# What is metagenomics and why is it a big deal?

BIONOMICS-MMlab
MULTIPARTITE PARASITIC INTERACTIONS

**Maliheh Mehrshad**

**Department of Aquatic Sciences and Assessment,** SLU

# Who am I?

BIONOMICS-MMlab
MULTIPARTITE PARASITIC INTERACTIONS

**www.bionomics-mmlab.com**
**Pushing the frontiers of multipartite parasitic interactions research**

## Current bionomics-mmlab members

Zahra Goodarzi – PhD student
Dr. Vinicius Silva Kavagutti – PostDoc
Dr. Vesna Grujcic – PostDoc
Lauren Davies – PhD student

## Former lab members

Maryan Resaei-some – PostDoc
Fanny Persson – Field assistant
Armand Stoe – MSc student
Remco Hoogervorst – Internship student
Ezhilarasan Mani Ezhilan – MSc student

# What is metagenomics and why is it a big deal?

BIONOMICS-MMlab
MULTIPARTITE PARASITIC INTERACTIONS

**Maliheh Mehrshad**
**Department of Aquatic Sciences and Assessment,** SLU

# Let’s answer some questions on menti!

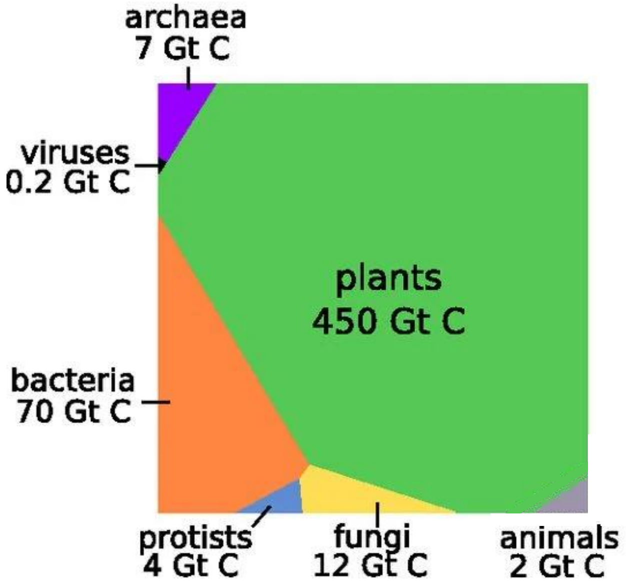
- Join at menti.com
- Use code 74440277

# Microbes Matter!

Life on Earth has been microscopic for much of its ~4 billion year history.

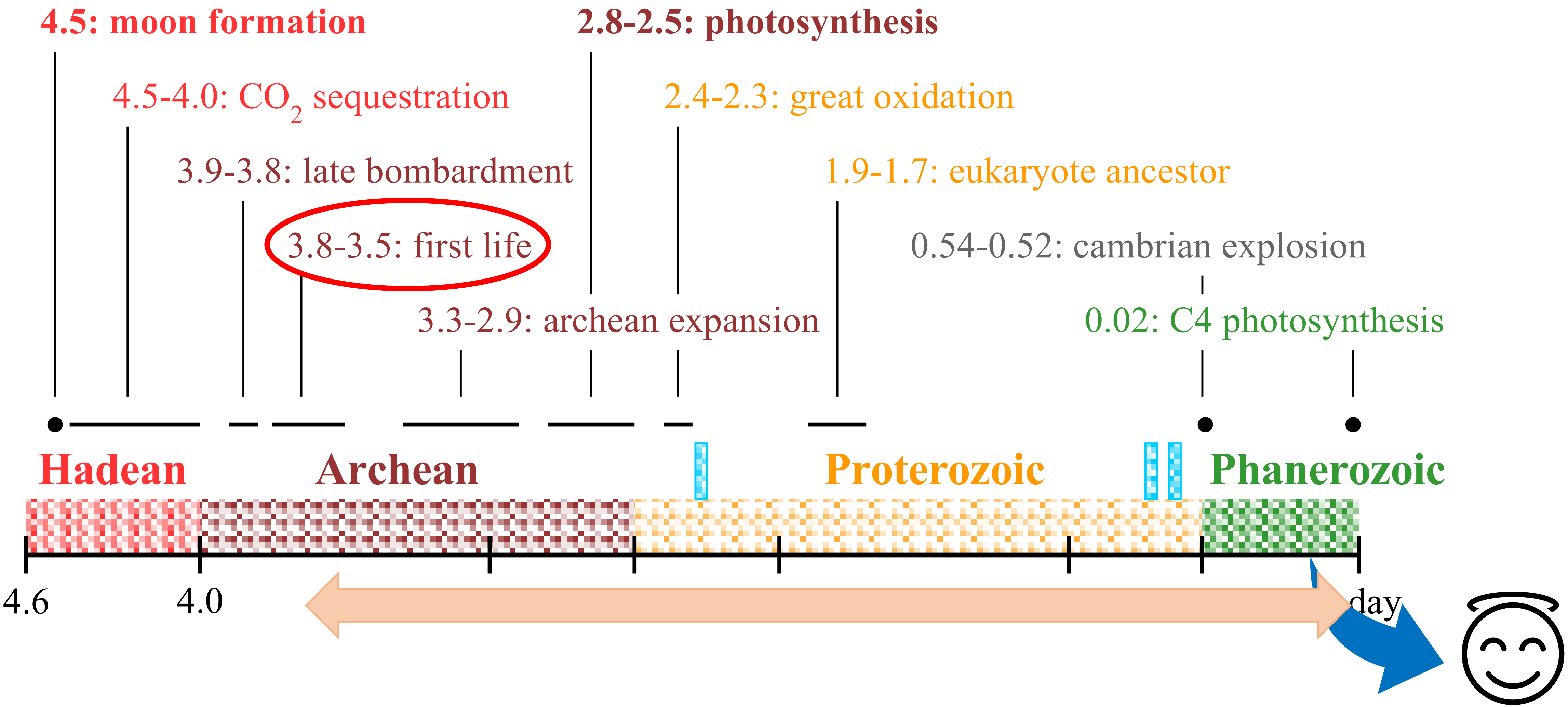
The metabolic activity of these organisms has left its mark.

- Great oxygenation event
- Photosynthesis
- Lignin and cellulose degradation
- Cycling of elements (C, N, S, Fe, …)
- Greenhouse gas sink/emission
- Interact with plants and animals
- …



Bar-On *et.al. PNAS*, 2018

# Life on the planetary timescale



**4.5: moon formation**

4.5-4.0: $CO_2$ sequestration

3.9-3.8: late bombardment

3.8-3.5: first life

3.3-2.9: archean expansion

**2.8-2.5: photosynthesis**

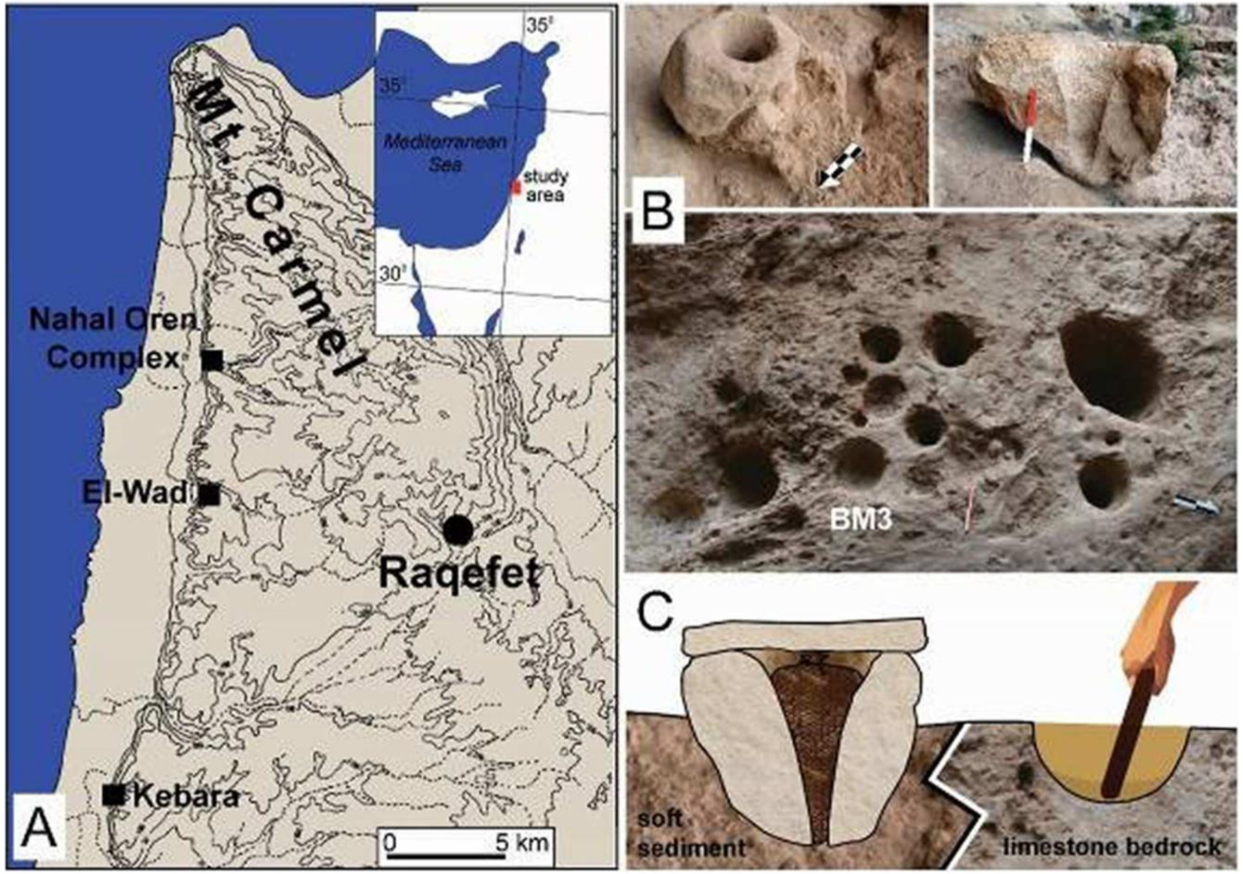
2.4-2.3: great oxidation

1.9-1.7: eukaryote ancestor

0.54-0.52: cambrian explosion

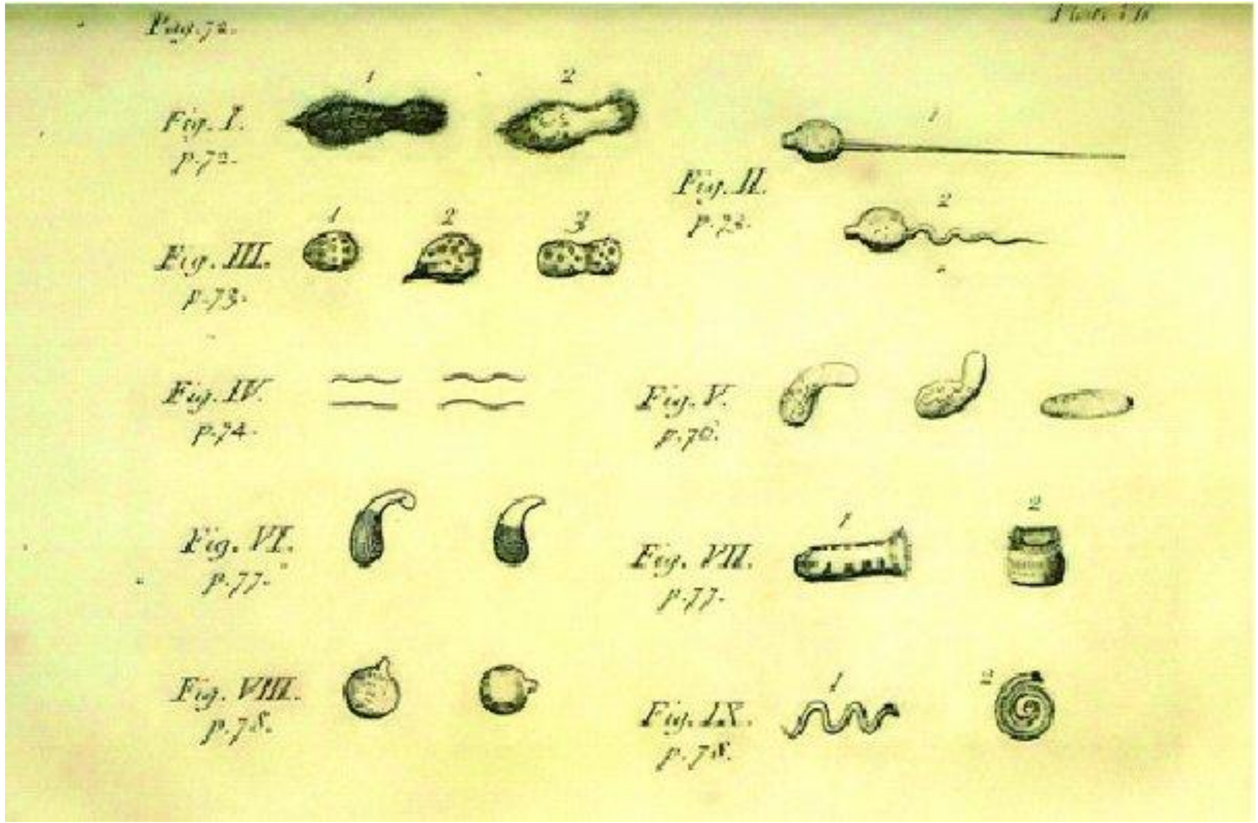
0.02: C4 photosynthesis

# We knew microbial ecology before knowing microbes



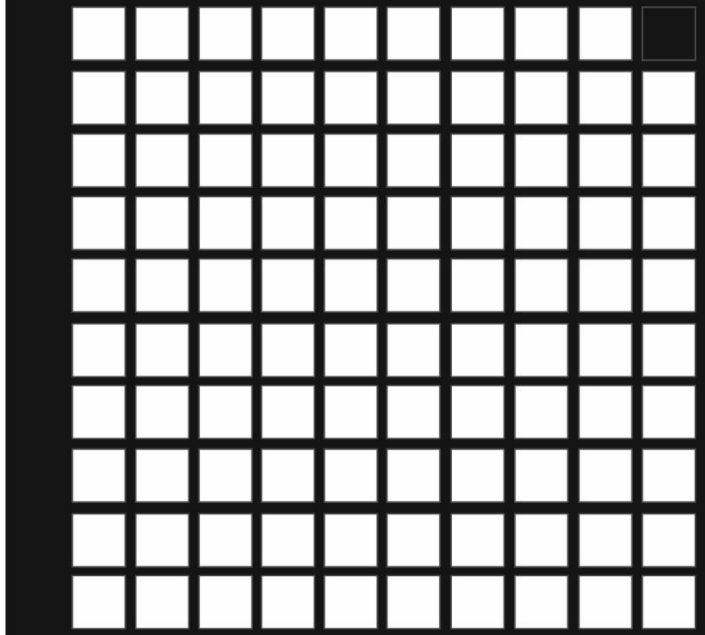
Fermented beverage and food storage in 13,000 year-old stone mortars

# The Unseen Majority



**1676**

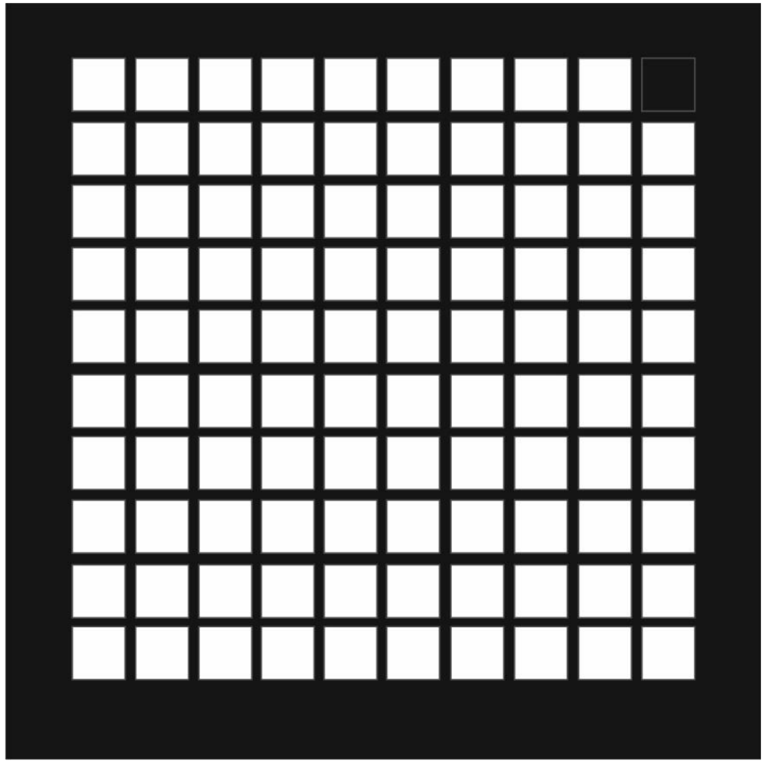
# The Unseen Majority: Great plate count anomaly



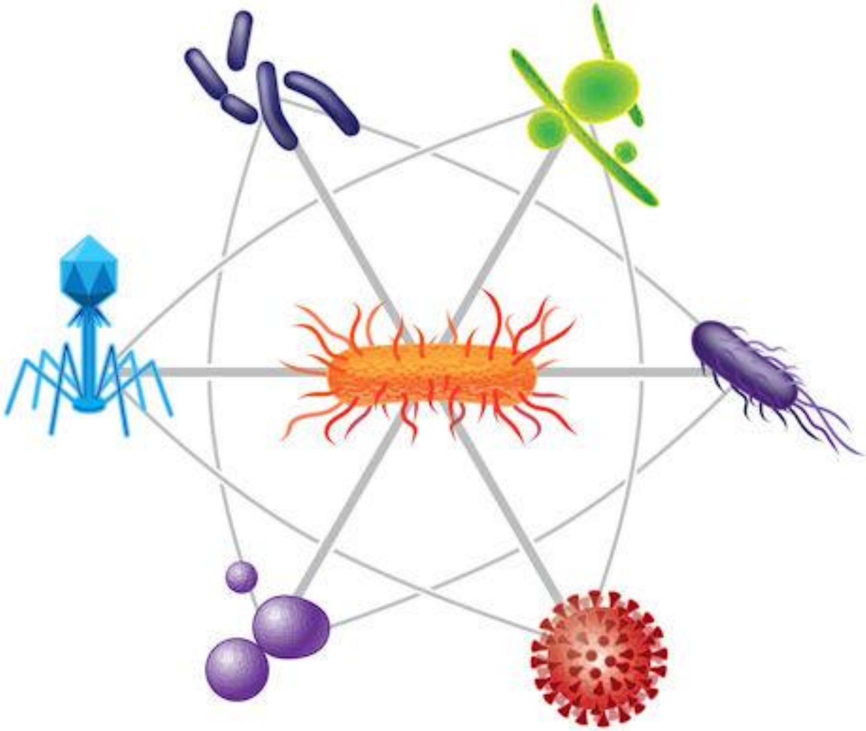
99 %

1985

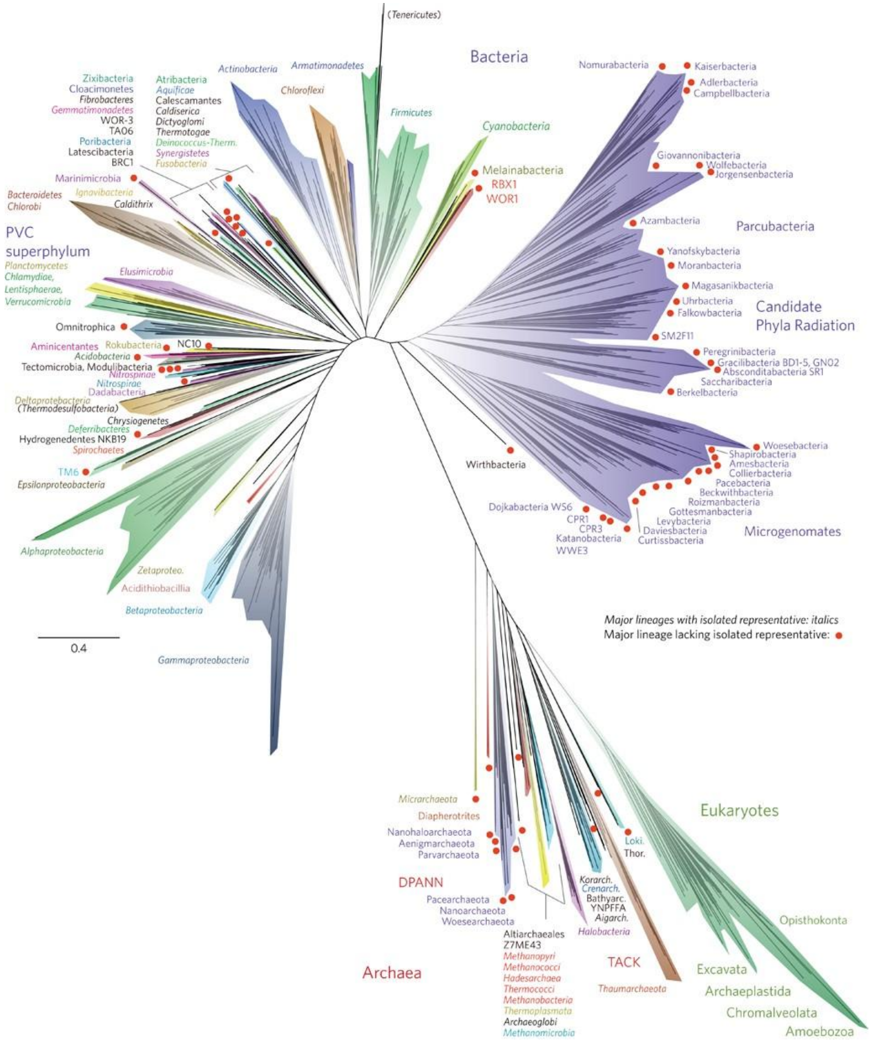
# The Unseen Majority: Great plate count anomaly



1985

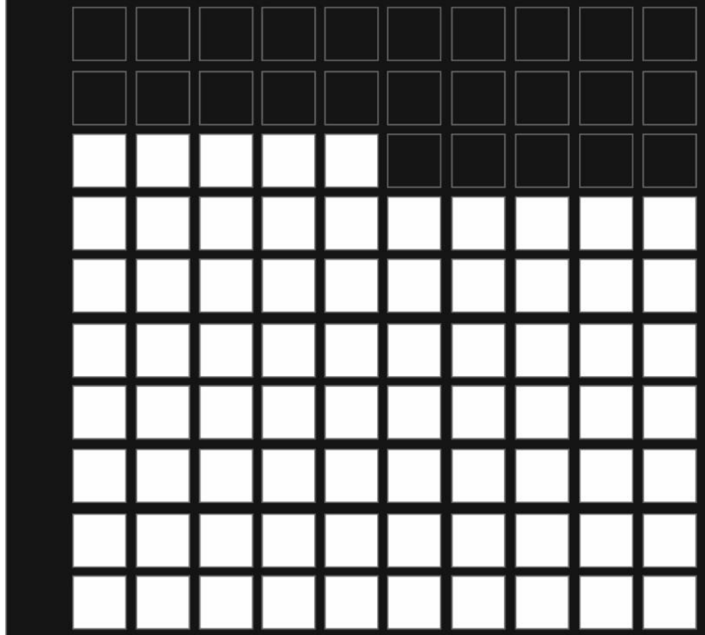


# The Unseen Majority: metagenomics revolution



**Hug et al., *Nat microbiol*, 2016**

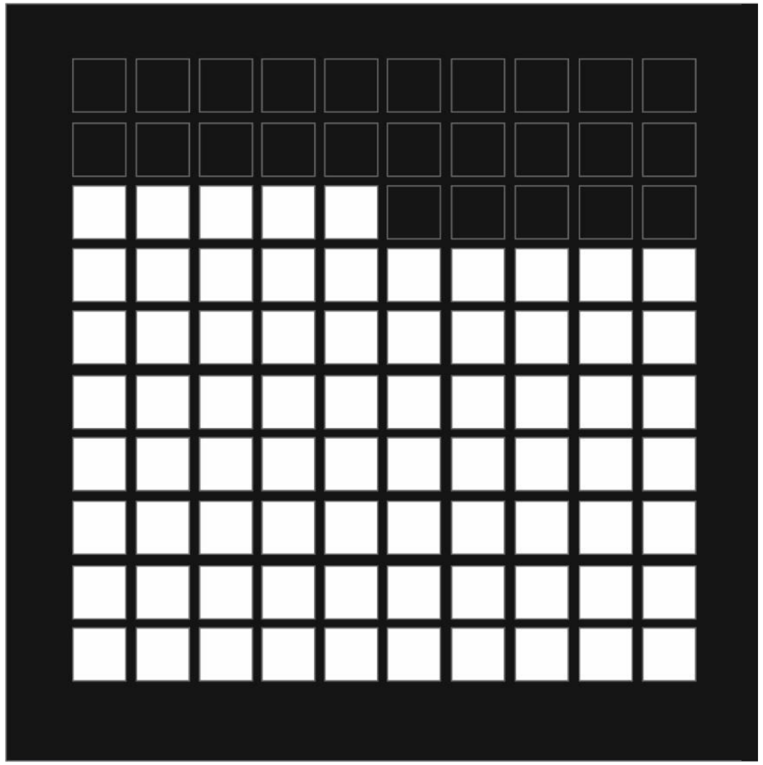
# The Unseen Majority: metagenomics revolution



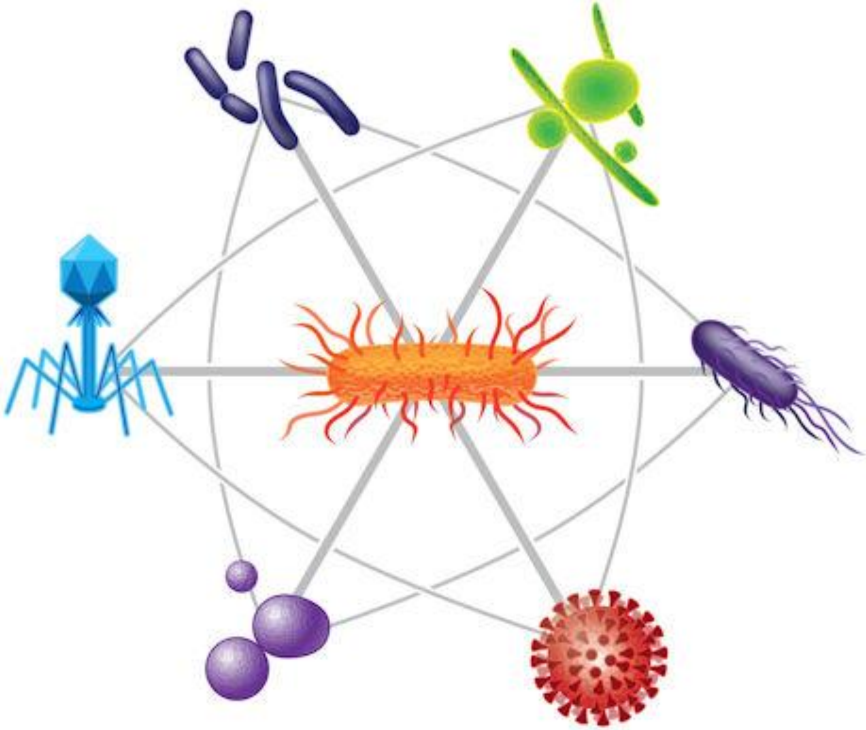
75 %

Now

# The Unseen Majority: metagenomics revolution



Now

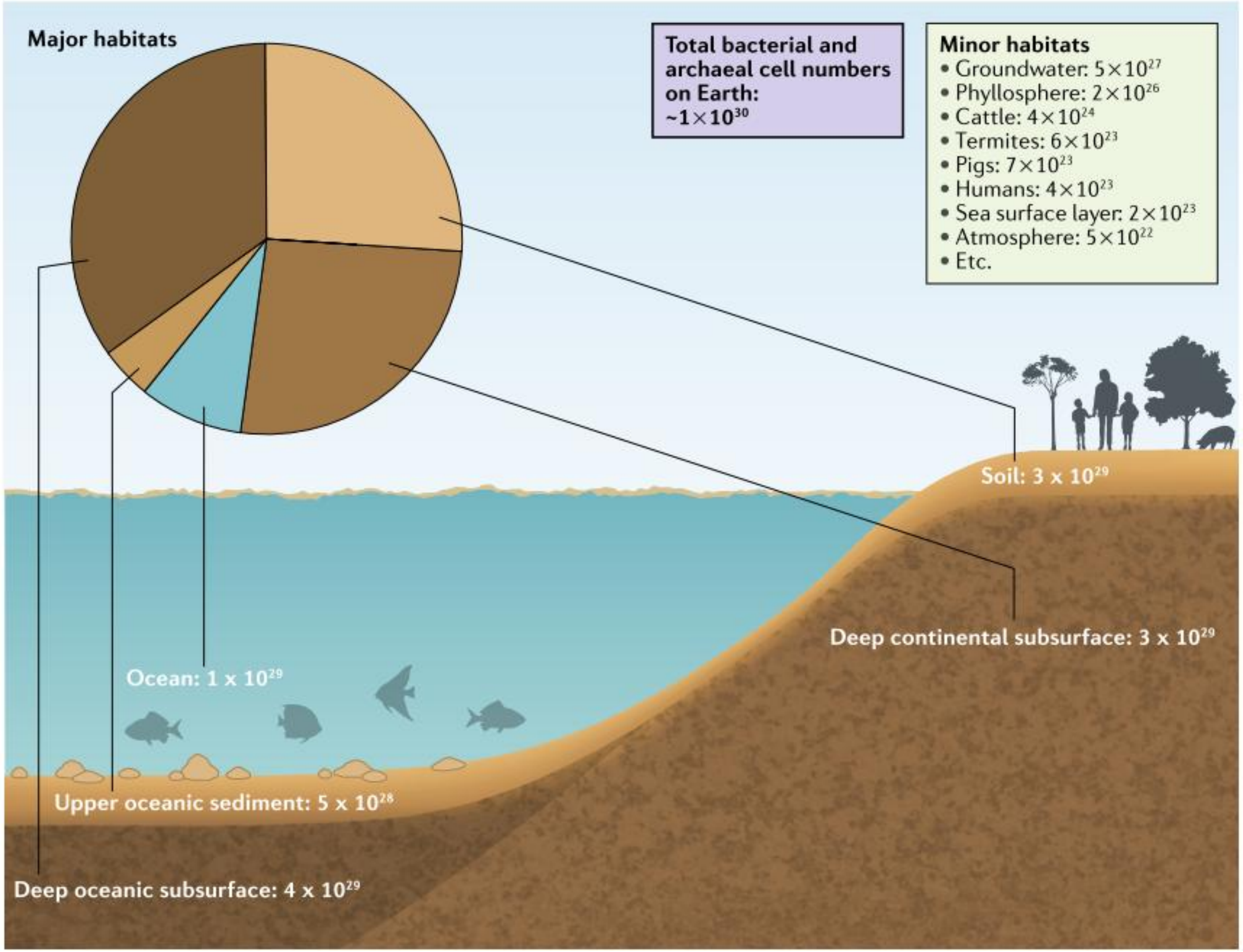


This enormous biomass is distributed in microscopic cells

~1.2 x $10^{30}$ bacterial/archaeal cells exist in the “big five” habitats of Earth

if all the 1 × $10^{31}$ viruses on earth were laid end to end, they would stretch for 100 million light years

Astronomical numbers!



