

University of Naples Federico II

Environmental Metagenomic

aa 2020-2021

METHODS IN MICROBIAL DIVERSITY

Donato Giovannelli

donato.giovannelli@unina.it • [@d_giovannelli](https://twitter.com/@d_giovannelli) • <http://dgiovannelli.github.io>



General (first order) questions in studying microbial diversity

Who's there?

What are they doing?

Who's doing what?

How are they doing it?

To what extent?

Milos shallow water hydrothermal vent biofilm



Higher order questions in studying microbial diversity

How is diversity influenced by environmental conditions?

How is functional diversity influenced by environmental conditions?

How is diversity and functioning influenced by species interactions?

How are metabolic rates influenced by environmental conditions and species interactions?

Milos shallow water hydrothermal vent biofilm

Approaches to Microbial Diversity

Each time we approach the study of microbial diversity we are following a similar workflow generally consisting of:

- 1- Definition of the study question and study design*
- 2- Sample collection*
- 3- Sample preservation*
- 4- Sample processing and data acquisition*
- 5- Data analysis*
- 6- Interpretation of results*

Approaches to Microbial Diversity

Each time we approach the study of microbial diversity we are following a similar workflow generally consisting of:

1- Definition of the study question and study design

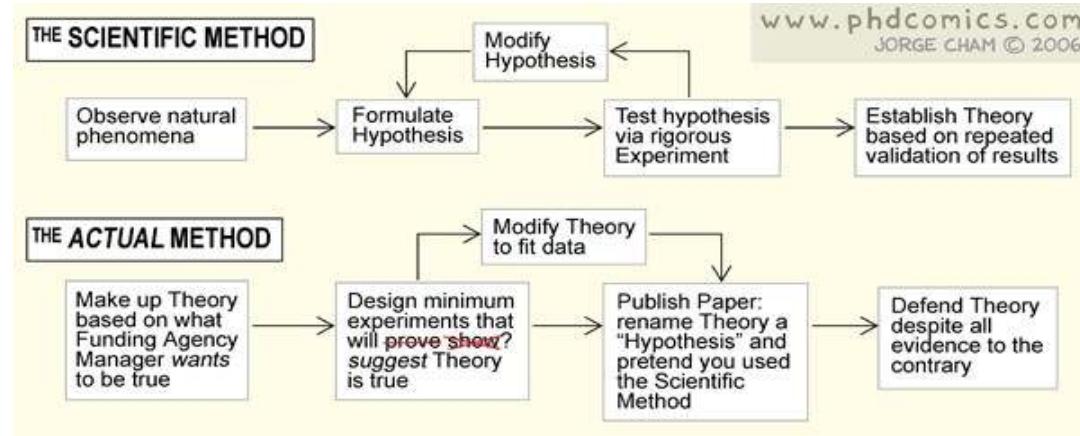
2- Sample collection

3- Sample preservation

4- Sample processing and data acquisition

5- Data analysis

6- Interpretation of results



THE SCIENTIFIC METHOD

www.phdcomics.com
JORGE CHAM © 2006

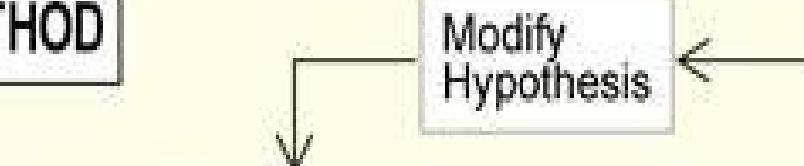
Observe natural phenomena

Formulate Hypothesis

Test hypothesis via rigorous Experiment

Establish Theory based on repeated validation of results

Modify Hypothesis



THE SCIENTIFIC METHOD

www.phdcomics.com
JORGE CHAM © 2006

Observe natural phenomena

Formulate Hypothesis

Test hypothesis via rigorous Experiment

Establish Theory based on repeated validation of results

Modify Hypothesis

THE ACTUAL METHOD

Make up Theory based on what Funding Agency Manager wants to be true

Design minimum experiments that will prove ~~show?~~ suggest Theory is true

Publish Paper: rename Theory a "Hypothesis" and pretend you used the Scientific Method

Defend Theory despite all evidence to the contrary

Modify Theory to fit data

Richard Feynman on Scientific Method (1964)



<https://www.youtube.com/watch?v=0KmimDq4cSU>

Problem in studying microbial diversity

Spatial heterogeneity – most methods need several hundreds milliliters of waters or few grams of sediments

Inability to culture – Yet most of the microbial diversity is uncultured, with big implications for what we can understand

Taxonomic ambiguities of microbes – Difficult definition of “species”, high genomic plasticity, horizontal gene transfer

Technical bias of chosen methods – Each techniques has its own bias, that need to be considered while making inferences

sampling for Microbial Diversity

Sampling an environment can be done in several different ways depending on the scientific question, the environment under investigation and the technique to be used

Generally speaking samples can be collected of (sea)water, specific fluids, sediments/soils, aerosols, biofilms or flora/fauna

The difficulties and technological hurdles associated with collecting these sample largely depend on the environment investigated

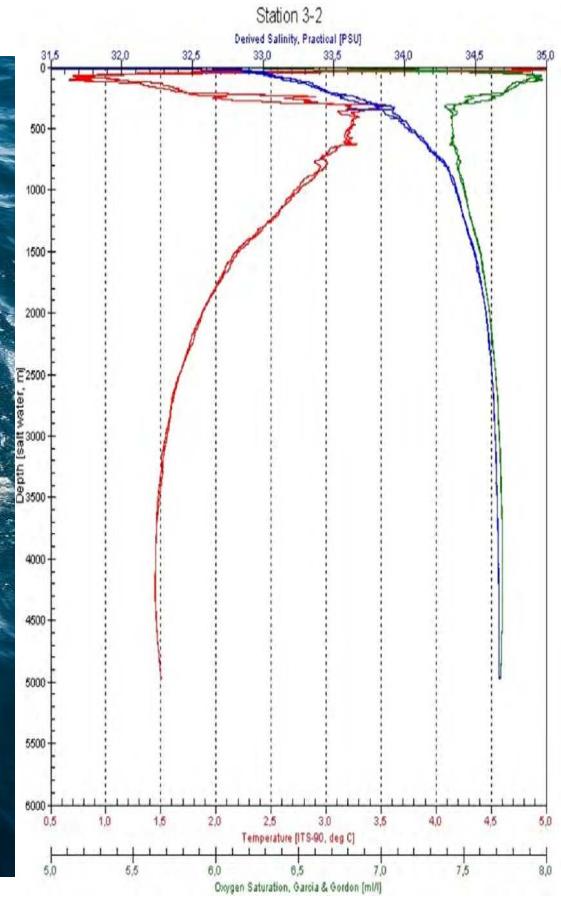
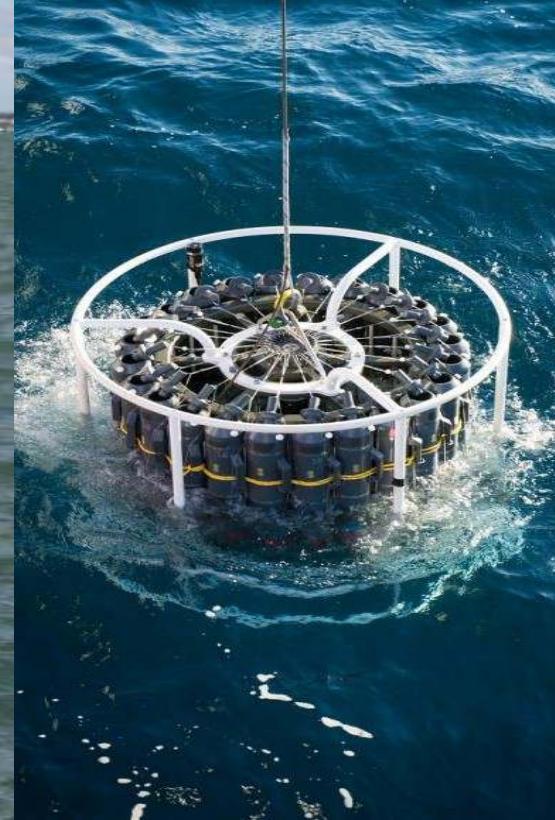
The need for specific sample preservation protocols might also complicate the sampling procedure

sampling (sea)water

Seawater can be sampled in several different ways. The principal techniques are the use of hand-operated sterile bottles or containers, collection of known volumes using pumps and using the Niskin bottles, often mounted on a CTD-Rosette sampler



sampling (sea)water



sampling (sea)water



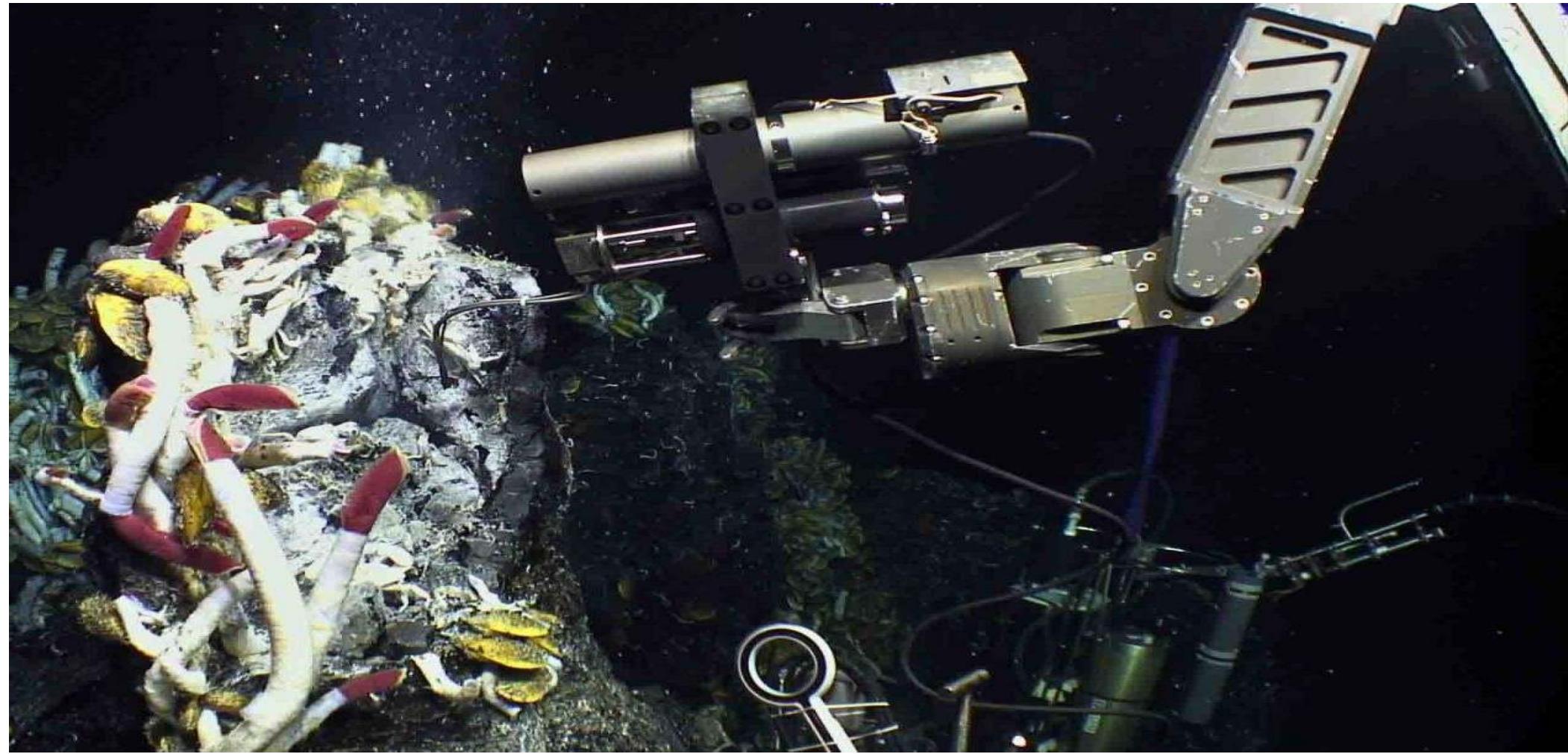
sampling Fluids

Fluids, like sediment pore fluids, or fluids seeping from a cold seeps or hydrothermal vent can be sampled using different approaches, that are specific to each case.

