACAGGGGAGACUUGC
UCUCUGCUGACGAGCGCCGACGGGUGCGUAACACGUAGGAUUCUGCCUGGUAGUGGGGAUAGCCGGAGAAUUCGGGA
UUAUACCGCAUACGCCCAAGGGGAAACGAGGGGAUCCGUUCUUUCCGGAGAGUGGGGUACUUGCGCUAUUGGAUGAGCC
UGCGUCGGAUAGCUUGGUUGGGGAAAGCCGGGUACCAAGGCAACGAUCCGUAGCUGGUAGAGGGGUAGAUACAGCCAC
ACUGGGACUGAGACACGCCAGACCUUACGGGAGGCAGCAGUGGGAAUAUUGCACAAUUGGGGAACCCUGAUGCAG
CCAUGCCGCGUGUGUGAAGAAGGCUCUAGGGUUGUAAGACACUUUACCGAGGAGGAAGGUUGUUGAUUAACUACAC
AGCUGUGACGUUAUCUGCAGAAGAACCCGCUACUCCUGCCAGCAGCCGCGGUAAUACGGAGGGUGCGAGCGUUA
UCGGAAUUAUCUGGGCAGAACGCGUGCGUAGCGCUGGUAGUAGUGUAGAAGCCCAGGUAAUCCUGGGAACUGC
AUCCCAAACAGGGUAGCGUAGAGCGGAAGAGGAGUGGGAAUUCUGGUAGCGGUAGAAGUAGAUAAUAGGAAGG
AAACACAGUGGGCAAGGGCAGACUCUGGUAGACACUGACUGCCAGGUAGCGUAGGGGAGCAAACAGGAUUAAGUA
CCCCUGGUAGUCACGCCGUAACGAUGCUACUGUUGUCCGGAGUCUUGUACUGGUAGCGAAGCUAACCGCAUAG
UAGACCAGCCUGGGGAGUACGGCCGAAGGUAAAACUAAUUGACGGGGGCCGACAAGCGGUGGAGCAU
UUGGUUUAAUUCGAAGCAACCGGAAGAACCUUACCCUGGUAGACAUCCUGGAAUCCUUUAGAGAUAGAGGAGUGCCU
CGGGAACCGCAGAGACAGGGUGCUGCAUGGUGCUGUCAGCUCUGUGUGAGAUGUUGGGGUAAAGUCCCGCAACGAGCGCA
ACCCUUGUCCUAGGUUACCGCAGCUUAGGGGACACUGCCAGGUAGACUGCCGGAGACUACCCGGAGGAAGGUUGGGGAC
GACGUCAAGUCAUAGGCCUUAACGGCAGGGGUACACAGCUGGUACAUAGGUAGCGUUAACAGACGGUUGCCAAGCCGGA
GGUGGAGCUAACUGAGAAGAGUGCAUCGUAGUCCGGAUUUGAGUCUGCAACUCGACUCCAGUAAGUGCGGAAUCGUAGU
AUCGUGAAUCAGAAUGUCACGGGUGAAUACGUUCCGGGCCUUGUACACACCGCCGUACACCAUAGGGAGUGGUUGUC
CAGAAGUGGUAGCCUAACCUUCUCCCCUAGGGGGAGGGCAUCACACGGAGUGAUUCAUGACUGGGGUGAAGUG
UAACAAGGUAGCCUAGGGGAACCUUGGGCUGGUACCCUAAUUAUG
```

# Biases in Databases



## Loripes lacteus gill symbiont

### General information

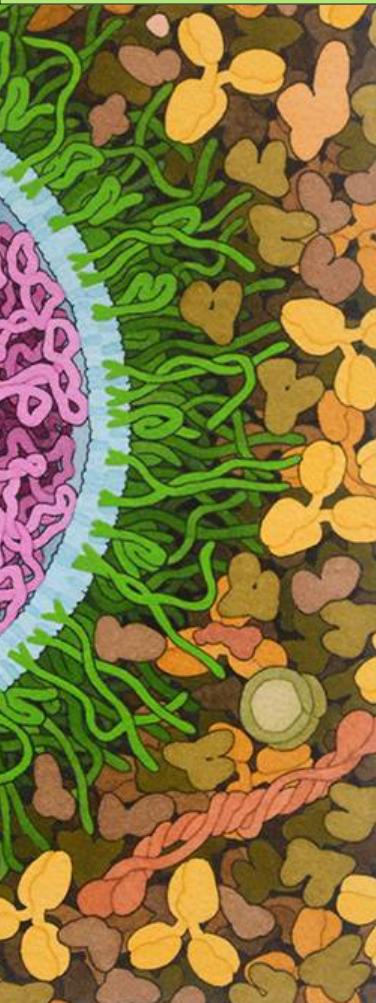
Accession number	GQ853555 ENA ▾
Organism name	<i>Loripes lacteus gill symbiont</i>
Obsolete names	--
Strain	--
Isolate	--
Clone	1A
Length	1569
Date created	2010-09-07
Date modified	2010-09-07
Version	1
Auto Classified	no
Classification Reference	--
Reference Identity	--

### Taxonomy

SILVA	Bacteria ▾ Proteobacteria ▾ Gammaproteobacteria ▾ Pseudomonadales ▾ Endozoicomonadaceae ▾ Endozoicomonas ▾
SILVA Ref	Bacteria ▾ Proteobacteria ▾ Gammaproteobacteria ▾ Pseudomonadales ▾ Endozoicomonadaceae ▾ Endozoicomonas ▾
SILVA Ref NR	Bacteria ▾ Proteobacteria ▾ Gammaproteobacteria ▾ Pseudomonadales ▾ Endozoicomonadaceae ▾ Endozoicomonas ▾
LTP	Unclassified ▾
EMBL-EBI/ENA	Bacteria ▾ Proteobacteria ▾ Gammaproteobacteria ▾ sulfur-oxidizing symbionts ▾
RDP	Bacteria ▾ "Proteobacteria" ▾ Gammaproteobacteria ▾ Oceanospirillales ▾ Hahellaceae ▾ Endozoicomonas ▾
GTDB	Unclassified ▾

Digging deeper with associated online metadata

# Biases in Databases



## Loripes lacteus gill symbiont

### General information

Accession number	GQ853555 ENA ▶
Organism name	<i>Loripes lacteus gill symbiont</i>
Obsolete names	--
Strain	--
Isolate	--
Clone	1A
Length	1569
Date created	2010-09-07
Date modified	2010-09-07
Version	1