Major habitats

Total bacterial and archaeal cell numbers on Earth: ~1×$10^{30}$

Minor habitats
- Groundwater: 5×$10^{27}$
- Phyllosphere: 2×$10^{26}$
- Cattle: 4×$10^{24}$
- Termites: 6×$10^{23}$
- Pigs: 7×$10^{23}$
- Humans: 4×$10^{23}$
- Sea surface layer: 2×$10^{23}$
- Atmosphere: 5×$10^{22}$
- Etc.

Soil: 3 x $10^{29}$

Deep continental subsurface: 3 x $10^{29}$

Ocean: 1 x $10^{29}$

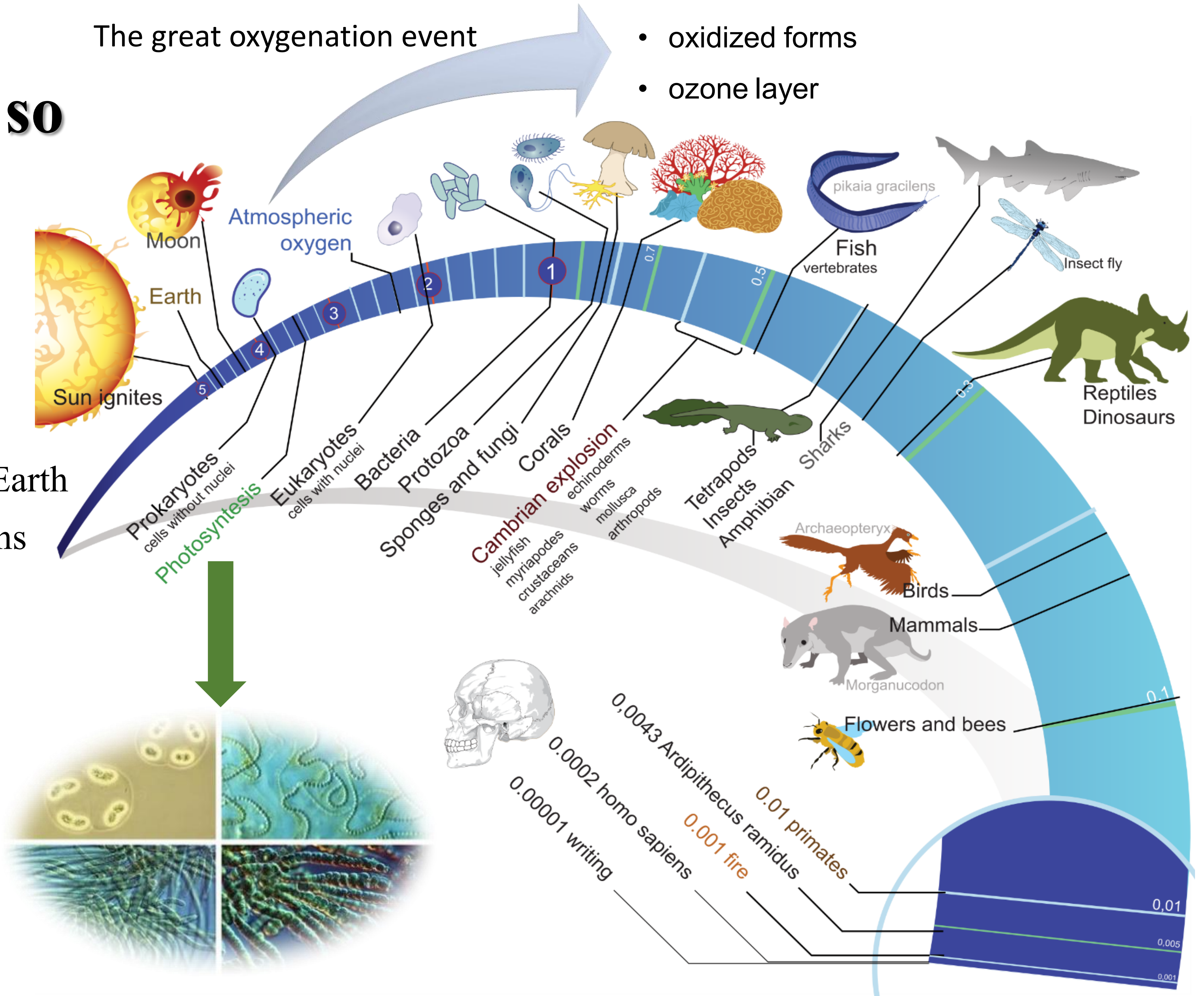
Upper oceanic sediment: 5 x $10^{28}$

Deep oceanic subsurface: 4 x $10^{29}$

(Flemming *et.al.* 2018. *Nat Rev Microbiol*)

# Why are microbes so diverse?

- Evolved early
- Initially access to all habitats on Earth
- Survided a large span of conditions
- More habitas, more niches
- Short generation times
- Inter-species gene transfer



# Microbes in carbon cycle

Microbes transfer an enormous flow of carbon through

Trophic interaction

Metabolism

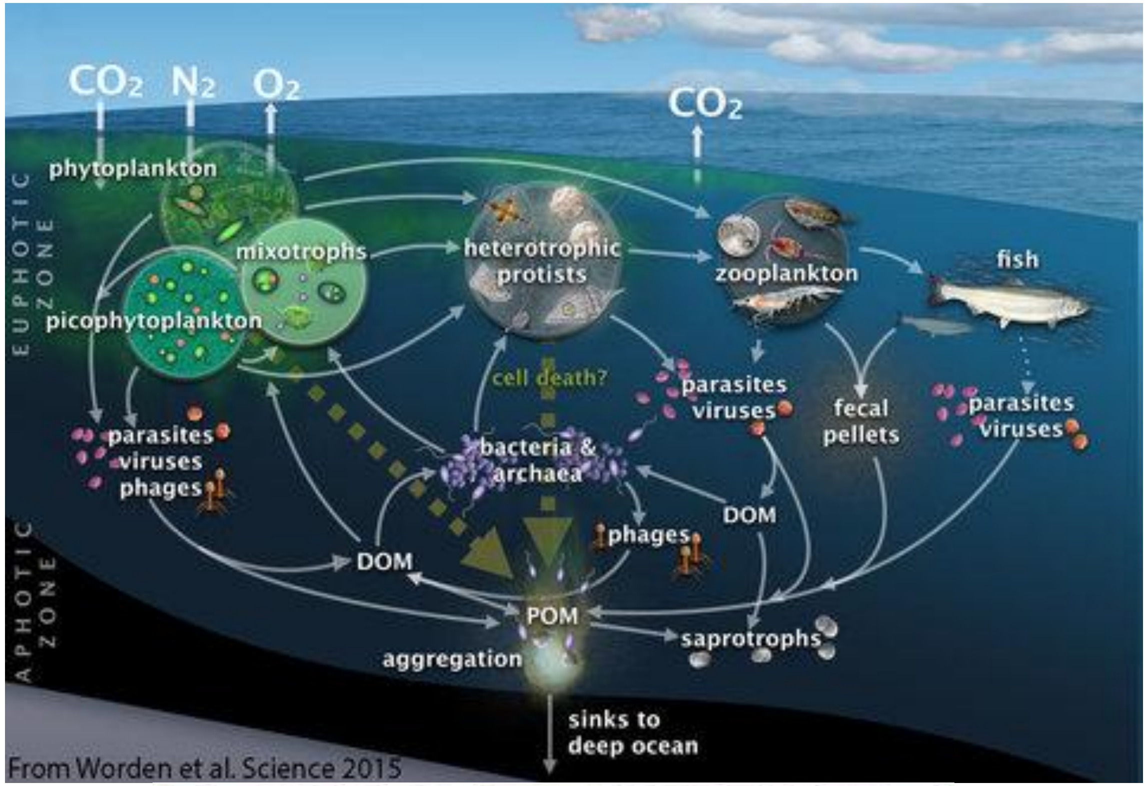
Greenhouse gas sink/emission control

Methane

Nitrous Oxide

CO2

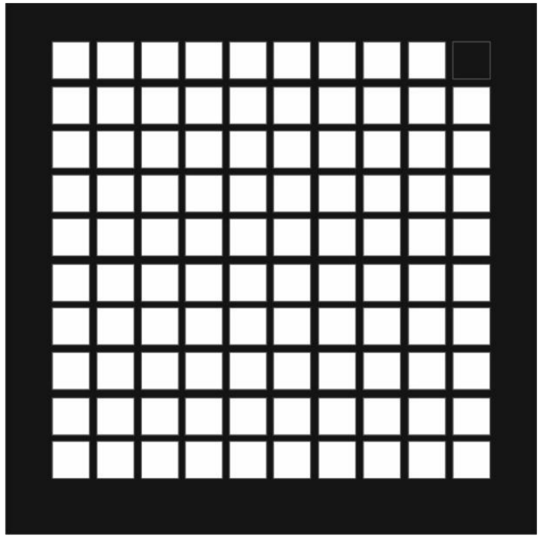
**Climate impact**

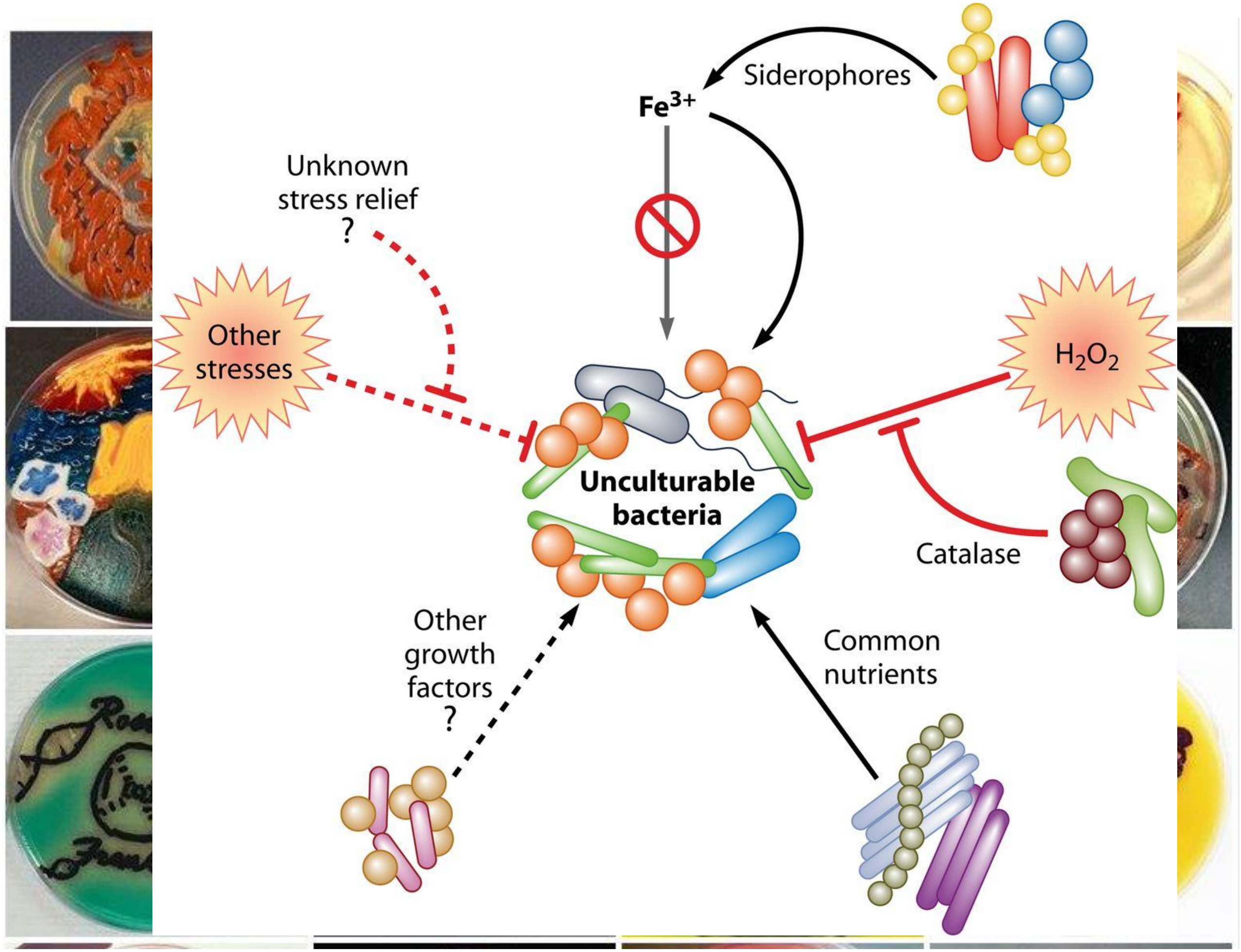


From Worden et al. Science 2015

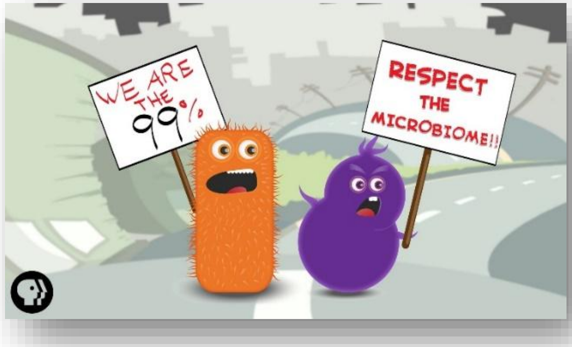
# Challenge: Great plate count anomaly

Using conventional cultivation techniques only **0.1-1%** of prokaryotes are Culturable in laboratory conditions.

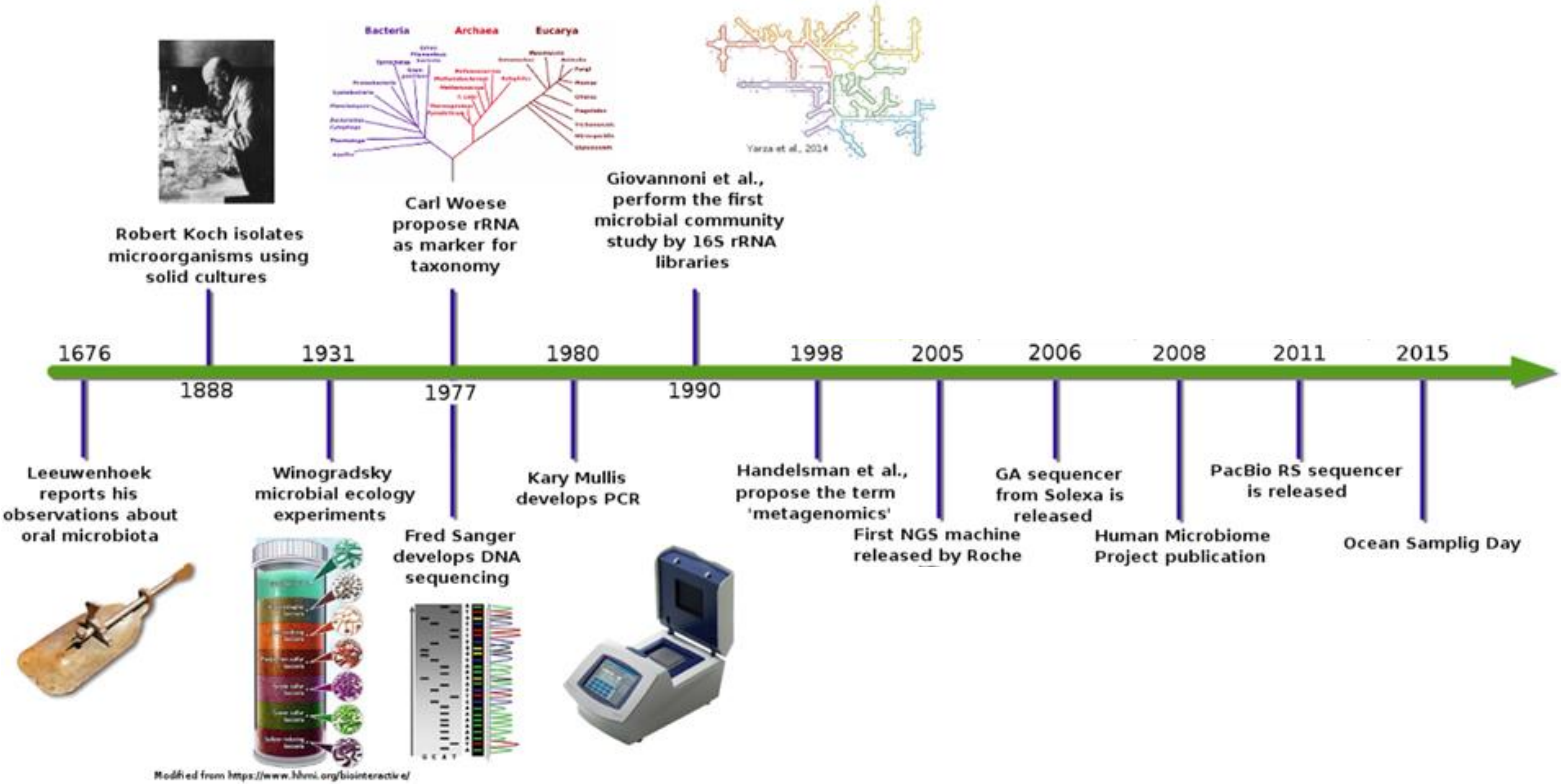




# Microbiome analysis



Desired efficiency anything more that **0.1-1%** !



1676 — Leeuwenhoek reports his observations about oral microbiota

1888 — Robert Koch isolates microorganisms using solid cultures

1931 — Winogradsky microbial ecology experiments

1977 — Carl Woese propose rRNA as marker for taxonomy

1977 — Fred Sanger develops DNA sequencing

1980 — Kary Mullis develops PCR

1990 — Giovannoni et al., perform the first microbial community study by 16S rRNA libraries

1998 — Handelsman et al., propose the term 'metagenomics'

2005 — First NGS machine released by Roche

2006 — GA sequencer from Solexa is released

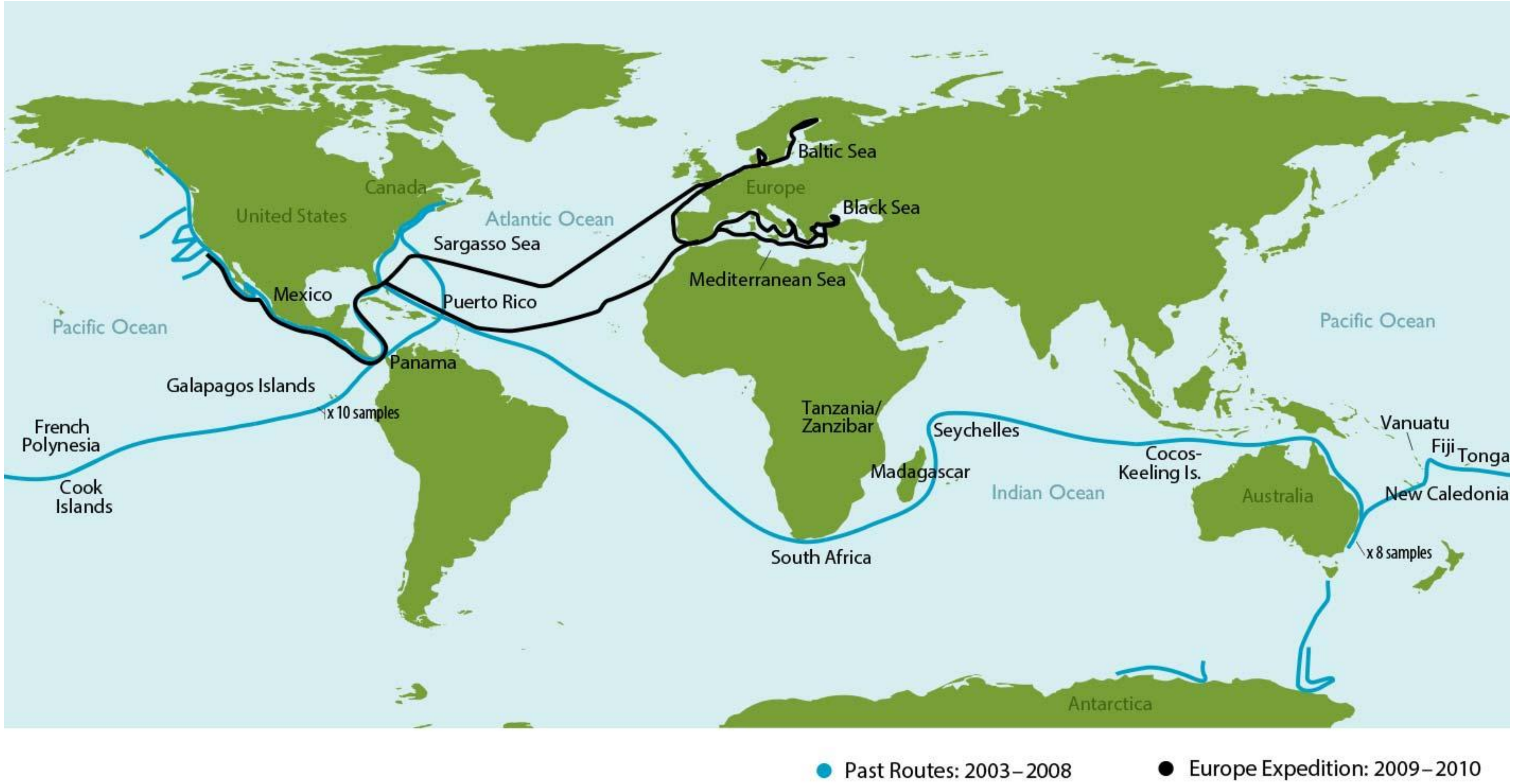
2008 — Human Microbiome Project publication