Common approaches include centrifugation of sediments (pore fluids), core suction (pore fluids), pumps (fluids) and syringes (pore fluids and fluids)



sampling Fluids



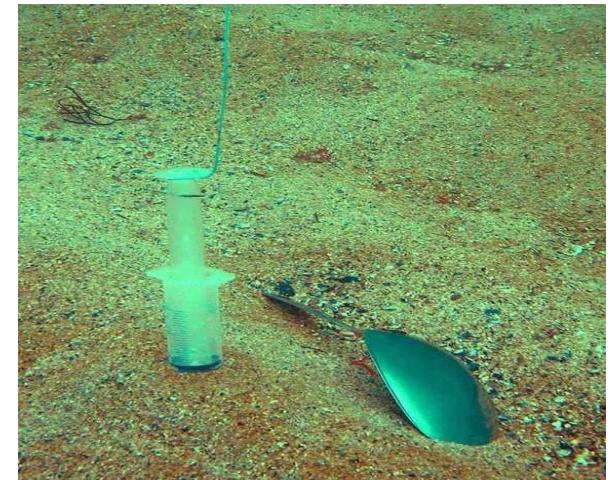
sampling *Biofilms*

Biofilms can be sampled through a syringe, a push core, or directly swabbing the surface

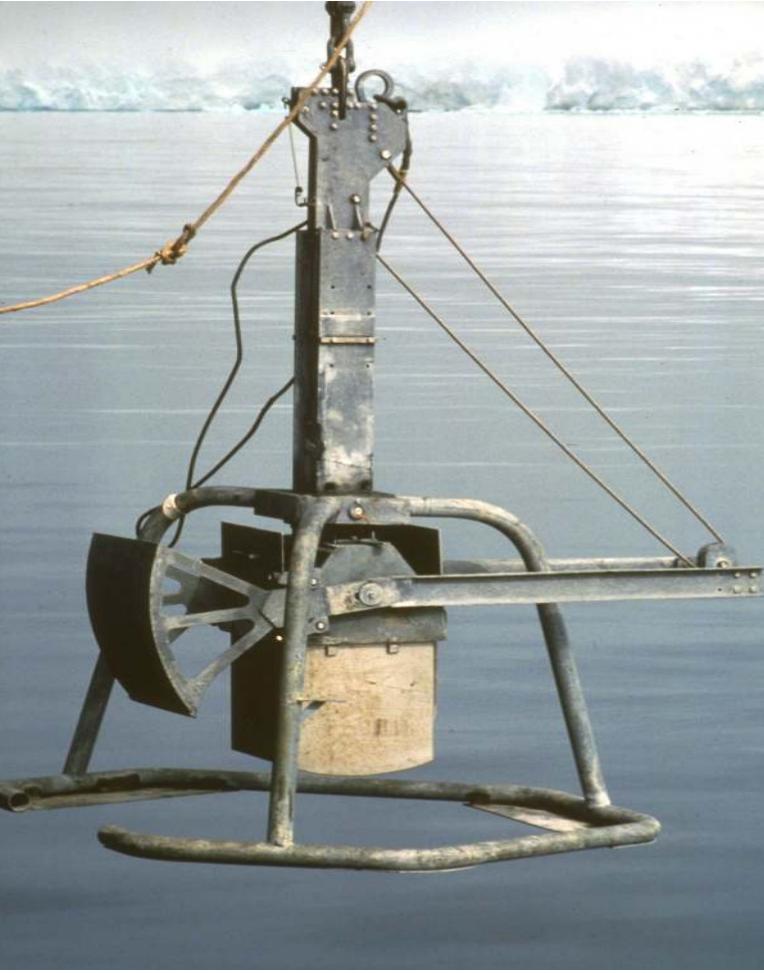
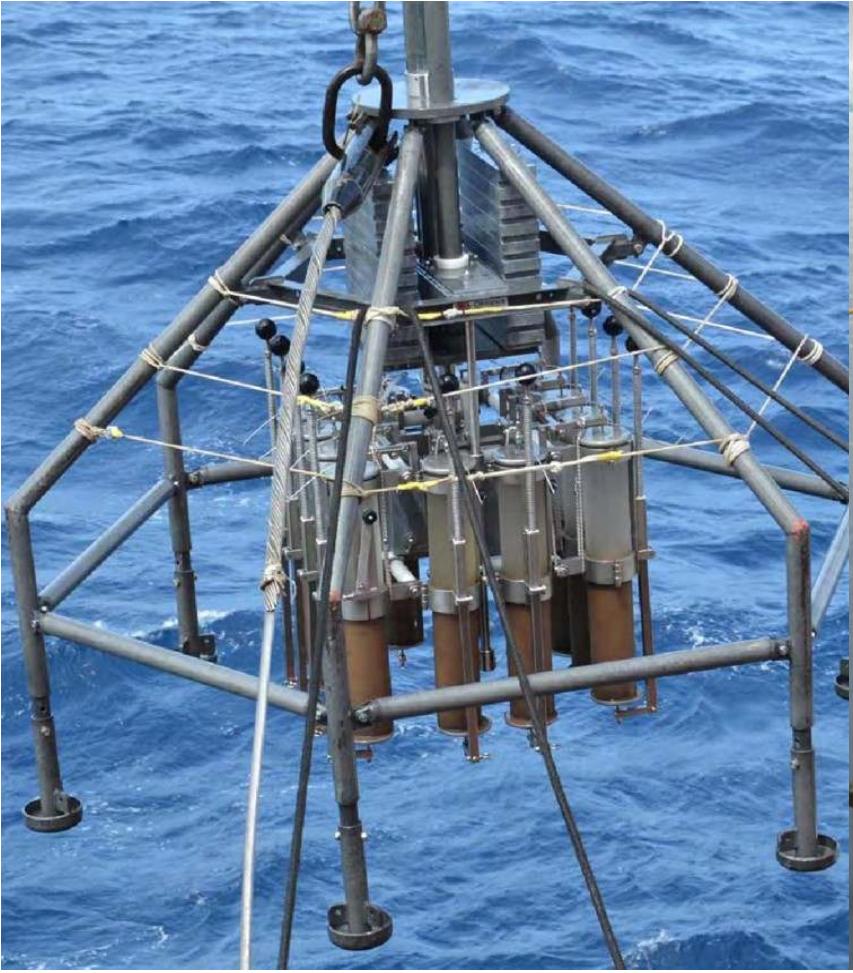


sampling **Sediments/Soils**

Sediments can be sampled in a variety of ways. The simplest way is to use a sterile container (like a falcon tube), however the most used method is a push core. Beyond that the use of multicorers, boxcorers, gravity corers and drilling is also used, depending on the purpose and location

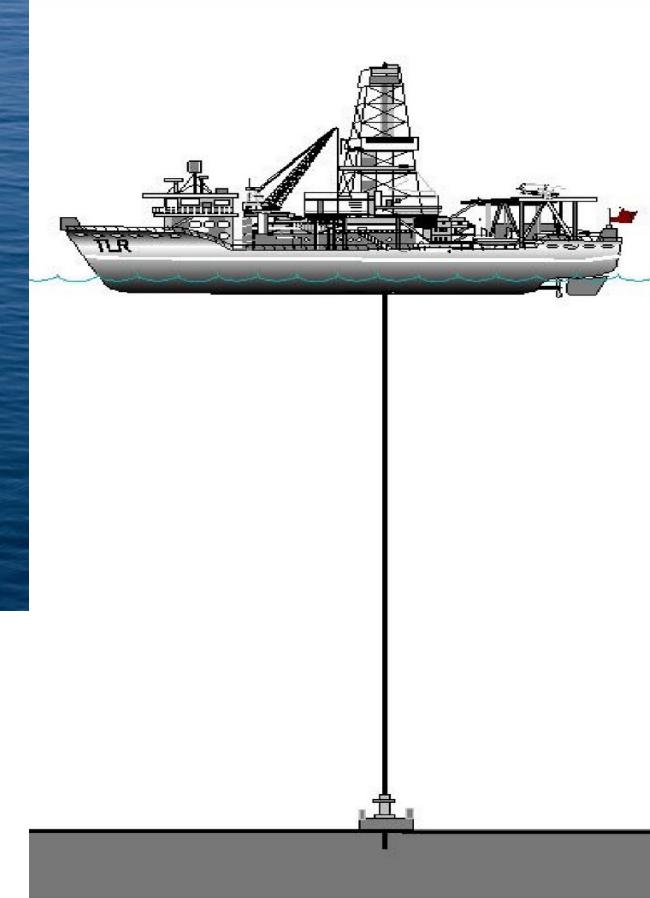
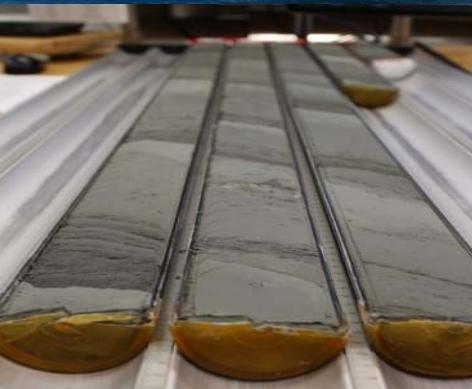


sampling Sediments



sampling Sediments

研究航海



sampling Soils



sample *Preservation*

Preservation is highly dependent upon the type of downstream analysis that needs to be carried out. There are some common preservation strategies that include the modulation of temperature or the addition of specific preservatives. These are often used in combinations

- Refrigeration at +4°C (culturing, enrichment, live specimens)
- Freezing at -20°C (viral counts, chemistry, DNA)
- Freezing at -80°C or liq-N₂ (DNA, RNA, proteins)
- Preserving solutions like formaldehyde, gluteraldehyde, RNA later, PBS)



sample Analysis

The technique used depends on the scientific question and the analytical principles used. We can divide techniques on different ways:

Culture-dependent and **culture-independent** techniques, refers to a grouping of techniques based on the use of culturing as a basic step in the investigation. Since the majority of the microbes is currently uncultured, the decision of using culture-dependent vs independent techniques is very important

Qualitative, semi-quantitative and **quantitative**, refers to the type of information obtained from the analysis. Often semi-quantitative and quantitative techniques are all referred as quantitative, although not correct. Bias in the chose technique usually impair our ability to obtain true quantitative data

Chemical, biochemical, molecular, isotopic, or a combination refers to the analytical principles used by each technique

Culture-dependent vs culture-independent

Given our current inability to grow a large portion of marine microbes, the choice between culture dependent and independent technique is very important

Culture-independent techniques allow you to probe the natural diversity of microbial communities **avoiding the bottleneck of culturing**. Despite this the inferences about functional diversity are entirely dependent upon information obtained from pure microbial cultures

Culture-dependent approaches, while highly selective for a small subset of microorganisms, have the advantage of resulting in a **new model system** that can be used for probing functional diversity using physiology, genetics and biochemistry

Some approach are an hybrid between the two techniques, for example combining enrichments with molecular tools

Culture-dependent vs culture-independent

Given our current inability to grow a large portion of marine microbes, the choice between culture dependent and independent technique is very important

Culture-independent techniques allow you to probe the natural diversity of microbial communities **avoiding the bottleneck of culturing**. Despite this the inferences about functional diversity are entirely dependent upon information obtained from pure microbial cultures

Culture-dependent approaches, while highly selective for a small subset of microorganisms, have the advantage of resulting in a **new model system** that can be used for probing functional diversity using physiology, genetics and biochemistry

Some approach are an hybrid between the two techniques, for example combining enrichments with molecular tools

A successful approach to the study of microbial diversity requires both approaches combined



General questions in studying microbial diversity

Who's there?

What are they doing?

Who's doing what?

How are they doing it?

How much?



General questions in studying microbial diversity

Who's there?



What are they doing?



Who's doing what?



How are they doing it?



How much?