Auto Classified	no
Classification Reference	--
Reference Identity	--

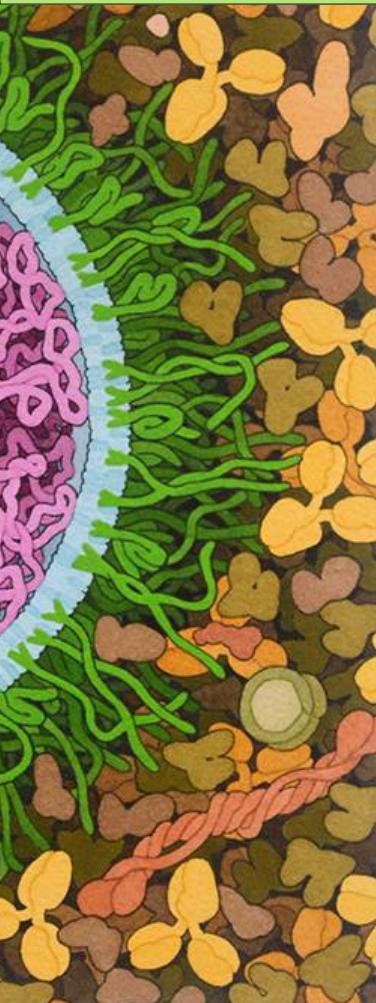
### Taxonomy

SILVA	Bacteria▶Proteobacteria▶Gammaproteobacteria▶Pseudomonadales▶Endozoicomonadaceae▶Endozoicomonas▶
SILVA Ref	Bacteria▶Proteobacteria▶Gammaproteobacteria▶Pseudomonadales▶Endozoicomonadaceae▶Endozoicomonas▶
SILVA Ref NR	Bacteria▶Proteobacteria▶Gammaproteobacteria▶Pseudomonadales▶Endozoicomonadaceae▶Endozoicomonas▶
LTP	Unclassified▶
EMBL-EBI/ENA	Bacteria▶Proteobacteria▶Gammaproteobacteria▶sulfur-oxidizing symbionts ▶
RDP	Bacteria▶"Proteobacteria"▶Gammaproteobacteria▶Oceanospirillales▶Hahellaceae▶Endozoicomonas▶
GTDB	Unclassified▶

The date :

This sequence is old

# Biases in Databases



## Loripes lacteus gill symbiont

### General information

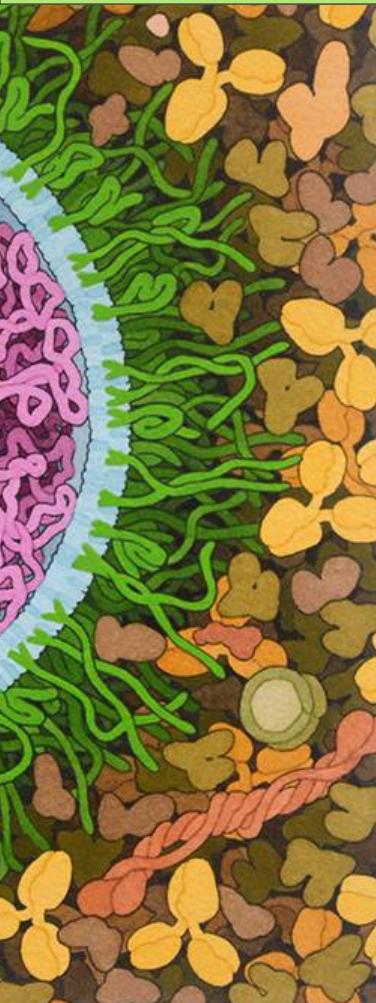
Accession number	GQ853555 ENA ▾
Organism name	<i>Loripes lacteus gill symbiont</i>
Obsolete names	--
Strain	--
Isolate	--
Clone	1A
Length	1569
Date created	2010-09-07
Date modified	2010-09-07
Version	1
Auto Classified	no
Classification Reference	--
Reference Identity	--

It was not corrected

### Taxonomy

SILVA	Bacteria ▾ Proteobacteria ▾ Gammaproteobacteria ▾ Pseudomonadales ▾ Endozoicomonadaceae ▾ Endozoicomonas ▾
SILVA Ref	Bacteria ▾ Proteobacteria ▾ Gammaproteobacteria ▾ Pseudomonadales ▾ Endozoicomonadaceae ▾ Endozoicomonas ▾
SILVA Ref NR	Bacteria ▾ Proteobacteria ▾ Gammaproteobacteria ▾ Pseudomonadales ▾ Endozoicomonadaceae ▾ Endozoicomonas ▾
LTP	Unclassified ▾
EMBL-EBI/ENA	Bacteria ▾ Proteobacteria ▾ Gammaproteobacteria ▾ sulfur-oxidizing symbionts ▾
RDP	Bacteria ▾ "Proteobacteria" ▾ Gammaproteobacteria ▾ Oceanospirillales ▾ Hahellaceae ▾ Endozoicomonas ▾
GTDB	Unclassified ▾

# Biases in Databases



## Loripes lacteus gill symbiont

### General information

Accession number	GQ853555 ENA ▾
Organism name	<i>Loripes lacteus gill symbiont</i>
Obsolete names	--
Strain	--
Isolate	--
Clone	1A
Length	1569
Date created	2010-09-07
Date modified	2010-09-07
Version	1
Auto Classified	no
Classification Reference	--
Reference Identity	--

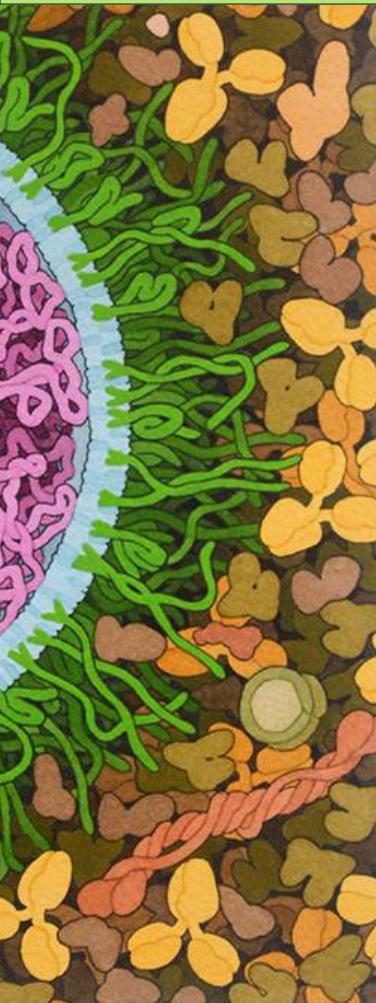
### Taxonomy

SILVA	Bacteria ▾ Proteobacteria ▾ Gammaproteobacteria ▾ Pseudomonadales ▾ Endozoicomonadaceae ▾ Endozoicomonas ▾
SILVA Ref	Bacteria ▾ Proteobacteria ▾ Gammaproteobacteria ▾ Pseudomonadales ▾ Endozoicomonadaceae ▾ Endozoicomonas ▾
SILVA Ref NR	Bacteria ▾ Proteobacteria ▾ Gammaproteobacteria ▾ Pseudomonadales ▾ Endozoicomonadaceae ▾ Endozoicomonas ▾
LTP	Unclassified ▾
EMBL-EBI/ENA	Bacteria ▾ Proteobacteria ▾ Gammaproteobacteria ▾ sulfur-oxidizing symbionts ▾
RDP	Bacteria ▾ "Proteobacteria" ▾ Gammaproteobacteria ▾ Oceanospirillales ▾ Hahellaceae ▾ Endozoicomonas ▾
GTDB	Unclassified ▾

Nor automatically  
Classed with an  
Algorithm

Human did this.

