

MICROBIOLOGY OF EXTREME ENVIRONMENTS

STUDYING MICROBIAL DIVERSITY IN EXTREME ENVIRONMENTS

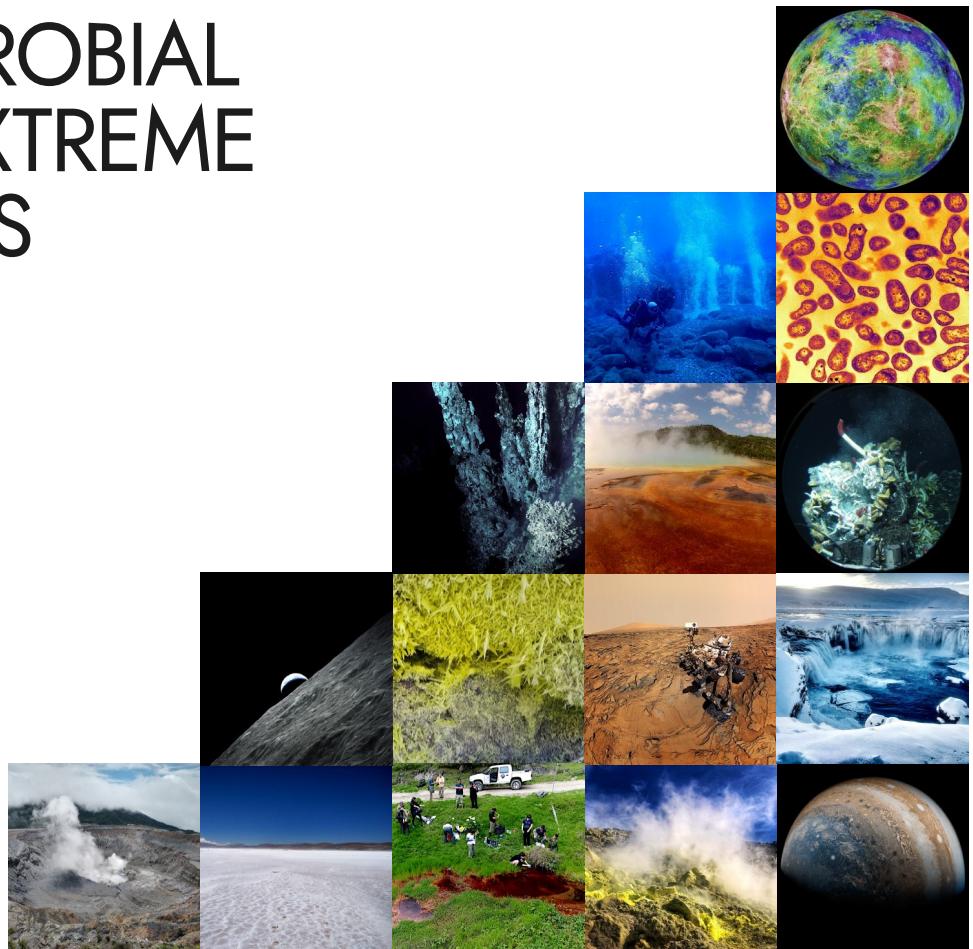
Donato Giovannelli

donato.giovannelli@unina.it

www.donatogiovannelli.com

[@d_giovannelli](https://twitter.com/d_giovannelli)

[@donatogiovannelli](https://www.instagram.com/donatogiovannelli)



A vertical strip on the left edge of the slide shows a close-up of a ship's metal railing against a dark blue, slightly choppy sea.

Culture-independent approaches



Culture-independent approaches

There are a number of culture independent approaches to study microbial diversity, and can be distinguished based either the approach (microscopy, molecular, biochemical) or based on the type of information they provide

Earlier culture-independent approaches were based on different microscopy techniques, from optic and phase contrast to transmission and scanning electron microscopy

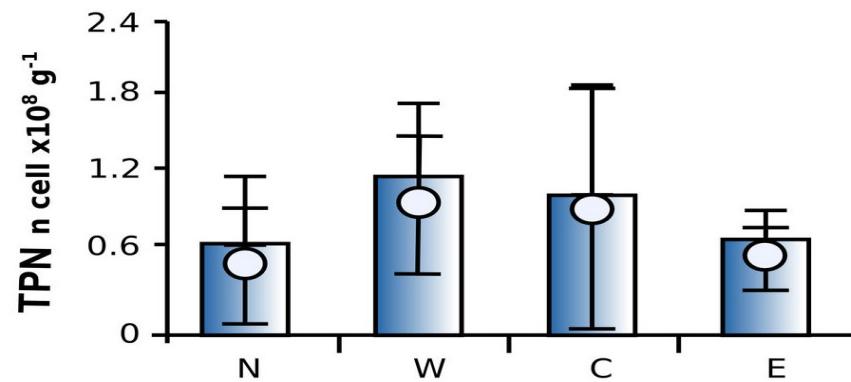
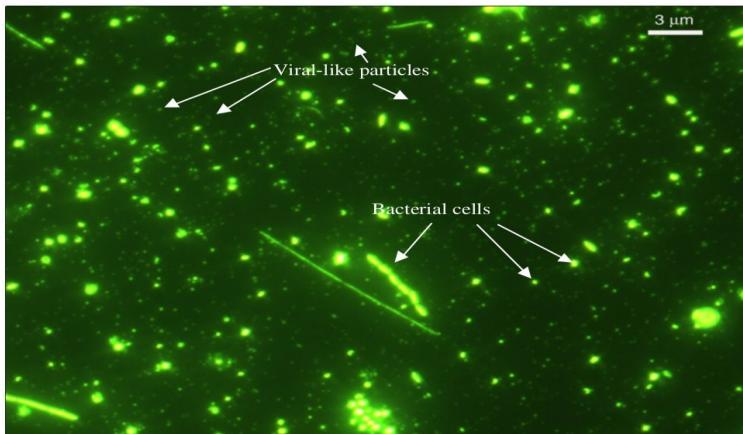
The big leap in culture-independent techniques started in the seventies with the evolution of molecular biology techniques. These allowed to probe the microbial world in new ways (mainly DNA at the time), bringing upon us a big revolution

The establishment of a third domain of life (the Archaea) is a direct result of the application of molecular biology techniques to the study of microbial diversity

Epifluorescence Microscopy

Counting microbes in natural samples can be achieved by using DNA staining chemicals that fluoresce under UV light. The staining is aspecific to any DNA (double or single strands depending on the dye)

These approaches have lead to the “Great Plate Count Anomaly” paradox, and are used to obtain quantitative data on population abundance. Data on microbial abundance can be also obtained with flow cytometry techniques, depending on the sample matrix

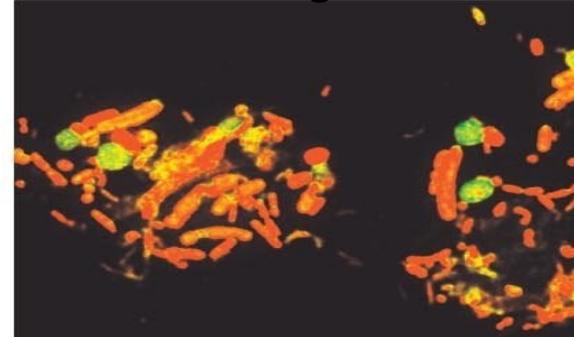


Epifluorescence Microscopy

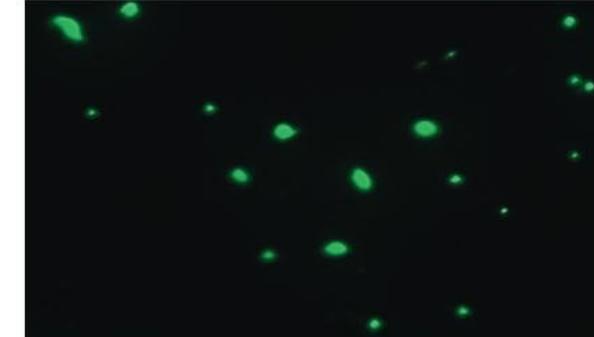
DAPI stain



Acridine orange stain



SYBR Green stain



(a)

© 2015 Pearson Education, Inc.

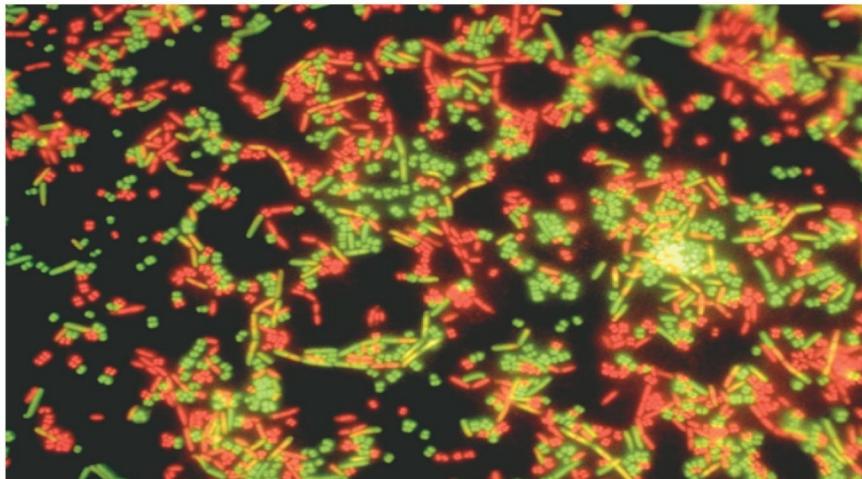
Marc Mussman and Michael Wagner

(b)

Marc Mussman and Michael Wagner

(c)

Wilm Martins-Habanna



Molecular Probes, Inc., Eugene, OR

Viability staining: Live (green) and red (dead) cells are revealed. The technique is based on the integrity of the cytoplasmic membrane.