2011 — PacBio RS sequencer is released

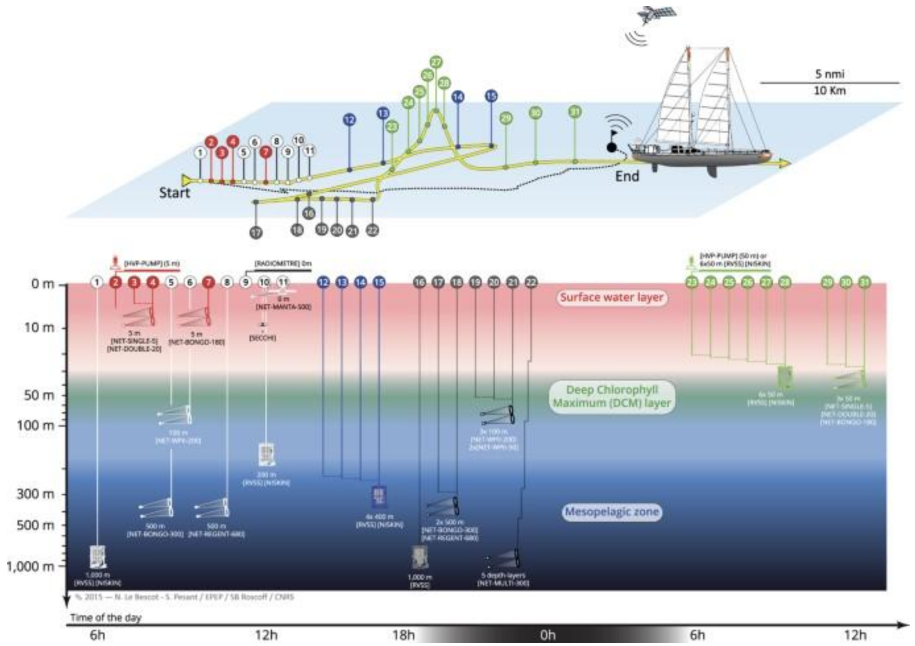
2015 — Ocean Samplig Day

Modified from https://www.hhmi.org/biointeractive/

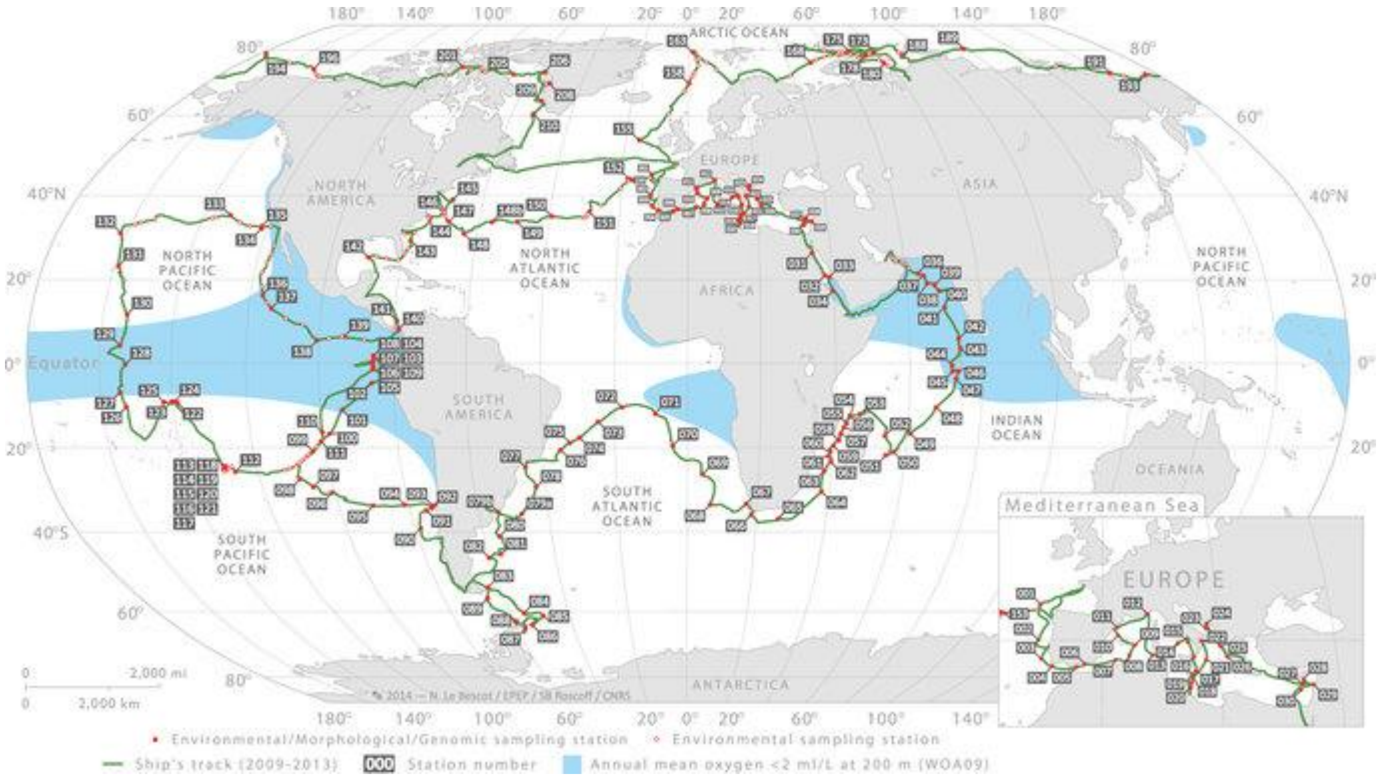
# Global Ocean Sampling



# Tara Ocean Sampling



% 2015 — N. Le Bescot - S. Pesant / EPEP / SB Roscoff / CNRS

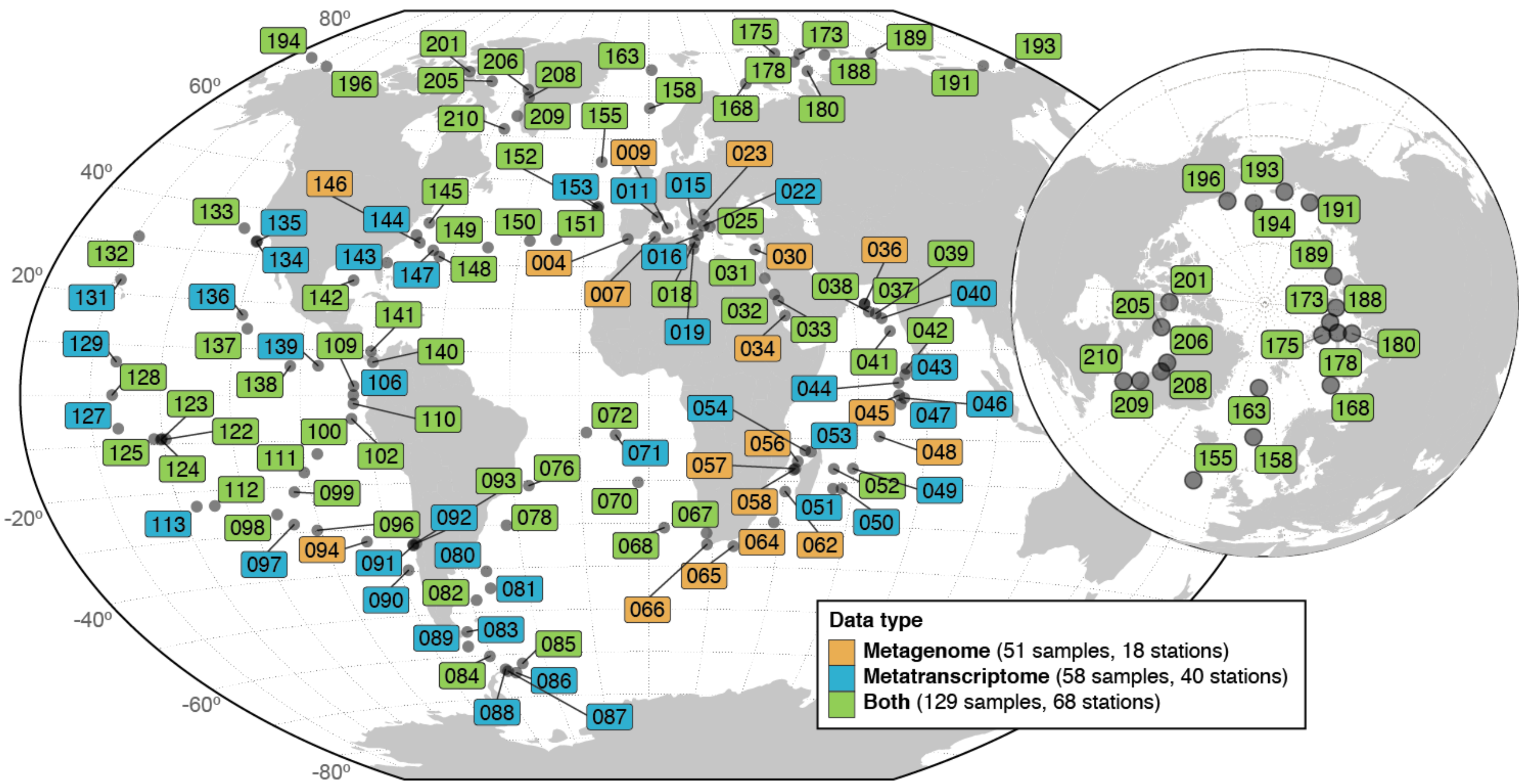


% 2014 — N. Le Bescot / EPEP / SB Roscoff / CNRS

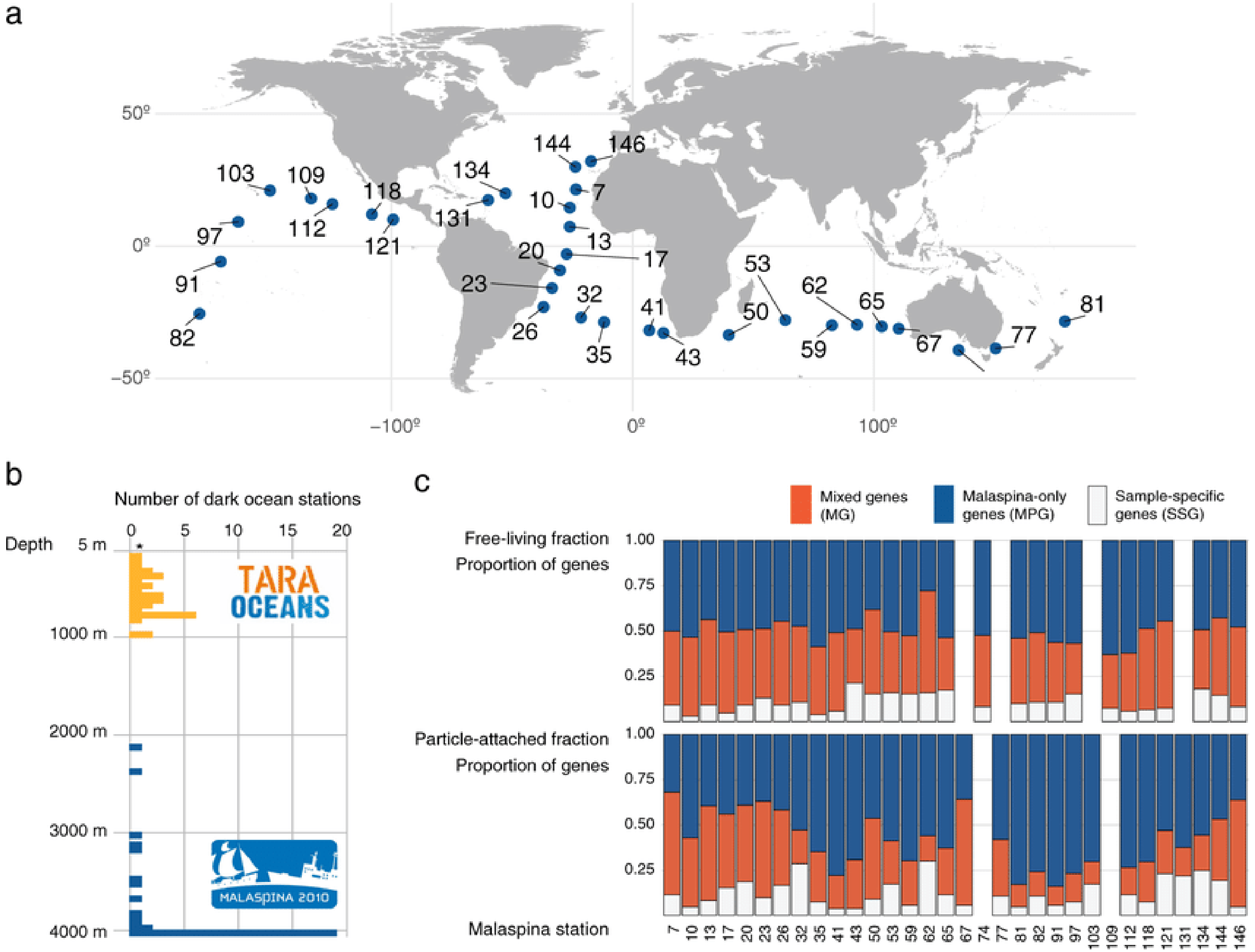
• Environmental/Morphological/Genomic sampling station ∘ Environmental sampling station

— Ship's track (2009-2013) 000 Station number 🟦 Annual mean oxygen <2 ml/L at 200 m (WOA09)

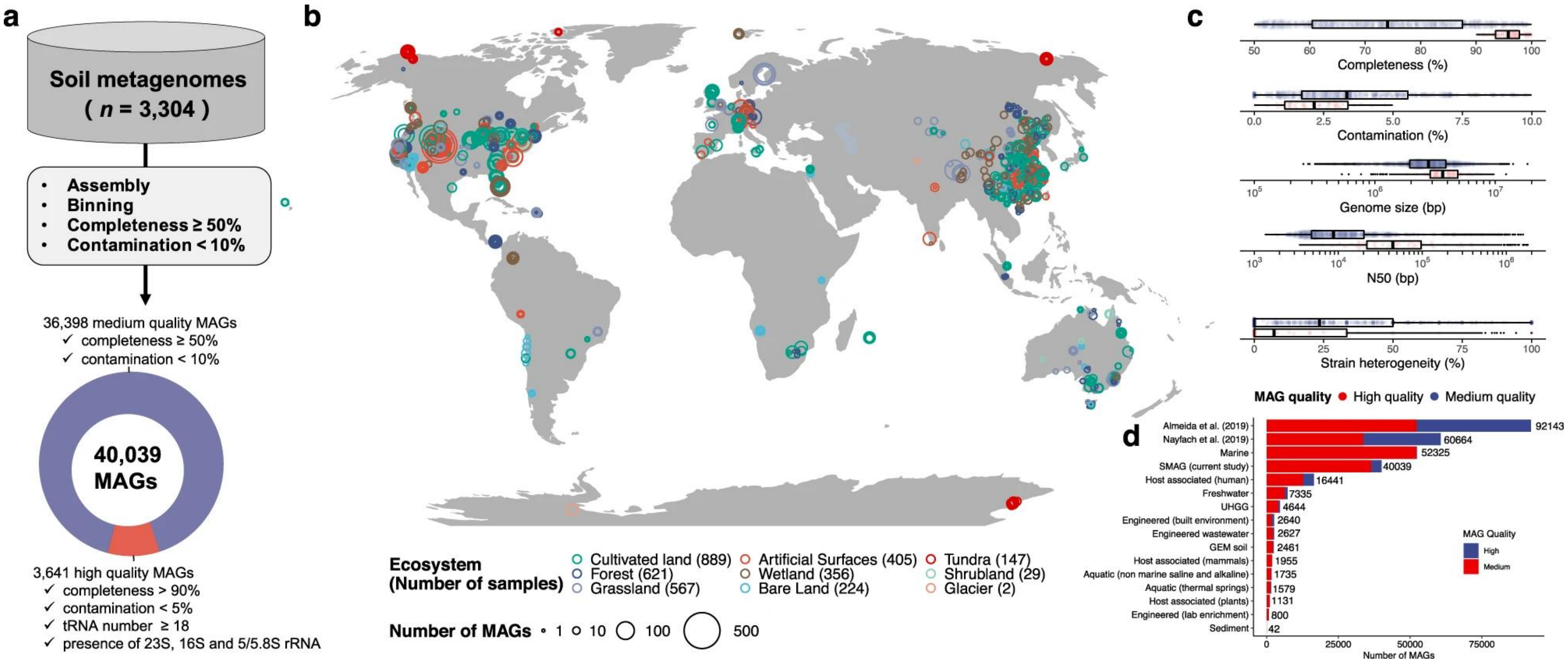
# Tara Ocean Sampling



# Malaspina Sampling



# Soil metagenomes



Soil metagenomes ( *n* = 3,304 )

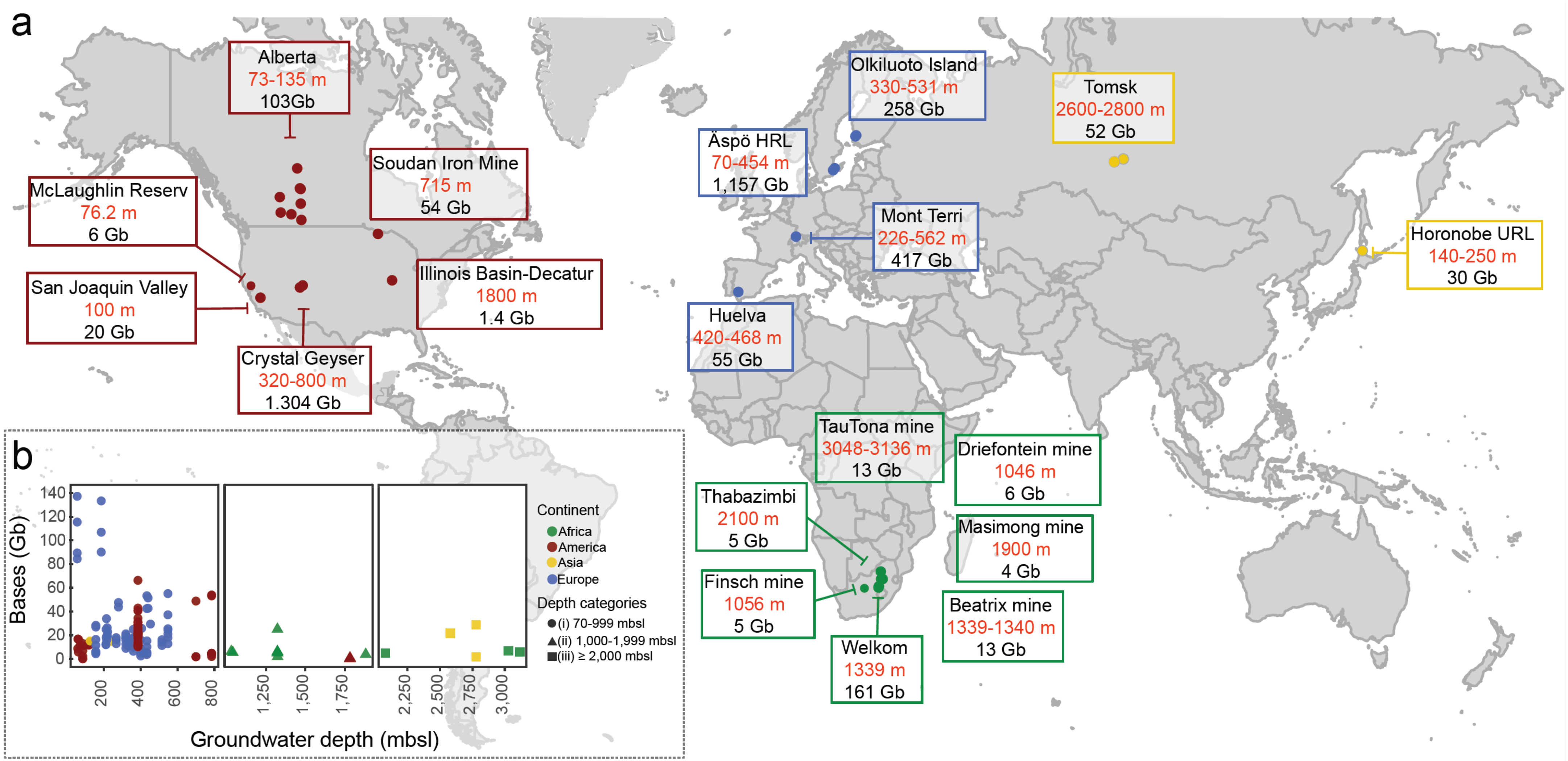
- Assembly
- Binning
- Completeness ≥ 50%
- Contamination < 10%

36,398 medium quality MAGs
✓ completeness ≥ 50%
✓ contamination < 10%

40,039 MAGs

3,641 high quality MAGs
✓ completeness > 90%
✓ contamination < 5%
✓ tRNA number ≥ 18
✓ presence of 23S, 16S and 5/5.8S rRNA

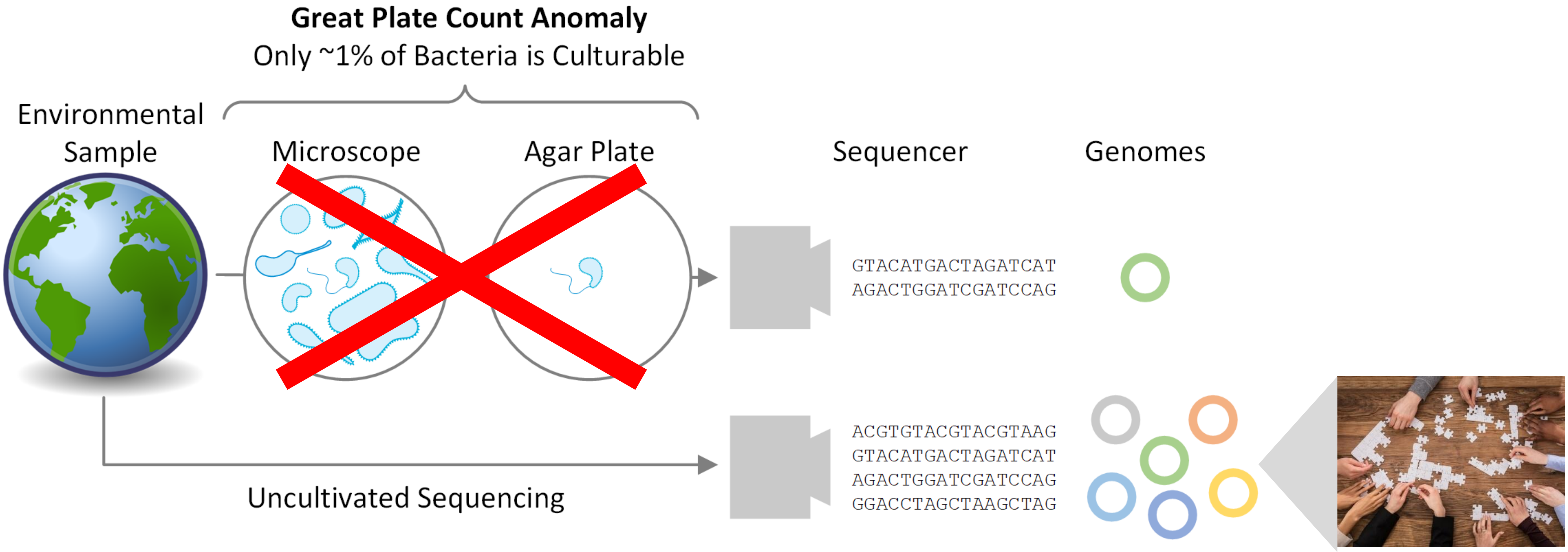
# Deep groundwater metagenomes



# How is Metagenomics done and what does it tell us?

# How is Metagenomics done and what does it tell us?

# Capturing the unseen Majority or their footprint…



**Great Plate Count Anomaly**

Only ~1% of Bacteria is Culturable

Environmental Sample

Microscope

Agar Plate

Sequencer

Genomes

GTACATGACTAGATCAT
AGACTGGATCGATCCAG

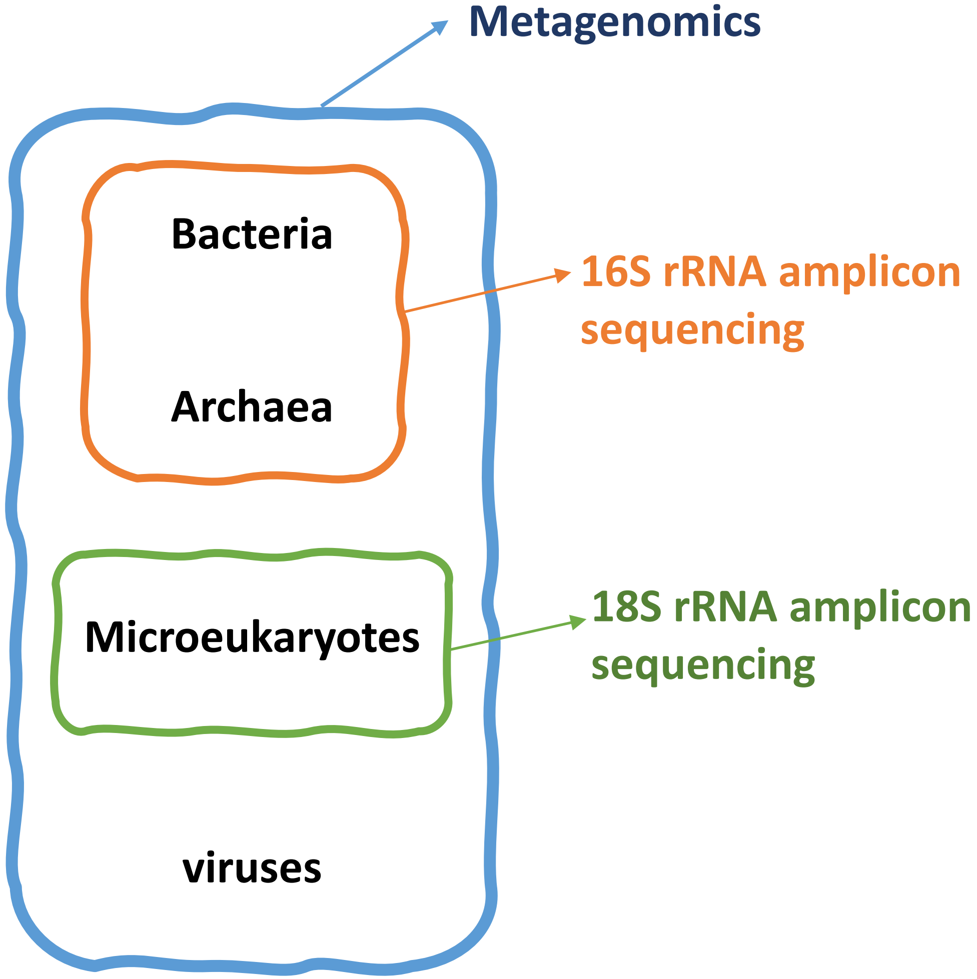
ACGTGTACGTACGTAAG
GTACATGACTAGATCAT
AGACTGGATCGATCCAG
GGACCTAGCTAAGCTAG

Uncultivated Sequencing

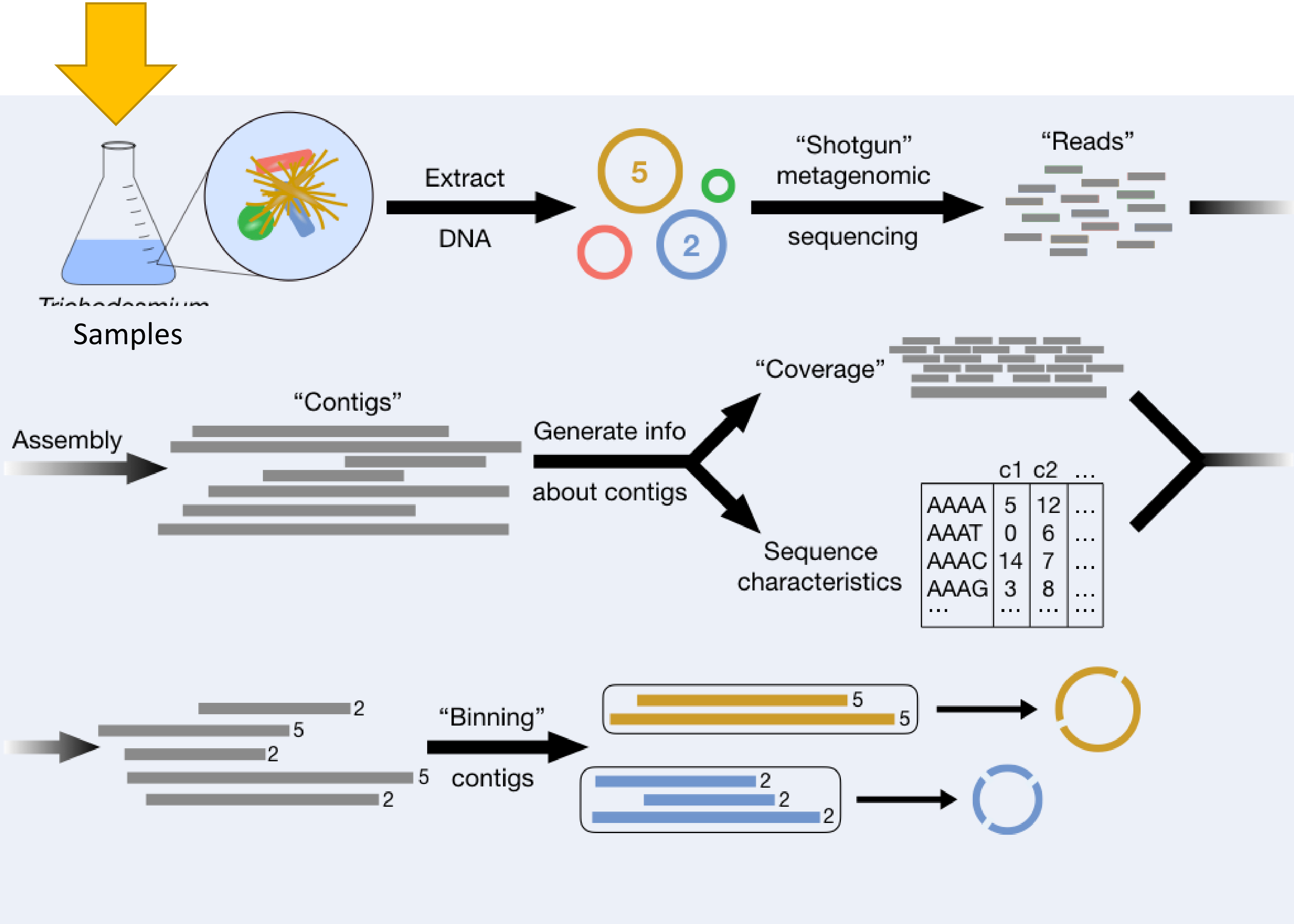
Parks *et.al.* NatMicrobiol, 2017

# Amplicon vs. Metagenomics

- Less complex
  - Better coverage
  - More samples
- Extensive database
- Same fragment
  - Comparable phylogenetic info
- Qualitative
- PCR and primer bias
- Limited phylogenetic info
- Limited functional information

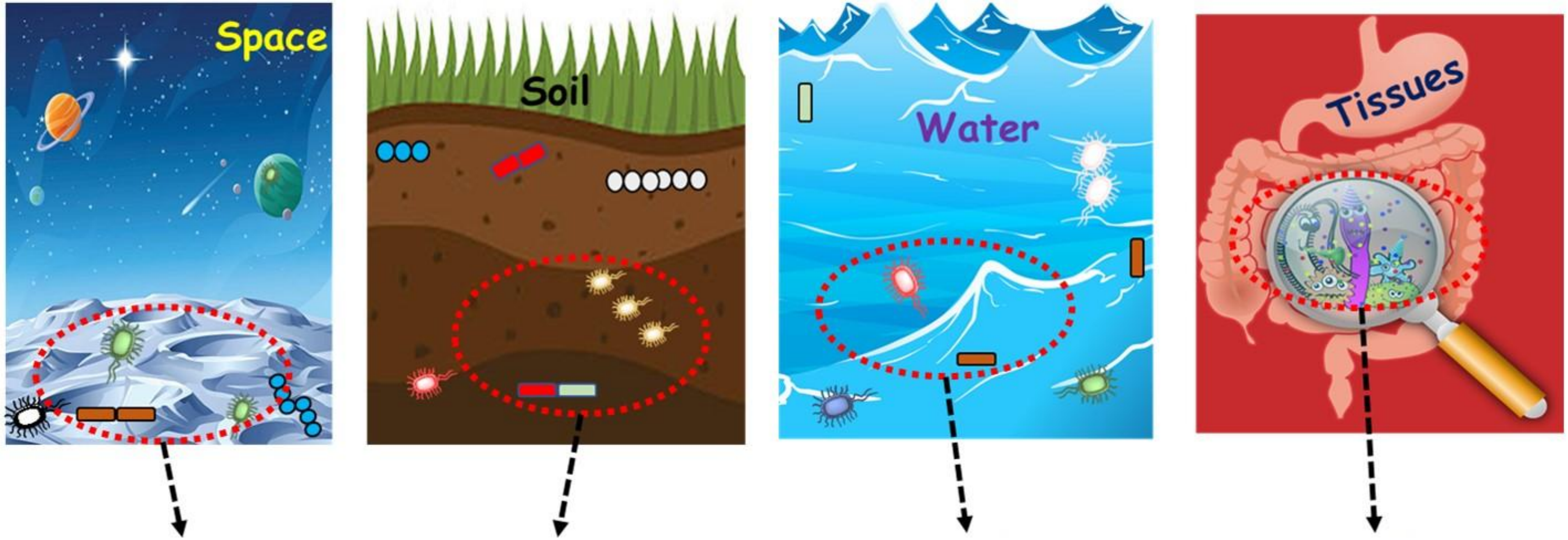


# Metagenomics Workflow



| | c1 | c2 | … |
|---|---|---|---|
| AAAA | 5 | 12 | … |
| AAAT | 0 | 6 | … |
| AAAC | 14 | 7 | … |
| AAAG | 3 | 8 | … |
| … | … | … | … |