# Biases in Databases



Loripes lacteus gill symbiont

General information

Accession number GQ853555 ENA ▶

Organism name *Loripes lacteus gill symbiont*

Obsolete names

Strain

Isolate

Clone

Length

Date created

Date modified

Version

Author

Comments

SILVA Bacteria▶Proteobacteria▶Gammaproteobacteria▶Pseudomonadales▶Endozoicomonadaceae▶Endozoicomonas▶

SILVA Ref Bacteria▶Proteobacteria▶Gammaproteobacteria▶Pseudomonadales▶Endozoicomonadaceae▶Endozoicomonas▶

SILVA Ref NR Bacteria▶Proteobacteria▶Gammaproteobacteria▶Pseudomonadales▶Endozoicomonadaceae▶Endozoicomonas▶

LTP Unclassified▶

EMBL-EBI/ENA Bacteria▶Proteobacteria▶Gammaproteobacteria▶sulfur-oxidizing symbionts ▶

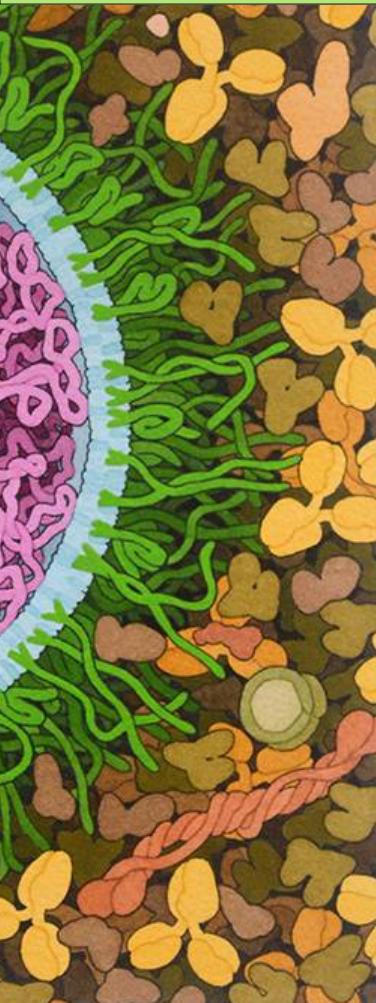
RDP Bacteria▶"Proteobacteria"▶Gammaproteobacteria▶Oceanospirillales▶Hahellaceae▶Endozoicomonas▶

GTDB Unclassified▶

!

Human bad !

# Biases in Databases



## Loripes lacteus gill symbiont

### General information

Accession number	GQ853555 ENA ▶
Organism name	<i>Loripes lacteus gill symbiont</i>
Obsolete names	--
Strain	--
Isolate	--
Clone	1A
Length	1569
Date created	2010-09-07
Date modified	2010-09-07
Version	1
Auto Classified	no
Classification Reference	--
Reference Identity	--

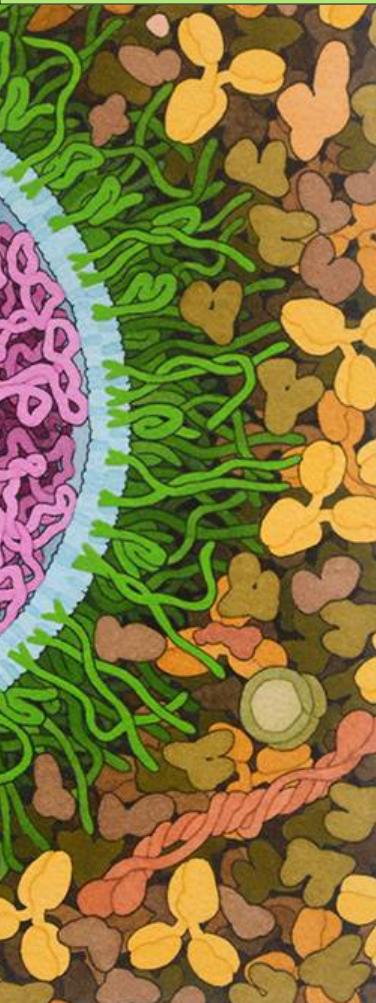
### Taxonomy

SILVA	Bacteria▶Proteobacteria▶Gammaproteobacteria▶Pseudomonadales▶Endozoicomonadaceae▶Endozoicomonas▶
SILVA Ref	Bacteria▶Proteobacteria▶Gammaproteobacteria▶Pseudomonadales▶Endozoicomonadaceae▶Endozoicomonas▶
SILVA Ref NR	Bacteria▶Proteobacteria▶Gammaproteobacteria▶Pseudomonadales▶Endozoicomonadaceae▶Endozoicomonas▶
LTP	Unclassified▶
EMBL-EBI/ENA	Bacteria▶Proteobacteria▶Gammaproteobacteria▶sulfur-oxidizing symbionts ▶
RDP	Bacteria▶"Proteobacteria"▶Gammaproteobacteria▶Oceanospirillales▶Hahellaceae▶Endozoicomonas▶
GTDB	Unclassified▶

Do you trust them ?

Who are they ?