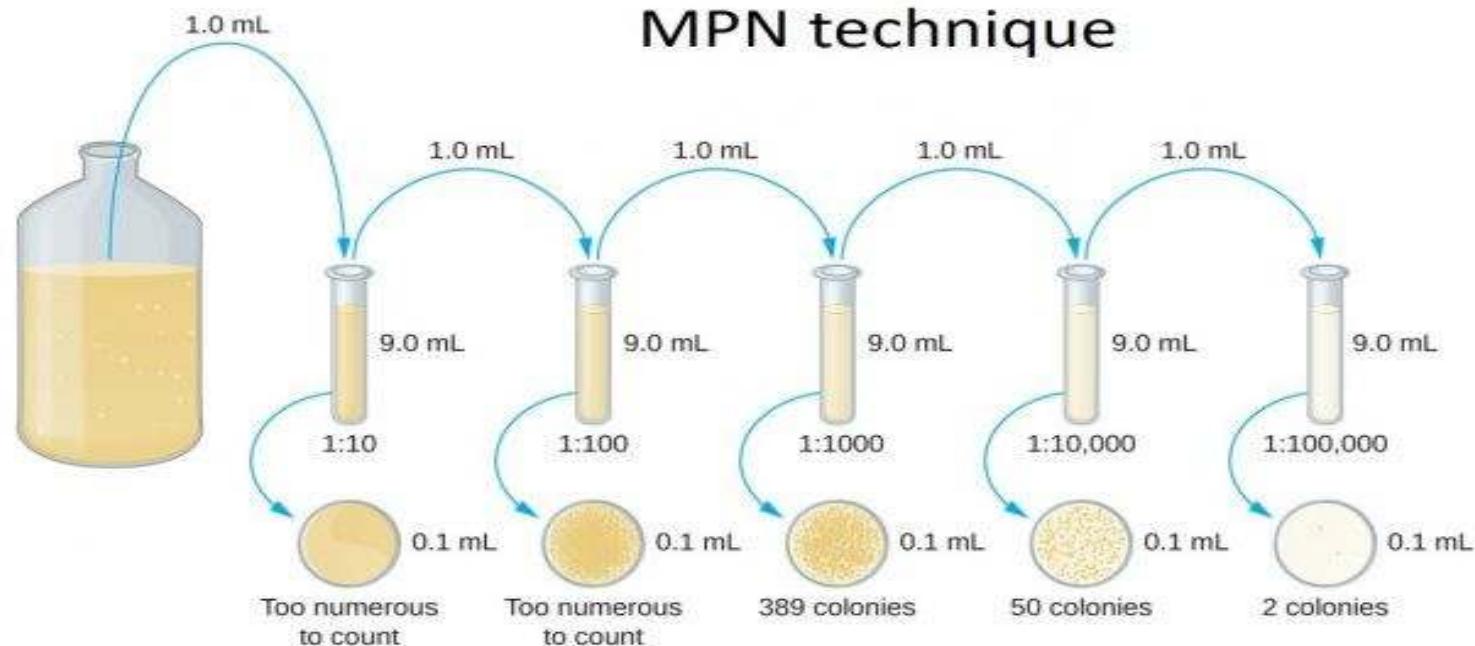


Culture-dependent approaches



Most Probable Numbers

Most Probable Numbers (MPN) techniques relies on our ability to selectively culture specific trophic groups. A serial dilution of a sample is made on a selective media (e.g. thiosulfate oxidation), and the viable colonies for each dilution are counted on a plate. The number of original cells for that specific metabolism are then back calculated





Most Probable Numbers

Most Probable Numbers (MPN) techniques relies on our ability to selectively culture specific trophic groups. A serial dilution of a sample is made on a selective media (e.g. thiosulfate oxidation), and the viable colonies for each dilution are counted on a plate. The number of original cells for that specific metabolism are then back calculated

PROS: Gives you a quantitative assessment of the number of active cells performing a given metabolism

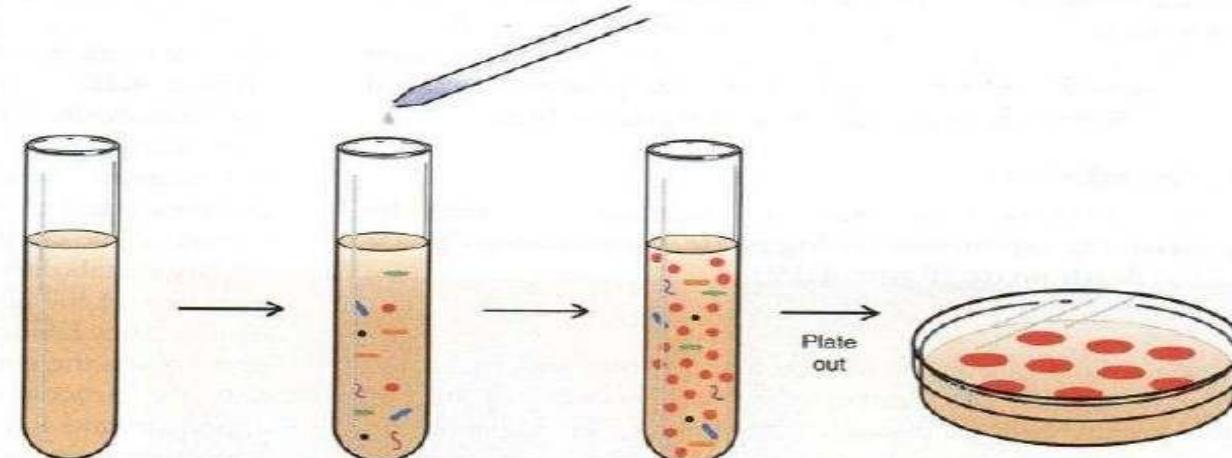
CONS: Useful only for well defined metabolisms and taxonomic groups; bias toward cultured groups; the amount of cells that can perform a given metabolism is usually underestimated



Enrichments and Isolation



Defined media for a specific metabolism of interest is inoculated with the original sample, following the enrichment for microbial groups able to grow, pure cultures are then isolated and characterized



Medium contains select nutrient sources chosen because few bacteria, other than the organism of interest, can use them.

Sample that contains a wide variety of organisms, including the organism of interest, is added to the medium.

Organism of interest can multiply, whereas most others cannot.

Enriched sample is plated onto appropriate agar medium. A pure culture is obtained by selecting a single colony of the organism of interest.



Enrichments and Isolation



Defined media for a specific metabolism of interest is inoculated with the original sample, following the enrichment for microbial groups able to grow, pure cultures are then isolated and characterized

While the initial enrichment and isolation can be based on different techniques (different types of incubators, cell sorting, co-cultures), pure cultures are the **ONLY officially approved approach** that can lead to the description of a **new microbial species**

New species names are officially valid only if they appear on the journal of the International Committee on the Systematic of Prokaryotes (ICSP) in the **International Journal of Systematic and Evolutionary Microbiology (IJSEM)**

Approved name appear as **Genus specie sp. nov.** in their first appearance, and **Genus species** from then on. Not officially approved species appear instead as "**Candidatus Genus species**" or "**Ca. Genus species**"



Enrichments and Isolation



Officially described species are deposited in at least two culture collections, from where they are available to the community for analysis

Pure cultures give a model organism for **manipulative experiments, physiological, biochemical** and **genetic studies**, allowing for new microbial functions to be characterized

Being “Uncultured” is **an operational definition**, not an intrinsic attribute of the organism. The ultimate goal is to culture the uncultured

Recently the community has started referring to uncultured microbes as “microbial dark matter”, a term that while great for science communication purpose should be avoided within the community (more at
<http://merenlab.org/2017/06/22/microbial-dark-matter>)

Even a single isolate from a previously uncultured group can dramatically change our view of the ecological role of the group

example *Thermovibrio ammonificans*

International Journal of Systematic and Evolutionary Microbiology (2004), 54, 175–181

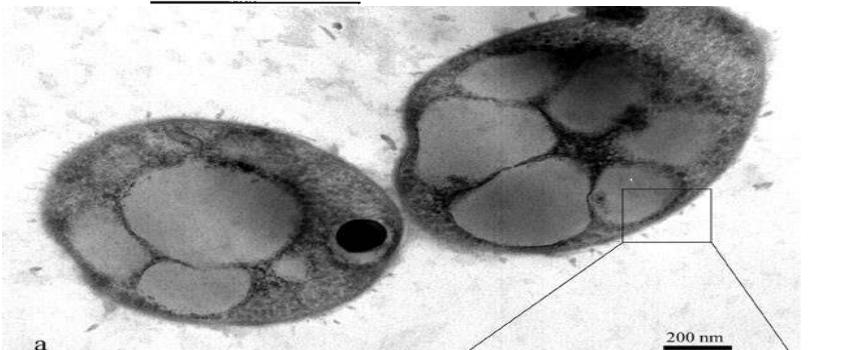
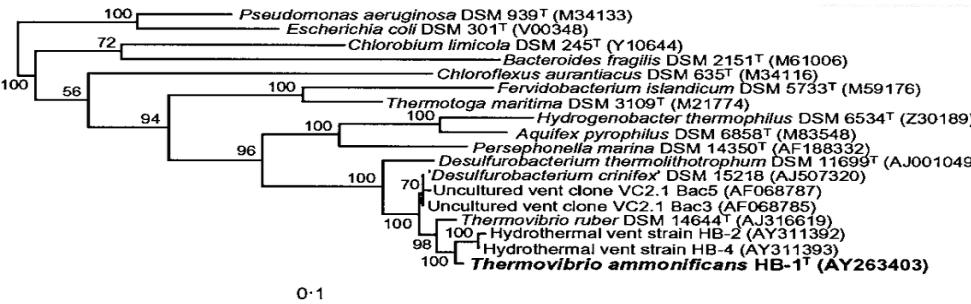
DOI 10.1099/ijs.0.02781-0



research papers

Thermovibrio ammonificans sp. nov.,
a thermophilic, chemolithotrophic,
nitrate-ammonifying bacterium from
deep-sea hydrothermal vents

Costantino Vetriani,^{1,2} Mark D. Speck,² Susan V. Ellor,² Richard A. Lutz²
and Valentin Starovoytov³



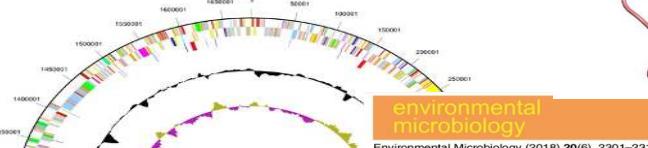
Thermovibrio ammonificans 4: DVIVVGGGPGAGFNAVKAVRSLYPEKRVLVLNDR--DLQIPECSIPY---VVAAGL-LKPEDNRY-PVKKL-AELG
Thermococcus kodakarensis 3: DVVVIGGSAGGLTAASIAKRFYPDKSVLVIKKED-VSMIPCGIPIY---IFGTLRSVEDDVLL--PTERFLKPLG
Nautilia profundicola 2: DIVVIGGGVGSGVVAANAKRFYPSKVVVIEKKNP-KKLIPECGIPIY---IFDFTY-GIDDDLMH--LEKKLKKFN
Pyrococcus furiosus 4: KVVIIGGGAAAGMSAASRVKRLRPEWDVKVFATEWVSHAPCGIPIY---VVEGI-APKEKLMHYPPEVFIKKRG
Pyrococcus horikoshii 7: KVVIIGGGAAAGMSAASRVKRLRPEWDVKVFATEWVSHAPCGIPIY---VVEGI-APKEKLMHYPPEVFIKKRG
Shewanella loihica 2: KILIIIGGVAGGASAARARRLSETAEIIMFERGEYVSFANGLPY---HSIGEIAORSALVLOTPESFKARFN
Aquifex aeolicus 1: KVVIIVGNGPAAASAVEAFRKVDRDSEIIILSDEEFPTYA PNCMEN---VIRDD-ISKEALFYKGGEKFYEKYR
Hydrogenobacter thermophilus 1: KVVIIVGTMAGSALVDELLKTDPSLEHLHFGEK---SPYRNRYLTDVDLSGK-KLPSQLLFKSQYQRF-EEGG
Thermocrinis ruber 1: KVVIIGAGMAGSALVDELLSLEPSLEIHLFGEER---TPYRNRYLTDLLAGA-KLPSQLLLLKSQKF-EEGG

Sciences (2012) 7:82–90

DOI:10.4056

**Complete genome sequence of *Thermovibrio ammon*
*HB-1*T, a thermophilic, chemolithoautotrophic bacter
isolated from a deep-sea hydrothermal vent**