# What to think of when designing a experiment…



**Optimized sample collection, preparation, and DNA extraction**

What is GIGO?
The quality of information coming out cannot be better than the quality of information that went in.
GARBAGE IN
GARBAGE OUT
GIGO is used in IT and mathematics
Garbage In, Garbage Out

& Metagenomics

Scientific question

Sample type

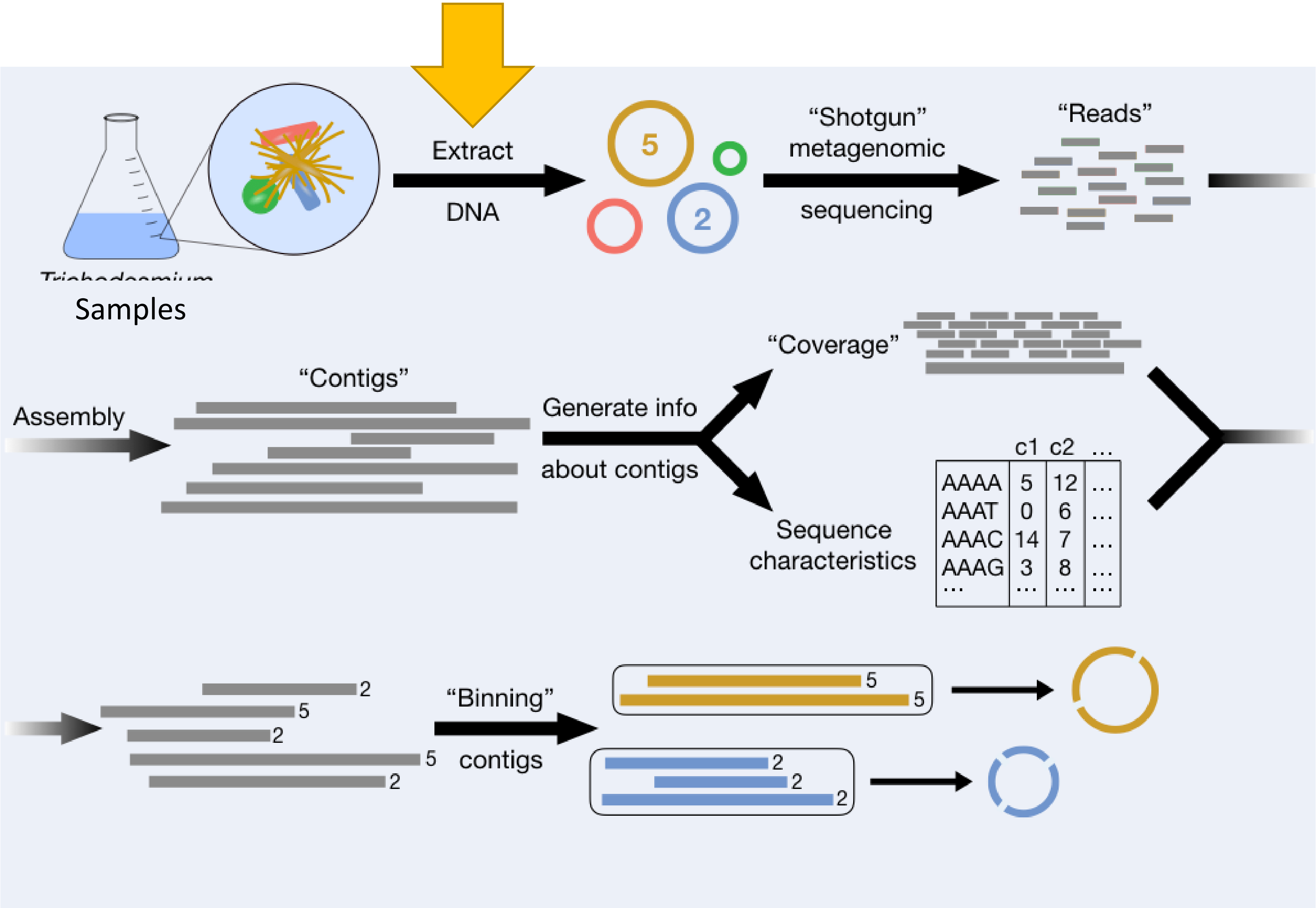
Sample should be representative

Remember all metagenomics values are RELATIVE

Process samples fast

and preserve DNA properly

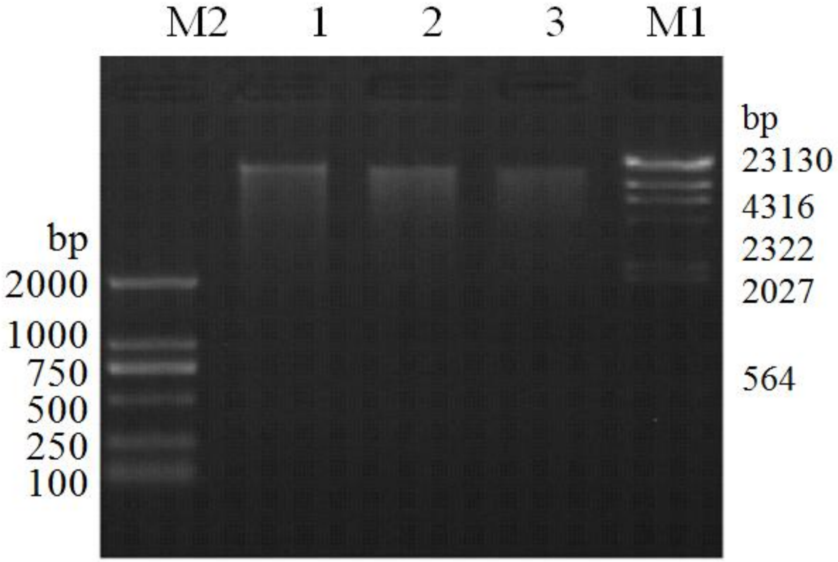
# Metagenomics Workflow



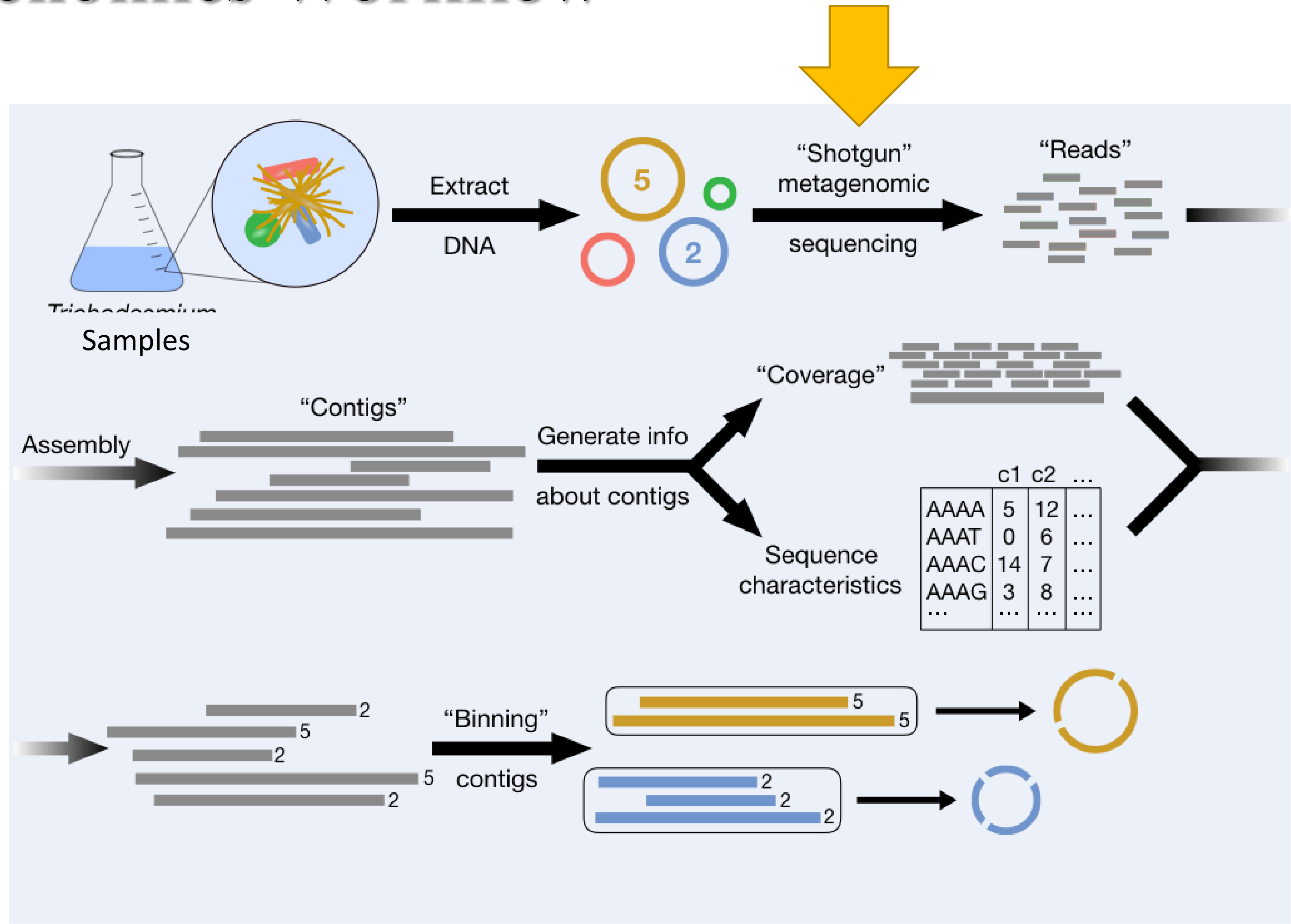
|  | c1 | c2 | … |
|---|---|---|---|
| AAAA | 5 | 12 | … |
| AAAT | 0 | 6 | … |
| AAAC | 14 | 7 | … |
| AAAG | 3 | 8 | … |
| … | … | … | … |

# DNA recovery method impacts the output

- Choice of DNA extraction method
  - Consistent method for all experiments we want to compare
- DNA extraction quality
  - Gel electrophoresis
    - The integrity and size of genomic DNA
  - Spectrophotometry
    - Pure DNA has an A260/A280 ratio of 1.7–1.9
    - DNA concentration has been determined using nanodrop
  - Fluorometry concentration measurements



# Metagenomics Workflow



Samples

Extract DNA

"Shotgun" metagenomic sequencing

"Reads"

Assembly

"Contigs"

Generate info about contigs

"Coverage"

Sequence characteristics

|  | c1 | c2 | … |
|---|---|---|---|
| AAAA | 5 | 12 | … |
| AAAT | 0 | 6 | … |
| AAAC | 14 | 7 | … |
| AAAG | 3 | 8 | … |
| … | … | … | … |

"Binning" contigs

# Sequencing

Short read

illumina®

Long read

PACBIO®

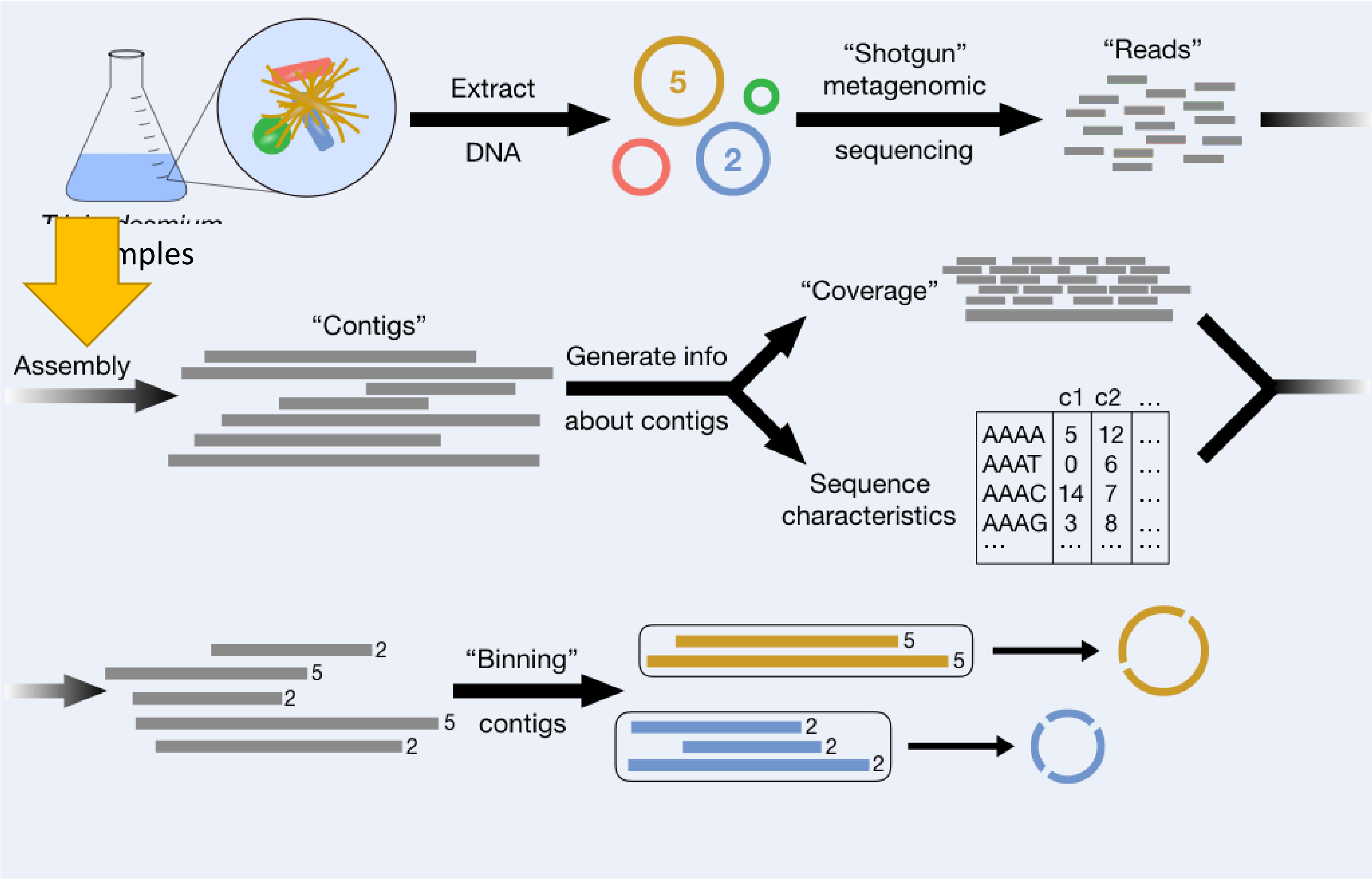
Oxford NANOPORE Technologies

**Consider the sequencing quality, read length, and price** ❗

**Deeper sequencing = higher resolution also = computationally intensive**

**Let’s have a break and come back in 20 min**

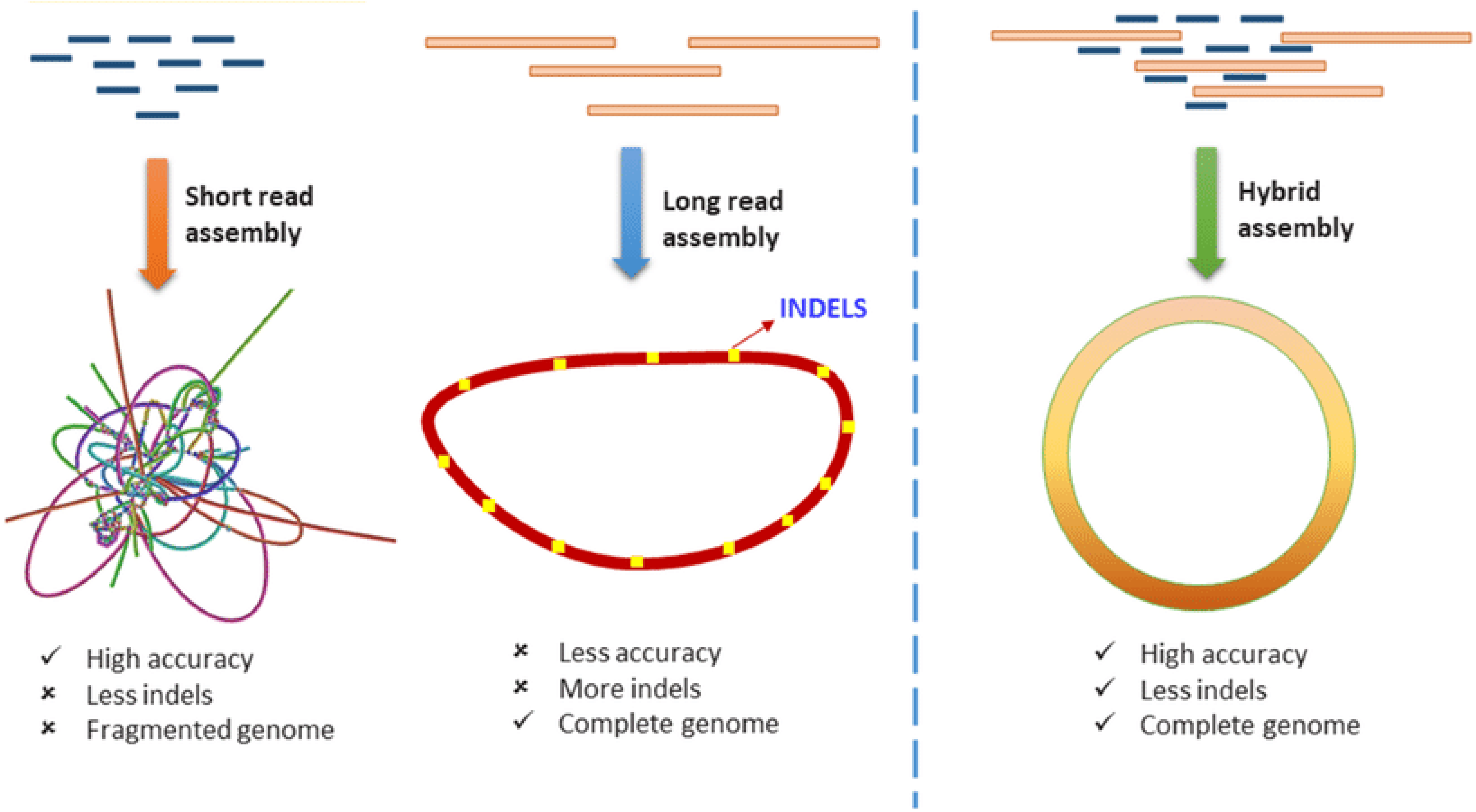
# Metagenomics Workflow



| | c1 | c2 | … |
|---|---|---|---|
| AAAA | 5 | 12 | … |
| AAAT | 0 | 6 | … |
| AAAC | 14 | 7 | … |
| AAAG | 3 | 8 | … |
| … | … | … | … |

# Sequence analysis

## Assembly



Short read assembly

- ✓ High accuracy
- ✗ Less indels
- ✗ Fragmented genome

Long read assembly

- ✗ Less accuracy
- ✗ More indels
- ✓ Complete genome

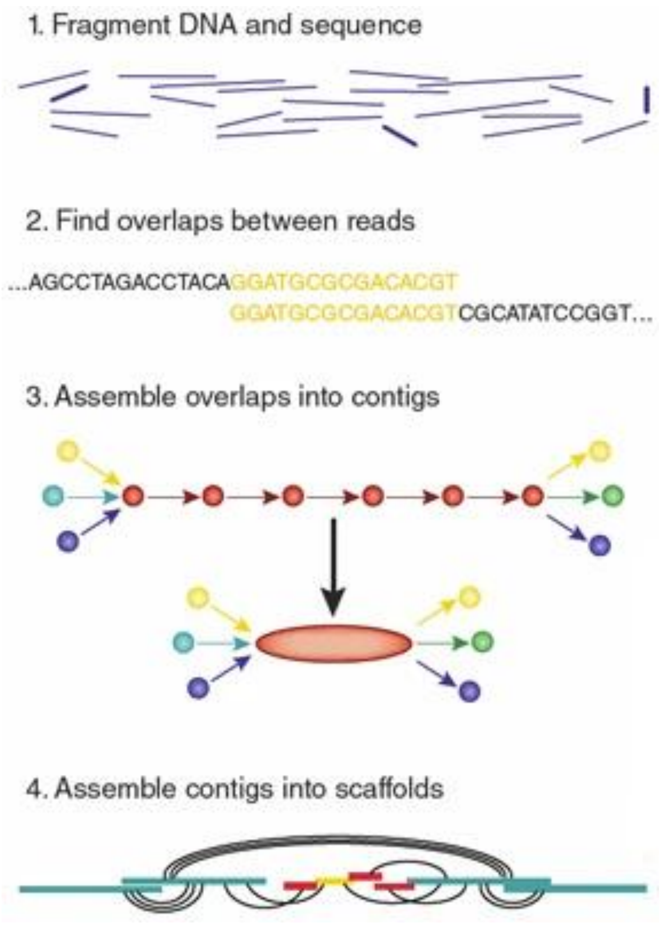
Hybrid assembly

- ✓ High accuracy
- ✓ Less indels
- ✓ Complete genome

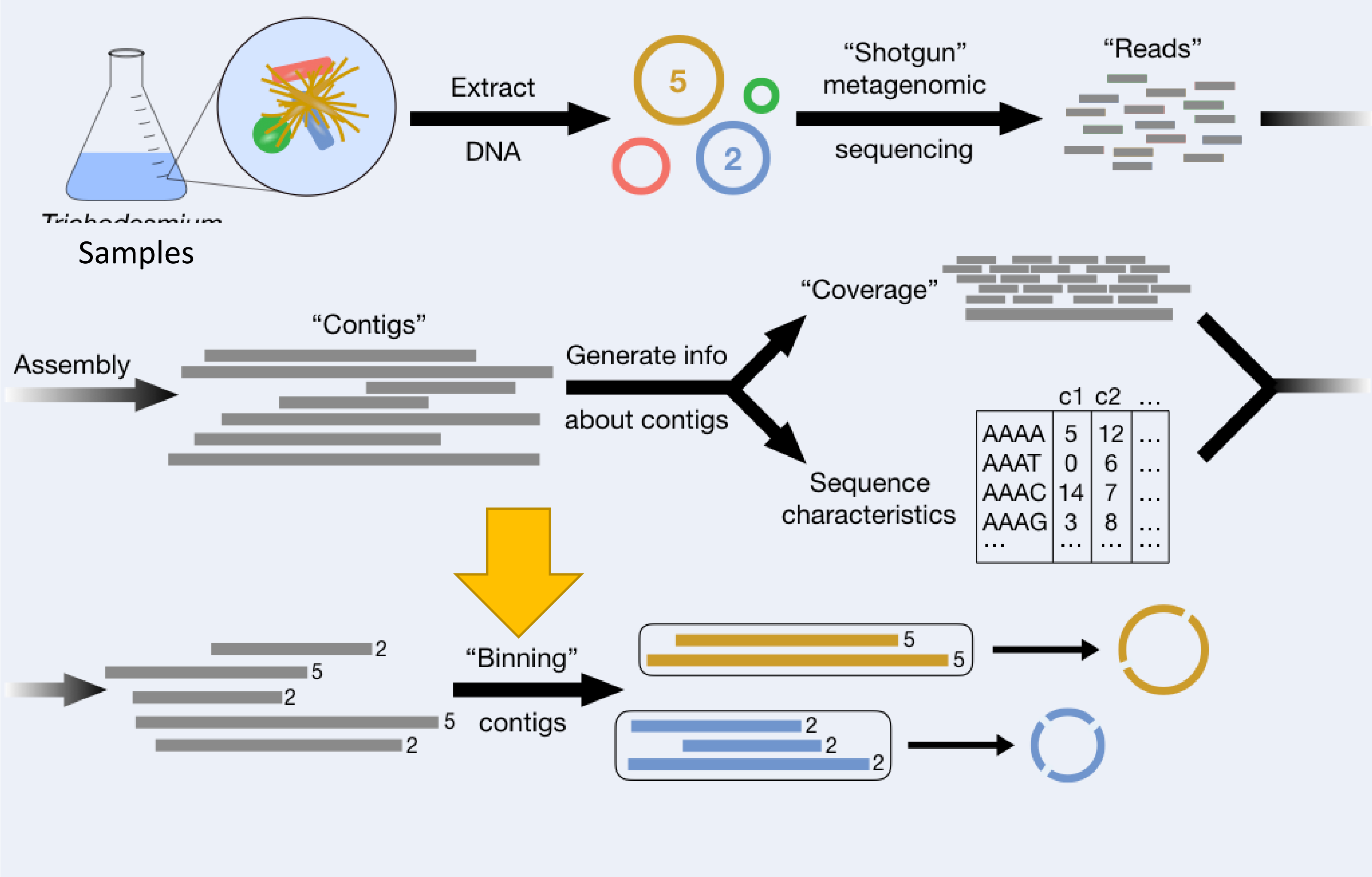
# Assembly tools

- **MEGAHIT:** makes use of succinct *de Bruijn* graphs (SdBG; Bowe *et al*., 2012), which are compressed representation of *de Bruijn* graphs.

- **metaSPAdes:** first constructs the de Bruijn graph of all reads using SPAdes, transforms it into the assembly graph using various graph simplification procedures, and reconstructs paths in the assembly graph that correspond to long genomic fragments within a metagenome.


1. Fragment DNA and sequence
2. Find overlaps between reads
...AGCCTAGACCTACAGGATGCGCGACACGT
GGATGCGCGACACGTCGCATATCCGGT...
3. Assemble overlaps into contigs
4. Assemble contigs into scaffolds

# Metagenomics Workflow



Trichodesmium

Samples

Extract DNA

"Shotgun" metagenomic sequencing

"Reads"

Assembly

"Contigs"

Generate info about contigs

"Coverage"

Sequence characteristics

|  | c1 | c2 | … |
|---|---|---|---|
| AAAA | 5 | 12 | … |
| AAAT | 0 | 6 | … |
| AAAC | 14 | 7 | … |
| AAAG | 3 | 8 | … |
| … | … | … | … |

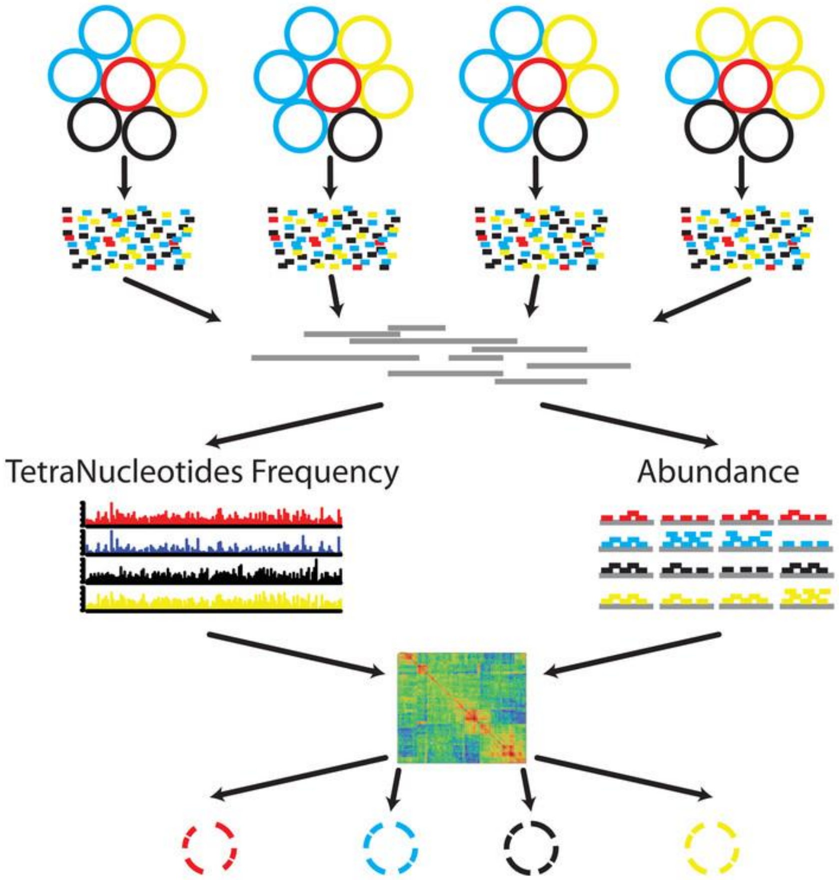
"Binning" contigs

# Sequence analysis

**Binning**

MAG (Metagenome assembled genome)

MAG quality check



**Preprocessing**

1 Samples from multiple sites or times

2 Metagenome libraries

3 Initial de-novo assembly using the combined library

**MetaBAT**

4 Calculate TNF for each contig

5 Calculate Abundance per library for each contig

6 Calculate the pairwise distance matrix using pre-trained probabilistic models

7 Forming genome bins iteratively

# Binning tools

- **MetaBAT2:** uses the same raw TNF and abundance (ABD) scores

- **CONCOCT:** does unsupervised binning of metagenomic contigs by using nucleotide composition - kmer frequencies - and coverage data for multiple samples.