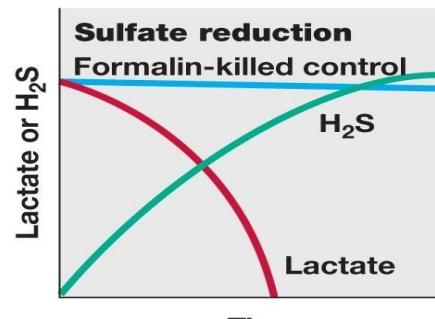


Activity measurements

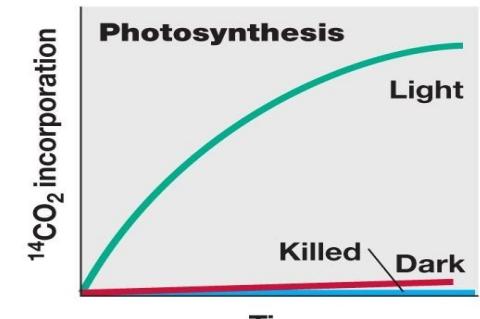
Specific metabolic activities can be measured by incubating natural samples with selected substrates. The resulting rates are often not representative of in situ rates, since the community is stimulated by substrate addition

Chemical substrates

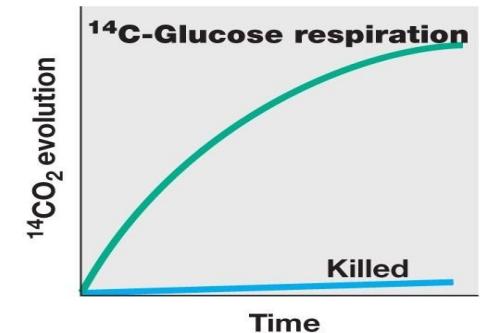
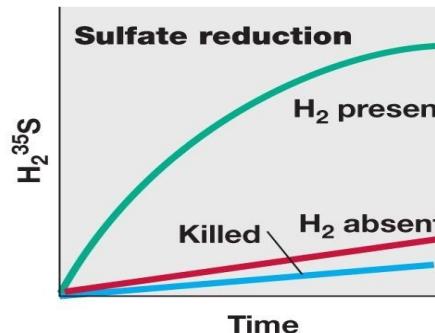
Stable or radioactive isotopes



(a)



(b)

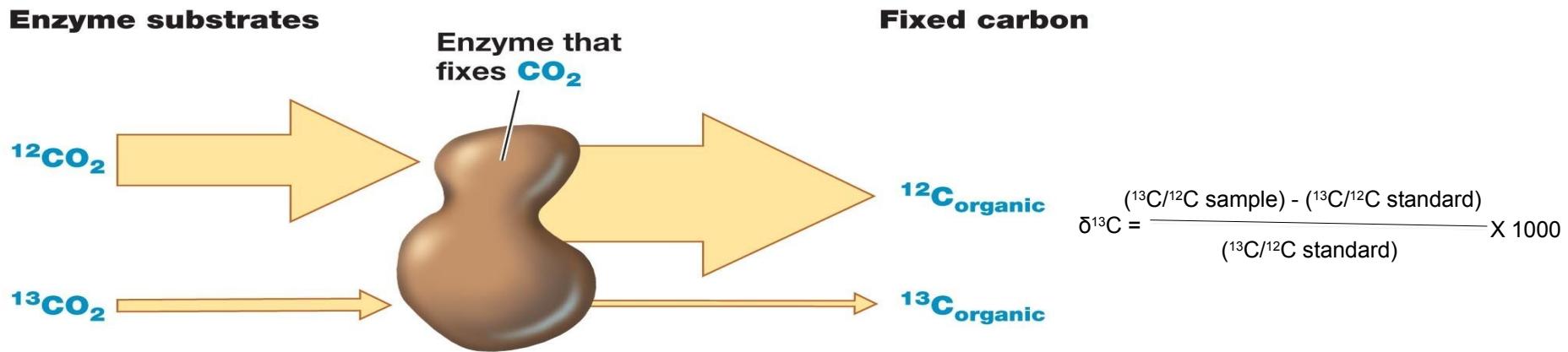




Stable isotopes

Enzymes can be more or less selective light isotopes during their activity, often leaving behind a specific isotopic signature as a result

Stable isotopes can thus be used not only for tracking rates, but also for linking identity (who) to function (what). They are often combined with other techniques such as FISH and sequencing

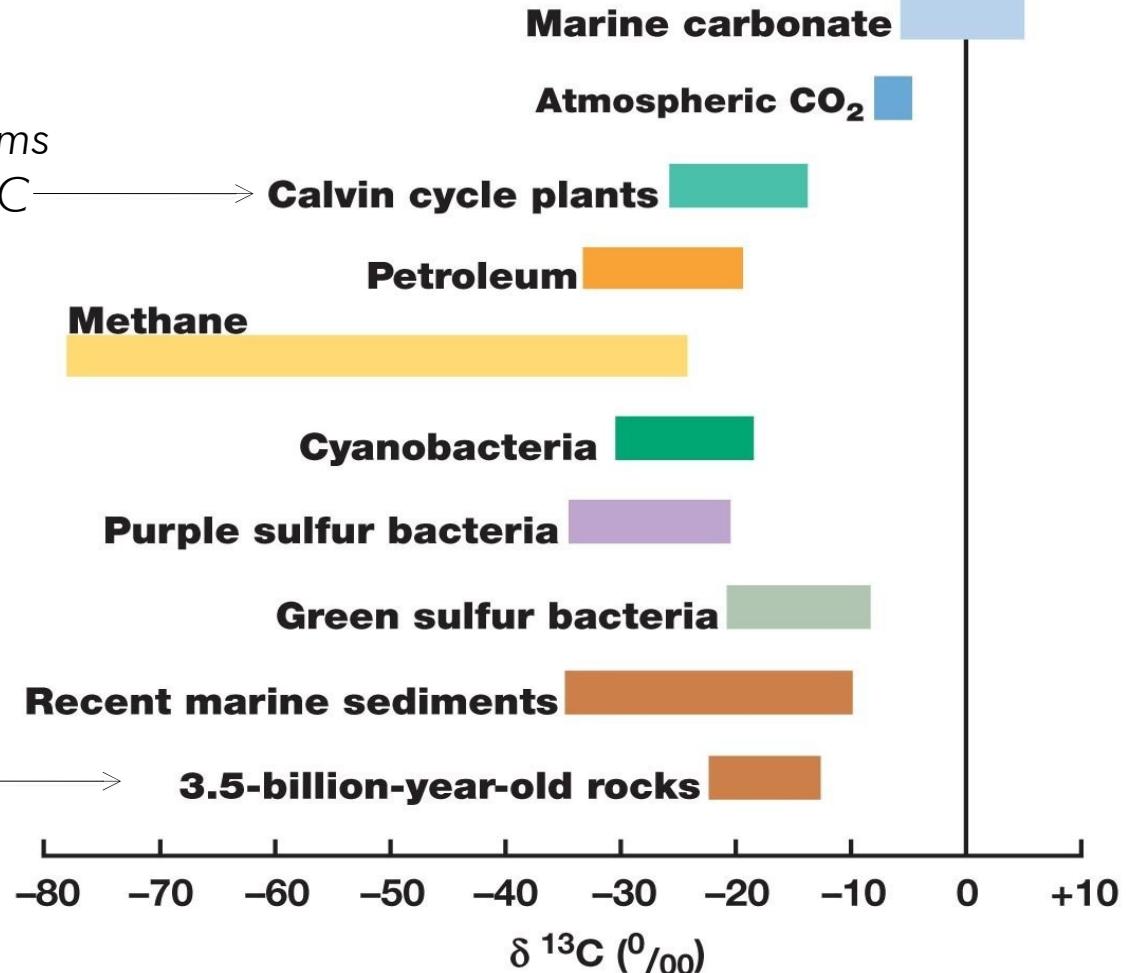


© 2015 Pearson Education, Inc.

Biological C isotopic fractionation in enzymes that fix CO_2 is calculated as ^{13}C depletion relative to a standard of geologic origin

example Carbon signature

Carbon fixed by autotrophic organisms using the CBB cycle is depleted in ^{13}C

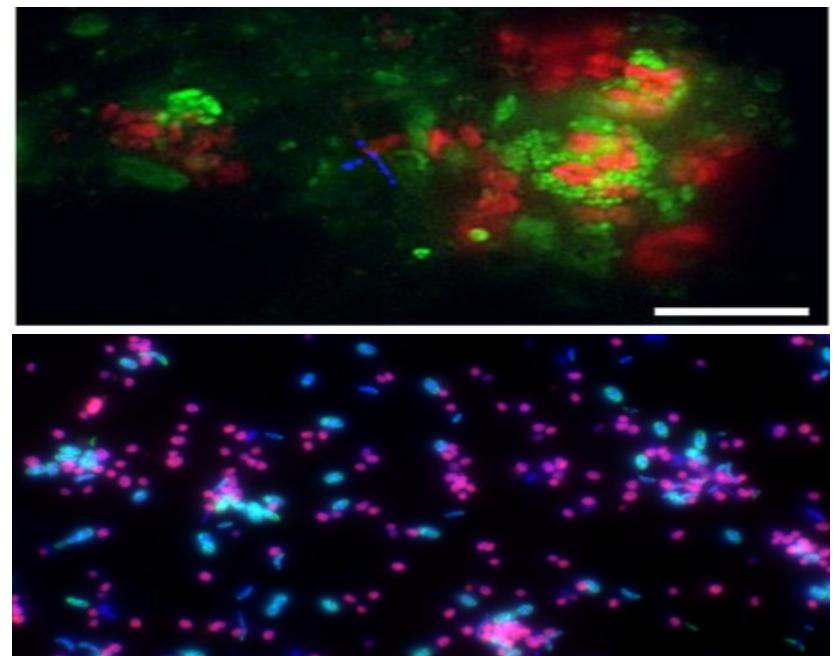
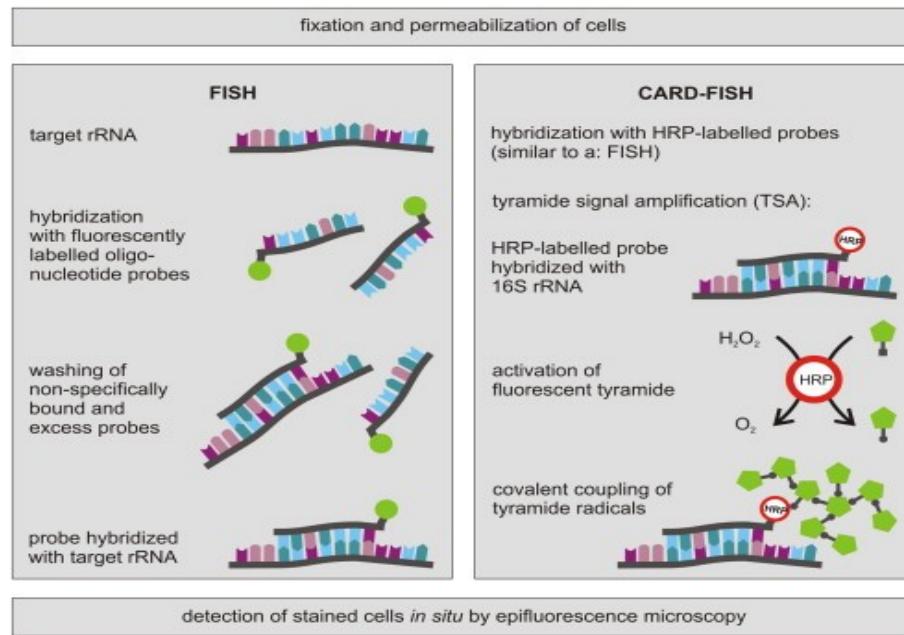


Organic Carbon in ancient rocks also shows evidence of isotopic fractionation



Fluorescent in situ hybridization (FISH)

FISH (like CARD-FISH and related techniques) rely on the use of fluorescent probes that bind selectively to specific DNA or RNA sequences, allowing to selectively color and count specific groups of microbes. The specificity of the probe is selected during the design phase. Target sequences need to be known in order to design the probes.