# Biases in Databases



## Loripes lacteus gill symbiont

### General information

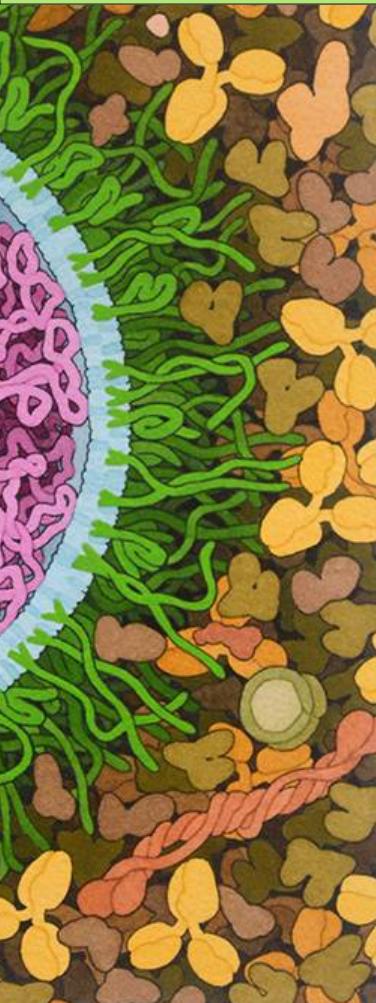
Accession number	GQ853555 <a href="#">ENA</a>
Organism name	<i>Loripes lacteus gill symbiont</i>
Obsolete names	--
Strain	--
Isolate	--
Clone	1A
Length	1569
Date created	2010-09-07
Date modified	2010-09-07
Version	1
Auto Classified	no
Classification Reference	--
Reference Identity	--

Different databases  
are agreeing with  
each-other

### Taxonomy

SILVA	Bacteria ➔ Proteobacteria ➔ Gammaproteobacteria ➔ Pseudomonadales ➔ Endozoicomonadaceae ➔ Endozoicomonas ➔
SILVA Ref	Bacteria ➔ Proteobacteria ➔ Gammaproteobacteria ➔ Pseudomonadales ➔ Endozoicomonadaceae ➔ Endozoicomonas ➔
SILVA Ref NR	Bacteria ➔ Proteobacteria ➔ Gammaproteobacteria ➔ Pseudomonadales ➔ Endozoicomonadaceae ➔ Endozoicomonas ➔
LTP	Unclassified ➔
EMBL-EBI/ENA	Bacteria ➔ Proteobacteria ➔ Gammaproteobacteria ➔ sulfur-oxidizing symbionts ➔
RDP	Bacteria ➔ "Proteobacteria" ➔ Gammaproteobacteria ➔ Oceanospirillales ➔ Hahellaceae ➔ Endozoicomonas ➔
GTDB	Unclassified ➔

# Biases in Databases



## Loripes lacteus gill symbiont

### General information

Accession number	GQ853555 ENA ▶
Organism name	<i>Loripes lacteus gill symbiont</i>
Obsolete names	--
Strain	--
Isolate	--
Clone	1A
Length	1569
Date created	2010-09-07
Date modified	2010-09-07
Version	1
Auto Classified	no
Classification Reference	--
Reference Identity	--

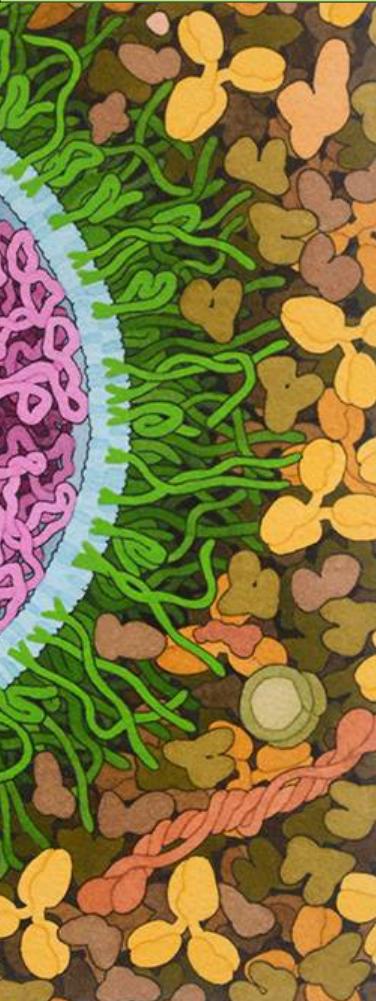
### Taxonomy

SILVA	Bacteria▶Proteobacteria▶Gammaproteobacteria▶Pseudomonadales▶Endozoicomonadaceae▶Endozoicomonas▶
SILVA Ref	Bacteria▶Proteobacteria▶Gammaproteobacteria▶Pseudomonadales▶Endozoicomonadaceae▶Endozoicomonas▶
SILVA Ref NR	Bacteria▶Proteobacteria▶Gammaproteobacteria▶Pseudomonadales▶Endozoicomonadaceae▶Endozoicomonas▶
LTP	Unclassified▶
EMBL-EBI/ENA	Bacteria▶Proteobacteria▶Gammaproteobacteria▶sulfur-oxidizing symbionts ▶
RDP	Bacteria▶"Proteobacteria"▶Gammaproteobacteria▶Oceanospirillales▶Hahellaceae▶Endozoicomonas▶
GTDB	Unclassified▶

Grab something

We'll dive in a strange  
world

# Biases in Databases



**ENA**  
European Nucleotide Archive

Home | Submit ▾ | Search ▾ | Rulespace | About ▾ | Support ▾

Sequence: GQ853555.1

Loripes lacteus gill symbiont clone 1A 16S ribosomal RNA gene, partial sequence.

**Organism:** Loripes lacteus gill symbiont

**Accession:** GQ853555

**Mol Type:** genomic DNA

**Topology:** LINEAR

**Base Count:** 1569

**Dataclass:** STD

**Tax Division:** PRO

**Host:** Loripes lacteus isolate 0

**Md5 Checksum:** 6790737f85f29414a39c65a525f26c12

**No explicit taxonomic attribution here**

Show Less

# Biases in Databases

## Navigation & Cross References

- Taxon:

Taxon:682670

- RFAM:

[RF00177](#) (Click here to see all ENA records for this RFAM ID),  
[RF01959](#) (Click here to see all ENA records for this RFAM ID)

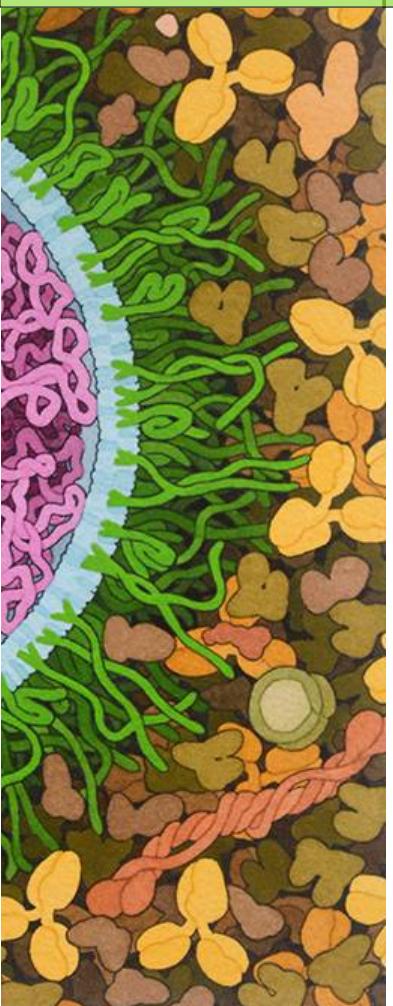
- SILVA-SSU:

[GQ853555](#) (Click here to see all ENA records for this SILVA-SSU ID)

**Let's keep digging**



# Biases in Databases



Taxon: 682670

Loripes lacteus gill symbiont [species]

## Lineage

Bacteria, Proteobacteria, Gammaproteobacteria, sulfur-oxidizing symbionts

Different taxonomic attribution, no source

## Related ENA Records



Result	Count	Bases	Taxon descendants count	Bases
Non-coding	8	12kB	8	12kB
Sequence	8	12kB	8	12kB

# Biases in Databases

## Navigation & Cross References

- Taxon:

[Taxon:682670](#)

- RFAM:

[RF00177](#) (Click here to see all ENA records for this RFAM ID),  
[RF01959](#) (Click here to see all ENA records for this RFAM ID)

- SILVA-SSU:

[GQ853555](#) (Click here to see all ENA records for this SILVA-SSU ID)

**Nothing really interesting  
for us right now in here  
neither**

?