- **MaxBin:** algorithm utilizes two different genomic features: tetranucleotide frequencies and scaffold coverage levels to populate the genomic bins using single-copy maker genes and an expectation-maximization algorithm.

# Genome/MAG quality check

- CheckM provides robust estimates of genome completeness and contamination by using collocated sets of genes that are ubiquitous and single-copy within a phylogenetic lineage.
- Assessment of genome quality using plots depicting key genomic characteristics (e.g., GC, coding density) which highlight sequences outside the expected distributions of a typical genome.
- CheckM also identifies genome bins that are likely candidates for merging based on marker set compatibility, similarity in genomic characteristics, and proximity within a reference genome tree.
- https://ecogenomics.github.io/CheckM/

# Genome completeness standards

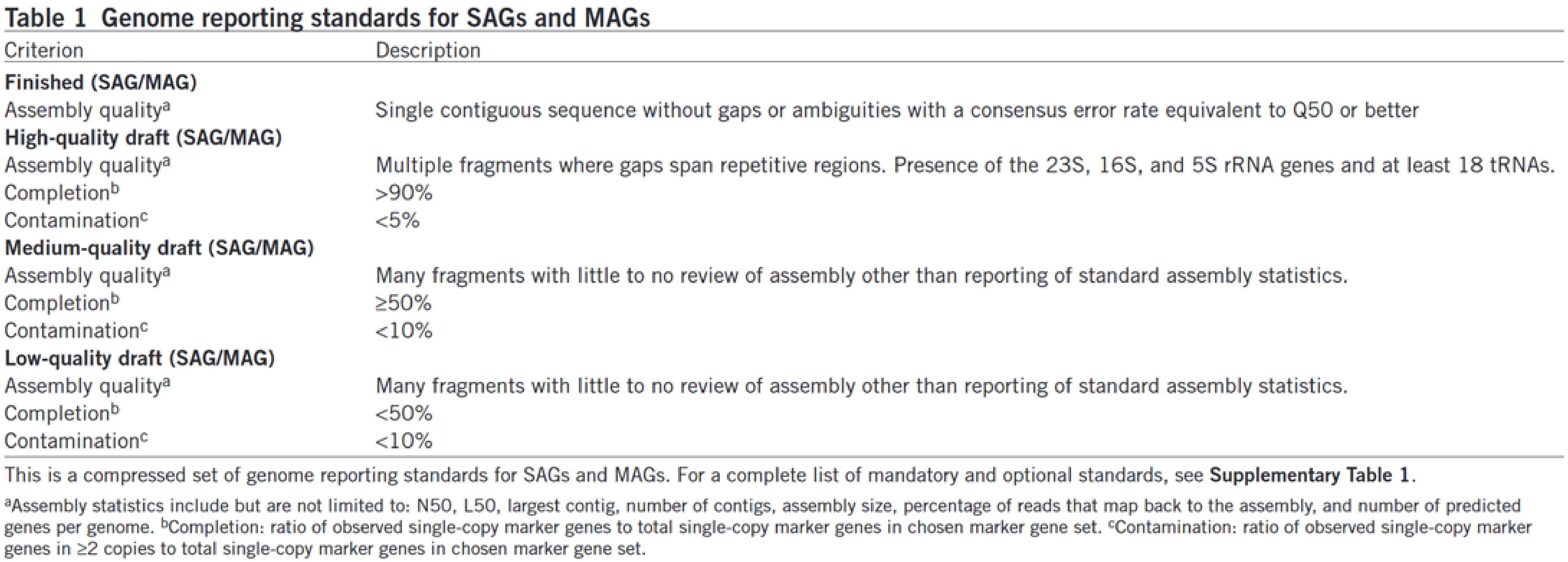


**Table 1 Genome reporting standards for SAGs and MAGs**

| Criterion | Description |
|---|---|
| **Finished (SAG/MAG)** | |
| Assembly quality[a] | Single contiguous sequence without gaps or ambiguities with a consensus error rate equivalent to Q50 or better |
| **High-quality draft (SAG/MAG)** | |
| Assembly quality[a] | Multiple fragments where gaps span repetitive regions. Presence of the 23S, 16S, and 5S rRNA genes and at least 18 tRNAs. |
| Completion[b] | >90% |
| Contamination[c] | <5% |
| **Medium-quality draft (SAG/MAG)** | |
| Assembly quality[a] | Many fragments with little to no review of assembly other than reporting of standard assembly statistics. |
| Completion[b] | ≥50% |
| Contamination[c] | <10% |
| **Low-quality draft (SAG/MAG)** | |
| Assembly quality[a] | Many fragments with little to no review of assembly other than reporting of standard assembly statistics. |
| Completion[b] | <50% |
| Contamination[c] | <10% |

This is a compressed set of genome reporting standards for SAGs and MAGs. For a complete list of mandatory and optional standards, see **Supplementary Table 1**.

[a] Assembly statistics include but are not limited to: N50, L50, largest contig, number of contigs, assembly size, percentage of reads that map back to the assembly, and number of predicted genes per genome. [b] Completion: ratio of observed single-copy marker genes to total single-copy marker genes in chosen marker gene set. [c] Contamination: ratio of observed single-copy marker genes in ≥2 copies to total single-copy marker genes in chosen marker gene set.

# Genome taxonomy

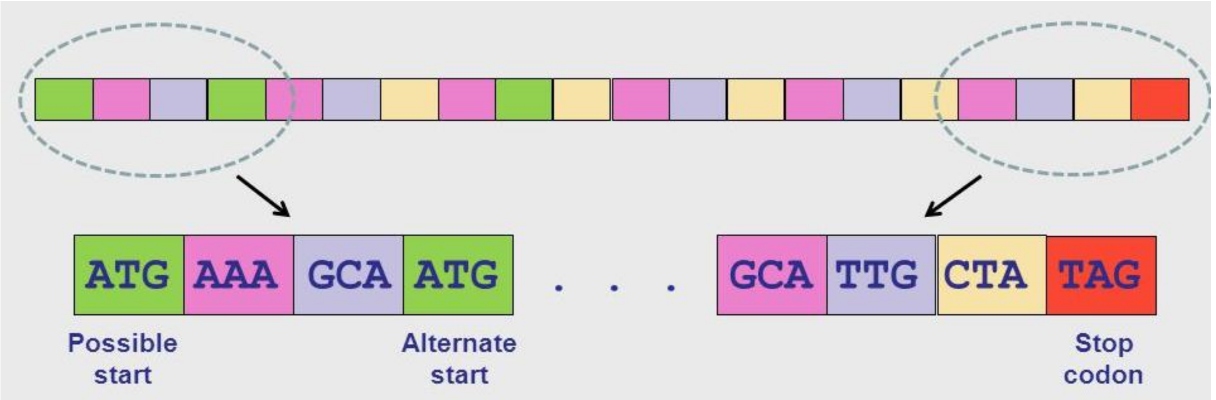
- Taxonomy and nomenclature
- https://gtdb.ecogenomic.org/
- https://ncbiinsights.ncbi.nlm.nih.gov/2021/12/10/ncbi-taxonomy-prokaryote-phyla-added/

# Sequence analysis

## Annotation

Gene prediction => ORF finding

Prodigal



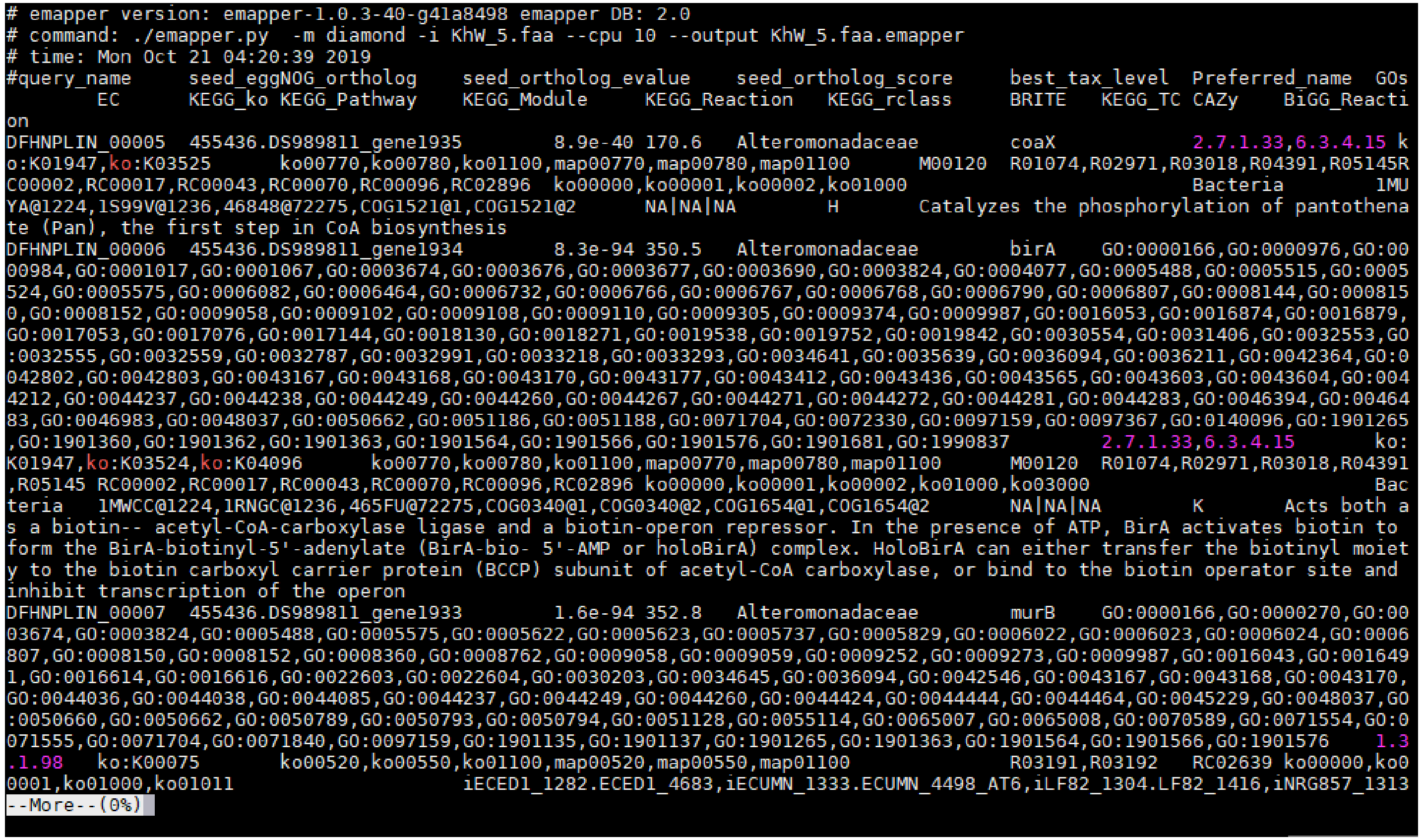
Function assignment

BLAST

HMM models

# Lists of IDs lists of names

- BioCyc
- KEGG
- Ensembl Bacteria
- Kbase
- IMG
- PATRIC



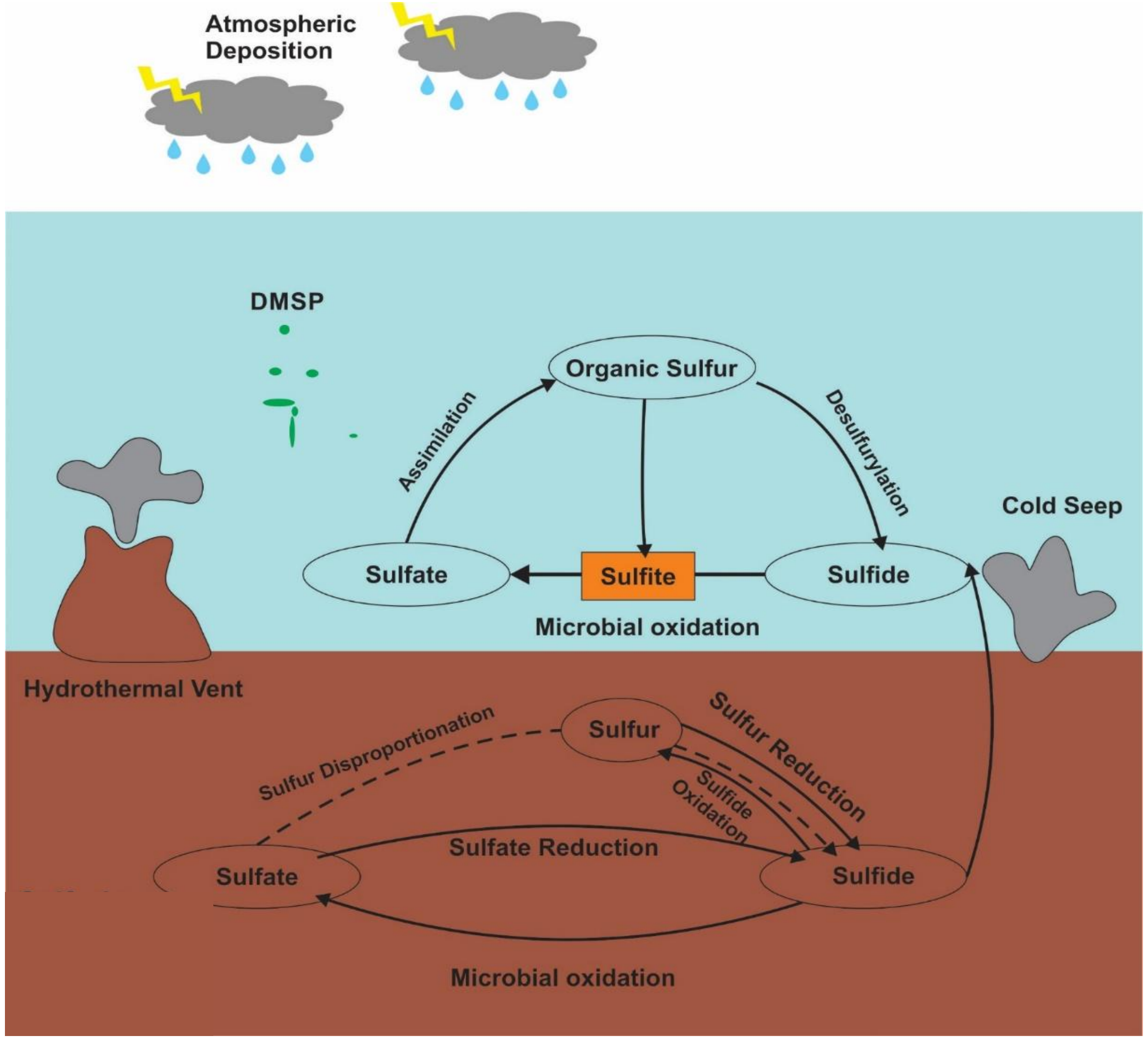
```
# emapper version: emapper-1.0.3-40-g41a8498 emapper DB: 2.0
# command: ./emapper.py  -m diamond -i KhW_5.faa --cpu 10 --output KhW_5.faa.emapper
# time: Mon Oct 21 04:20:39 2019
#query_name     seed_eggNOG_ortholog    seed_ortholog_evalue    seed_ortholog_score     best_tax_level  Preferred_name  GOs
        EC      KEGG_ko KEGG_Pathway    KEGG_Module     KEGG_Reaction   KEGG_rclass     BRITE   KEGG_TC CAZy    BiGG_Reacti
on
DFHNPLIN_00005  455436.DS989811_gene1935        8.9e-40 170.6   Alteromonadaceae        coaX            2.7.1.33,6.3.4.15 k
o:K01947,ko:K03525      ko00770,ko00780,ko01100,map00770,map00780,map01100      M00120  R01074,R02971,R03018,R04391,R05145R
C00002,RC00017,RC00043,RC00070,RC00096,RC02896  ko00000,ko00001,ko00002,ko01000                         Bacteria        1MU
YA@1224,1S99V@1236,46848@72275,COG1521@1,COG1521@2      NA|NA|NA        H       Catalyzes the phosphorylation of pantothena
te (Pan), the first step in CoA biosynthesis
DFHNPLIN_00006  455436.DS989811_gene1934        8.3e-94 350.5   Alteromonadaceae        birA    GO:0000166,GO:0000976,GO:00
00984,GO:0001017,GO:0001067,GO:0003674,GO:0003676,GO:0003677,GO:0003690,GO:0003824,GO:0004077,GO:0005488,GO:0005515,GO:0005
524,GO:0005575,GO:0006082,GO:0006464,GO:0006732,GO:0006766,GO:0006767,GO:0006768,GO:0006790,GO:0006807,GO:0008144,GO:000815
0,GO:0008152,GO:0009058,GO:0009102,GO:0009108,GO:0009110,GO:0009305,GO:0009374,GO:0009987,GO:0016053,GO:0016874,GO:0016879,
GO:0017053,GO:0017076,GO:0017144,GO:0018130,GO:0018271,GO:0019538,GO:0019752,GO:0019842,GO:0030554,GO:0031406,GO:0032553,GO
:0032555,GO:0032559,GO:0032787,GO:0032991,GO:0033218,GO:0033293,GO:0034641,GO:0035639,GO:0036094,GO:0036211,GO:0042364,GO:0
042802,GO:0042803,GO:0043167,GO:0043168,GO:0043170,GO:0043177,GO:0043412,GO:0043436,GO:0043565,GO:0043603,GO:0043604,GO:004
4212,GO:0044237,GO:0044238,GO:0044249,GO:0044260,GO:0044267,GO:0044271,GO:0044272,GO:0044281,GO:0044283,GO:0046394,GO:00464
83,GO:0046983,GO:0048037,GO:0050662,GO:0051186,GO:0051188,GO:0071704,GO:0072330,GO:0097159,GO:0097367,GO:0140096,GO:1901265
,GO:1901360,GO:1901362,GO:1901363,GO:1901564,GO:1901566,GO:1901576,GO:1901681,GO:1990837        2.7.1.33,6.3.4.15       ko:
K01947,ko:K03524,ko:K04096      ko00770,ko00780,ko01100,map00770,map00780,map01100      M00120  R01074,R02971,R03018,R04391
,R05145 RC00002,RC00017,RC00043,RC00070,RC00096,RC02896 ko00000,ko00001,ko00002,ko01000,ko03000                         Bac
teria   1MWCC@1224,1RNGC@1236,465FU@72275,COG0340@1,COG0340@2,COG1654@1,COG1654@2       NA|NA|NA        K       Acts both a
s a biotin-- acetyl-CoA-carboxylase ligase and a biotin-operon repressor. In the presence of ATP, BirA activates biotin to
form the BirA-biotinyl-5'-adenylate (BirA-bio- 5'-AMP or holoBirA) complex. HoloBirA can either transfer the biotinyl moiet
y to the biotin carboxyl carrier protein (BCCP) subunit of acetyl-CoA carboxylase, or bind to the biotin operator site and
inhibit transcription of the operon
DFHNPLIN_00007  455436.DS989811_gene1933        1.6e-94 352.8   Alteromonadaceae        murB    GO:0000166,GO:0000270,GO:00
03674,GO:0003824,GO:0005488,GO:0005575,GO:0005622,GO:0005623,GO:0005737,GO:0005829,GO:0006022,GO:0006023,GO:0006024,GO:0006
807,GO:0008150,GO:0008152,GO:0008360,GO:0008762,GO:0009058,GO:0009059,GO:0009252,GO:0009273,GO:0009987,GO:0016043,GO:001649
1,GO:0016614,GO:0016616,GO:0022603,GO:0022604,GO:0030203,GO:0034645,GO:0036094,GO:0042546,GO:0043167,GO:0043168,GO:0043170,
GO:0044036,GO:0044038,GO:0044085,GO:0044237,GO:0044249,GO:0044260,GO:0044424,GO:0044444,GO:0044464,GO:0045229,GO:0048037,GO
:0050660,GO:0050662,GO:0050789,GO:0050793,GO:0050794,GO:0051128,GO:0055114,GO:0065007,GO:0065008,GO:0070589,GO:0071554,GO:0
071555,GO:0071704,GO:0071840,GO:0097159,GO:1901135,GO:1901137,GO:1901265,GO:1901363,GO:1901564,GO:1901566,GO:1901576    1.3
.1.98   ko:K00075       ko00520,ko00550,ko01100,map00520,map00550,map01100                      R03191,R03192   RC02639 ko00000,ko0
0001,ko01000,ko01011                            iECED1_1282.ECED1_4683,iECUMN_1333.ECUMN_4498_AT6,iLF82_1304.LF82_1416,iNRG857_1313
--More--(0%)
```

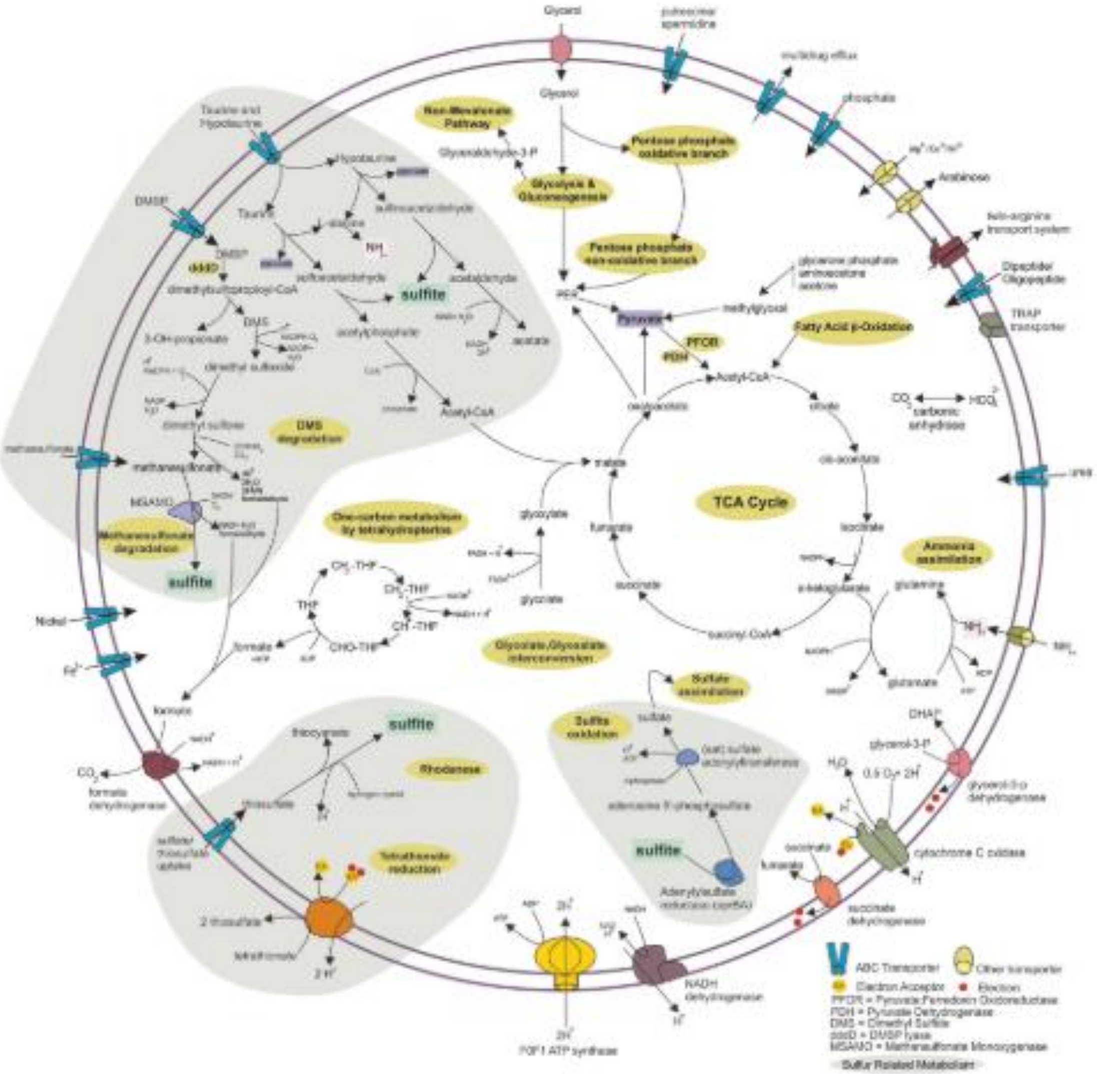
# KEGG

- The Kyoto Encyclopedia of Genes and Genomes is a resource for understanding high-level functions of a biological system from molecular-level information.
-
- Tools for analysis of large-scale molecular datasets generated by high-throughput experimental technologies.