Fluorescent in situ hybridization (FISH)

FISH (like CARD-FISH and related techniques) rely on the use of fluorescent probes that bind selectively to specific DNA or RNA sequences, allowing to selectively color and count specific groups of microbes. The specificity of the probe is selected during the design phase. Target sequences need to be known in order to design the probes.

FISH and related techniques are very powerful, and can be used to obtain qualitative and quantitative data that include the spatial distribution and interactions, and are widely used also to study microbial symbioses

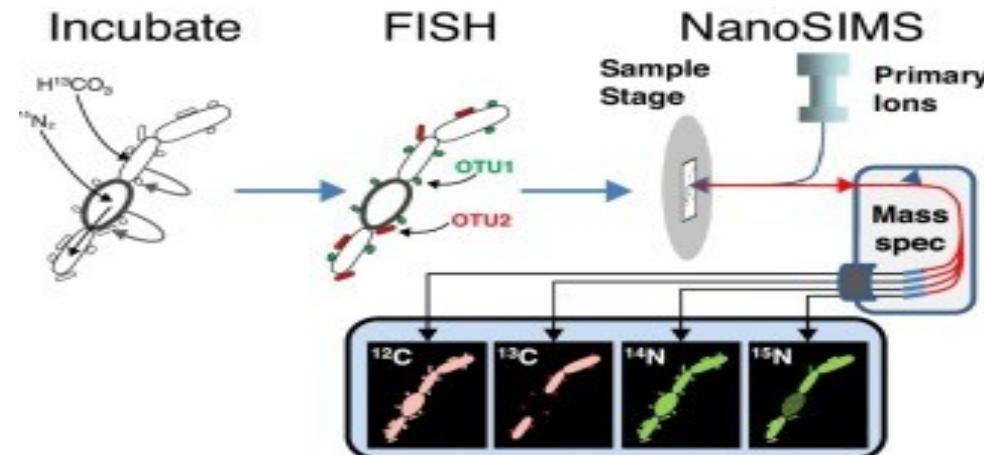
They can be combined with a number of other techniques, such as confocal laser microscopy, SIMS and stable and radioactive isotopes



nano-SIMS

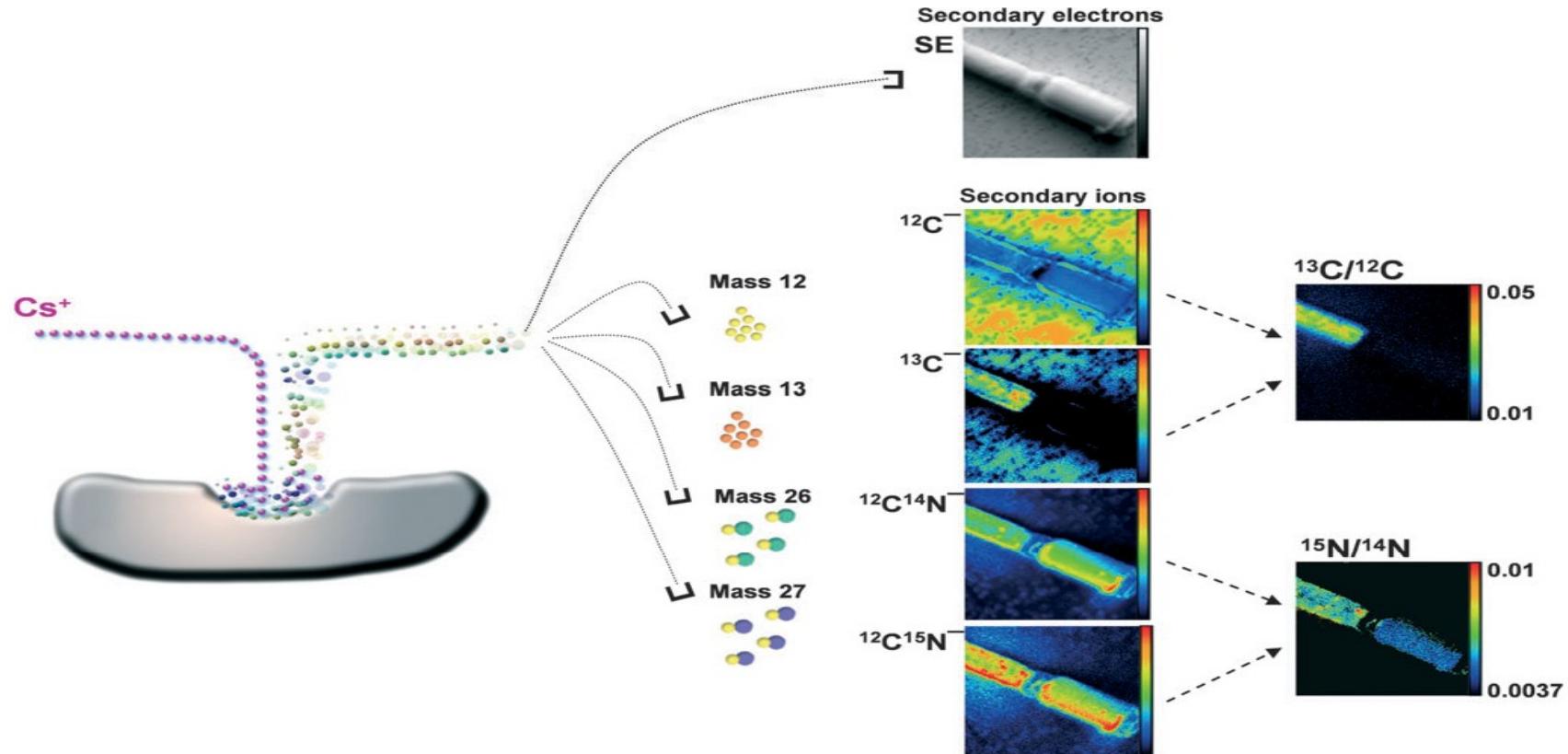
Nano-SIMS (nanoscale Secondary Ion Emission Mass Spectrometry) is a relatively recent techniques that combines stable isotope incubations with FISH and spatially resolved mass spectrometry

A natural sample is incubated using selected substrates enriched in a stable isotope (for example ^{13}C CO₂). The cells are then extracted and imaged using a combination of epifluorescence microscopy and probed for their incorporation of the label





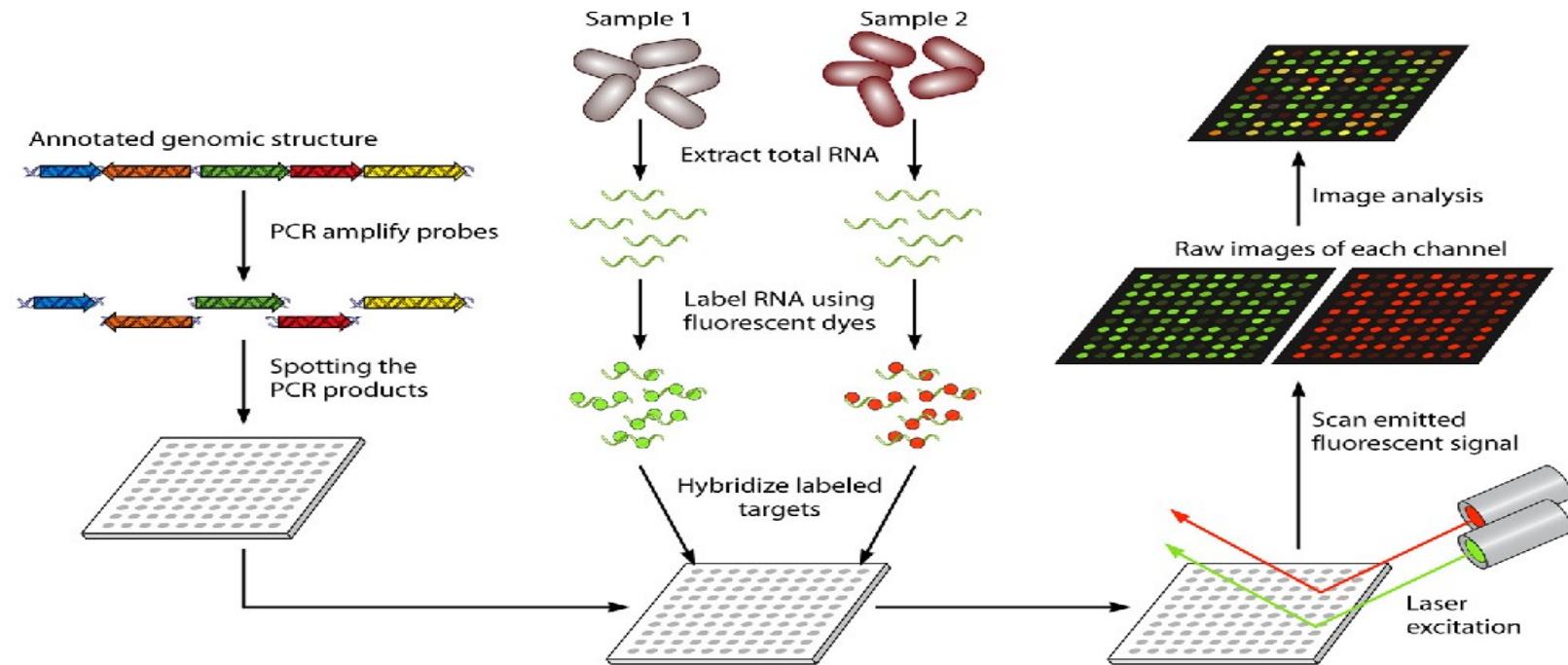
nano-SIMS





Microarray

Microarrays (such as the Geochip) are also based on our ability to design specific probes to visualize the presence of specific genes (DNA) or transcripts (RNA). They are limited by our ability to design probes since the sequence must be known