Donato Giovannelli^{1,2*}, Jessica Ricci^{1,2}, Illeana Pérez-Rodríguez^{1,2}, Michael Hügler¹,
O'Brien¹, Ramaydalis Keddis^{1,2}, Ashley Grosche^{1,2}, Lynne Goodwin¹, David Br¹,
Davenport¹, Chris Detter¹, James Han¹, Shunsheng Han¹, Natalia Ivanova²,
Land¹, Natalia Mikhailova³, Matt Nolan¹, Sam Pitluck¹, Roxanne Tapia¹, Tanja¹,
Costantino Vetriani^{1,2}

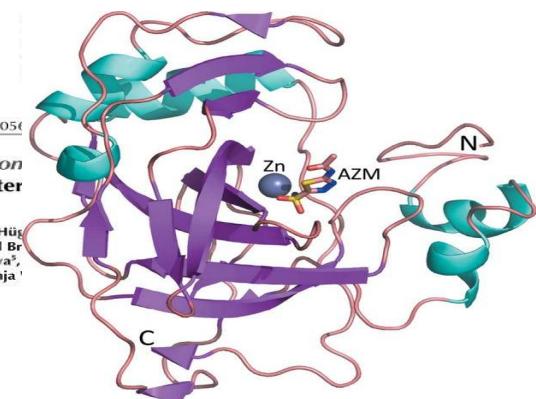


environmental
microbiology
Environmental Microbiology (2018) 20(6), 2301–2316



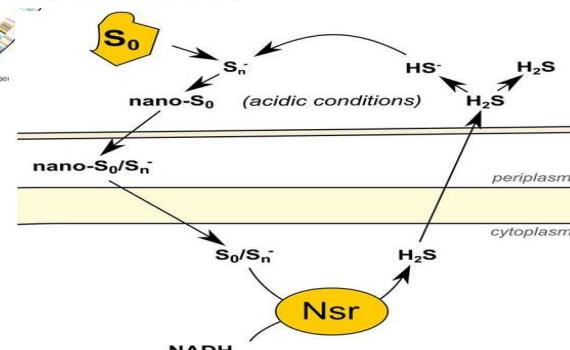
doi:10.1111/1462-2920.14280

The structure of a tetrameric α -carbonic anhydrase from *Thermovibrio ammonificans* reveals a core formed around intermolecular disulfides that contribute to its thermostability



Elemental sulfur reduction in the deep-sea vent thermophile, *Thermovibrio ammonificans*

Benjamin Jelen,¹ Donato Giovannelli^{1,2,3,4}
Paul G. Falkowski^{1,2,5} and Costantino Vetriani^{1,2,*}



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

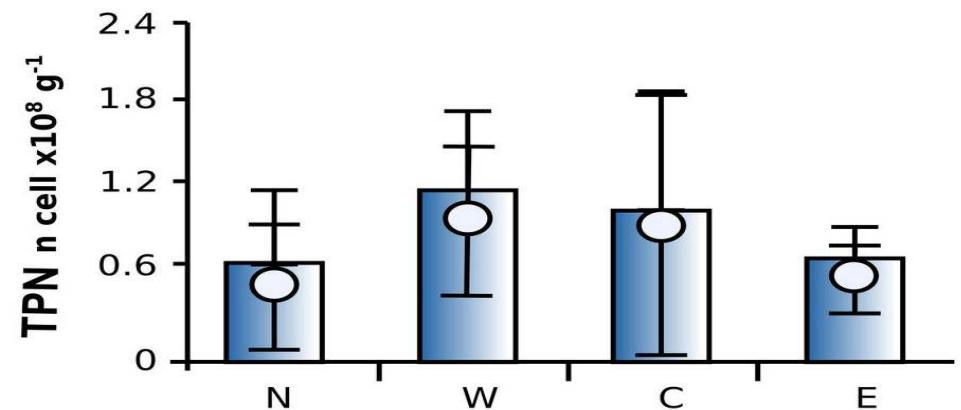
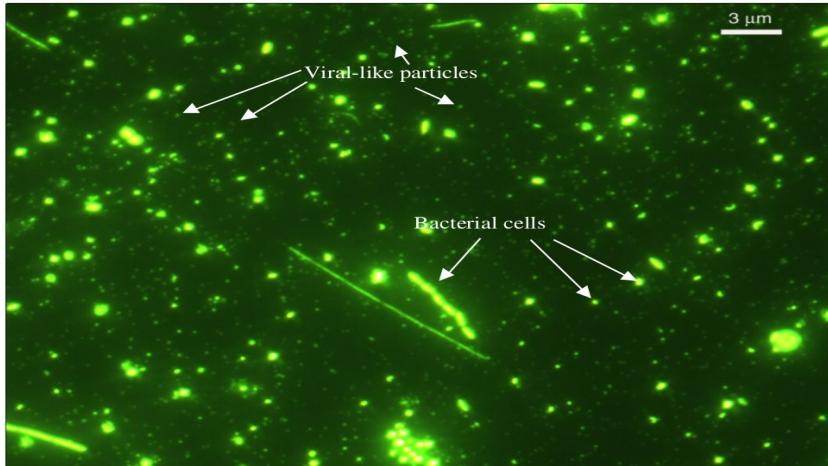
The big leap in culture-independent techniques started in the seventies with the evolution of molecular biology techniques. These allowed to probe the microbial world in new ways (mainly DNA at the time), bringing upon us a big revolution

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

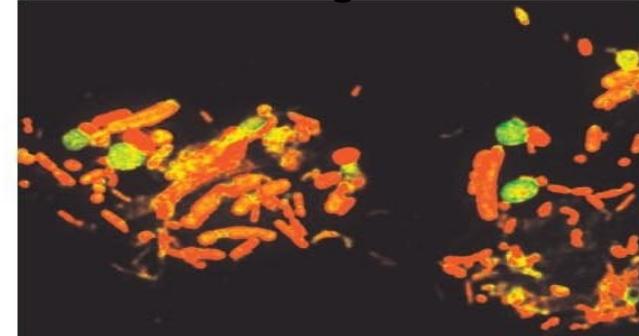


Marc Mussman and Michael Wagner

(a)

© 2015 Pearson Education, Inc.

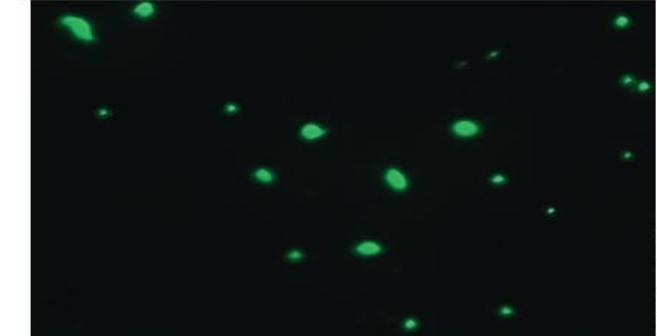
Acridine orange stain



Marc Mussman and Michael Wagner

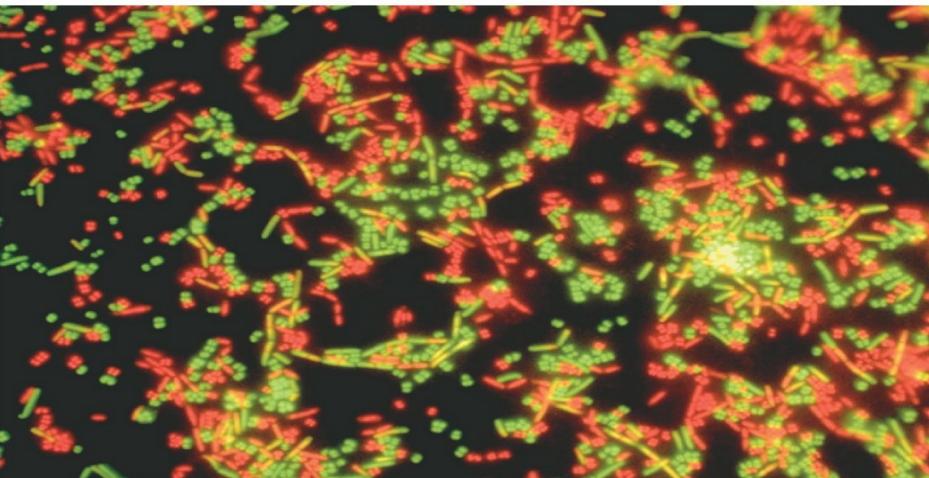
(b)

SYBR Green stain



Willim Martins-Habanna

(c)



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.

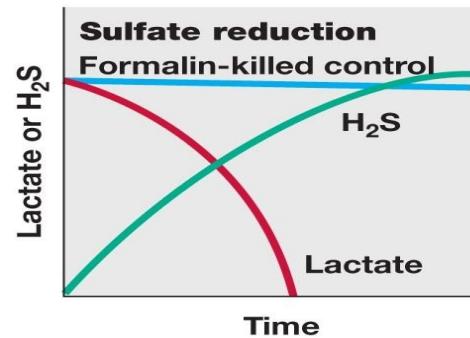
© 2015 Pearson Education, Inc.



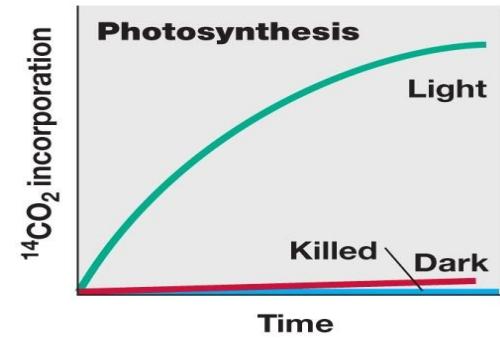
Activity measurements

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

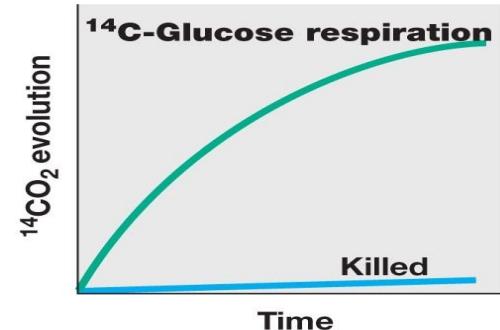
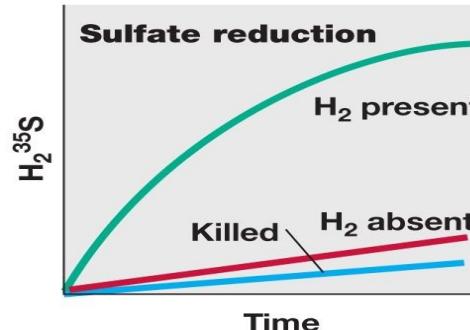


(a)



(b)