# Biases in Databases

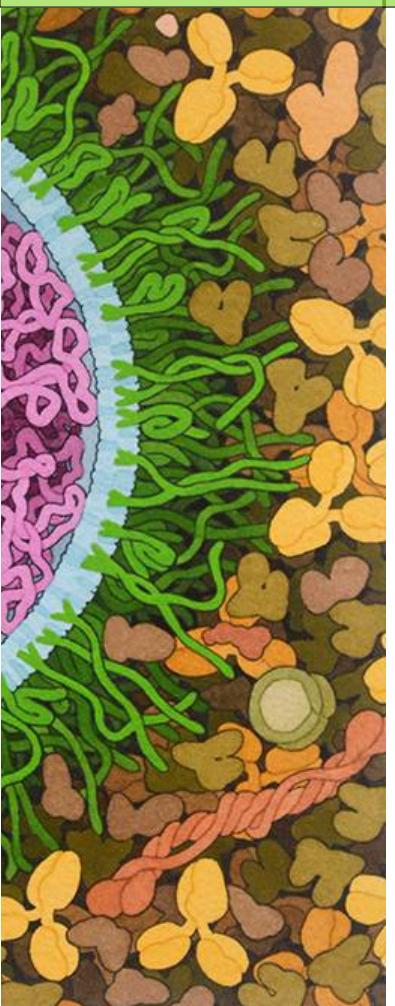
## Navigation & Cross References

- Taxon:  
[Taxon:682670](#)
- RFAM:  
[RF00177](#) (Click here to see all ENA records for this RFAM ID),  
[RF01959](#) (Click here to see all ENA records for this RFAM ID)
- SILVA-SSU:  
[GQ853555](#) (Click here to see all ENA records for this SILVA-SSU ID)

This is just the database  
we came from in the first  
place

?

# Biases in Databases



**NIH** National Library of Medicine  
*National Center for Biotechnology Information*

Search NCBI GQ853555.1

Back to the mothership

NCBI

Results found in 1 database

NUCLEOTIDE SEQUENCE

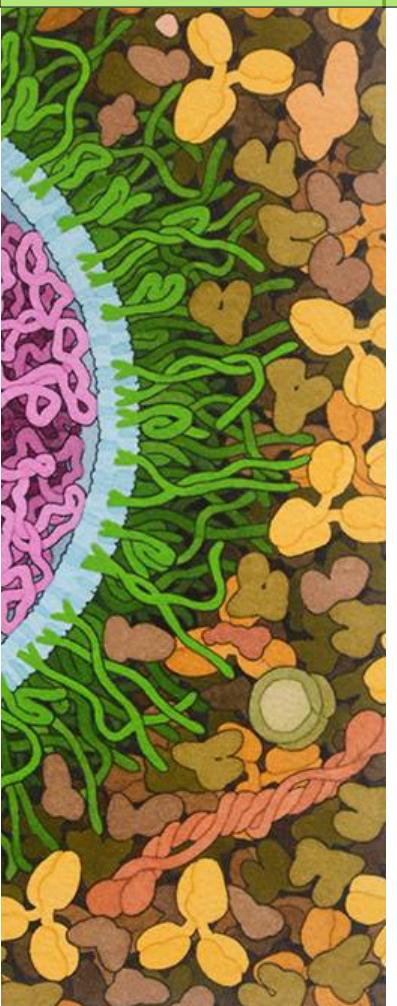
**Loripes lacteus gill symbiont clone 1A 16S ribosomal RNA gene, partial sequence**

Loripes lacteus gill symbiont  
1,569 bp genomic sequence  
GQ853555.1

**FASTA**

[Download](#) [BLAST](#)

# Biases in Databases



## Loripes lacteus gill symbiont clone 1A 16S ribosomal RNA gene, partial sequence

GenBank: GQ853555.1

[FASTA](#) [Graphics](#)

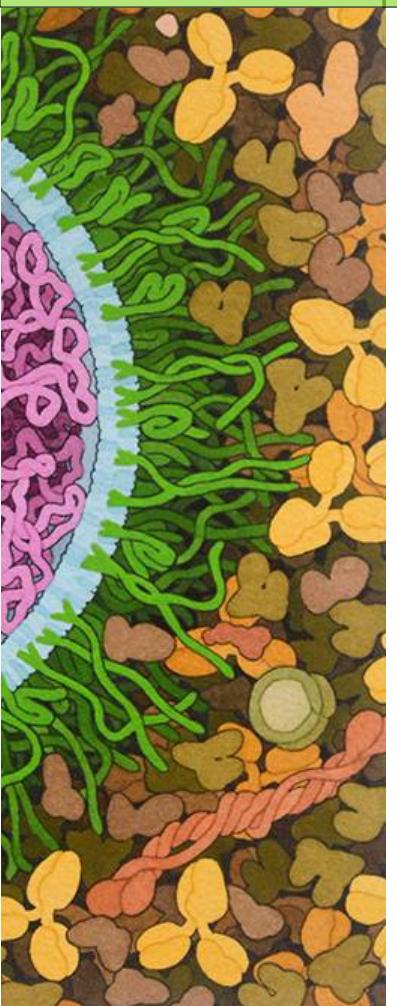
Go to:

LOCUS GQ853555 1569 bp DNA linear BCT 31-AUG-20  
DEFINITION Loripes lacteus gill symbiont clone 1A 16S ribosomal RNA gene, partial sequence.  
ACCESSION GQ853555  
VERSION GQ853555.1  
KEYWORDS .  
SOURCE Loripes lacteus gill symbiont  
ORGANISM [Loripes lacteus gill symbiont](#)  
Bacteria; Proteobacteria; Gammaproteobacteria; sulfur-oxidizing symbionts.  
REFERENCE 1 (bases 1 to 1569)  
AUTHORS Mausz,M., Schmitz-Esser,S. and Steiner,G.  
TITLE Identification and comparative analysis of the endosymbionts of Loripes lacteus and Anodontia fragilis (Bivalvia: Lucinidae)  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1569)  
AUTHORS Mausz,M., Schmitz-Esser,S. and Steiner,G.  