Culture-independent approaches: a detour on sequencing



Sequencing based approaches

DNA sequencing has revolutionized the study of microbial diversity in many ways. Current technologies and lowering prices allow us to probe the microbial world at an unprecedented rate

Sequencing can be divided based on the throughput in **single sequence** or **massive parallel sequencing** and base of the type of sequencing in **sequencing by synthesis** or more recently **sequencing by reading**

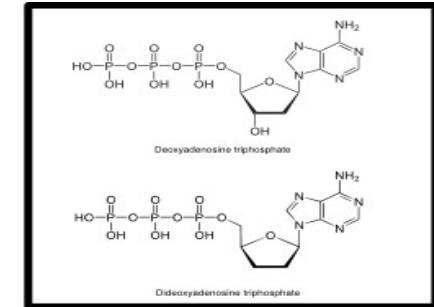
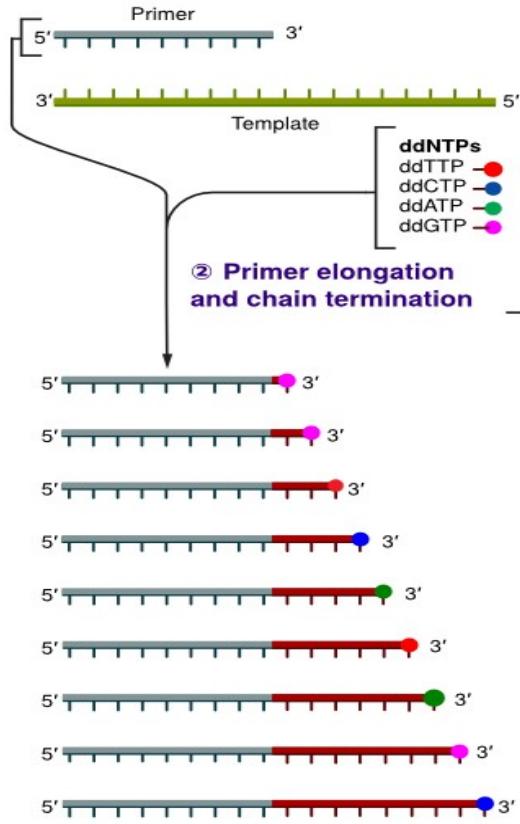
The choice of the type of sequencing to use is highly dependent on the study questions, and all the sequencing approaches have biases (like all the techniques used)

Current sequencing techniques result in millions of DNA reads to be generated, making **computer skills** a **MUST** for any biologist

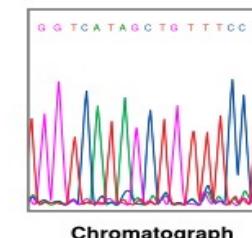
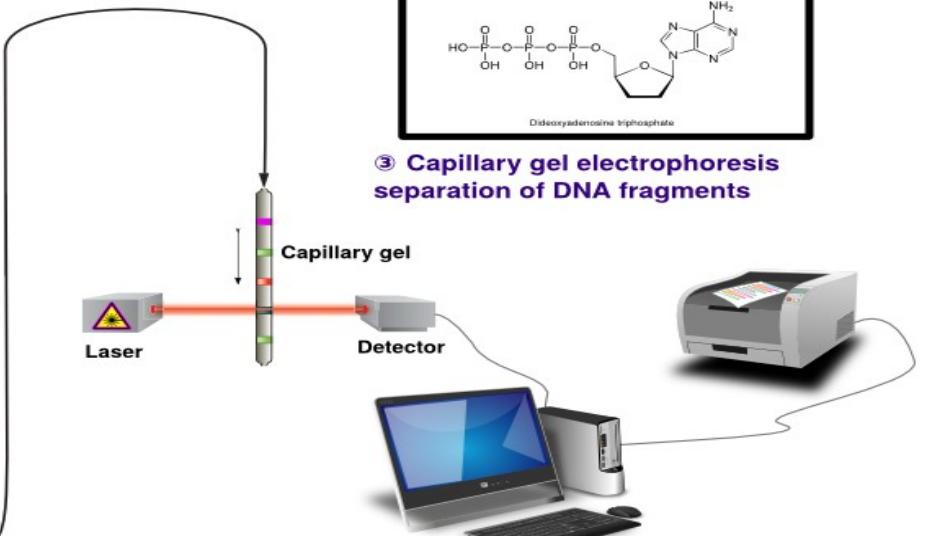
ABI Solid Dig Dye Terminator (Sanger)

① Reaction mixture

- Primer and DNA template ► DNA polymerase
- ddNTPs with flourochromes ► dNTPs (dATP, dCTP, dGTP, and dTTP)



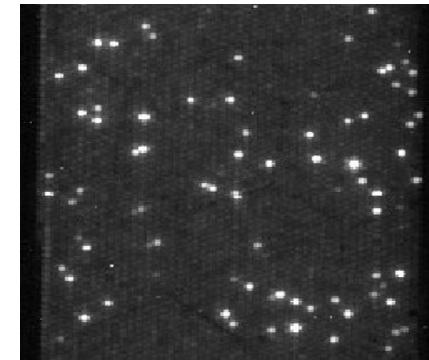
③ Capillary gel electrophoresis separation of DNA fragments



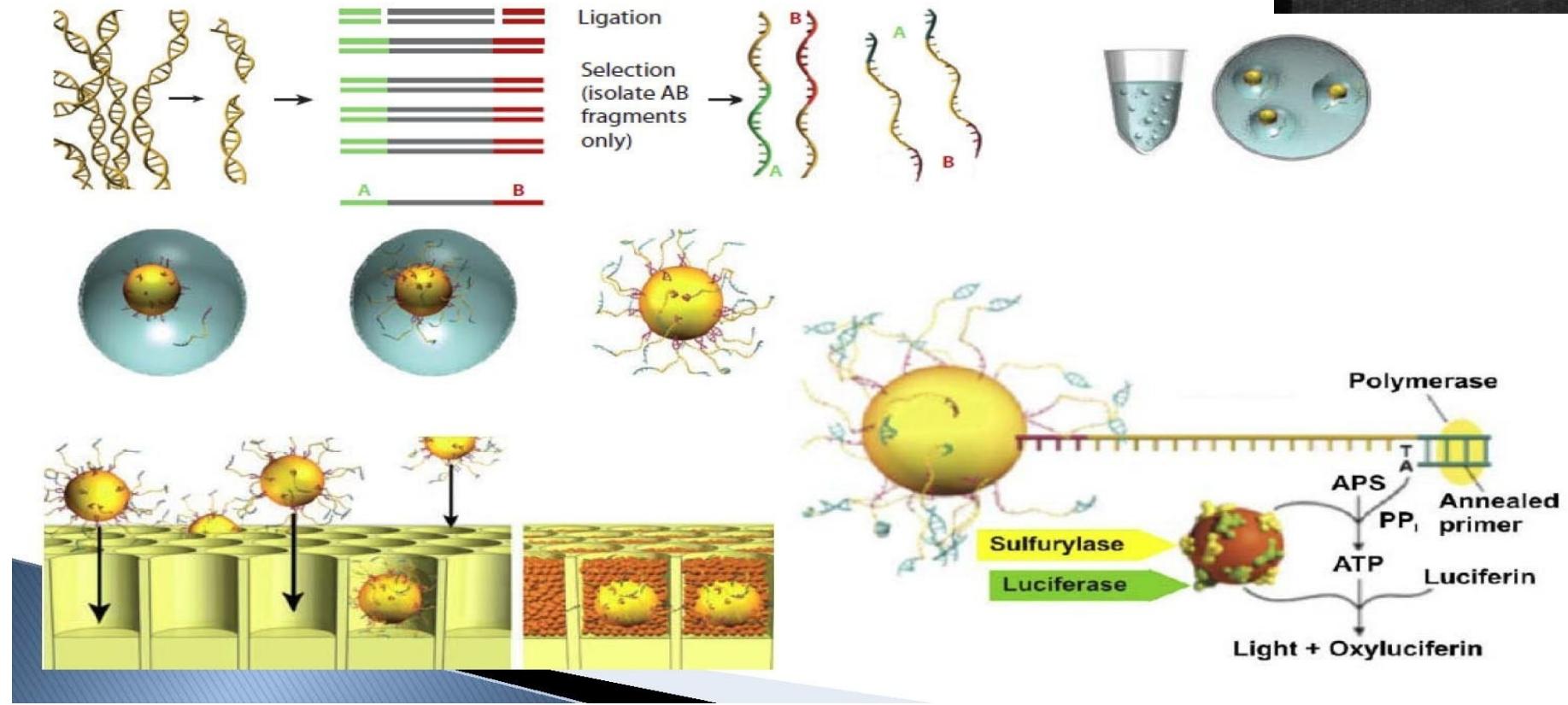
④ Laser detection of flourochromes and computational sequence analysis

"Sanger-sequencing" by Estevezj - Own work.