Stable or radioactive isotopes



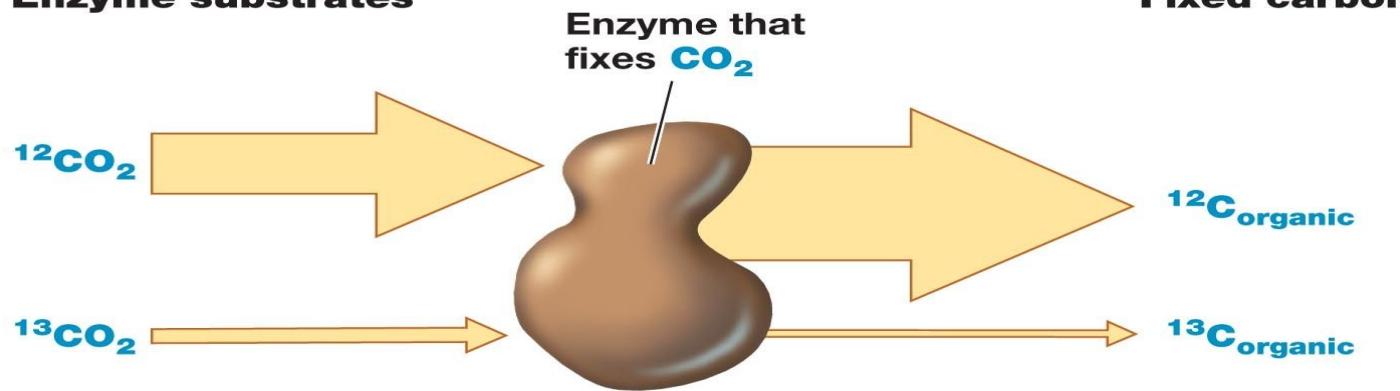


Stable isotopes

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

Enzyme substrates



Fixed carbon

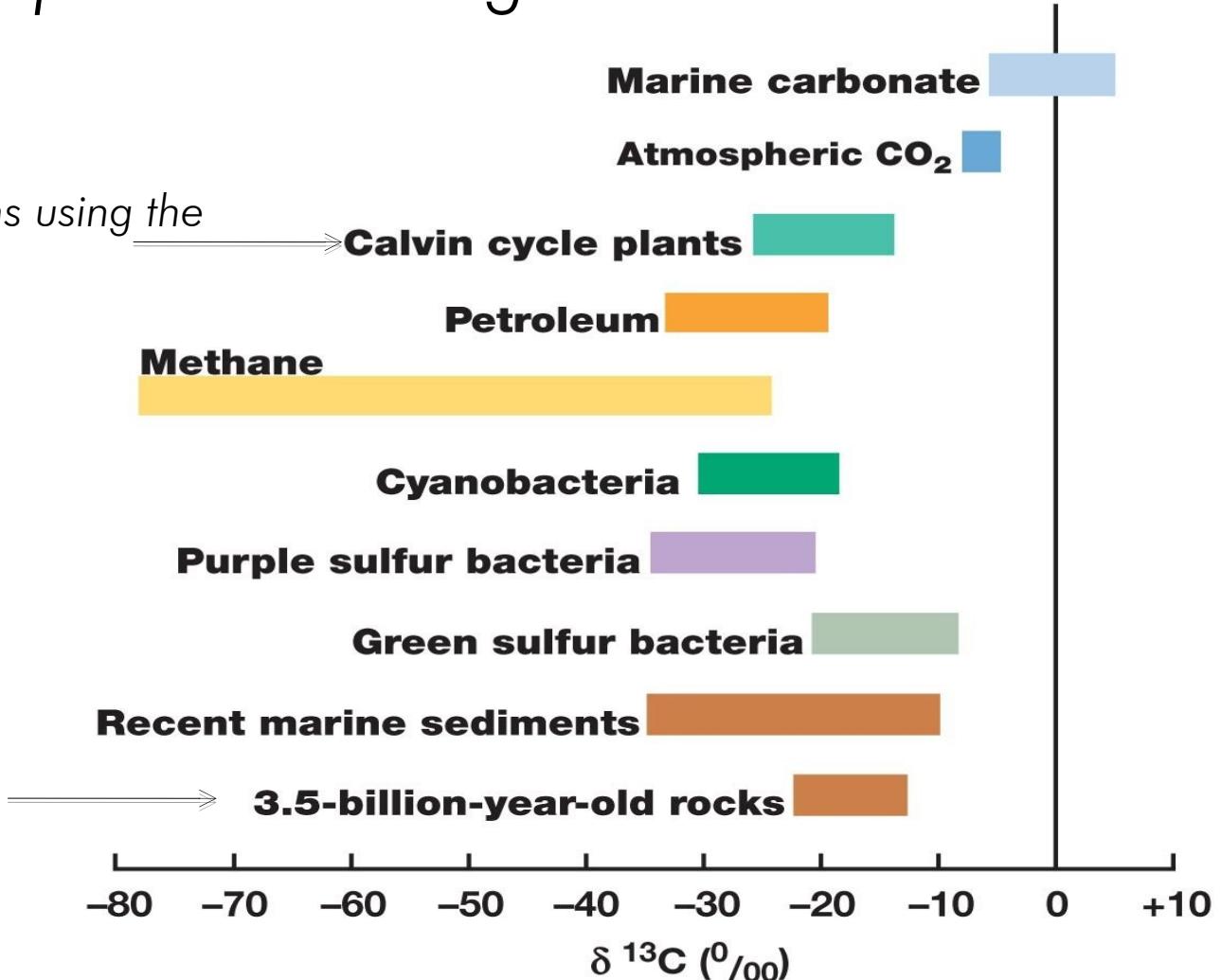
$$\delta^{13}\text{C} = \frac{(^{13}\text{C}/^{12}\text{C sample}) - (^{13}\text{C}/^{12}\text{C standard})}{(^{13}\text{C}/^{12}\text{C standard})} \times 1000$$

© 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

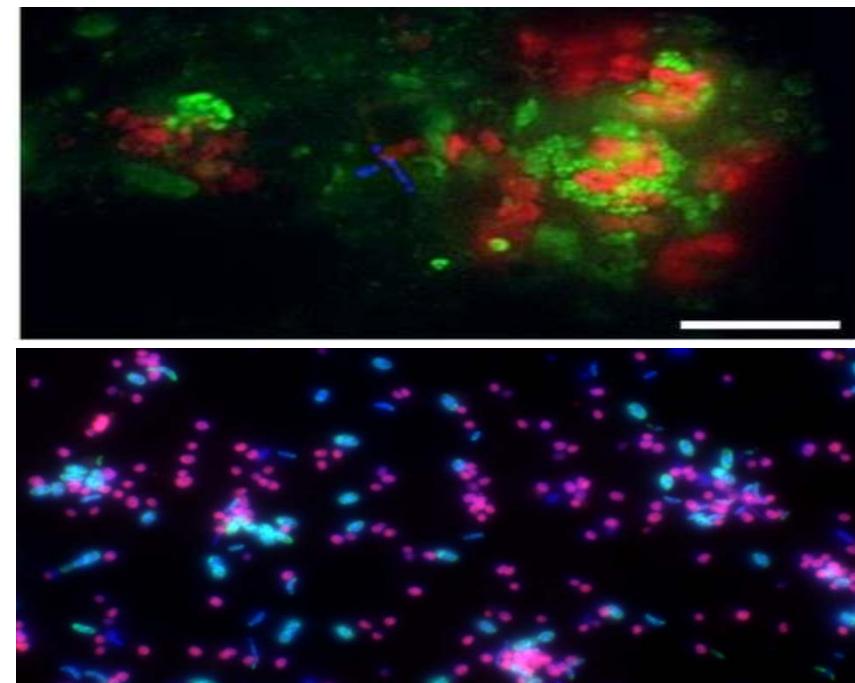
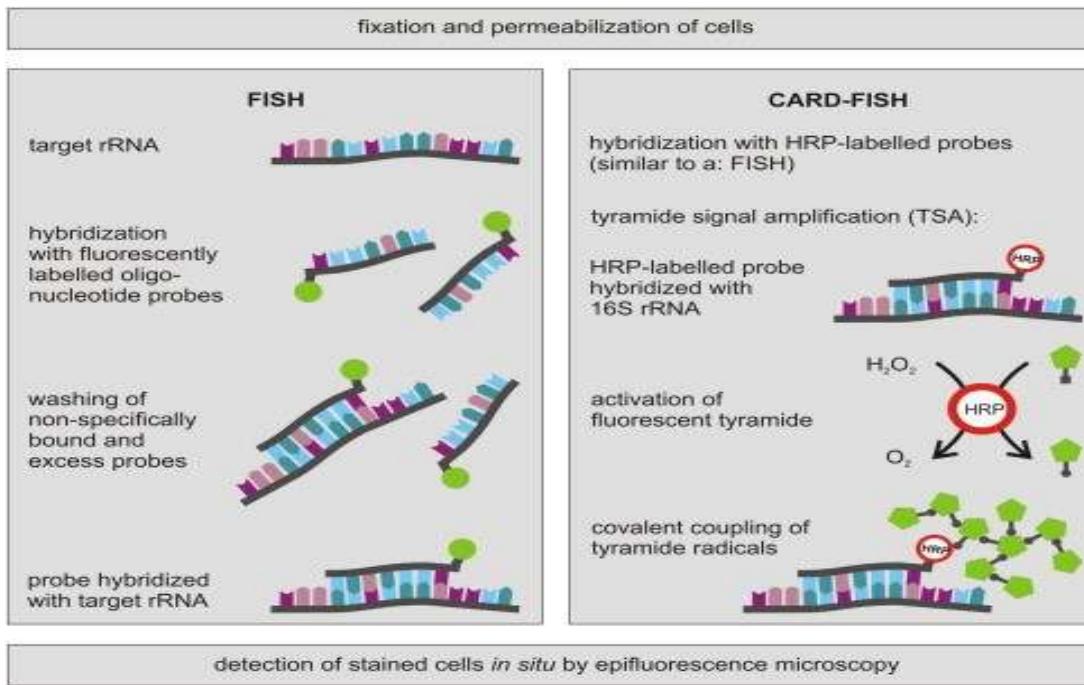
Carbon fixed by autotrophic organisms using the CBB cycle is depleted in ^{13}C



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.



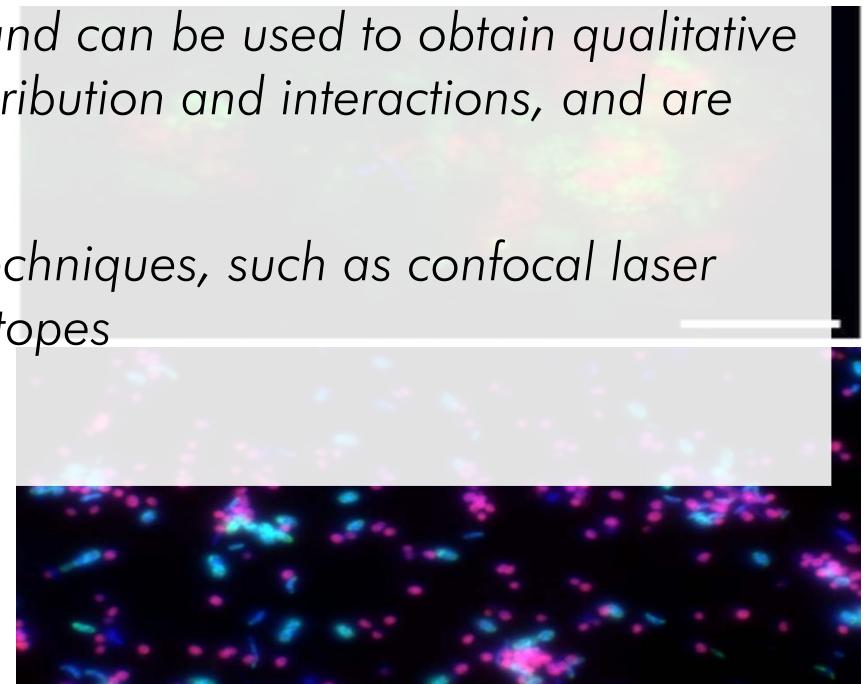
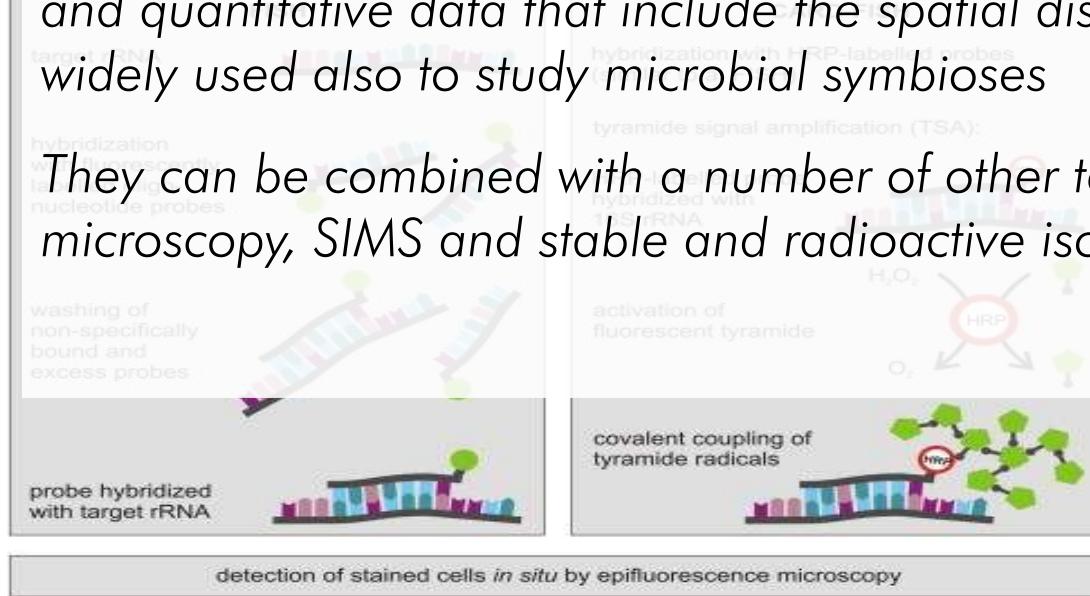