TITLE Direct Submission  
JOURNAL Submitted (26-AUG-2009) Department for Evolutionary Biology, University of Vienna, Institute for Organismal Systems Biology, Althanstrasse 14, Vienna A-3492, Austria  
FEATURES Location/Qualifiers  
Source 1..1569  
/organism="Loripes lacteus gill symbiont"  
/mol\_type="genomic DNA"  
/host="Loripes lacteus isolate 0"  
/db\_xref="taxon:682670"  
/clone="1A"  
rRNA <1..>1569  
/product="16S ribosomal RNA"  
ORIGIN  
1 gagagtttaa ttatggctca gattgaacgc tggcgccagg cctaacacat gcaagtcgag  
61 cggttaacagg gggagcttgc tctctgtcga cgagccgcg acgggtgcgt aacacgttgg  
121 aatctgcctg gtatgtgggg atagccccgg aaatccggta taaataccgc atacgccccca  
181 aaaaaaaaaaaaaaaa caaaaaaaaaaaaaatccca aaaaaaaaaaaaaatcccaatcccaatccca

Back to the mothership

NCBI

# Biases in Databases



## Loripes lacteus gill symbiont clone 1A 16S ribosomal RNA gene, partial sequence

GenBank: GQ853555.1

[FASTA](#) [Graphics](#)

Go to:

LOCUS GQ853555 1569 bp DNA linear BCT 31-AUG-20  
DEFINITION Loripes lacteus gill symbiont clone 1A 16S ribosomal RNA gene,  
partial sequence.

ACCESSION GQ853555  
VERSION GQ853555.1

KEYWORDS .

SOURCE Loripes lacteus gill symbiont

ORGANISM [Loripes lacteus gill symbiont](#)  
Bacteria; Proteobacteria; Gammaproteobacteria; sulfur-oxidizing  
symbionts.

REFERENCE 1 (bases 1 to 1569)

AUTHORS Mausz,M., Schmitz-Esser,S. and Steiner,G.  
TITLE Identification and comparative analysis of the endosymbionts of  
Loripes lacteus and Anodontia fragilis (Bivalvia: Lucinidae)  
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1569)

AUTHORS Mausz,M., Schmitz-Esser,S. and Steiner,G.  
TITLE Direct Submission  
JOURNAL Submitted (26-AUG-2009) Department for Evolutionary Biology,  
University of Vienna, Institute for Organismal Systems Biology,  
Althanstrasse 14, Vienna A-3492, Austria

FEATURES Location/Qualifiers

source 1..1569  
/organism="Loripes lacteus gill symbiont"  
/mol\_type="genomic DNA"  
/host="Loripes lacteus isolate 0"  
/db\_xref="taxon:682670"  
/clone="1A"

rRNA <1..>1569  
/product="16S ribosomal RNA"

ORIGIN

1 gagagtttaa ttatggctca gattgaacgc tggcgccagg cctaacacat gcaagtcgag  
61 cggttaacagg gggagcttgc tctctgtcga cgagccgcgg acgggtcggt aacacgttgg  
121 aatctgcctg gtatgtgggg atagccccgg aaatccggaa ttaataccgc atacgccccca  
181 aaaaaaaaaaaaaaaa caaaaaaaaaaaaaatccat ccacatccat tttataaaaacc

Too bad

# Biases in Databases

## Loripes lacteus gill symbiont clone 1A 16S ribosomal RNA gene, partial sequence

GenBank: GQ853555.1

## FASTA    Graphics

Go to:

LOCUS GQ853555 1569 bp DNA linear BCT 31-AUG-20  
DEFINITION *Loripes lacteus* gill symbiont clone 1A 16S ribosomal RNA gene, partial sequence.  
ACCESSION GQ853555

VERSION GQ853555.1  
KEYWORDS .  
SOURCE Loripes lacteus gill symbiont  
ORGANISM [Loripes lacteus gill symbiont](#)  
Bacteria; Proteobacteria; Gammaproteobacteria; sulfur-oxidizing  
symbionts.

REFERENCE 1 (bases 1 to 1569)  
AUTHORS Mausz,M., Schmitz-Esser,S. and Steiner,G.  
TITLE Identification and comparative analysis of the endosymbionts of  
*Loripes lacteus* and *Anodontia fragilis* (Bivalvia: Lucinidae)  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1569)

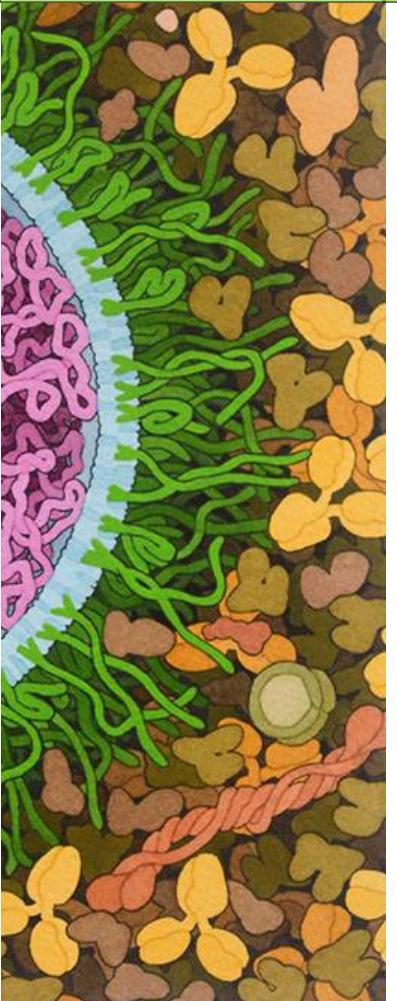
REFERENCE 2 (bases 1 to 1569)  
AUTHORS Mausz,M., Schmitz-Esser,S. and Steiner,G.  
TITLE Direct Submission  
JOURNAL Submitted (26-AUG-2009) Department for Evolutionary Biology,  
University of Vienna, Institute for Organismal Systems Biology,  
Althanstrasse 14, Vienna A-3492, Austria

FEATURES	Location/Qualifiers
source	1..1569 /organism="Loripes lacteus gill symbiont" /mol_type="genomic DNA" /host="Loripes lacteus isolate 0" /db_xref="taxon: <a href="#">682670</a> " /clone="1A"
rRNA	<1..>1569 /product="16S ribosomal RNA"

ORIGIN  
1 gagagtttga ttatggctca gattgaacgc tggcggcagg cctaacacat gcaagtgcag  
61 cgttagtcacagg gggtttgttc tcttcgtca cgaaatcgccg acgggttcgtc aacatcgtagg  
121 aatctgcctgt tgatgtgggg atagccggaa gaaatccggta ttataatccgc atacgc(ccca  
181 aaaaaaaaaaaaaaaaatggggatcc ttcttttcggaaatggggatcc ttgtttatcatggatggatcc

This mean they manually  
Edited the taxonomy  
And we can't go back  
To check their work

## Biases in Databases



This does not necessary  
means everything is garbage  
But maybe you should revisit  
the power you give to this information

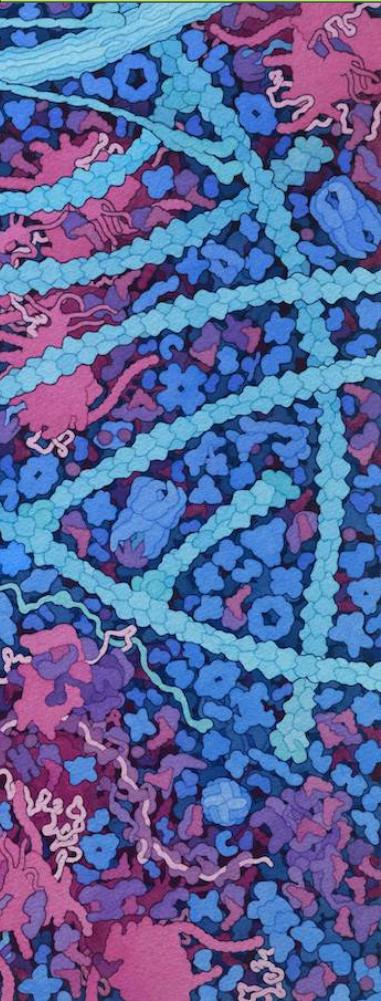
The answer is Cross-checking.