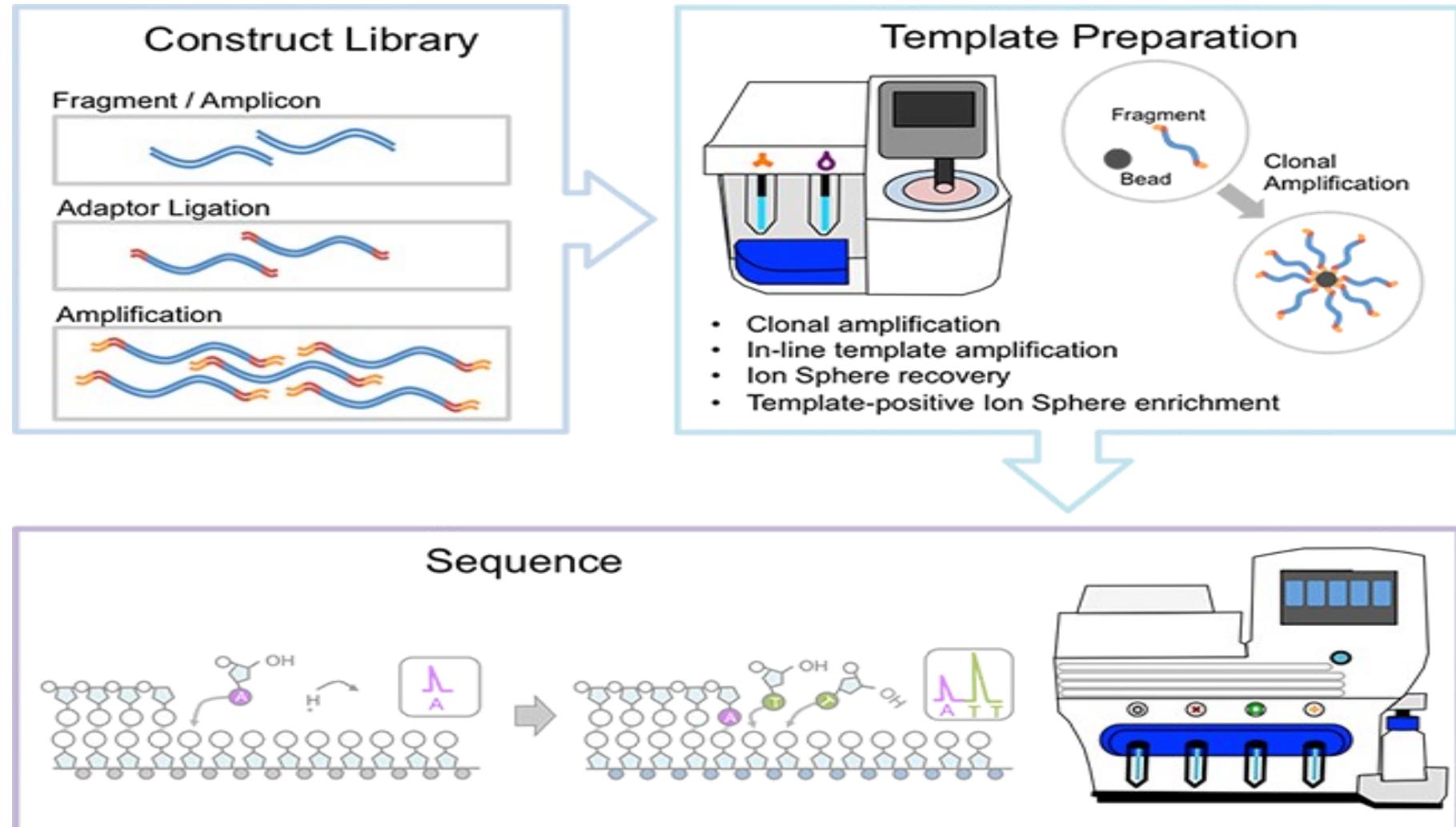
Licensed under Creative Commons Attribution-Share Alike 3.0 via Wikimedia Commons



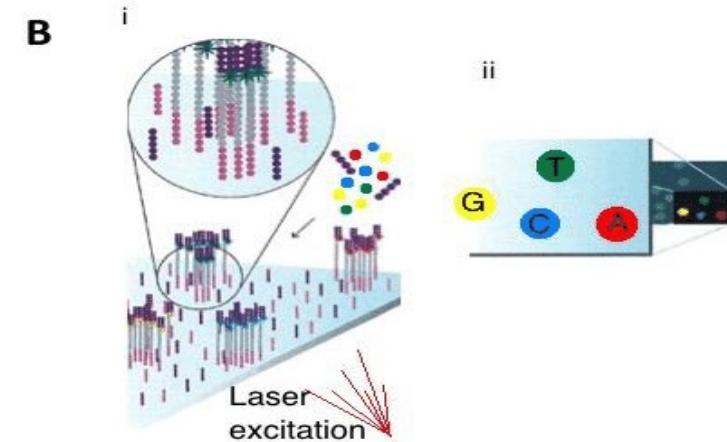
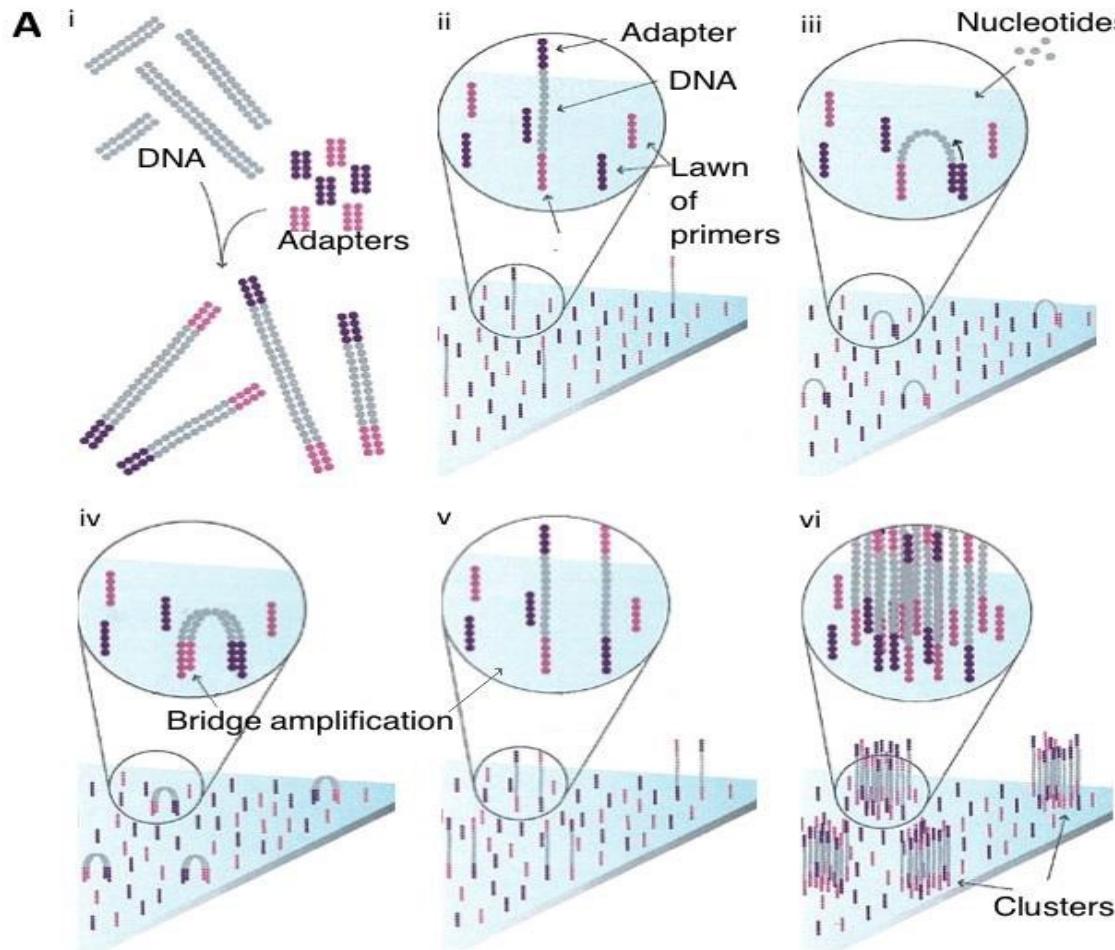
Pyrosequencing