Fluorescent in situ hybridization (FISH)

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

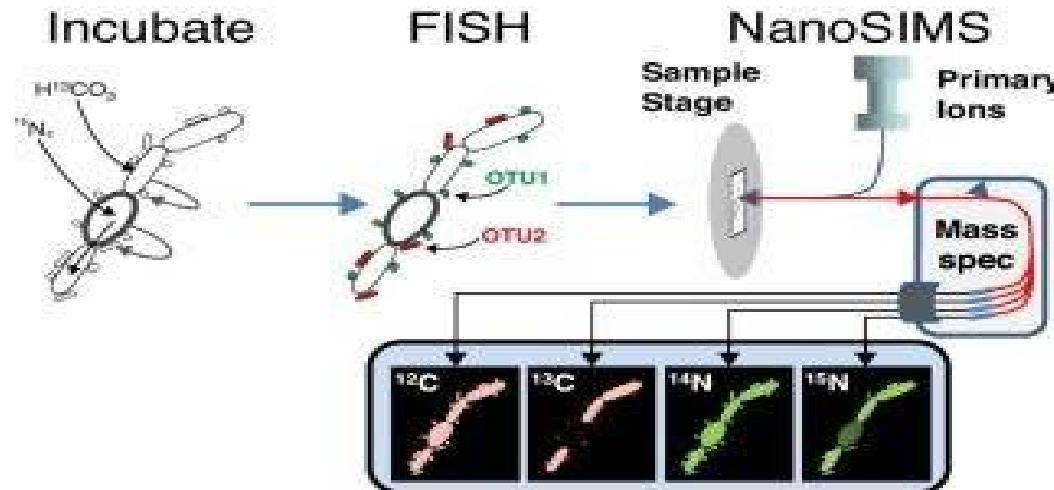




nano-SIMS

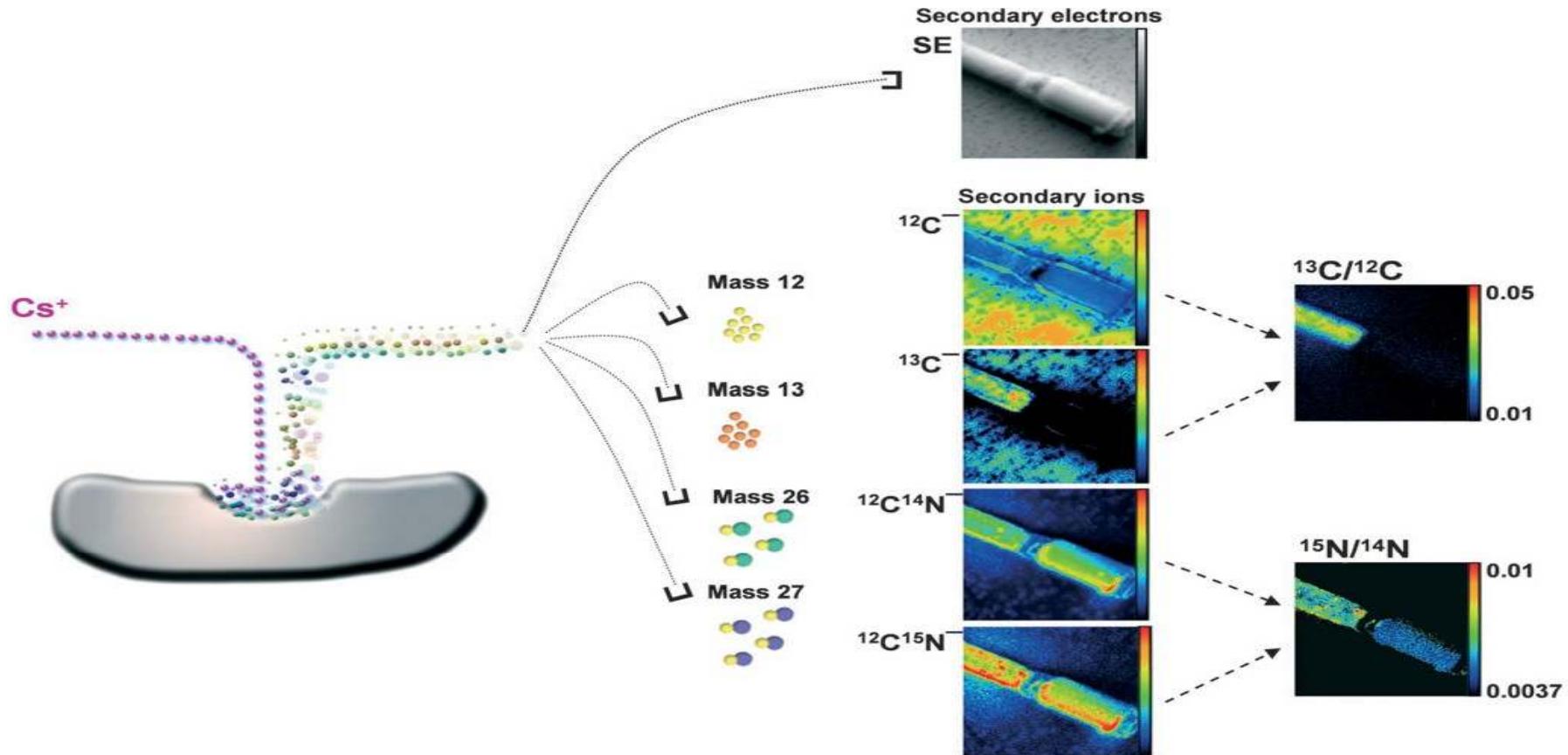
Nano-SIMS (*nanoscale Secondary Ion Emission Mass Spectrometry*) is a relatively recent technique that combines stable isotope incubations with FISH and spatially resolved mass spectrometry

A natural sample is *incubated* using selected substrates enriched in a stable isotope (for example ^{13}C CO_2). The cells are then *extracted* and *imaged* using a combination of epifluorescence microscopy and probed for their incorporation of the label





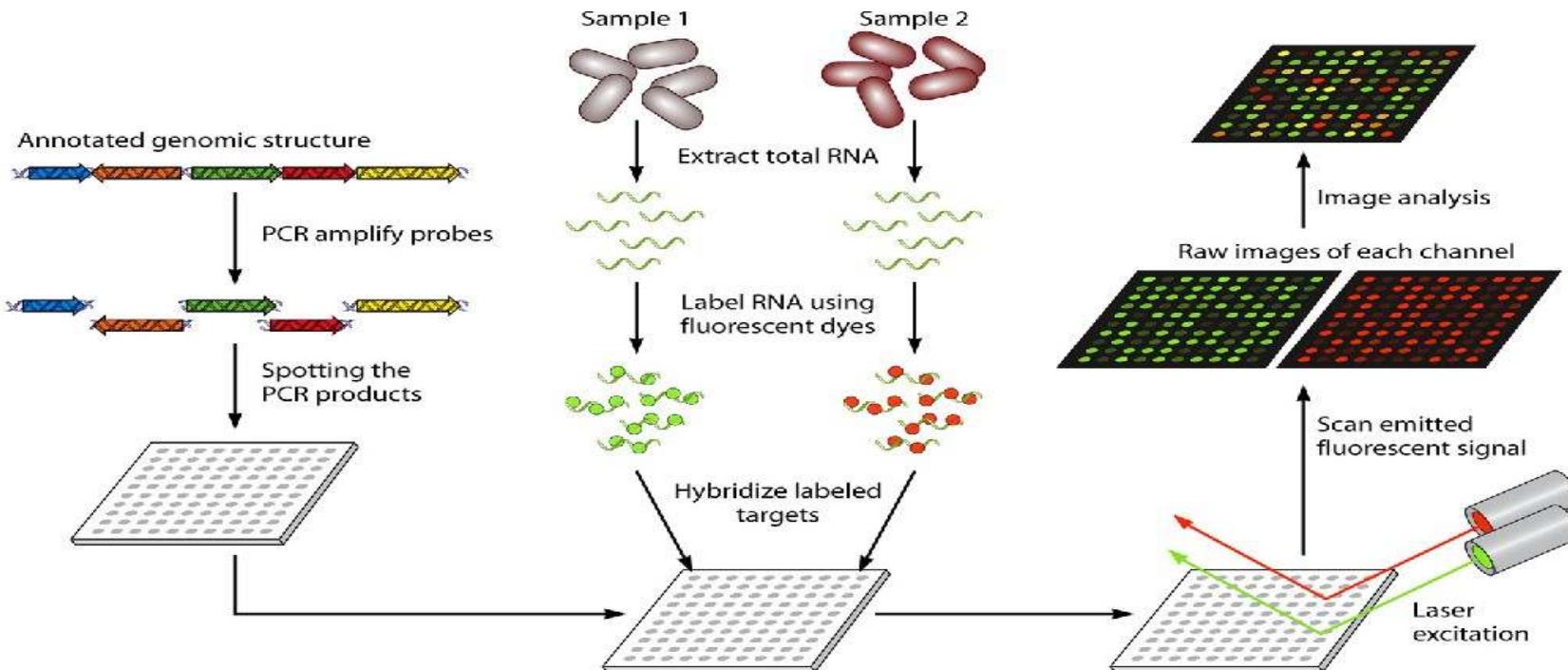
nano-SIMS





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**

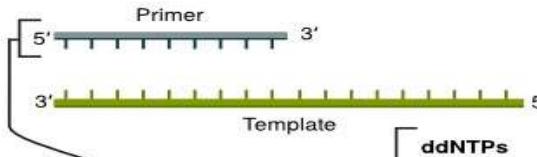
The choice of the type of sequencing to use is highly dependent on the study questions, and all the sequencing approaches have biases (like all the techniques used)

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

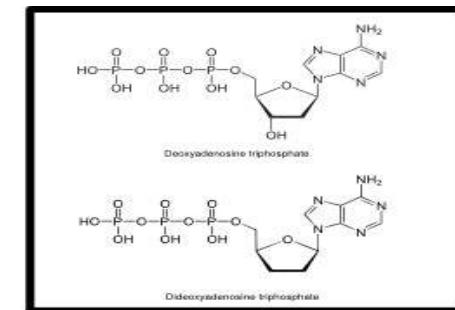
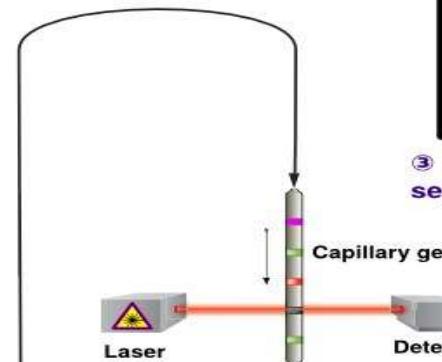
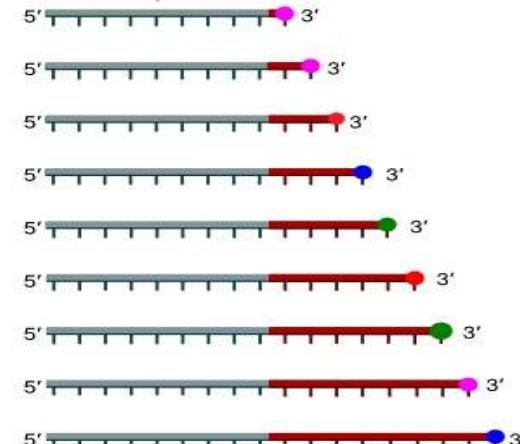
ABI Solid Dig Dye Terminator (Sanger)

① Reaction mixture

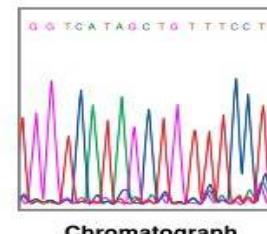
- Primer and DNA template
- DNA polymerase
- ddNTPs with fluorochromes
- dNTPs (dATP, dCTP, dGTP, and dTTP)