- Home page: https://www.kegg.jp/

# From Protein to metabolism …





# The chase takes time …

## The case of SAR202

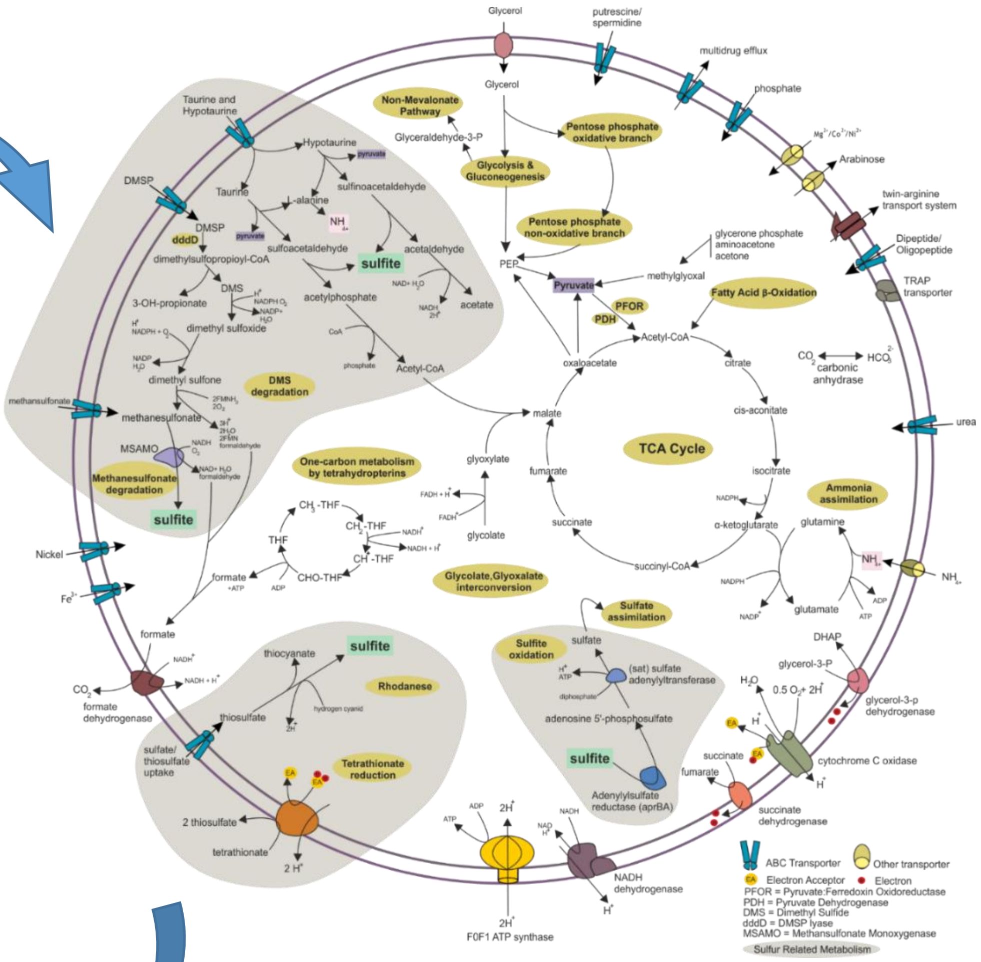
*Proc. Natl. Acad. Sci. USA*
Vol. 93, pp. 7979–7984, July 1996
Microbiology

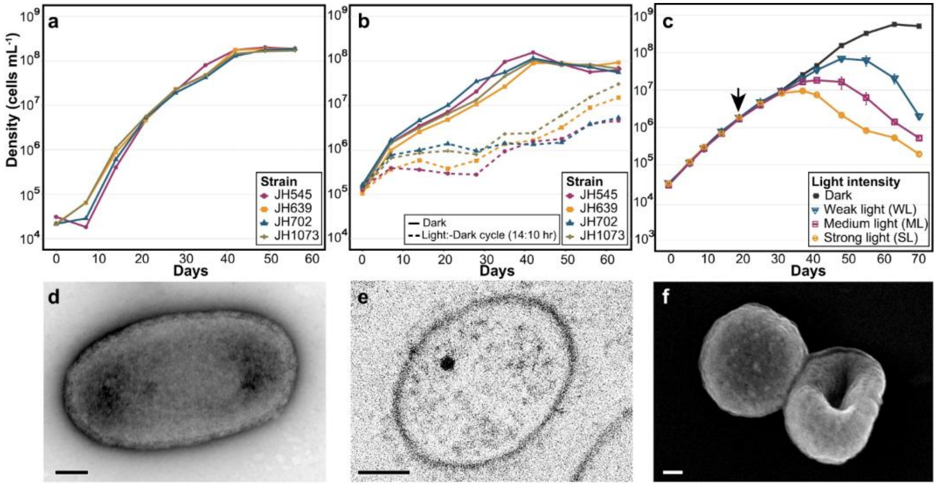
### 16S rRNA genes reveal stratified open ocean bacterioplankton populations related to the Green Non-Sulfur bacteria

(molecular ecology/phylogeny/thermophily)

STEPHEN J. GIOVANNONI*, MICHAEL S. RAPPÉ, KEVIN L. VERGIN, AND NANCI L. ADAIR

**1996**





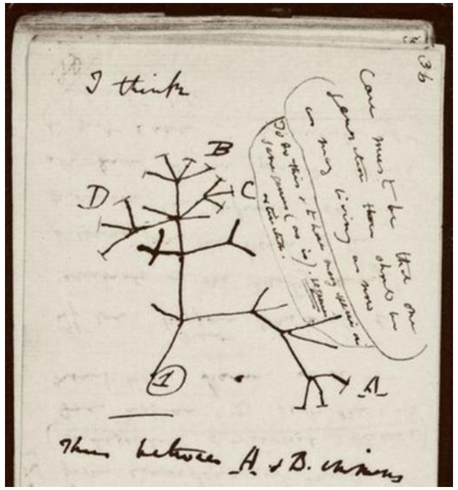
Lim et. al. *Nat. Commun*,

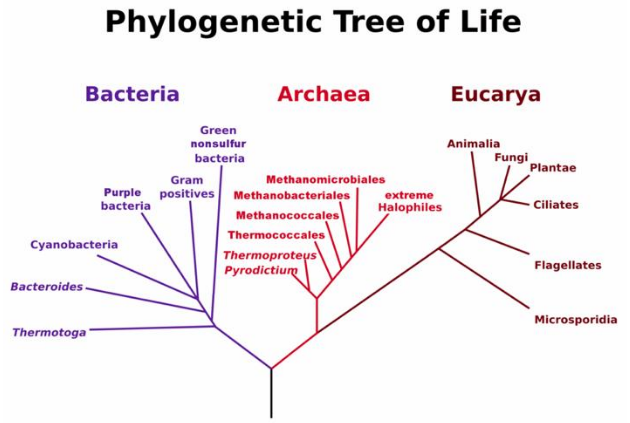
**2023**

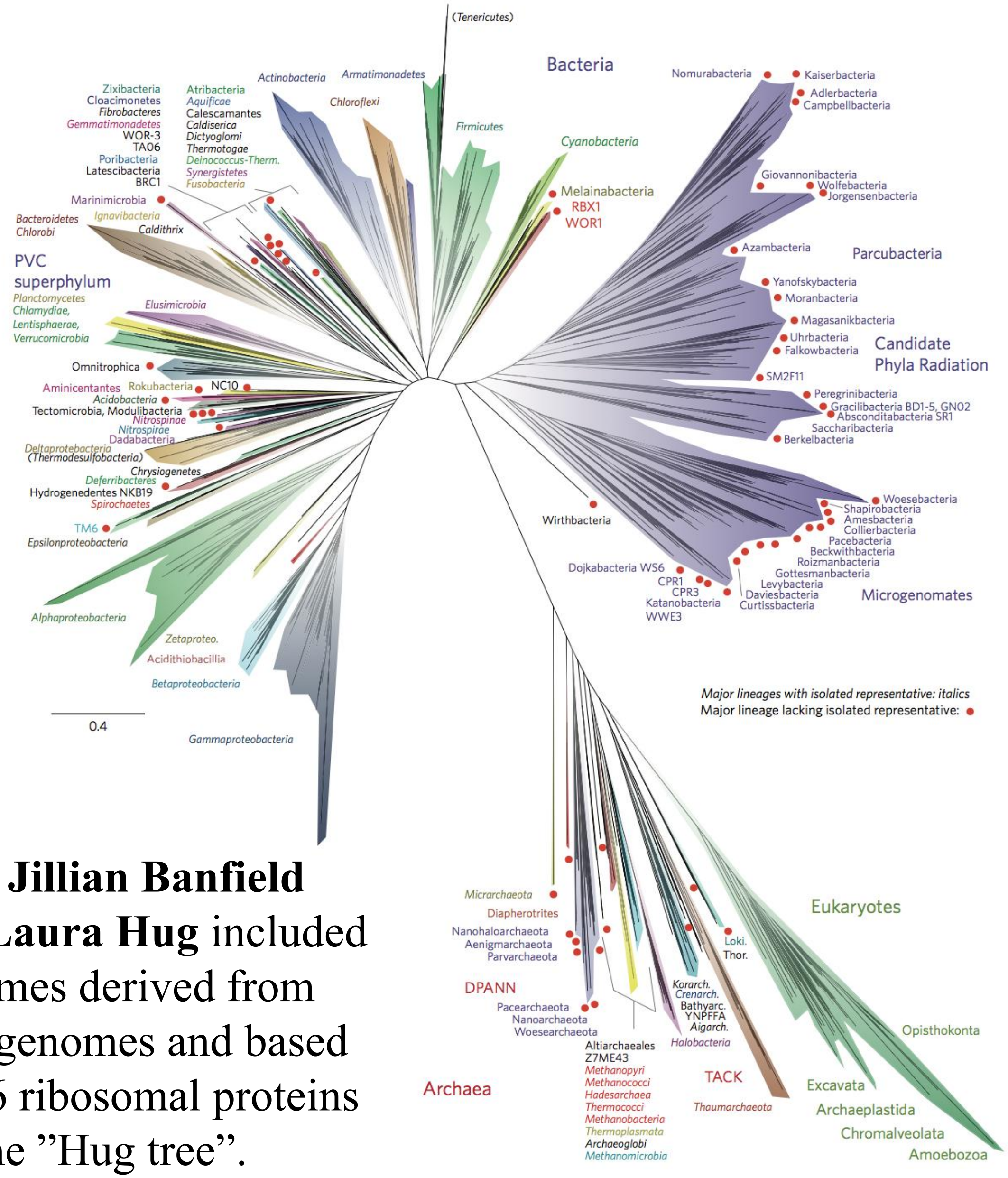
Mehrshad et. al. *ISME J*,

**2018**

# Tree of Life has evolved


36
I think
Case must be that one generation then should be as many living as now. To do this & to have many species in same genus (as is) requires extinction.
B
C
D
1
A
Thus between A & B. immens





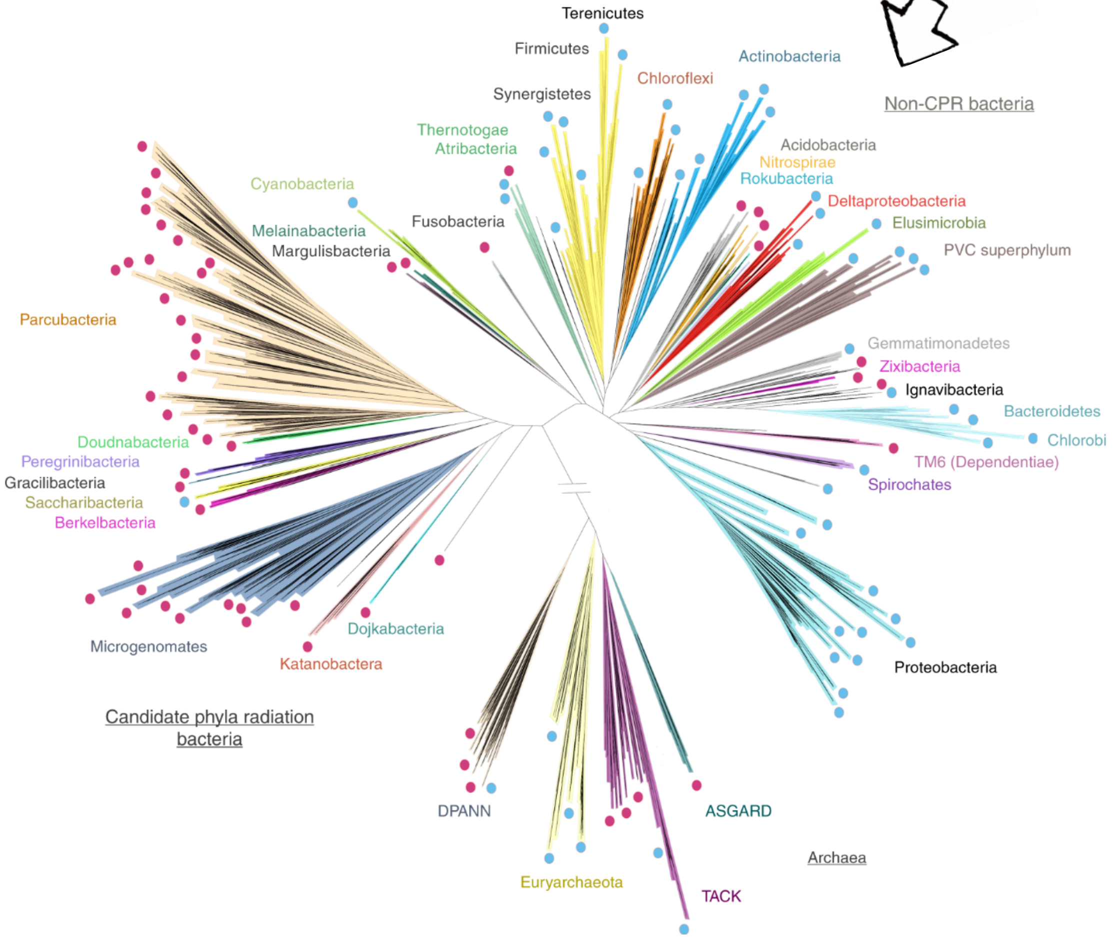
1837 **Charles Darwin**'s ideas on evolution, species descend from common ancestors and evolve over time.

1990 **Carl Woese** tree with LUCA and three domains. Based on rRNA gene. Later elaborated by **Norman Pace**.

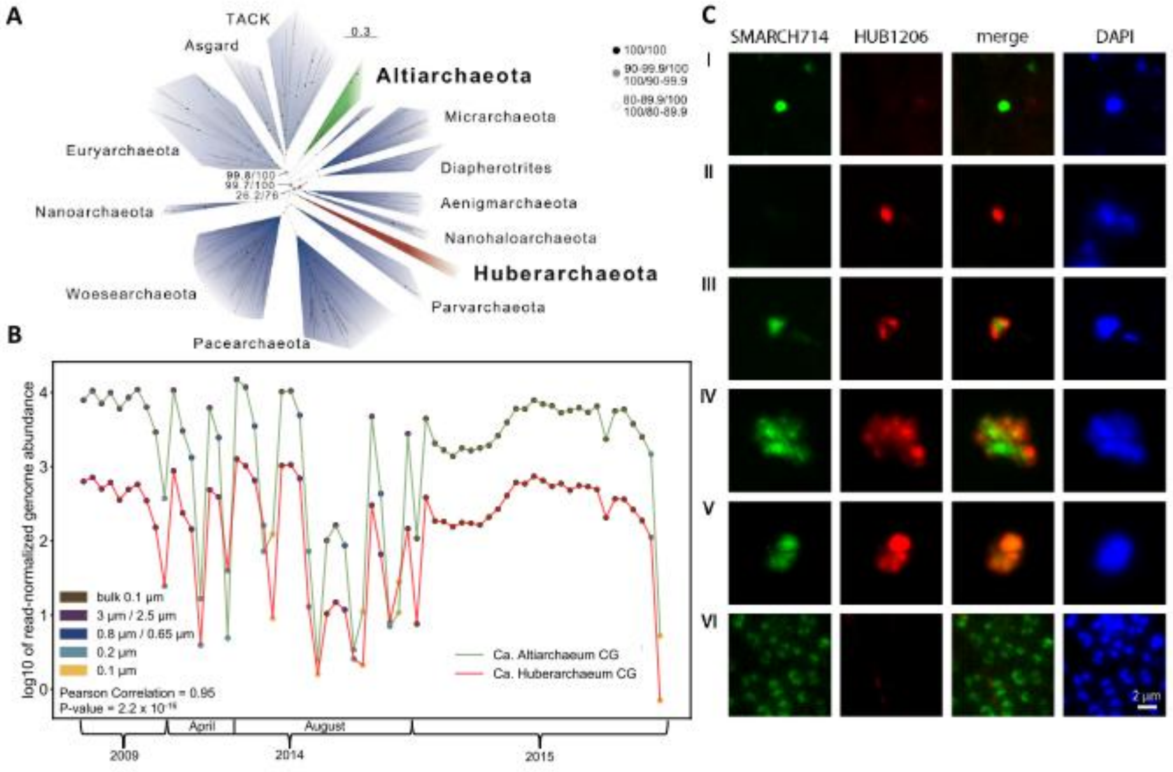
2016 **Jillian Banfield** and **Laura Hug** included genomes derived from metagenomes and based on 16 ribosomal proteins for the ”Hug tree”.

(Hug et al., 2016. Nature Microbiology)

# smallest genomes with epi-symbiotic lifestyle
# CPR & DPANN



Burstein et. al. *Nat Commun*,2019

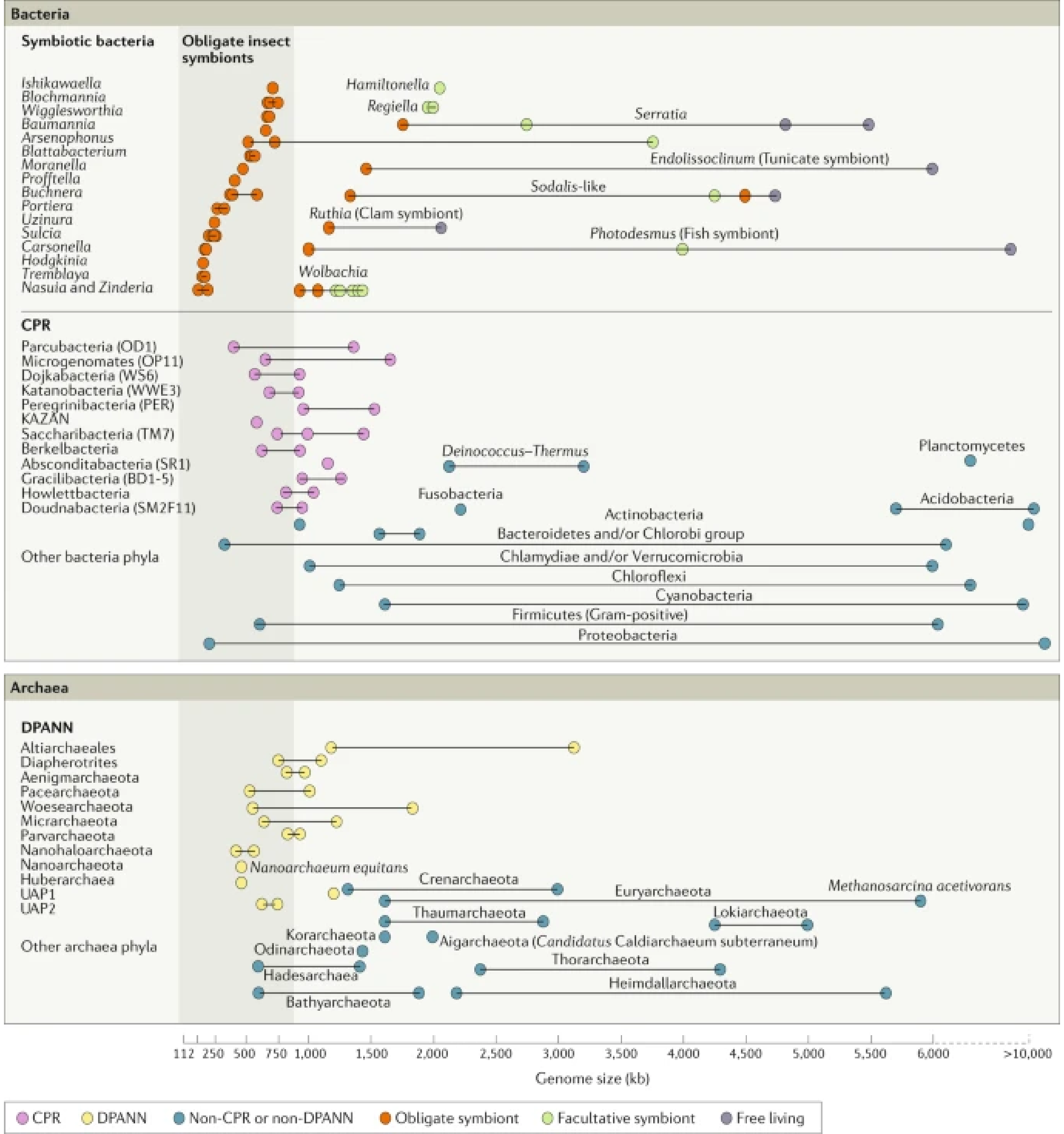


Schwank et. al. ISMEJ,2019

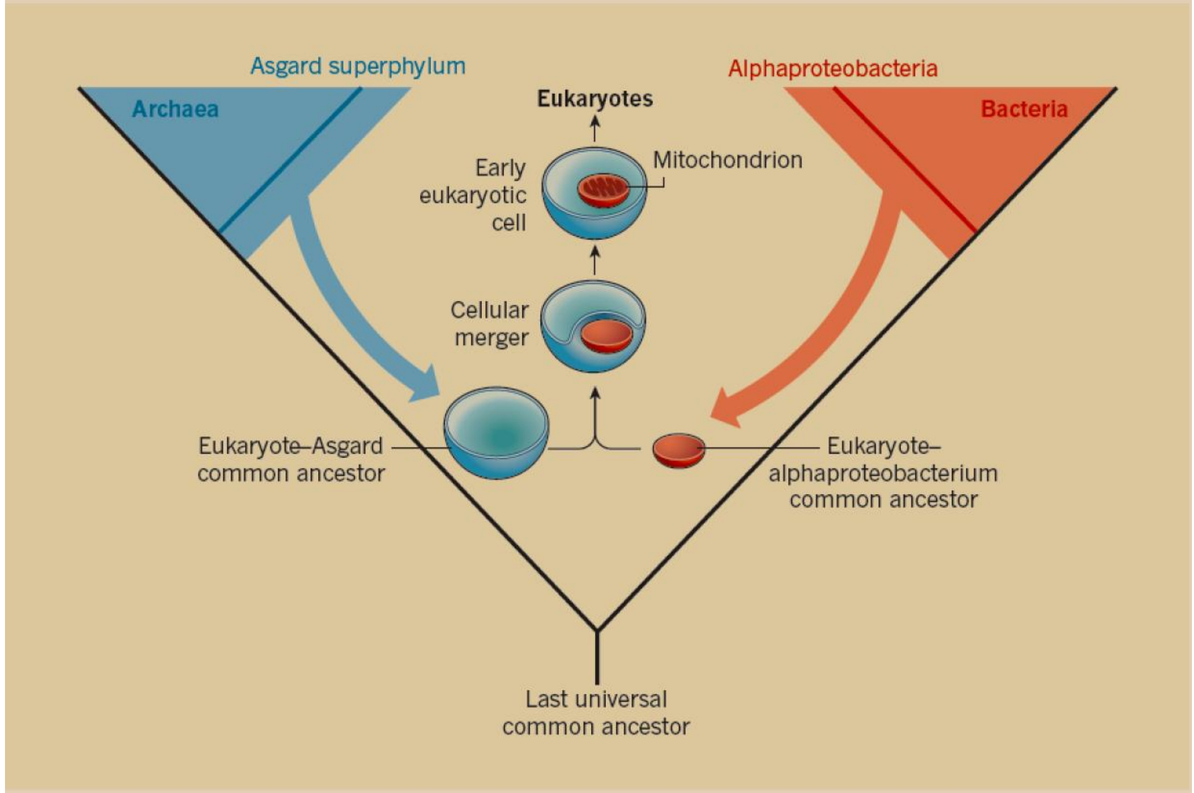
# Remarkable aspect of the tree of life

- Candidate phyla radiation (CPR)
- DPANN (an acronym of the names of the first included phyla, ‘Candidatus Diapherotrites’, ‘Candidatus Parvarchaeota’, ‘Candidatus Aenigmarchaeota’, Nanoarchaeota and ‘Candidatus Nanohaloarchaeota’)

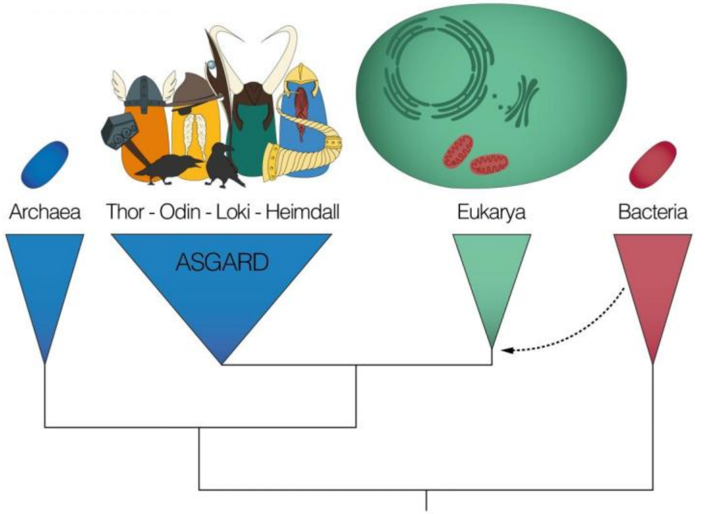
    - Small genomes
    - Small cell sizes
    - Notable gaps in core metabolic potential
    - Mostly symbiotic lifestyle
    - Their ecological role is not yet well understood



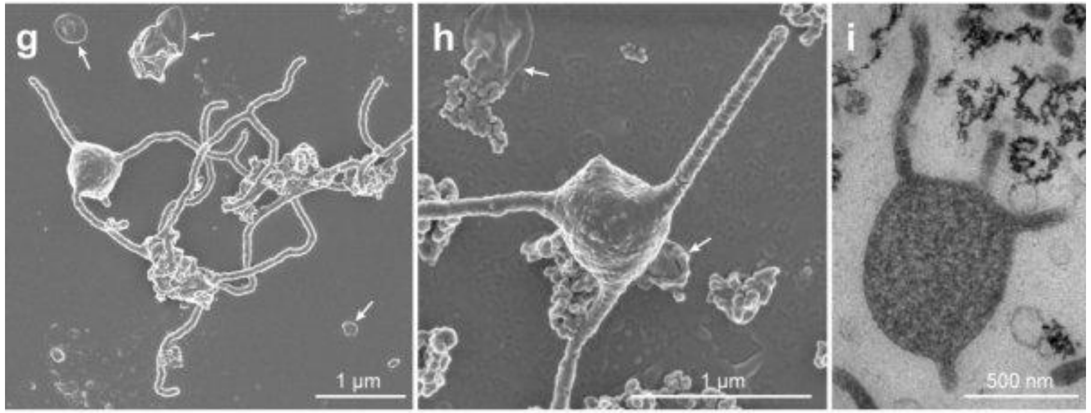
# Asgard archaea illuminate the origin of eukaryotic cellular comple



López-García et. al. *NatMicrobiol*,2019
[https://www.nature.com/articles/s41564-020-0710-4](https://www.nature.com/articles/s41564-020-0710-4)



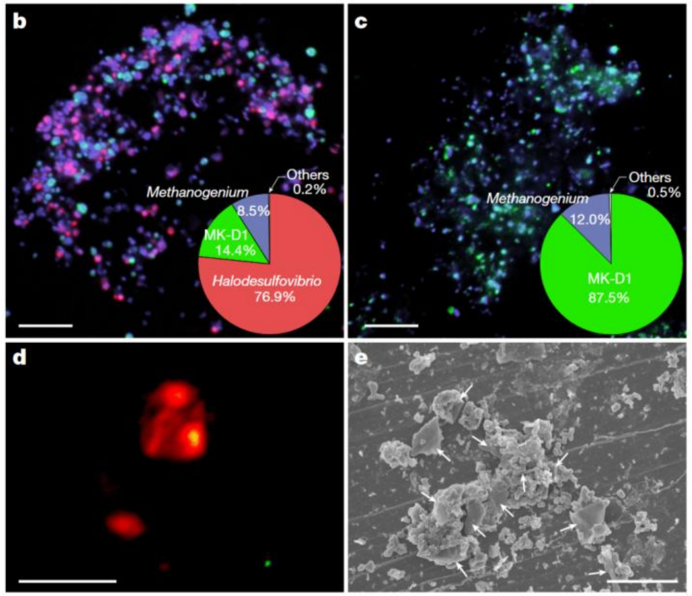
Spang et. al. *Nature*,2015



Imachi & Nobu et. al. *nature*,2020

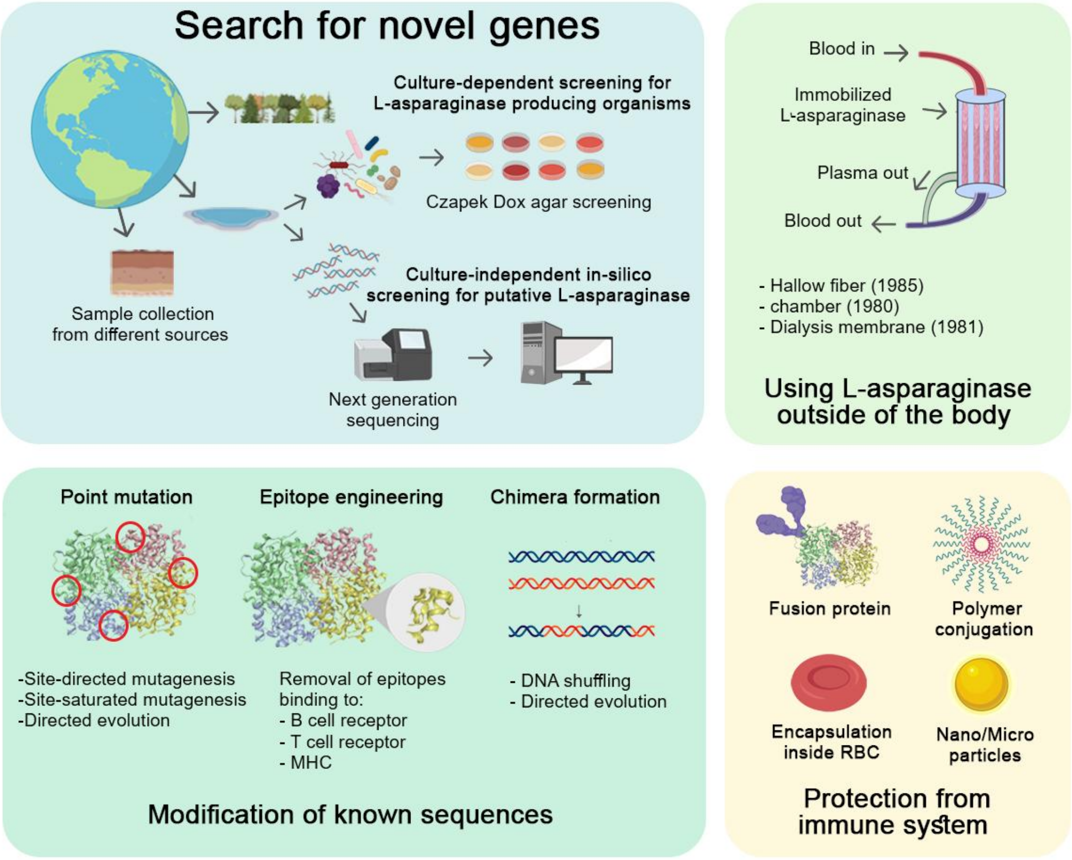
# ‘CandidatusPrometheoarchaeum syntrophicum’