Never blindly trust anything

*In regione caecorum rex est lucus*

Au royaume des aveugles les borgnes sont rois

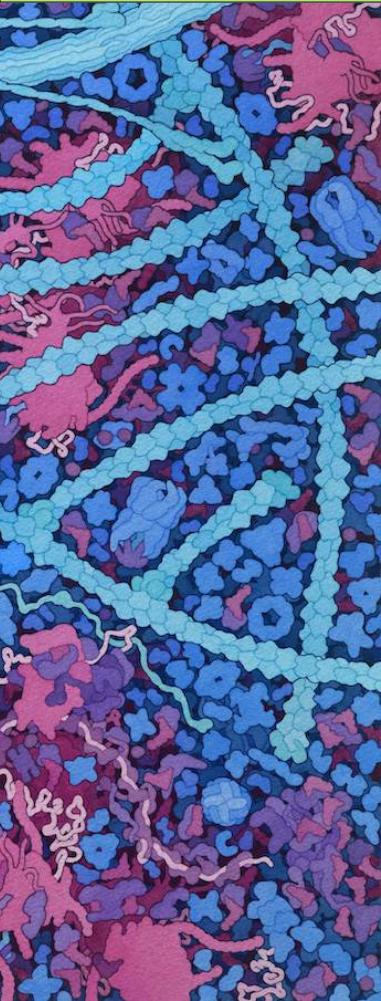
# The Future of Metabarcoding



**Metabarcoding is aging.**

**Shotgun metagenomics approaches show  
much more hope today**

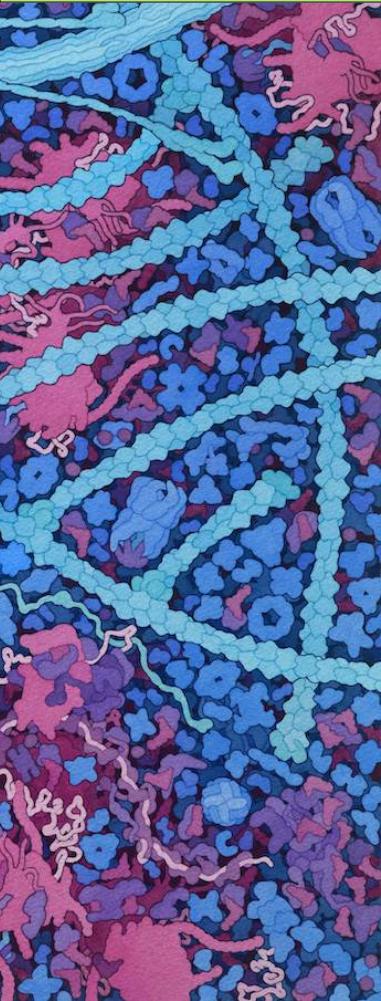
## Conclusions



Metabarcoding is a powerful technique to investigate living communities. It is versatile and cheap.

But it has weaknesses and limits that should be accounted for before using them and more importantly

**Before blindly believing them**

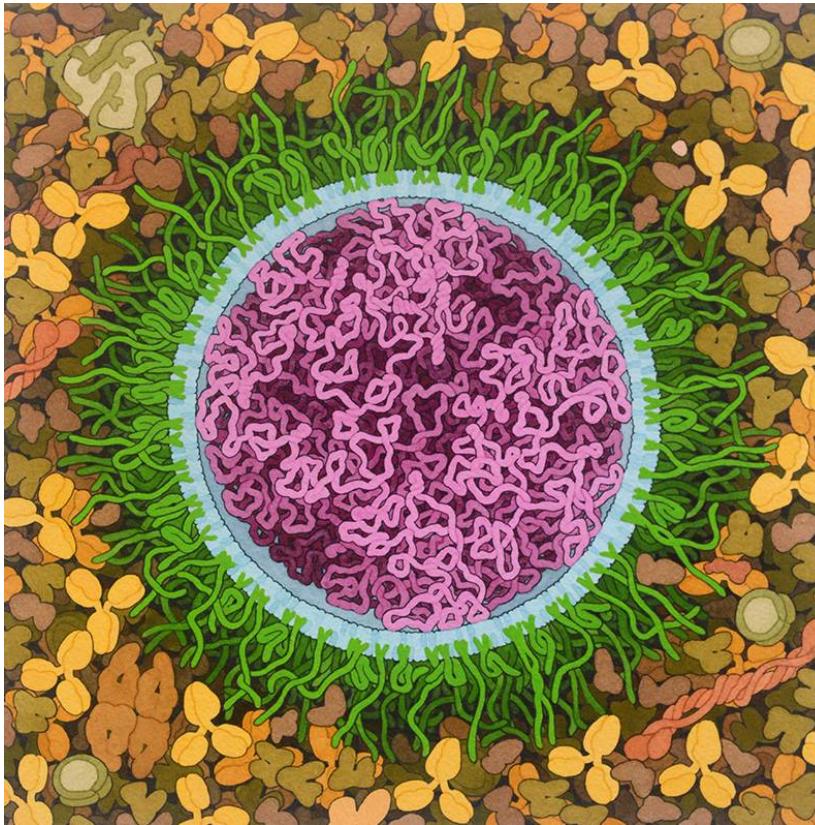
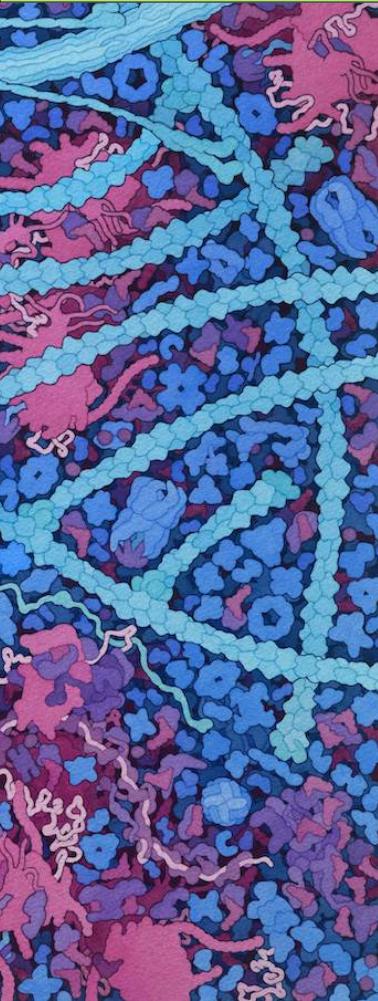


**Thank you for your attention**



**Any questions ?**

# Molecular Biology Artworks

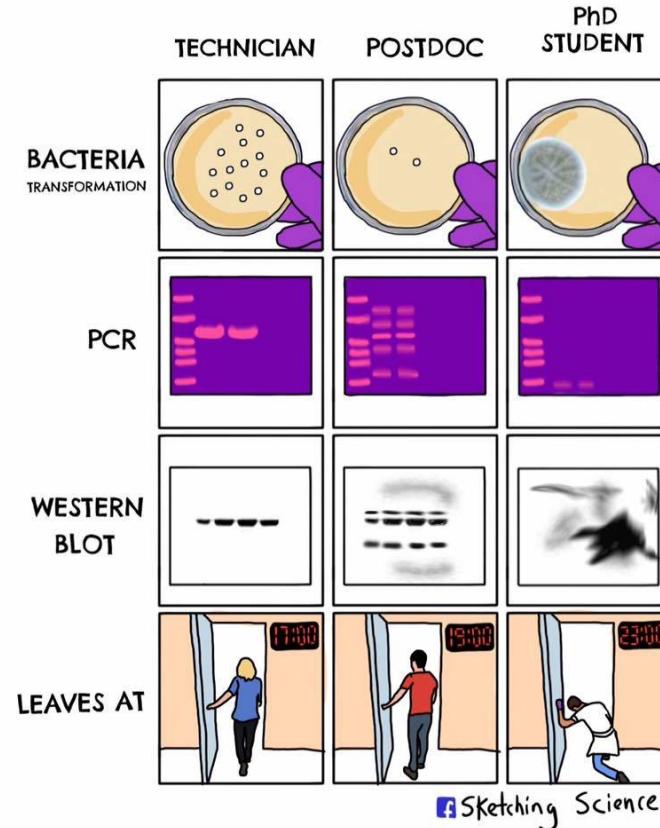
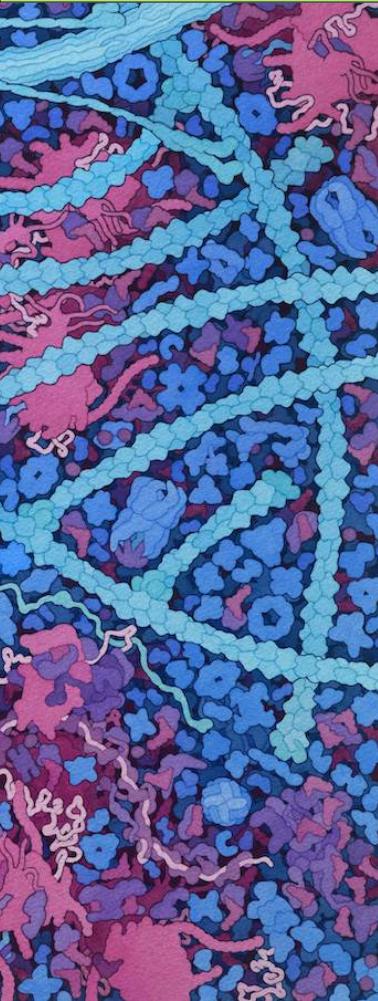


SARS-CoV2 mRNA Vaccine

**Molecular Landscapes**  
**by David S. Goodsell**

<https://pdb101.rcsb.org/sci-art/goodsell-gallery>

# Sketching Science



<https://sketchingscience.org/commissions/>