② Primer elongation and chain termination



③ Capillary gel electrophoresis separation of DNA fragments

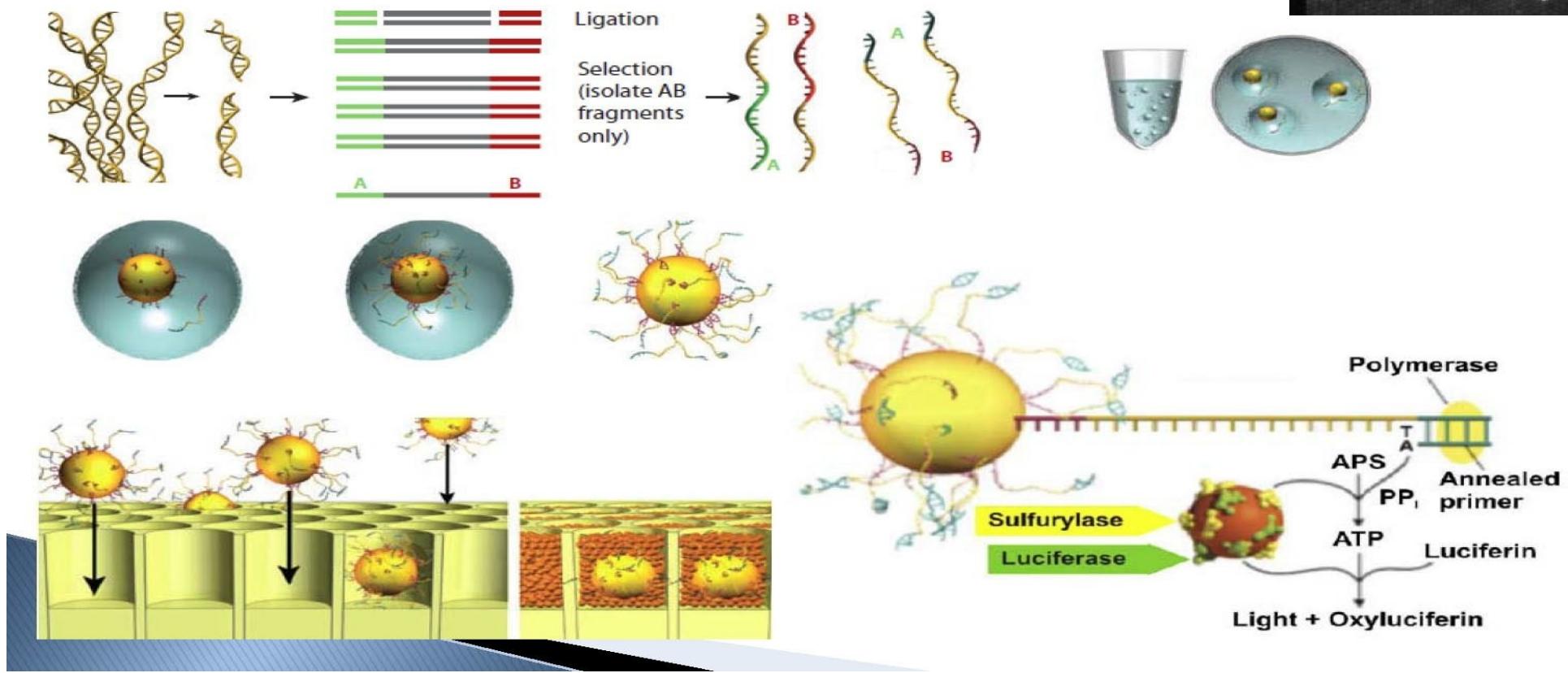
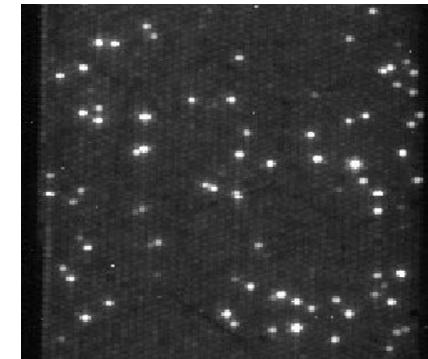


④ Laser detection of fluorochromes and computational sequence analysis

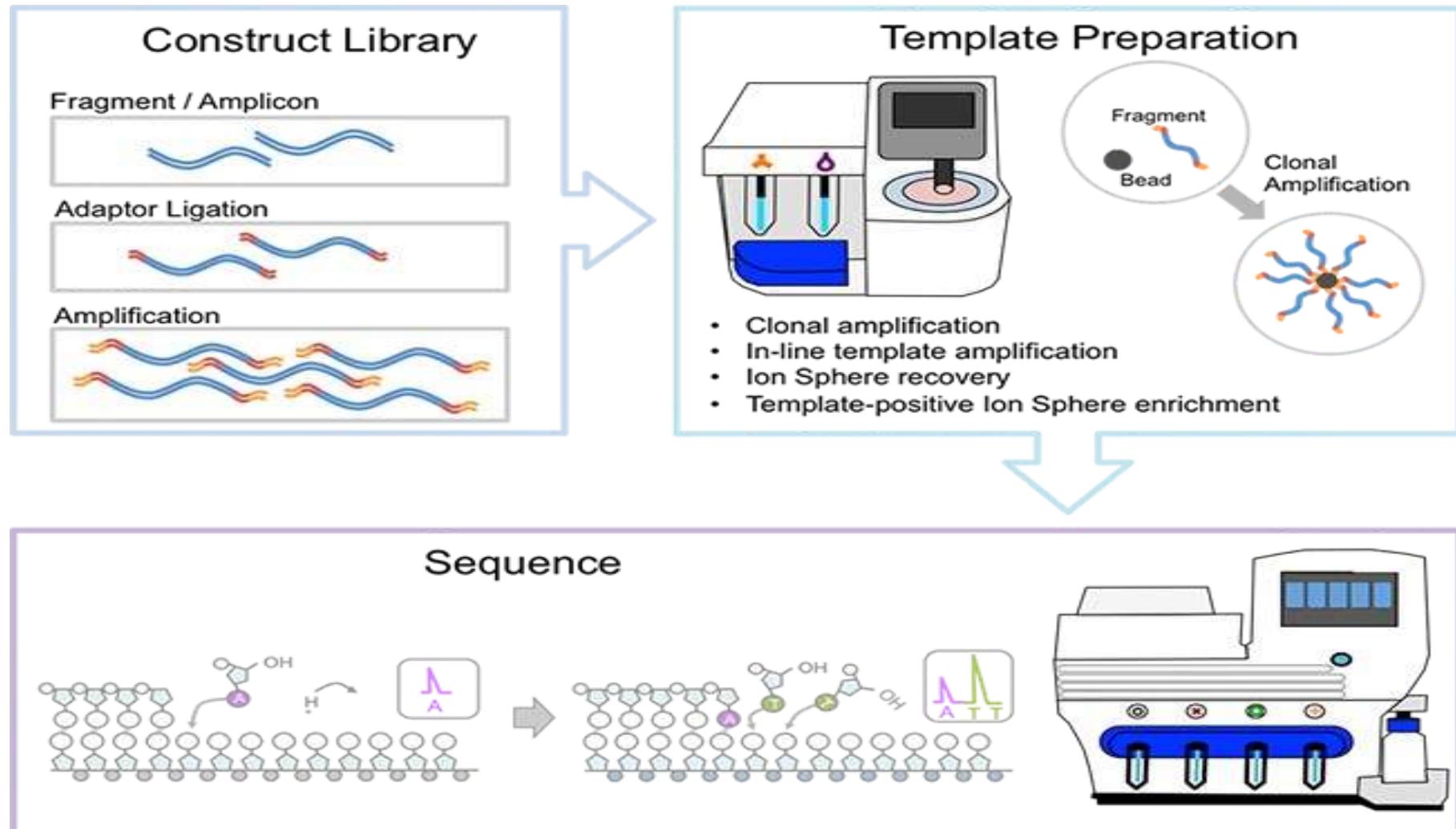
"Sanger-sequencing" by Estevezj - Own work.

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

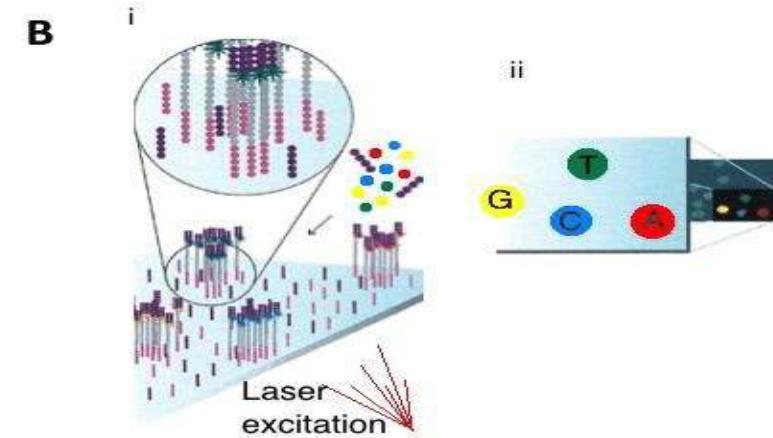
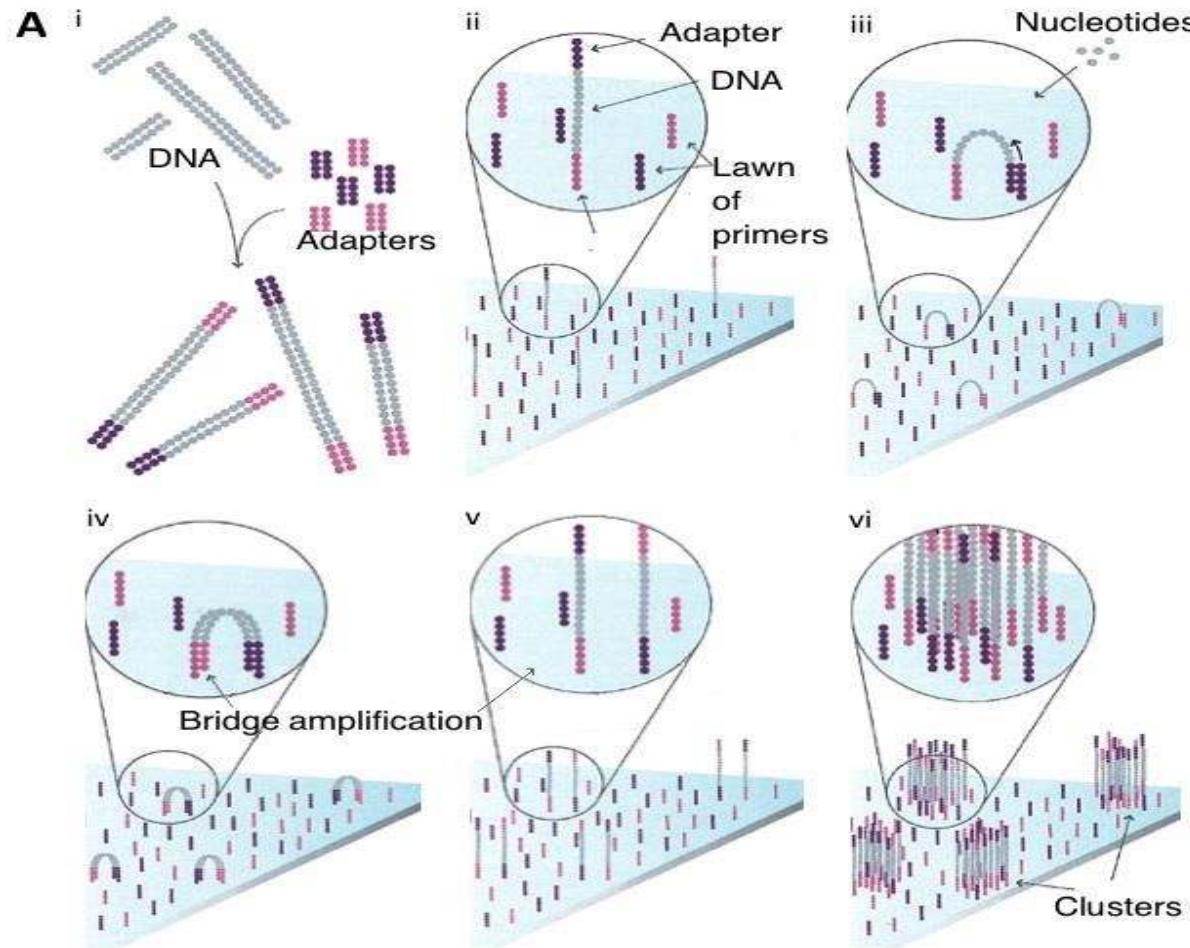
Pyrosequencing