IonTorrent Sequencing



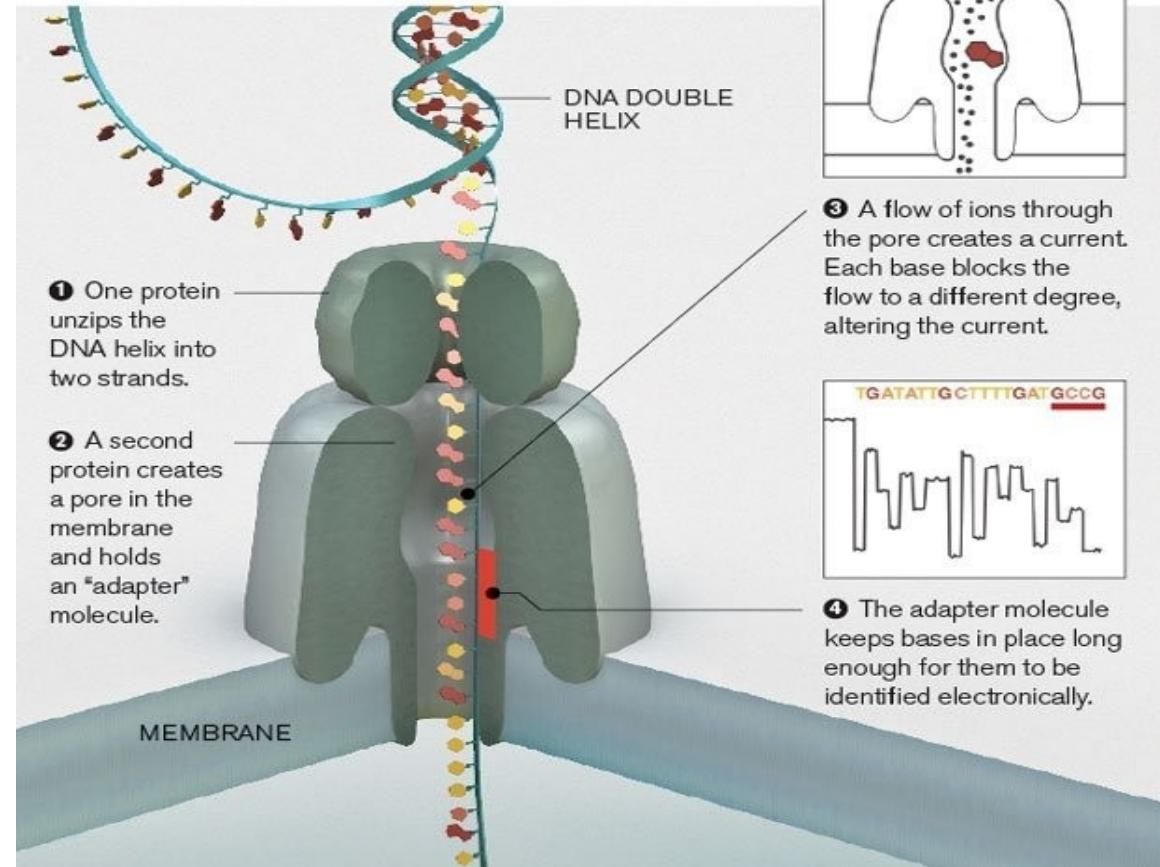
Illumina Sequencing



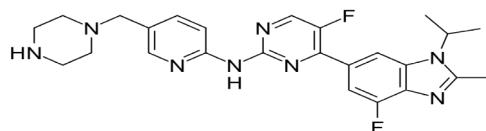
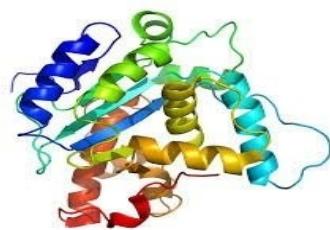
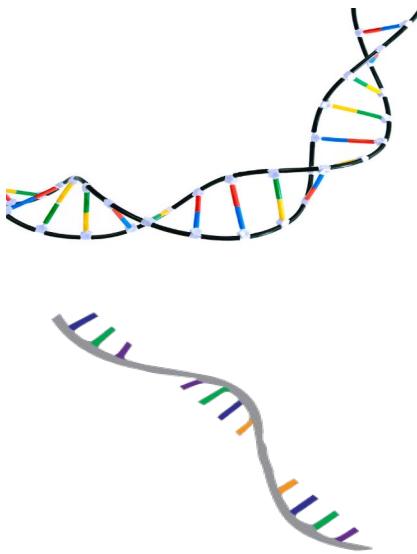
Nanopore Sequencing



DNA can be sequenced by threading it through a microscopic pore in a membrane. Bases are identified by the way they affect ions flowing through the pore from one side of the membrane to the other.



Information flow in biology



replication

GENES



transcription

TRANSCRIPTS



translation

PROTEIN



metabolism

METABOLITES



A primer on language

Moving from **gene** to **genome** to **metagenome** is a matter of scales. We go from the single gene to the whole genome to the genomes of the entire community

The suffix (**omic** and **meta**) are used also for the other words in the central dogma, so we have **transcriptomic**, **metatranscriptomic**, **metametabolomic**, etc..

The same suffix have been applied to a number of other words, like **metallomics**, **mobilome**, **secretome**, etc...

Not all the techniques referred to with omic suffix are about sequencing. Proteomic and metaproteomic require the use of mass spectrometry, metabolomic requires the use of mass spectromic and/or nuclear magnetic resonance. These techniques are usually combined with sequencing



The English-language neologism **omics** informally refers to a field of study in biology ending in -omics, such as genomics, proteomics or metabolomics. [...]. Omics aims at the **collective characterization and quantification of pools of biological molecules** that translate into the structure, function, and dynamics of an organism or organisms.

Wikipedia

Genetic – The
study of genes



Genomics – The
study of the
entire genome



Metagenomic –
The study of the
community
genomes
recovered from
environmental
samples



The English-language neologism **omics** informally refers to a field of study in biology ending in -omics, such as genomics, proteomics or metabolomics. [...]. Omics aims at the **collective characterization and quantification of pools of biological molecules** that translate into the structure, function, and dynamics of an organism or organisms.

Wikipedia

Genetic – The study of genes



Genomics – The study of the entire genome



Metagenomic – The study of the community genomes recovered from environmental samples

Transcriptomic

Metatranscriptomic

Proteomic

Metaproteomic

In Greek **meta** means “**trascendent**”. E.g. Metagenomic “trascends” the single organisms and look at the entire community's genomes.

A brief history of Metagenomic

Norman R. Pace

propose the idea of cloning DNA directly from environmental sample to analyze 16S rRNA diversity



1985

Jo Handelsman is the first to use the term METAGENOMIC referring to the analysis of community genomes

Mya Breitbart

used environmental shotgun sequencing to show the diversity of virus in seawater



1998

2002

Craig Venter leads the *Global Ocean Sampling Expedition* (GOS) to collect metagenomic samples throughout the journey

A pilot **GOS** project in the **Sargasso Sea** shows unprecedented bacterial diversity in seawater



2003

Robert Edwards

published sequences generated using pyrosequencing techniques



2005

Stephan Schuster publishes the first sequences generated using high-throughput sequencing.

2007



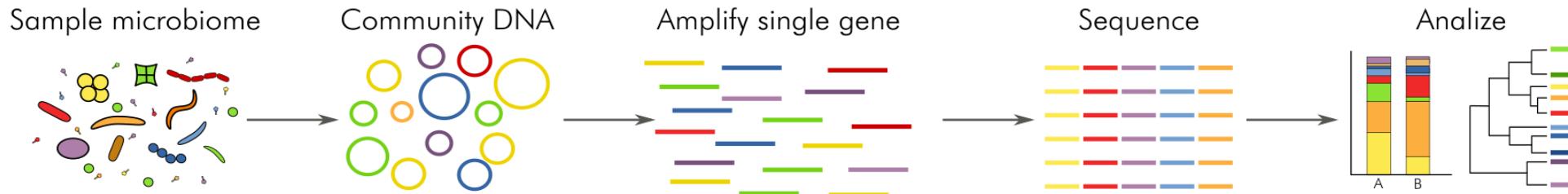


amplicon-sequencing

Many other genes can be investigated using this techniques, both genes with taxonomic relevance (18S rRNA, ITS, mCOI) and functional genes

If the starting material is RNA (instead of DNA) this has the benefit to link the results to the active fraction of the community

The reference database is of fundamental importance for amplicon sequencing approaches, as well as the choice of appropriate primers to be used at the amplification step



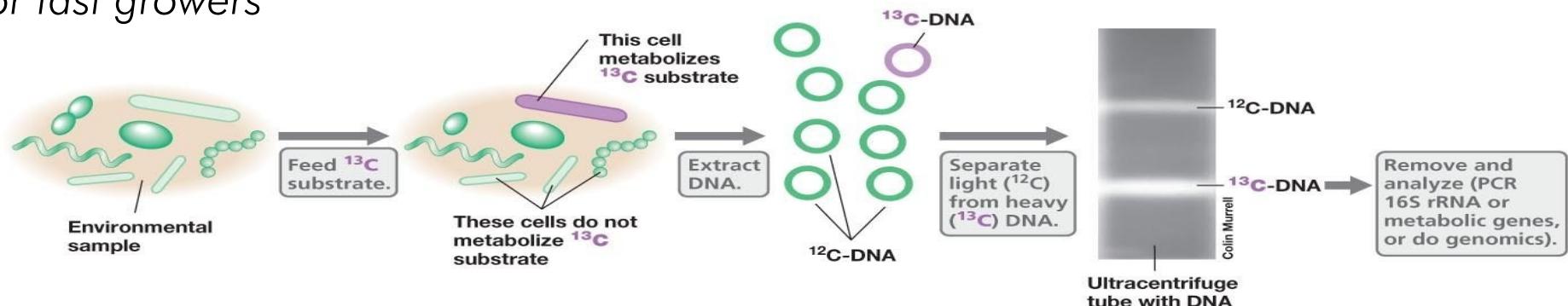


Stable Isotope Probing (SIP)

Combining stable isotopes incubation with sequencing, SIP approaches can be used to identify the fraction of the microbial community involved in specific metabolic pathways

If used in combination with amplicon sequencing, SIP has similar drawback, however SIP can be coupled to other techniques, such as (metagenomic, metatranscriptomic, etc...). It is a very powerful targeted technique if used properly linking identity to function

Main drawback is that typically you are stimulating the community and selecting for fast growers





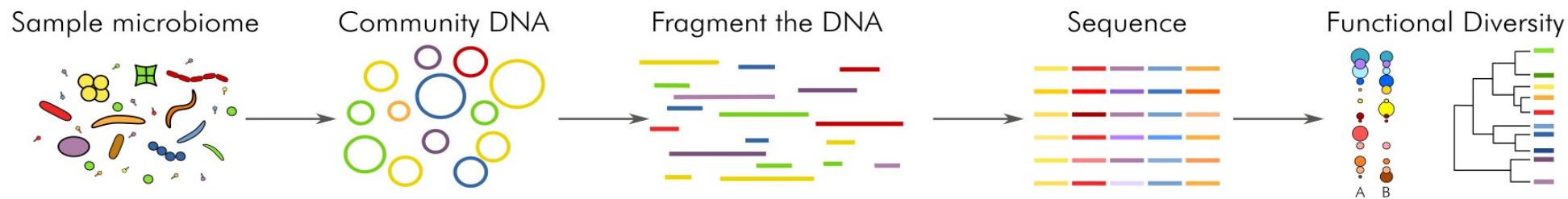
Metagenomic

Metagenomic, also called *shotgun metagenomic*, refers to the sequencing of the community DNA **without the need for amplification**. The result is a pool of sequences containing a sample of all the gene present in the community

This has practical implications, as it removes primer and amplification bias, potentially revealing the true diversity of the sample

Metagenomic allows to get at the genetic potential of the community, and see what functions are encoded into the community DNA

There are a number of downstream analytical approaches in metagenomic, and new get created every year