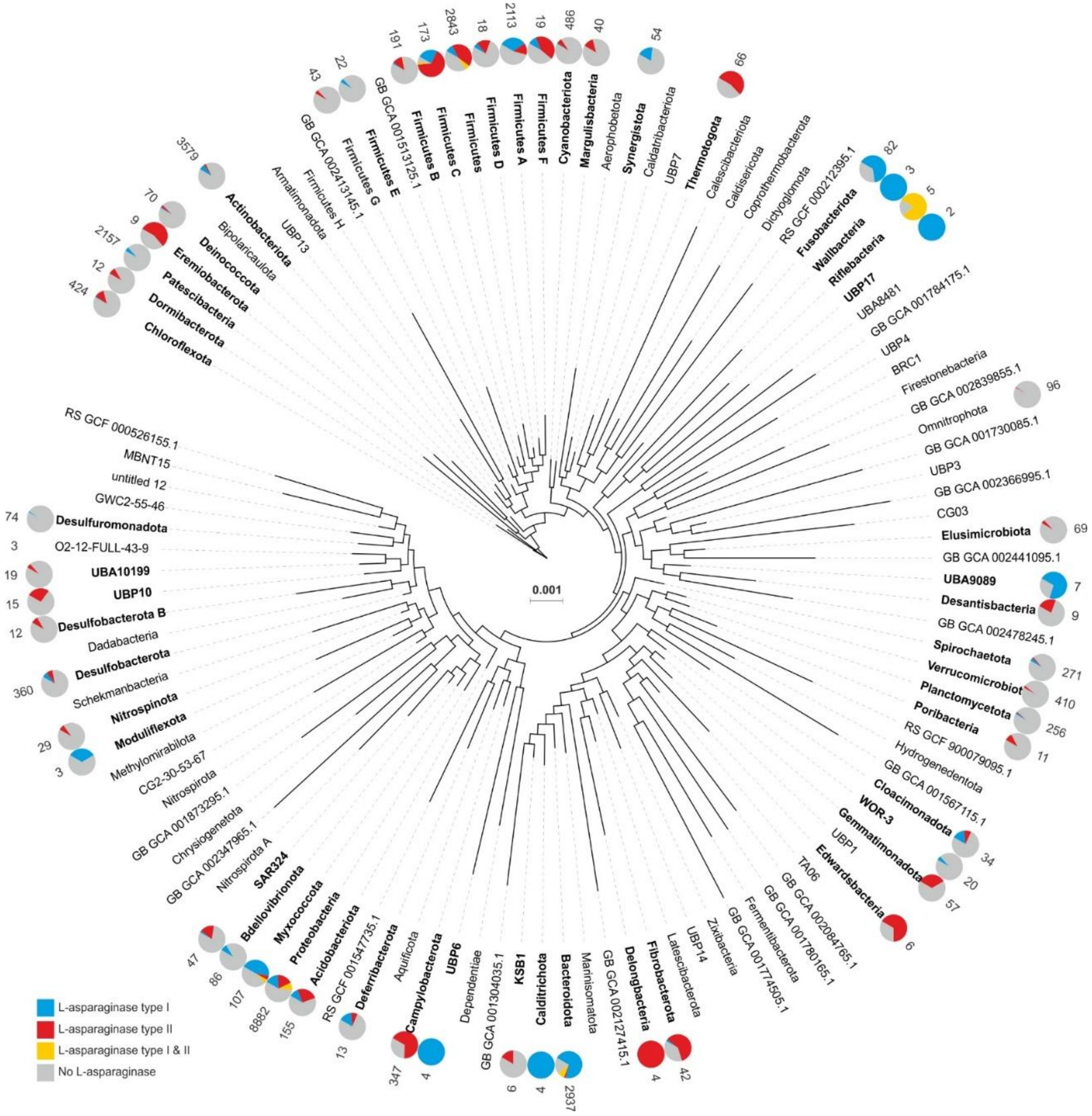
- Pure co-culture of the target archaeon MK-D1 and Methanogenium after a 12-year study
- From bioreactor-based pre-enrichment of deep-sea sediments to a final 7 years of in vitro enrichment.
- Extremely slow growth rate and low cell yield.
- The culture consistently had a **30–60-day lag phase** and required more than 3 months to reach full growth: around $10^5$ 16S rRNA gene copies $ml^{-1}$
- The doubling time was estimated to be approximately **14–25 days**.



- Imachi, H., *et al. Nature,* 2020

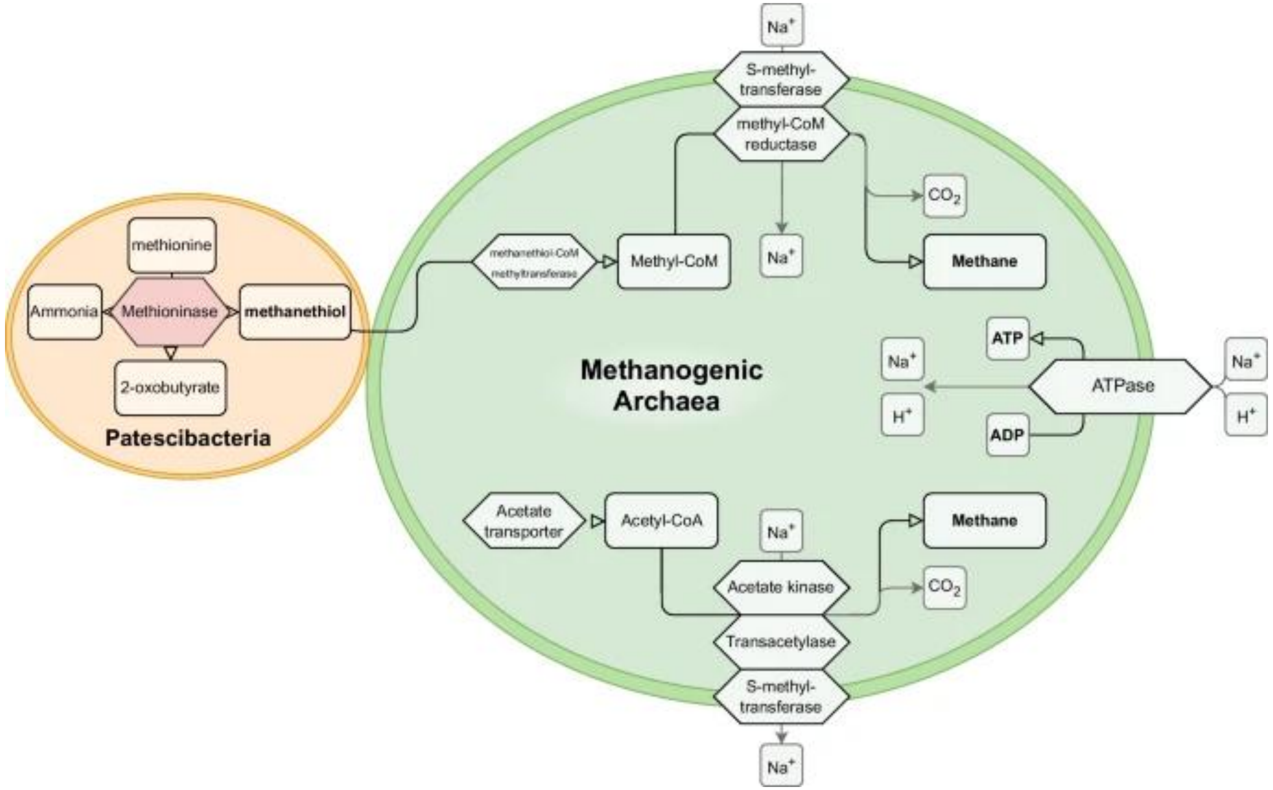
# L-asparaginase

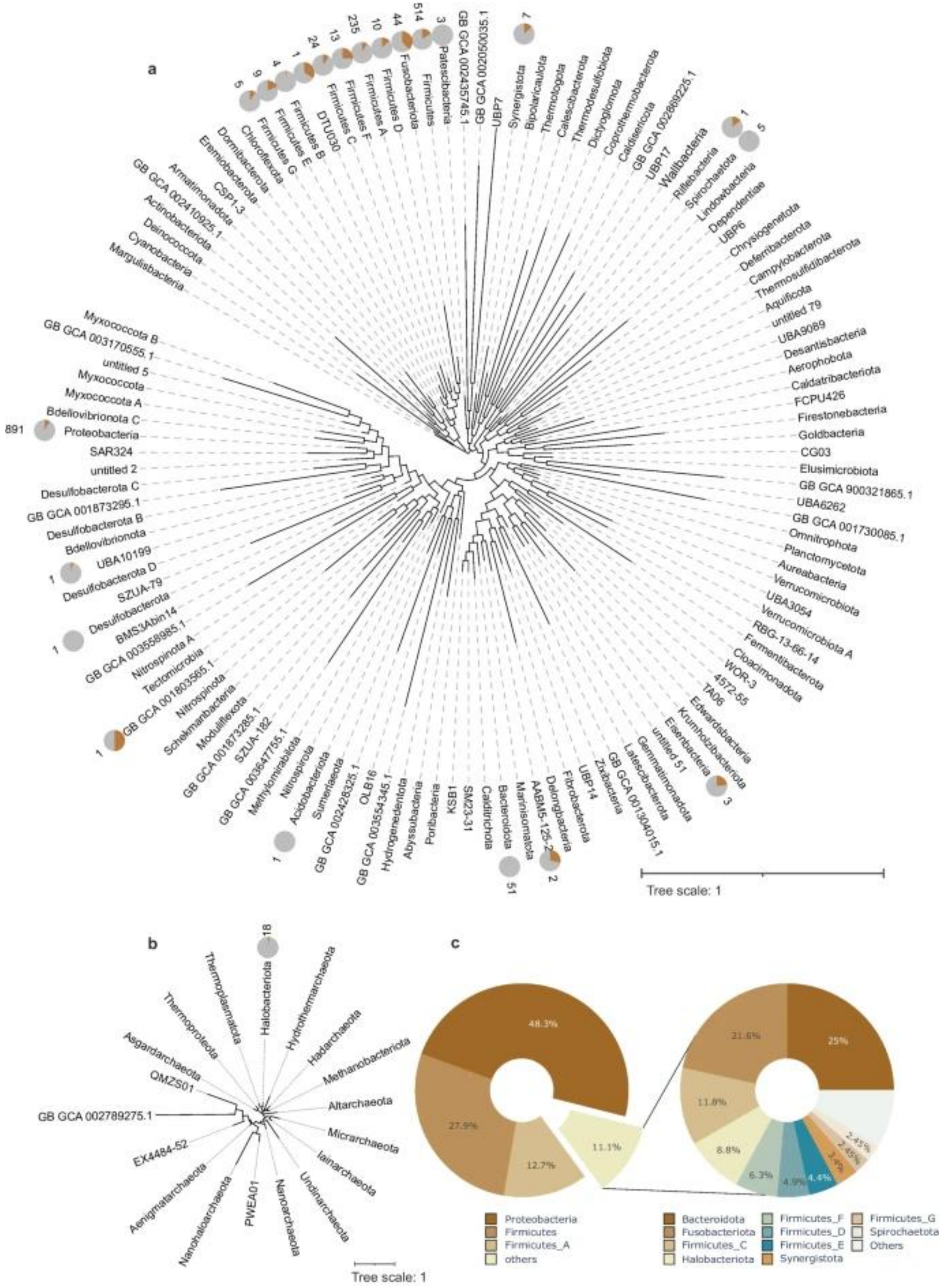

Search for novel genes
Culture-dependent screening for L-asparaginase producing organisms
Czapek Dox agar screening
Culture-independent in-silico screening for putative L-asparaginase
Sample collection from different sources
Next generation sequencing
Blood in
Immobilized L-asparaginase
Plasma out
Blood out
- Hallow fiber (1985)
- chamber (1980)
- Dialysis membrane (1981)
Using L-asparaginase outside of the body
Point mutation
Epitope engineering
Chimera formation
-Site-directed mutagenesis
-Site-saturated mutagenesis
-Directed evolution
Removal of epitopes binding to:
- B cell receptor
- T cell receptor
- MHC
- DNA shuffling
- Directed evolution
Modification of known sequences
Fusion protein
Polymer conjugation
Encapsulation inside RBC
Nano/Micro particles
Protection from immune system



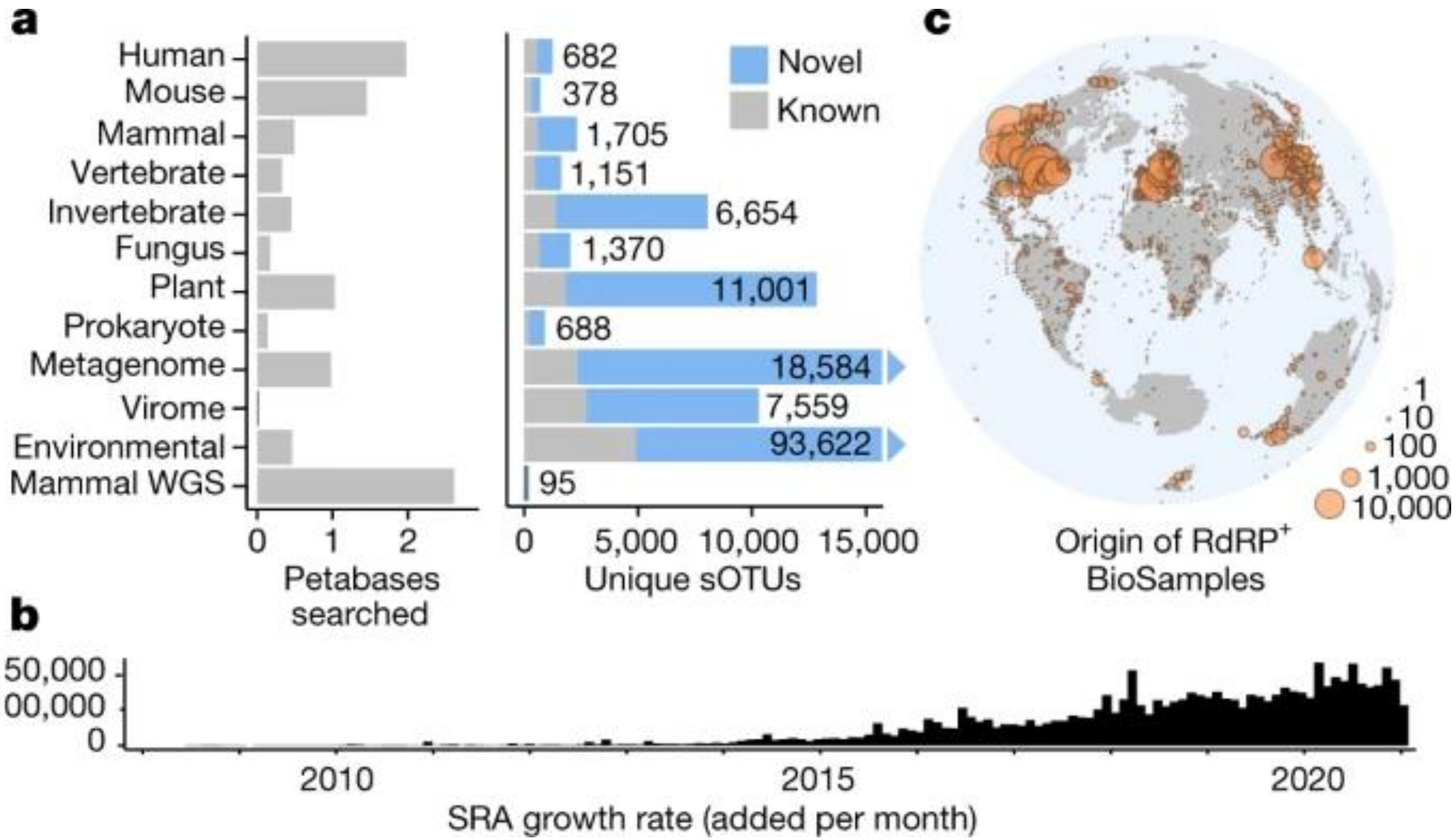
Sobat et. al. *iScience*, 2021

# Methioninase involved in Metabolic Syntrophy

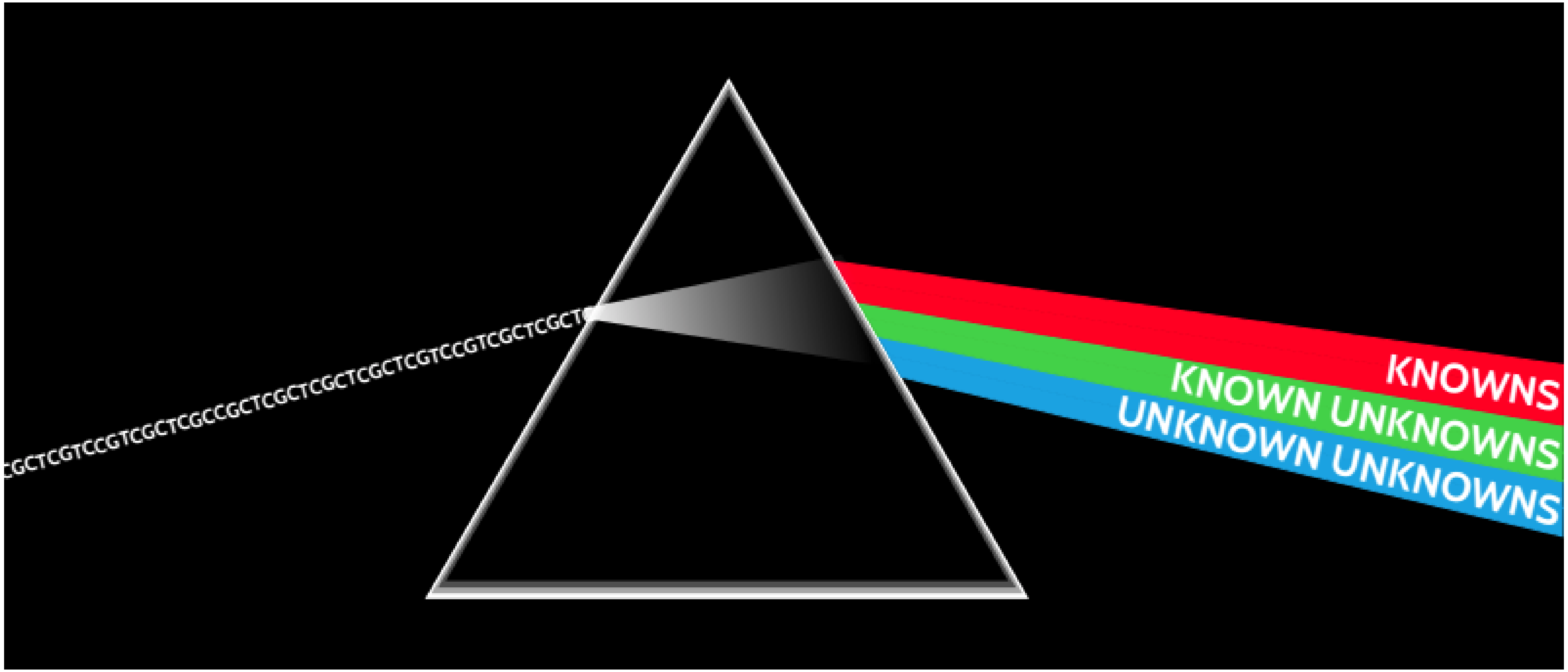




# Petabase-scale sequence alignment catalyses viral discovery



Edgar. *et al. Nature* **2022**.

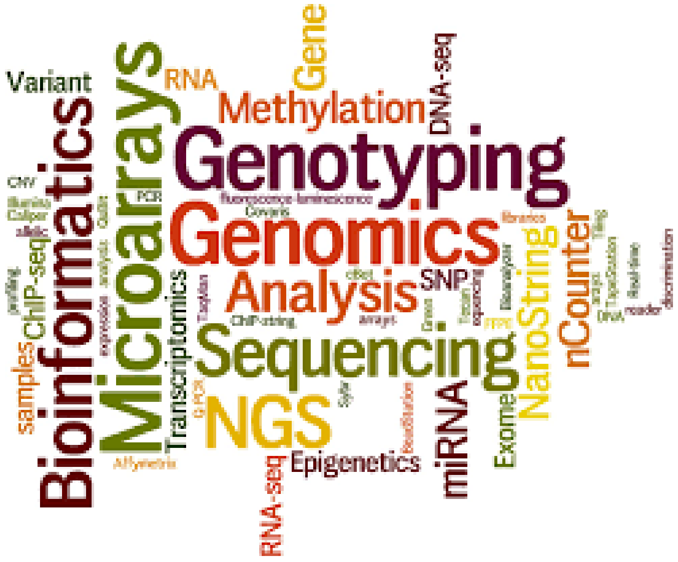

CGCTCGTCCGTCGCTCGCCGCTCGCTCGCTCGCTCGTCCGTCGCTCGCT
KNOWNS
KNOWN UNKNOWNS
UNKNOWN UNKNOWNS

# “multi-omics” approach to answer eco-evolutionary questions

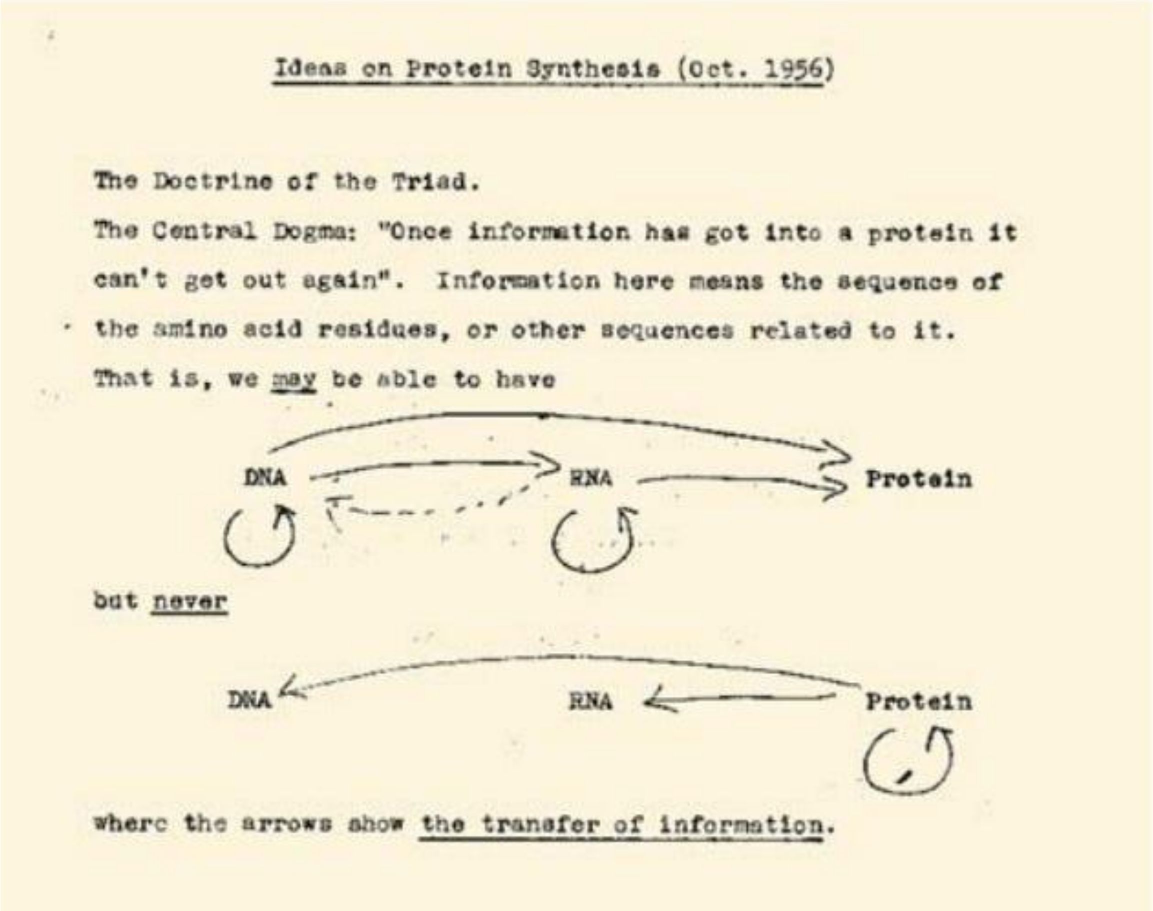
Single-cell amplified genomes

Metagenomics

Metatranscriptome



# The transfer of information



Ideas on Protein Synthesis (Oct. 1956)

The Doctrine of the Triad.

The Central Dogma: "Once information has got into a protein it can't get out again". Information here means the sequence of the amino acid residues, or other sequences related to it.

That is, we may be able to have

$$\begin{tikzcd}
DNA \arrow[r] \arrow[rr, bend left] \arrow[loop below] & RNA \arrow[r] \arrow[l, dashed, bend left] \arrow[loop below] & Protein
\end{tikzcd}$$

but never

$$\begin{tikzcd}
DNA & RNA & Protein \arrow[l] \arrow[ll, bend right] \arrow[loop below]
\end{tikzcd}$$

where the arrows show the transfer of information.

# Metatranscriptome

## What to consider?

- Sampling
- Sample processing
- RNA extraction
- Replicates
- Library preparation (rRNA depletion)

# Metatranscriptome

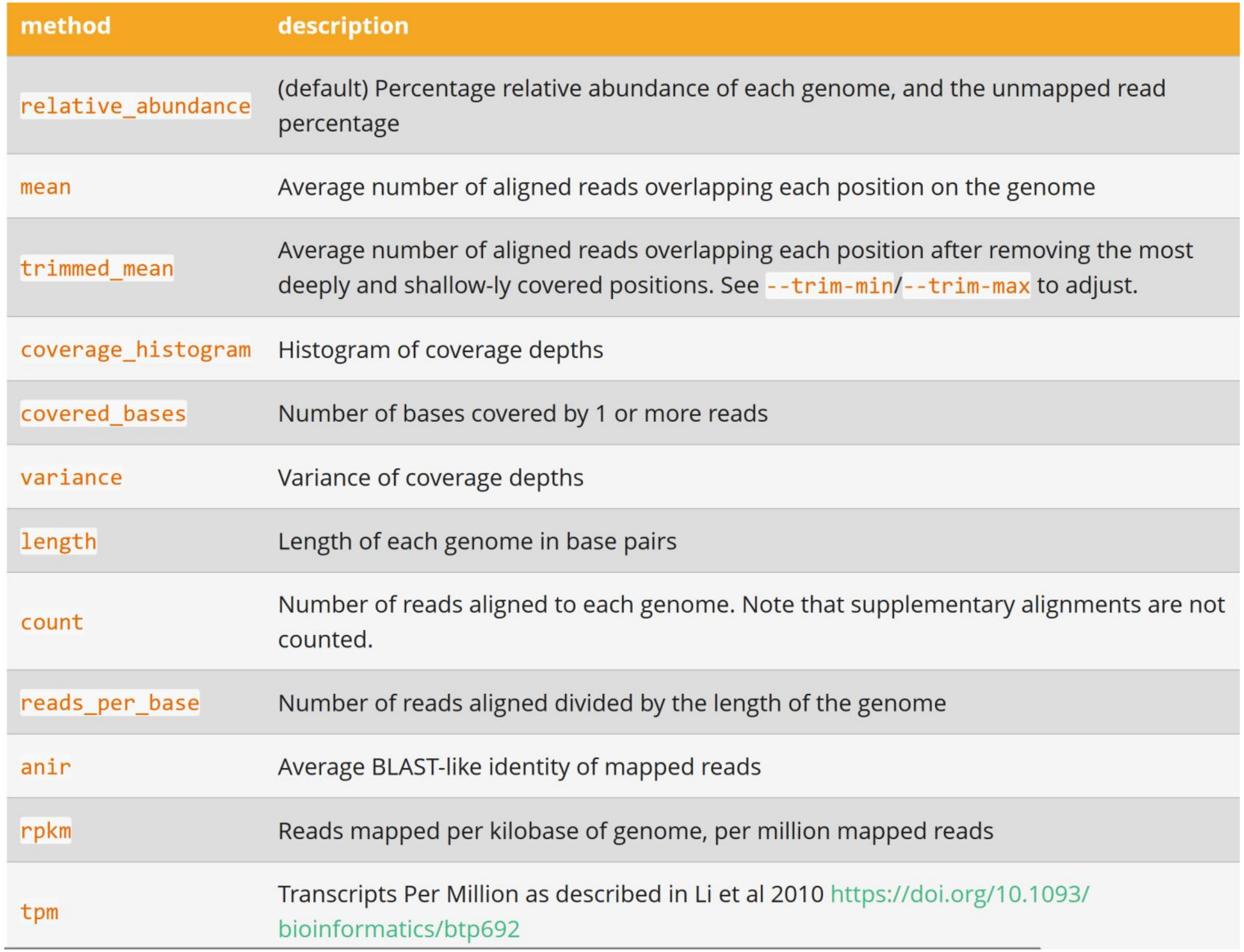
## What to consider?