## References

Fitch, Walter M., and Emanuel Margoliash. "Construction of Phylogenetic Trees." *Science*, vol. 155, no. 3760, American Association for the Advancement of Science, 1967, pp. 279–84, <http://www.jstor.org/stable/1720651>.

Ward, D., Weller, R. & Bateson, M. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* 345, 63–65 (1990). <https://doi.org/10.1038/345063a0>

bioinfo.genotoul.fr (Currently down :( )

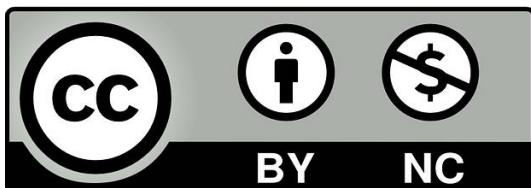
<https://www.france-bioinformatique.fr/en/home/>

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

<https://www.ncbi.nlm.nih.gov/>

<https://www.ebi.ac.uk/ena/browser/xref>

Don't hesitate to contact me if further questions or needing a hand for your project – arthur.cousson@ird.fr



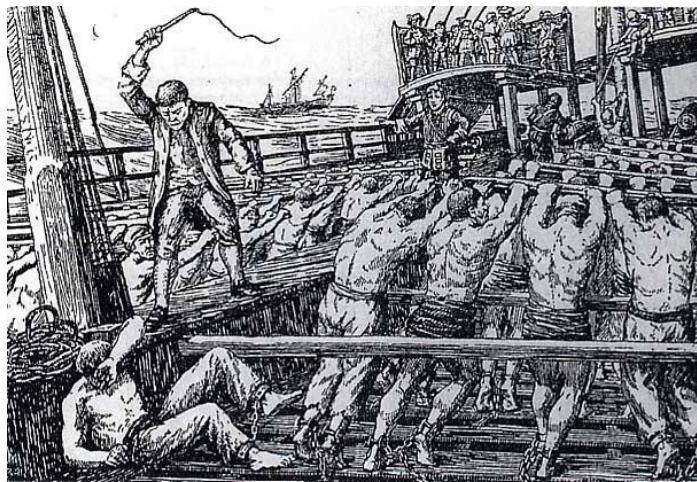
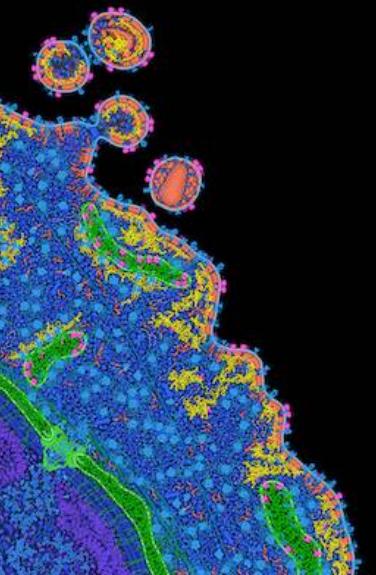
# TD – Metabarcoding – Bioinformatics – Epistemology

**Phyloseq  
Metabarcoding  
Microbiome analysis**

**Databases  
Epistemology  
And Rabbit hole**

**Nice plot cooking  
With R package  
ggplot2**

**Bash, Unix  
An introduction to  
Sequence manipulation**



**Choose your faith**