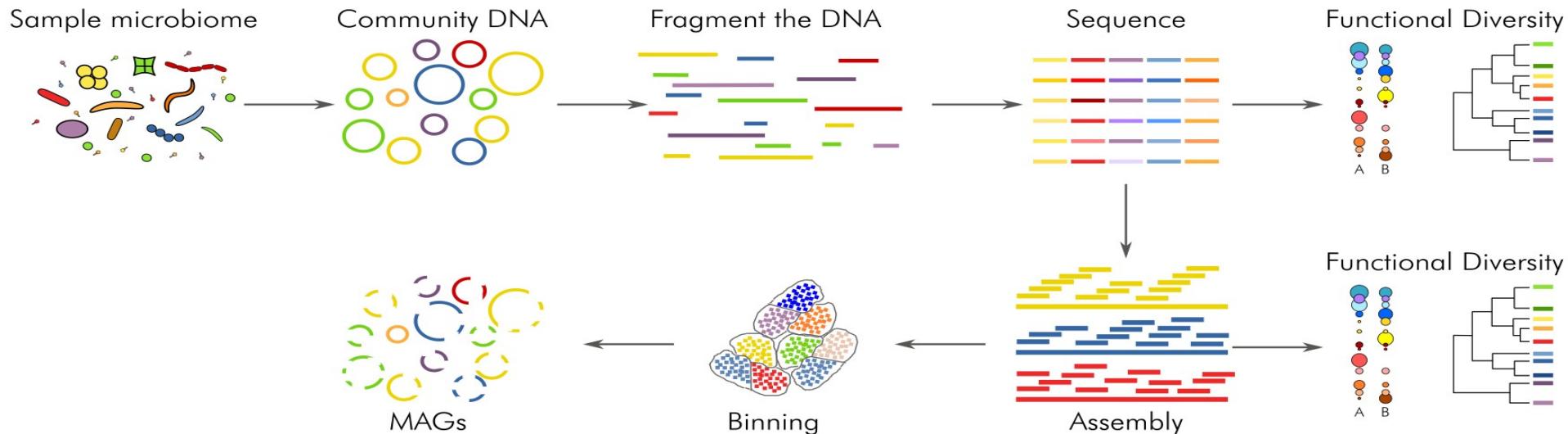




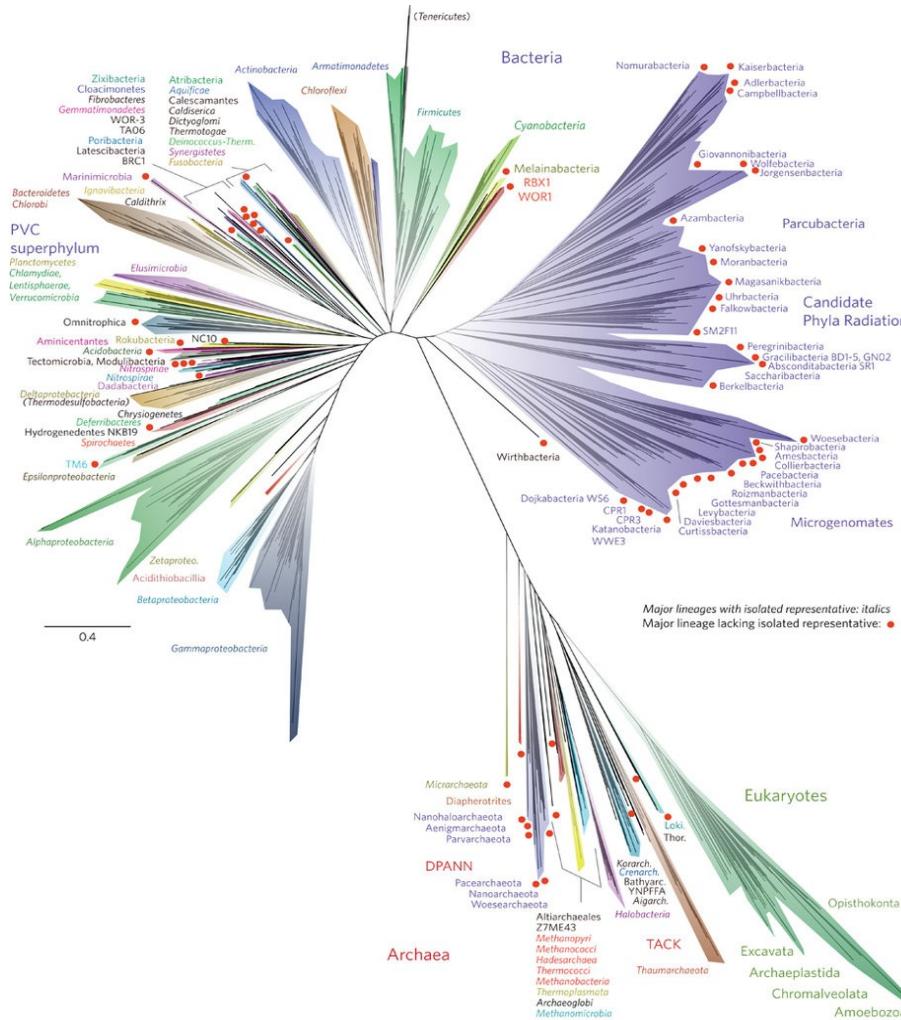
Metagenomic – MAGs

One of the main downstream possibility given by metagenomic is the reconstruction of complete genomes from the environmental sample

These approach, called **genome-resolved metagenomics** allow to draw a strong link between identity and function, and provides important information on the community taxonomic and functional structure. The result are called **Metagenome Assembled Genomes**



Metagenomic – MAGs



Hug et al 2016, Nat Microbiol

Metagenomic – MAGs

Meren Lab, University of Chicago
Microbial 'Omics: An introduction

<http://merenlab.org/momics/>



Metatranscriptomic

Metatranscriptomic refers to the sequencing of the community RNA **without the need for amplification (only reverse transcription to cDNA)**. The result is a pool of sequences containing a sample of all the **expressed** genes present in the community

Metatranscriptomic allows to get at the expressed functions of the community

It is usually linked to a metagenome, to which the reads are aligned and quantified against. One of the key problems, besides a large variability between replicates, is that up to 95% of a cell RNA is rRNA, with mRNA (the functions) being only 3-5%.

Some techniques are available for rRNA removal, but those are typically labourious and expensive



Metaproteomic

Metaproteomics refers to a number of different techniques, generally based on **high throughput liquid chromatography** (typically LC-MS/MS) to analyze the **total proteins** extracted from an environmental sample

Mass spectra are generally matched against a reference metagenome (or genomes) to find a match for the identified spectra

Metaproteomics allow to see the realized potential of the community, and to identify expressed proteins

It can be problematic to uniquely identify expressed proteins in complex samples



What?

Metametabolomic

Metametabolomics refers to the use of different techniques (like MS or NMR) to identify the metabolites from an environmental sample

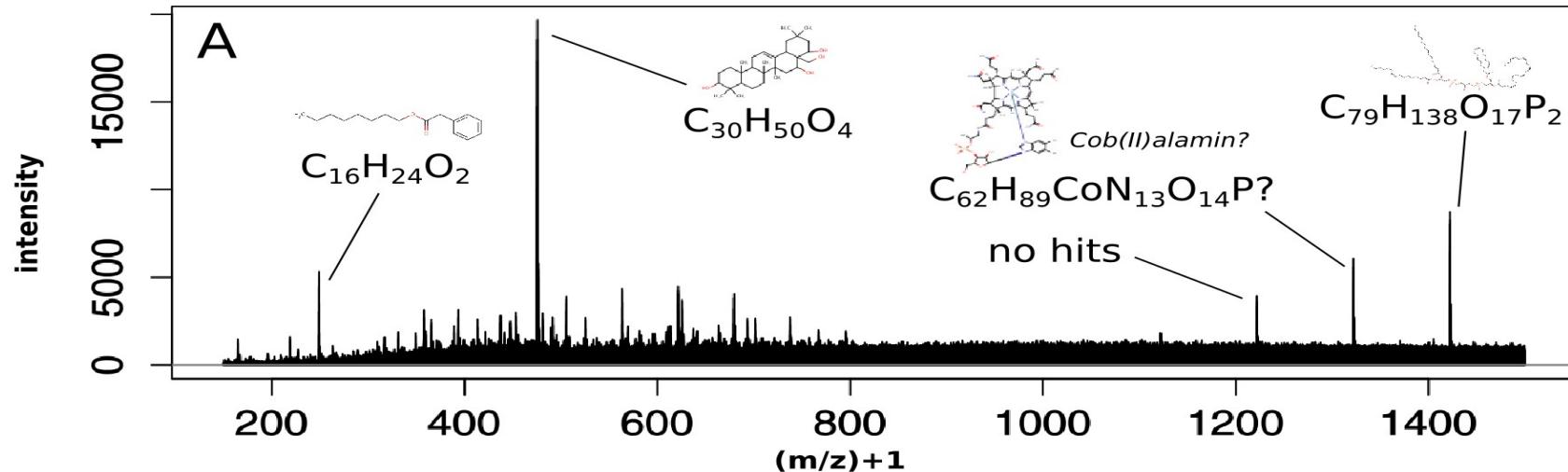
Metametabolomic give access to the metabolic products of a community

It is extremely challenging, as a single techniques cannot encompass the molecular diversity of metabolites, it is difficult to identify extracellular metabolites part of dissolved organic matter from intracellular metabolites and a single mass can refer to multiple compounds

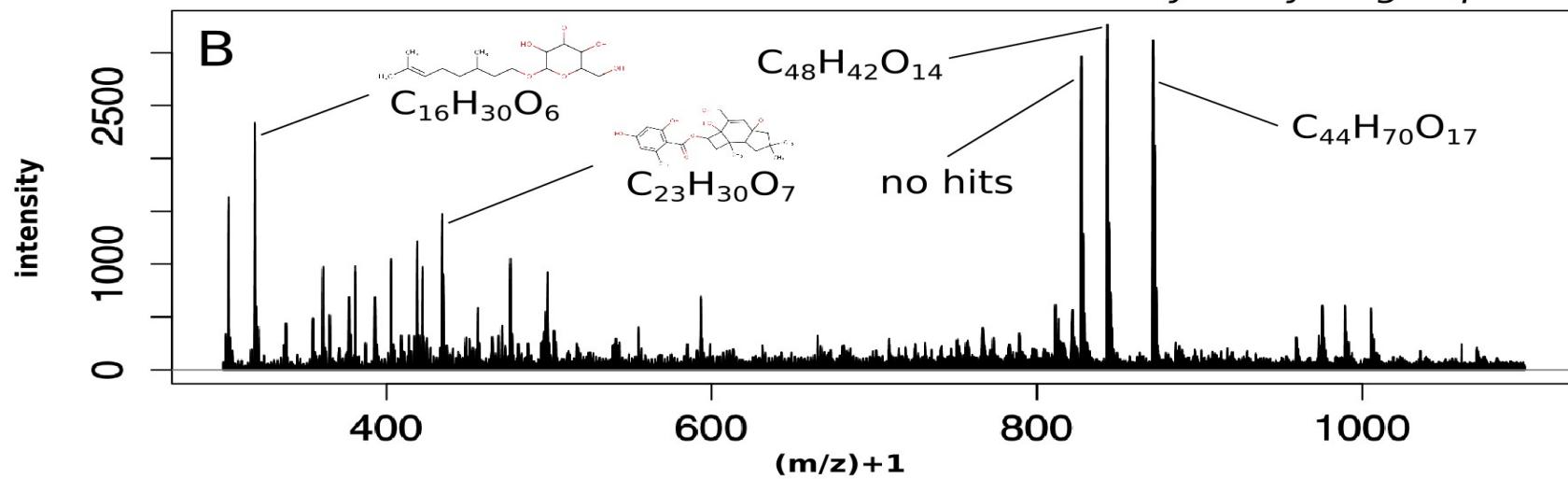
Approaches include looking either at specific compounds or using predicted metabolites from metagenomes as a template for mass searches