- Sampling
- Sample processing
- RNA extraction
- Replicates
- Library preparation (rRNA depletion)

## How to analyze?

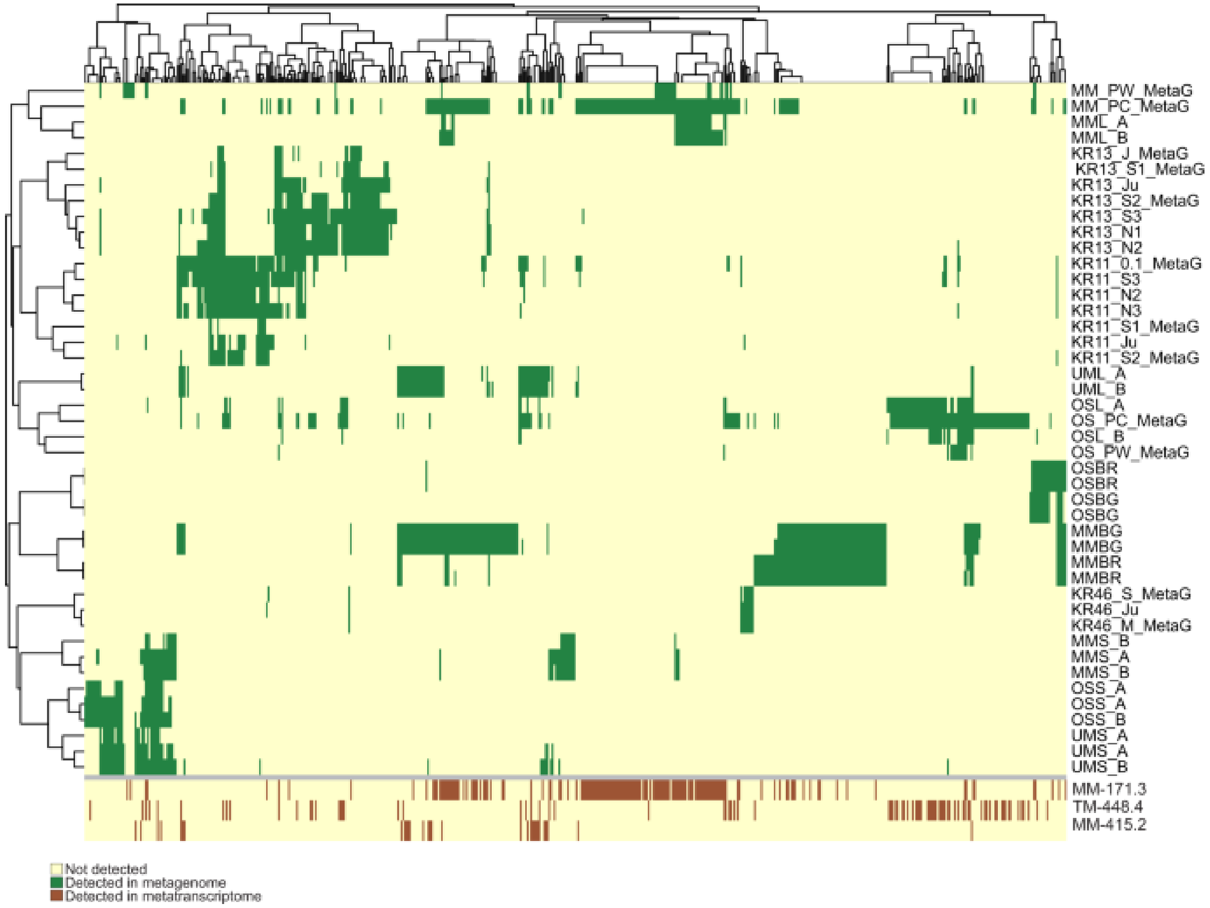
- Remove rRNA or not?
- Assemble or not?
- Gene-resolved metatranscriptomics
- How to normalize?

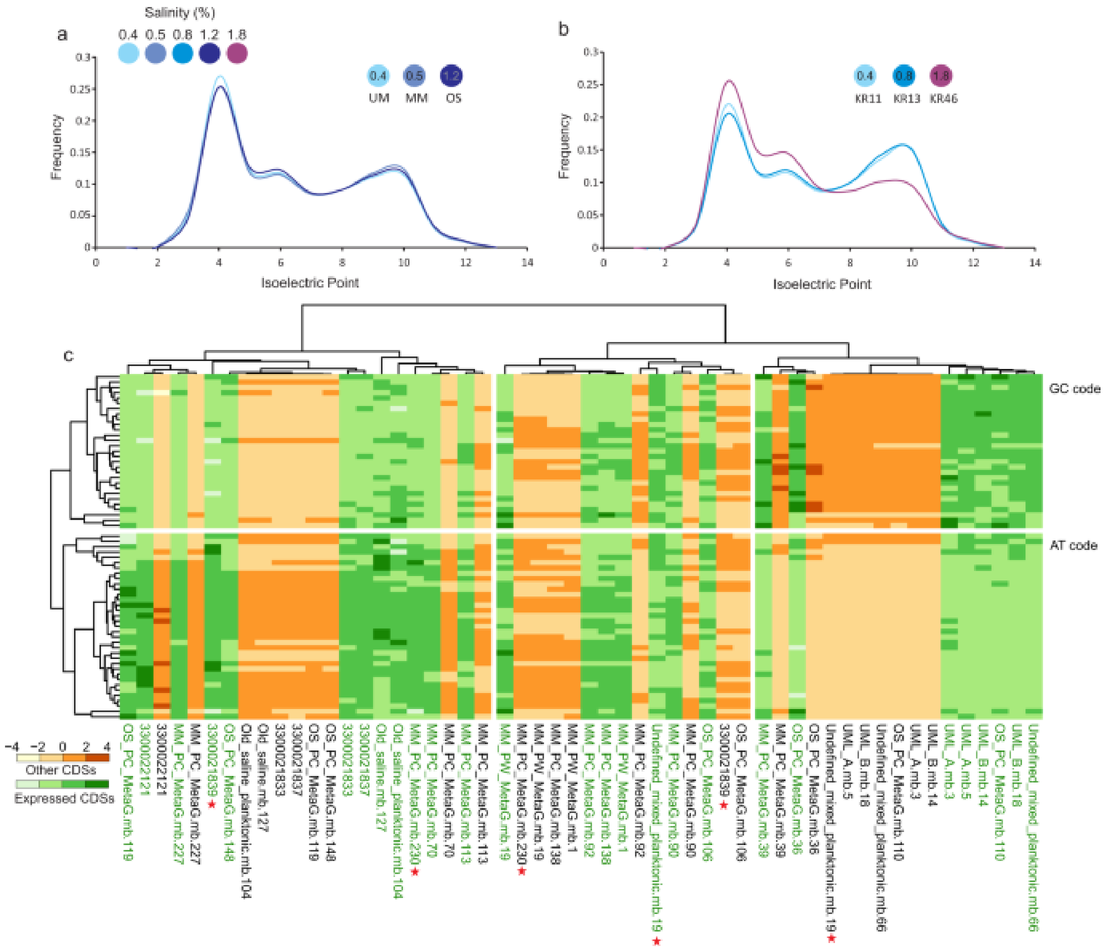
# Some normalization methods



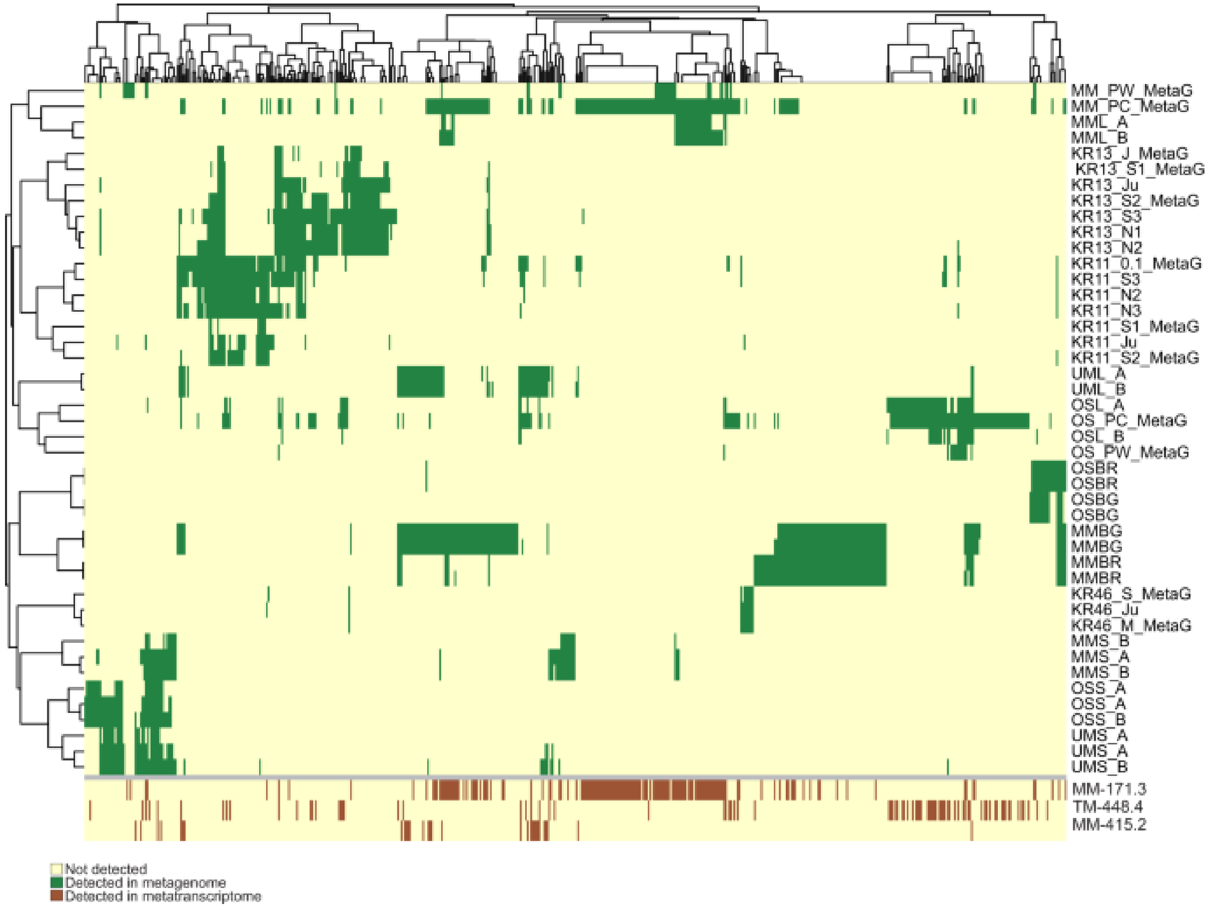
| method | description |
|---|---|
| `relative_abundance` | (default) Percentage relative abundance of each genome, and the unmapped read percentage |
| `mean` | Average number of aligned reads overlapping each position on the genome |
| `trimmed_mean` | Average number of aligned reads overlapping each position after removing the most deeply and shallow-ly covered positions. See `--trim-min`/`--trim-max` to adjust. |
| `coverage_histogram` | Histogram of coverage depths |
| `covered_bases` | Number of bases covered by 1 or more reads |
| `variance` | Variance of coverage depths |
| `length` | Length of each genome in base pairs |
| `count` | Number of reads aligned to each genome. Note that supplementary alignments are not counted. |
| `reads_per_base` | Number of reads aligned divided by the length of the genome |
| `anir` | Average BLAST-like identity of mapped reads |
| `rpkm` | Reads mapped per kilobase of genome, per million mapped reads |
| `tpm` | Transcripts Per Million as described in Li et al 2010 https://doi.org/10.1093/bioinformatics/btp692 |

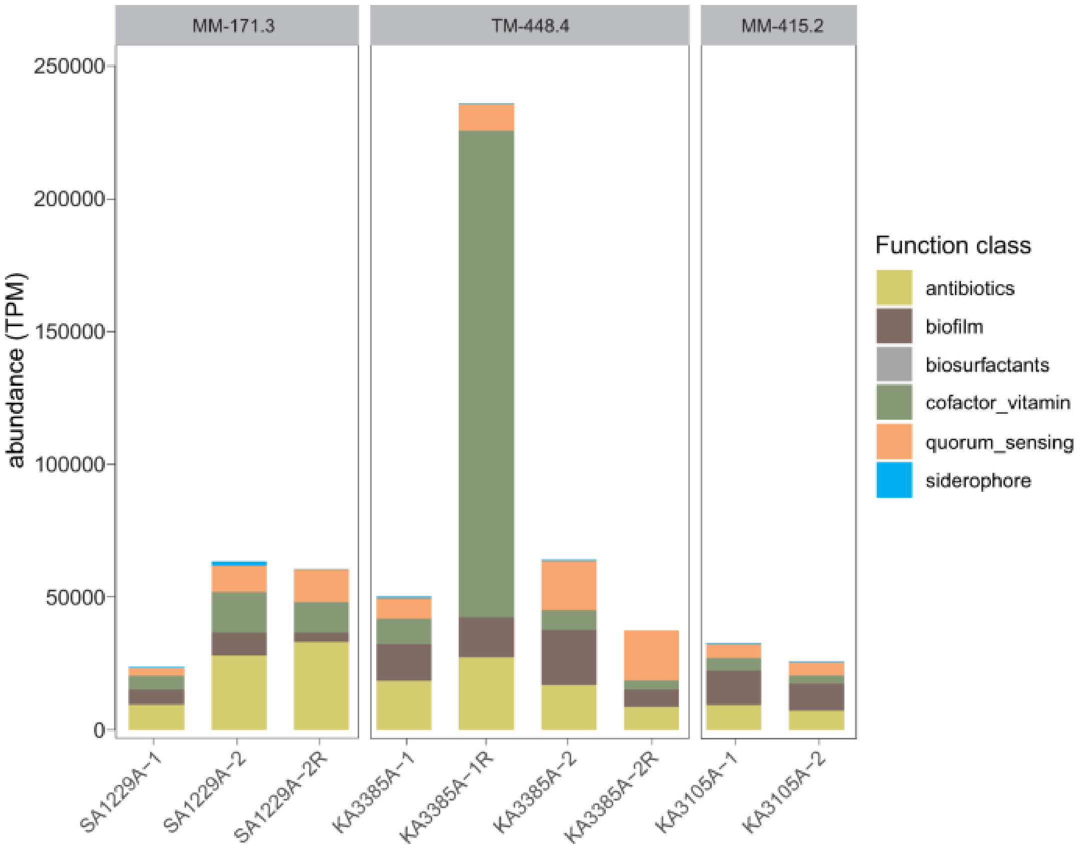
# The transfer of information





# The transfer of information



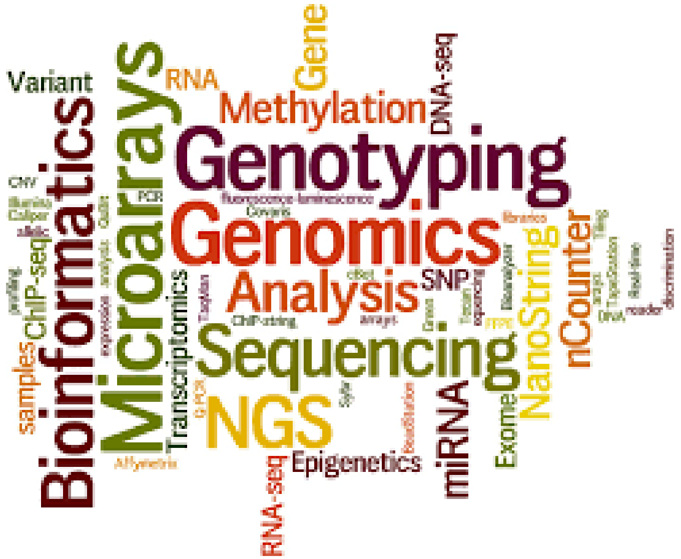


# “multi-omics” approach to answer eco-evolutionary questions

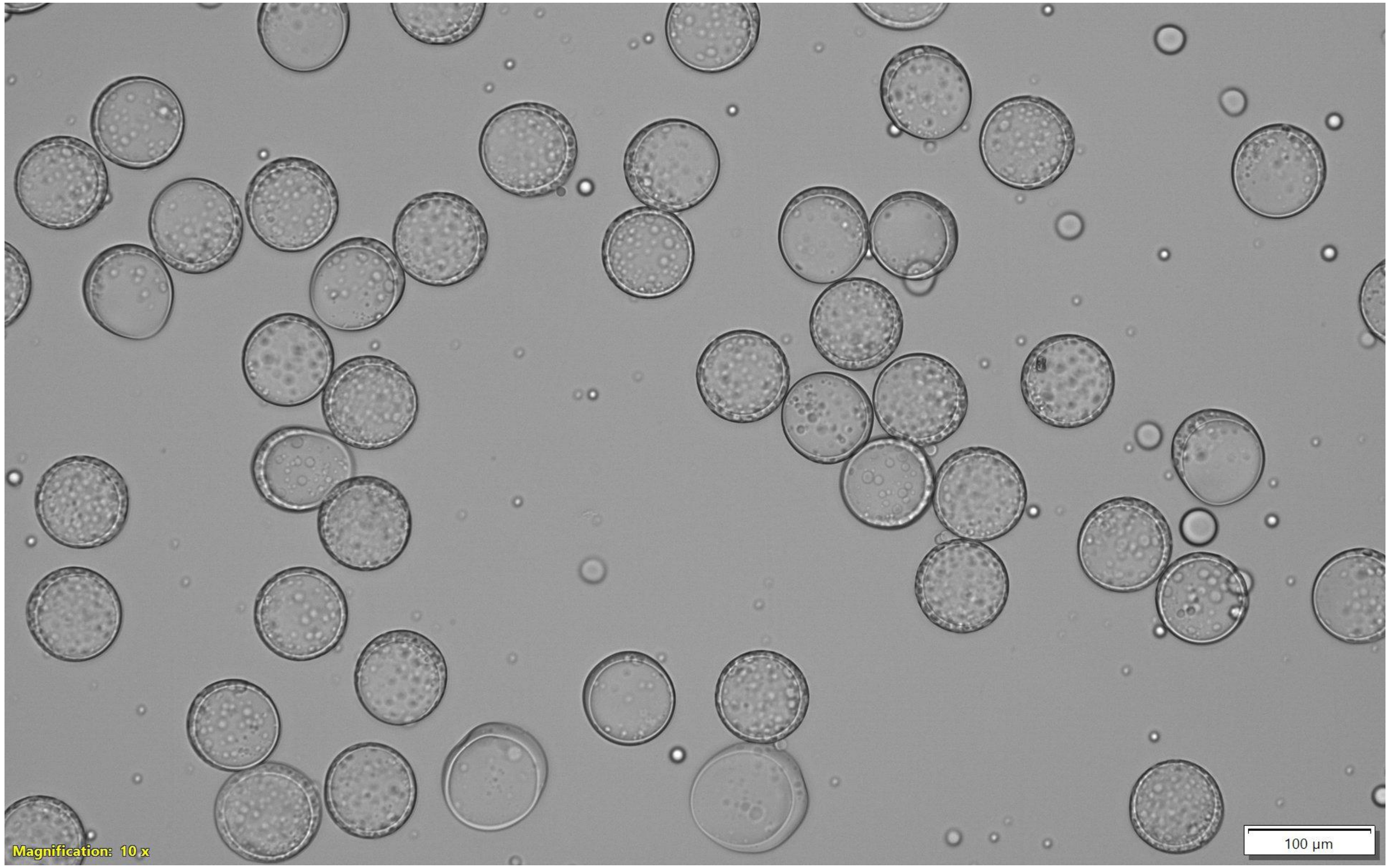
Single-cell amplified genomes

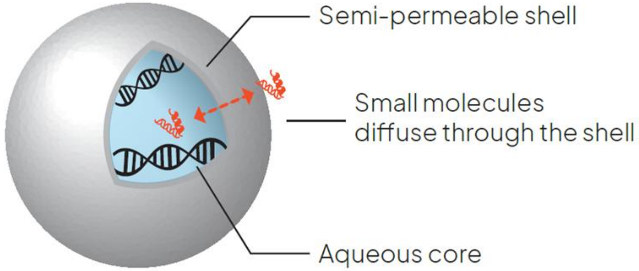
Metagenomics

Metatranscriptome



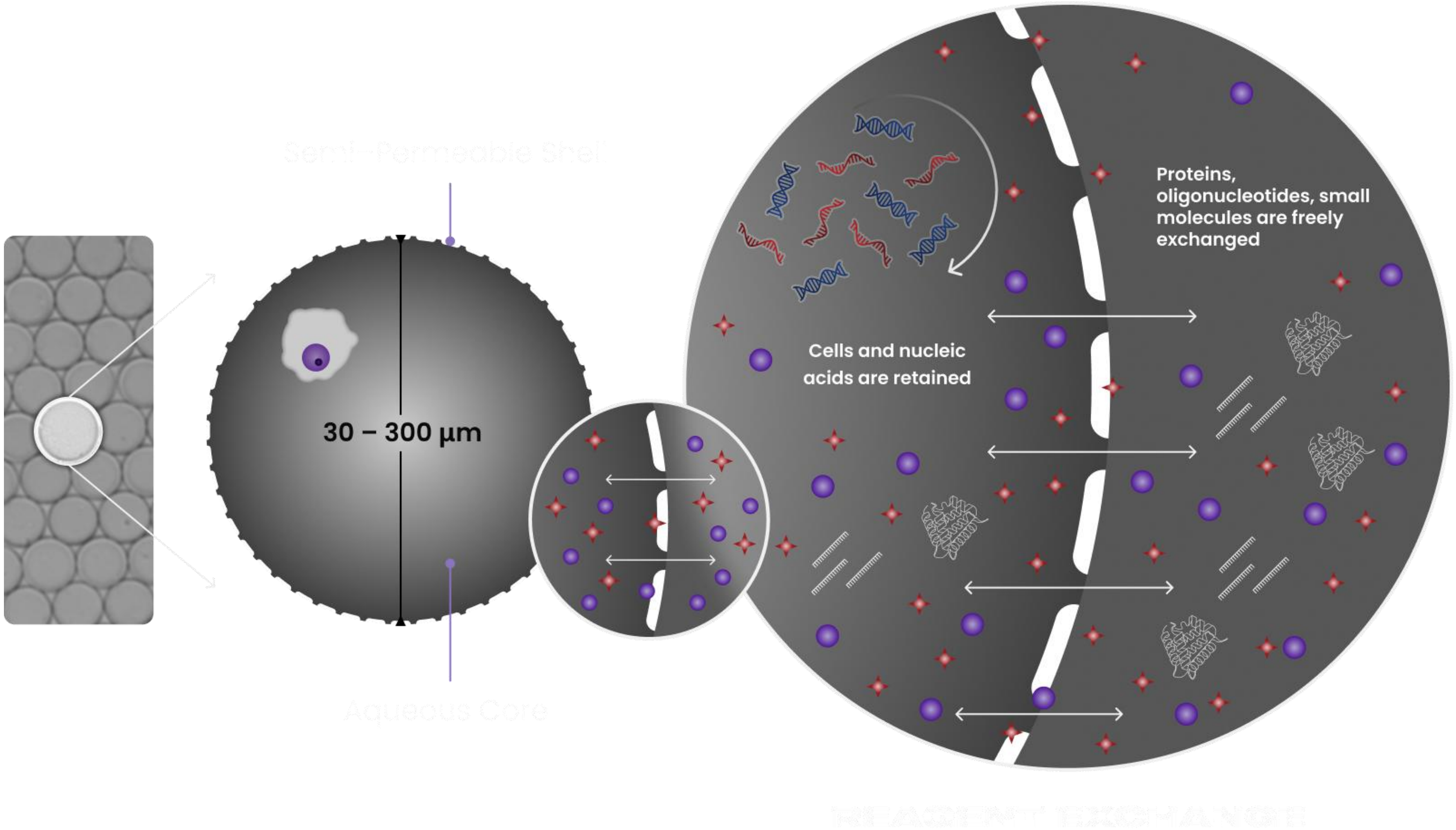
# Compartmentalized metagenomics





Atrandi BIOSCIENCES

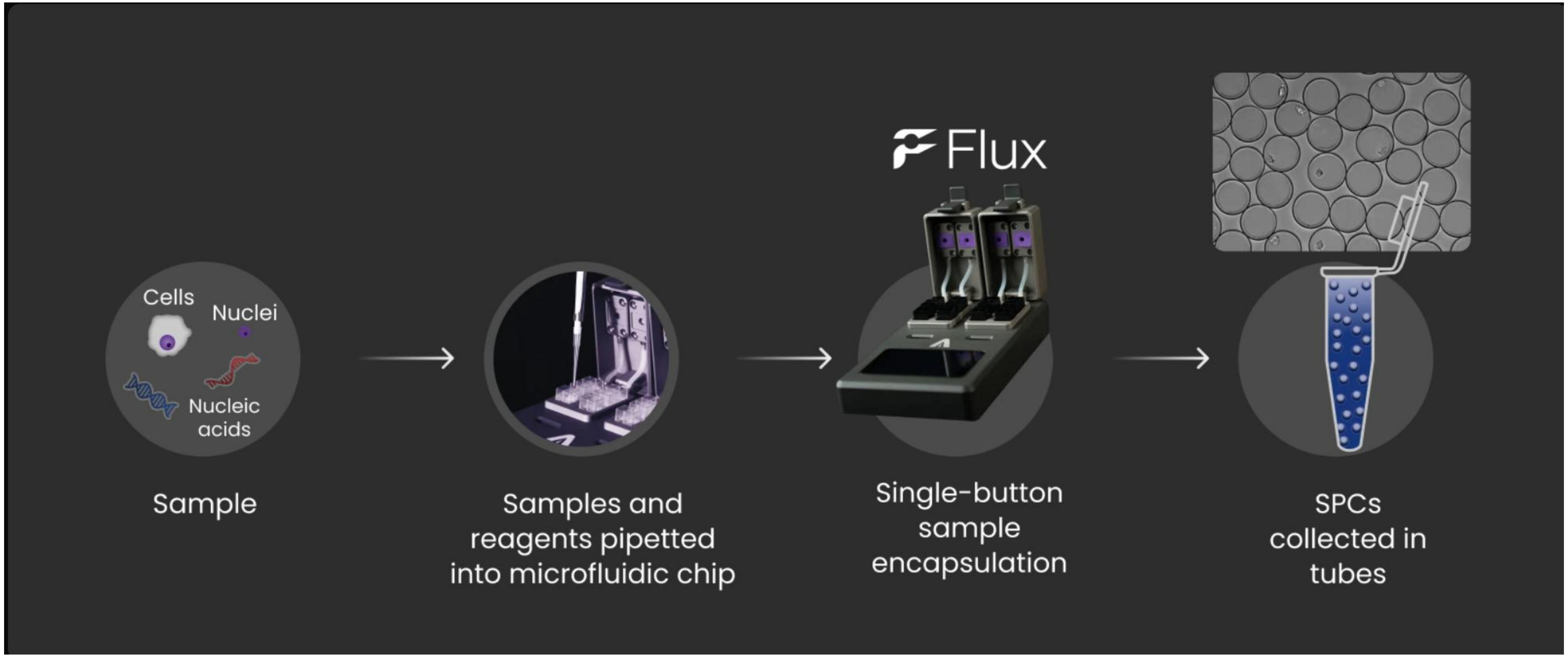
# Semi permeable capsules



Semi-Permeable Shell
30 – 300 µm
Aqueous Core
Proteins, oligonucleotides, small molecules are freely exchanged
Cells and nucleic acids are retained
REAGENT EXCHANGE

Atrandi
BIOSCIENCES

# Semi permeable capsules



Sample → Samples and reagents pipetted into microfluidic chip → Single-button sample encapsulation → SPCs collected in tubes

Atrandi BIOSCIENCES

# Removed unpublished results

# Metagenomics is it a big deal!?

BIONOMICS-MMlab
MULTIPARTITE PARASITIC INTERACTIONS

**Maliheh Mehrshad**

**Department of Aquatic Sciences and Assessment,** SLU