Themovibrio ammonificans



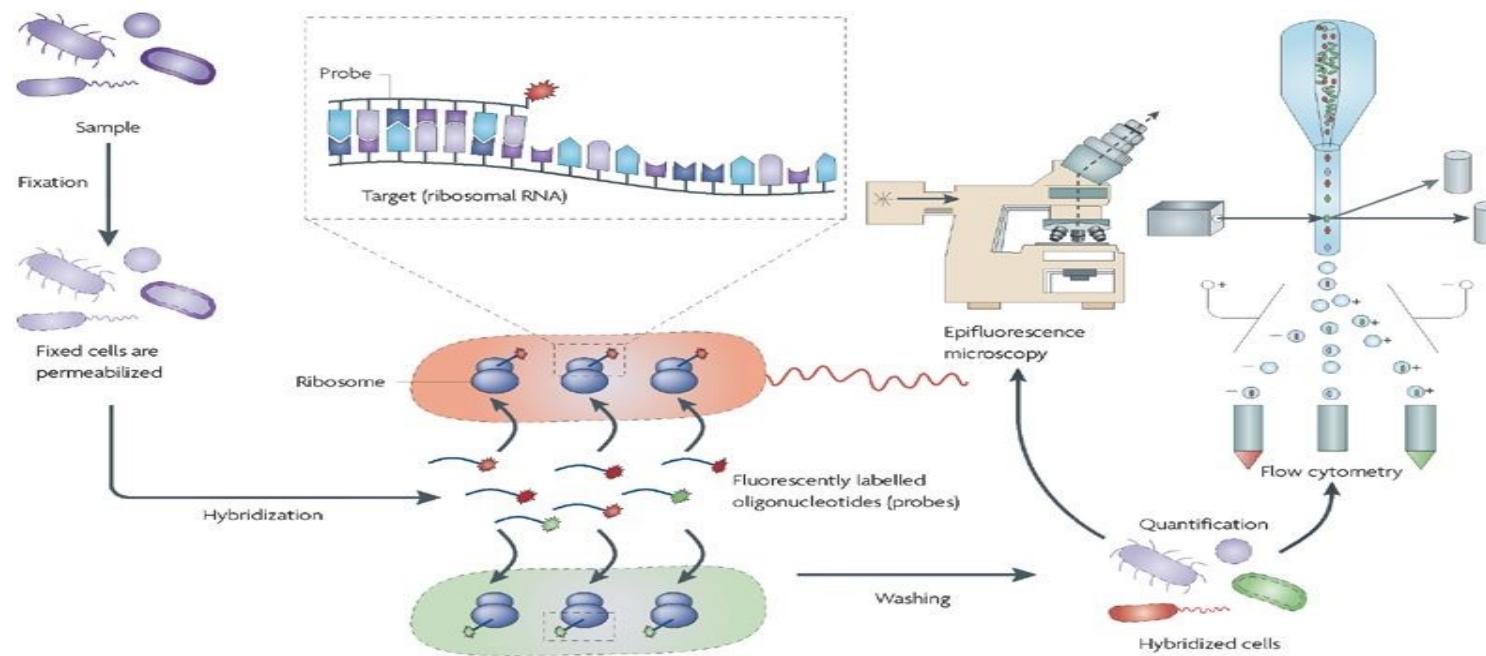
Phorcysia hydrogeniphila





Single cells Amplified Genomes

Single cells amplified genomes (SAGs) are obtained from cells that have been sorted using a Fluorescence Activated Cell Sorter (FACS). The DNA from these single cells is usually amplified using MDA and then sequenced. Draft genomes are usually obtained

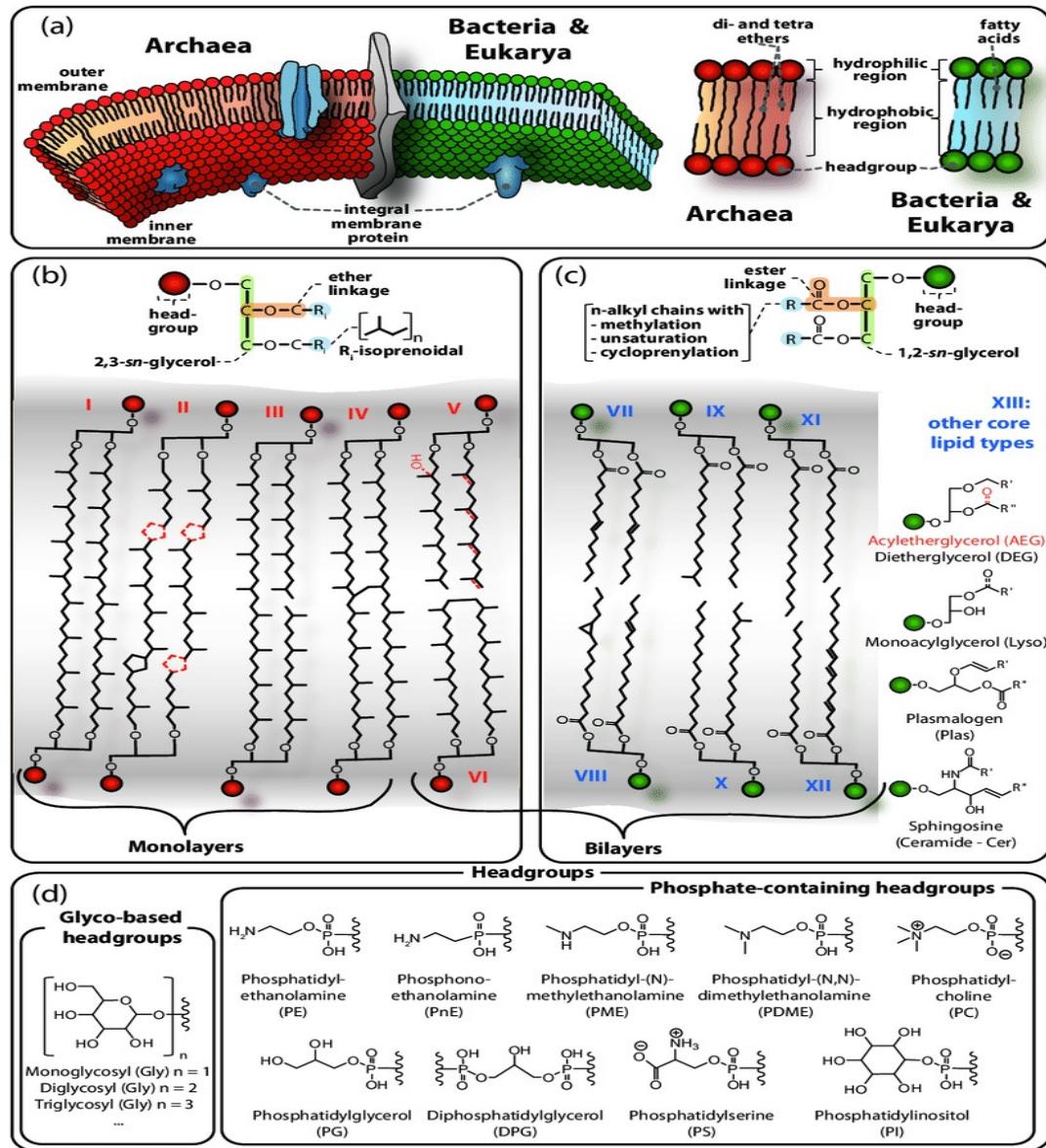




Lipids Biomarkers

Lipids are also used as biomarkers to identify groups of prokaryotes. Lipids, especially intact polar lipids, are an important source of information, especially in the sedimentary record, to reconstruct environmental information

Their resolutions depends on the group under investigation and the knowledge about their structure





Omics: the Good, the Bad and the Ugly

The Good: Omic techniques can be effective in providing a blue print of the ecosystem functioning. They can be combined with other techniques (e.g. Stable Isotope Probing) and can be applied to virtually every environment and microbial fraction (viruses, prokaryotes, microeukaryotes, protists). They can give us information on difficult or never cultivated microbes (like in *Riftia* example)



Omics: the Good, the Bad and the Ugly

The Good: Omic techniques can be effective in providing a blue print of the ecosystem functioning. They can be combined with other techniques (e.g. Stable Isotope Probing) and can be applied to virtually every environment and microbial fraction (viruses, prokaryotes, microeukaryotes, protists). They can give us information on difficult or never cultivated microbes (like in *Riftia* example)

The Bad: Omic technique can be technically challenging. The information obtained are as good as the database used to annotate the sequences. Some of the omics hide technical challenges



Omics: the Good, the Bad and the Ugly

The Good: Omic techniques can be effective in providing a blue print of the ecosystem functioning. They can be combined with other techniques (e.g. Stable Isotope Probing) and can be applied to virtually every environment and microbial fraction (viruses, prokaryotes, microeukaryotes, protists). They can give us information on difficult or never cultivated microbes (like in *Riftia* example)

The Bad: Omic technique can be technically challenging. The information obtained are as good as the database used to annotate the sequences. Some of the omics hide technical challenges

The Ugly: New diversity = unknown prokaryotes = unknown proteins = unknown functions. Hundred of Omics paper are published each year. The most common sentence are: "this is the first time that thisomic is applied to [...] environment", "we can conclude that 75% of the sampled diversity represent new, previously unknown species" or "of the sampled transcripts, 50% represents coding gene with unknown functions".

A vertical strip on the left edge of the slide shows a close-up of a ship's metal railing against a dark blue, slightly choppy ocean.

Thought exercise