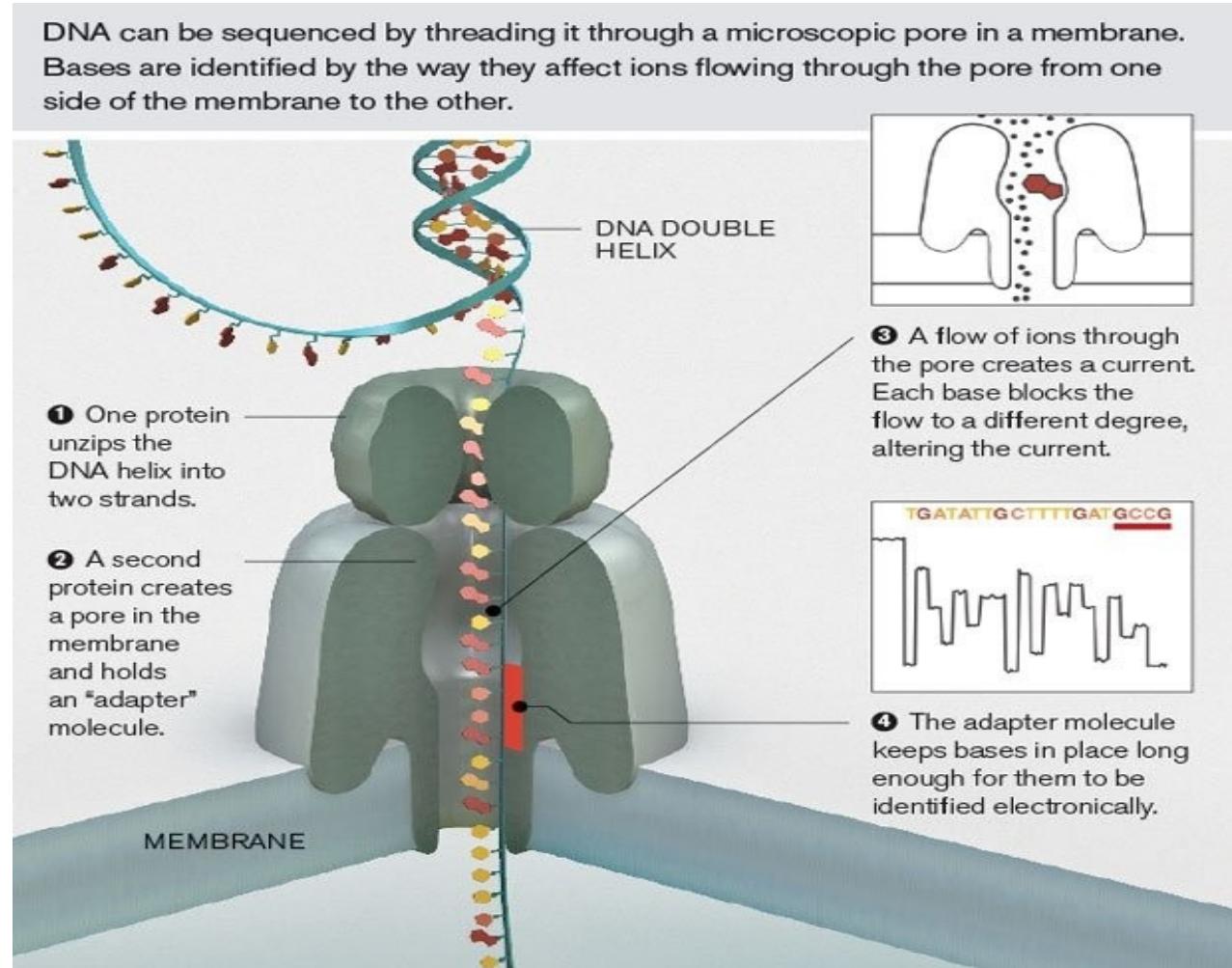
IonTorrent Sequencing



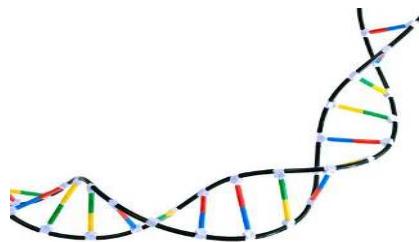
Illumina Sequencing



Nanopore Sequencing

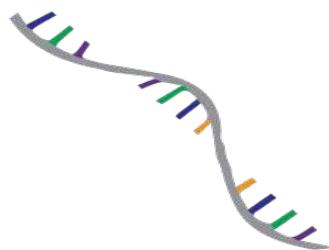


Information flow in biology



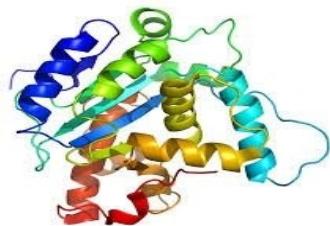
replication
DNA

GENES



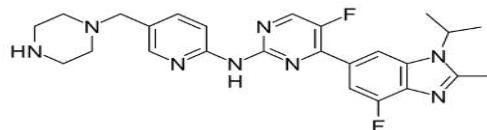
transcription
RNA

TRANSCRIPTS



translation
Protein

PROTEIN



metabolism
Metabolite

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

1998

2002

2003

2004

2005

2007



Mya Breitbart

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



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



Robert Edwards

published sequences generated using pyrosequencing techniques



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



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

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

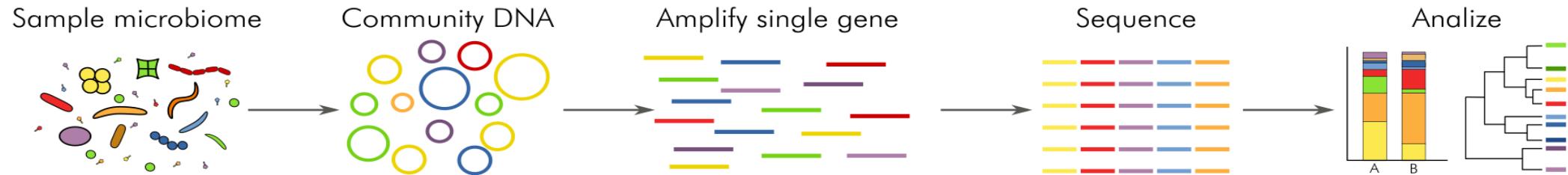


amplicon-sequencing 16S rRNA

Amplicon sequencing (sometime referred to as tag-amplicon sequencing) is a techniques that allows to sequence the 16S rRNA (or other target genes) in a high throughput way

It is de facto one of the most used techniques to investigate microbial taxonomic diversity. It is dependent on DNA amplification through PCR, so PCR primers (often designed on cultured microbes) and the PCR can introduce significant bias.

The tag suffix, refers to the ability to use known short DNA sequences (tags) to recognize sequences coming from different samples at the data analysis stage



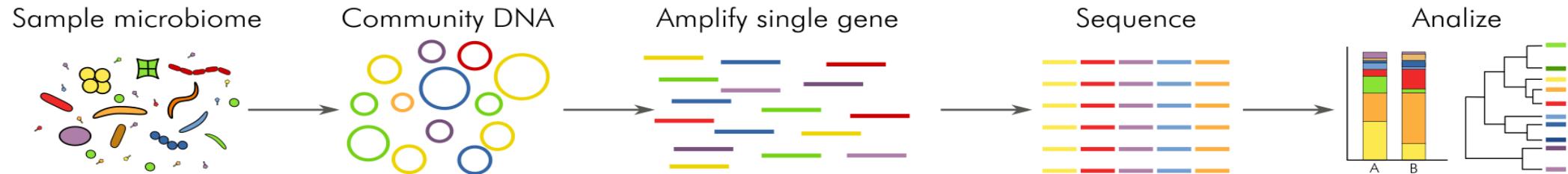


amplicon-sequencing

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



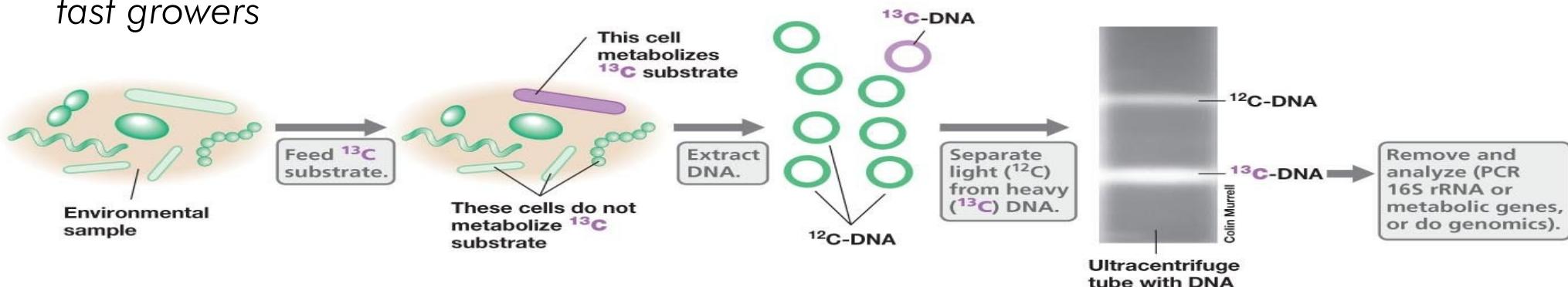


Stable Isotope Probing (SIP)

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





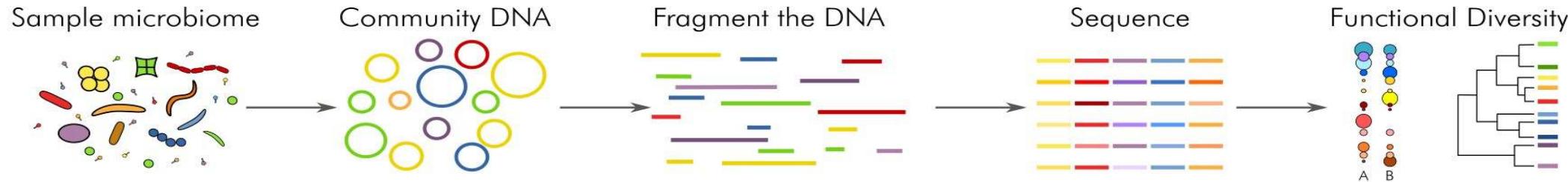
Metagenomic

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

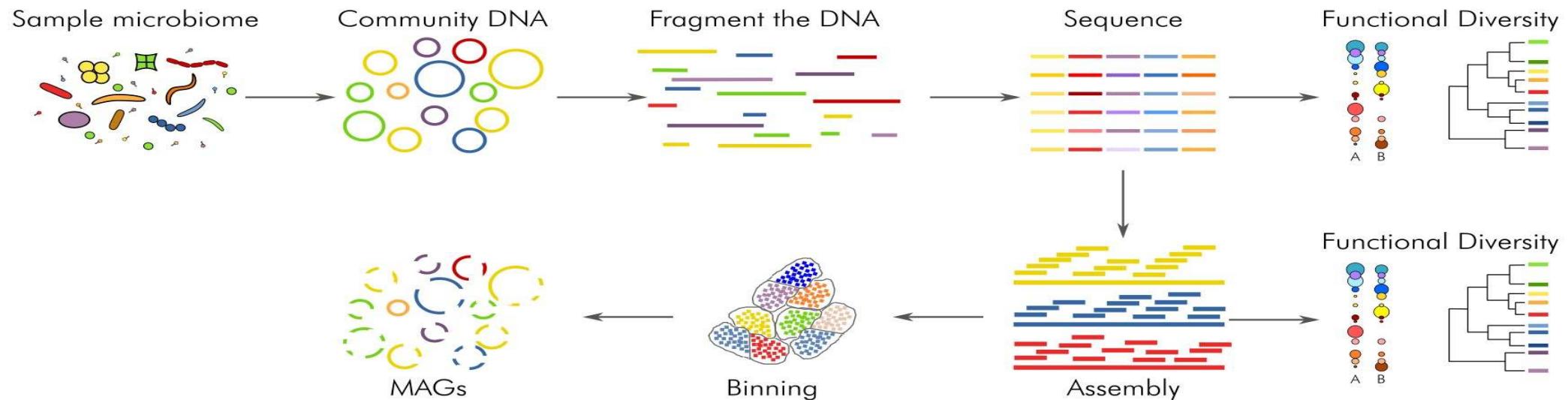




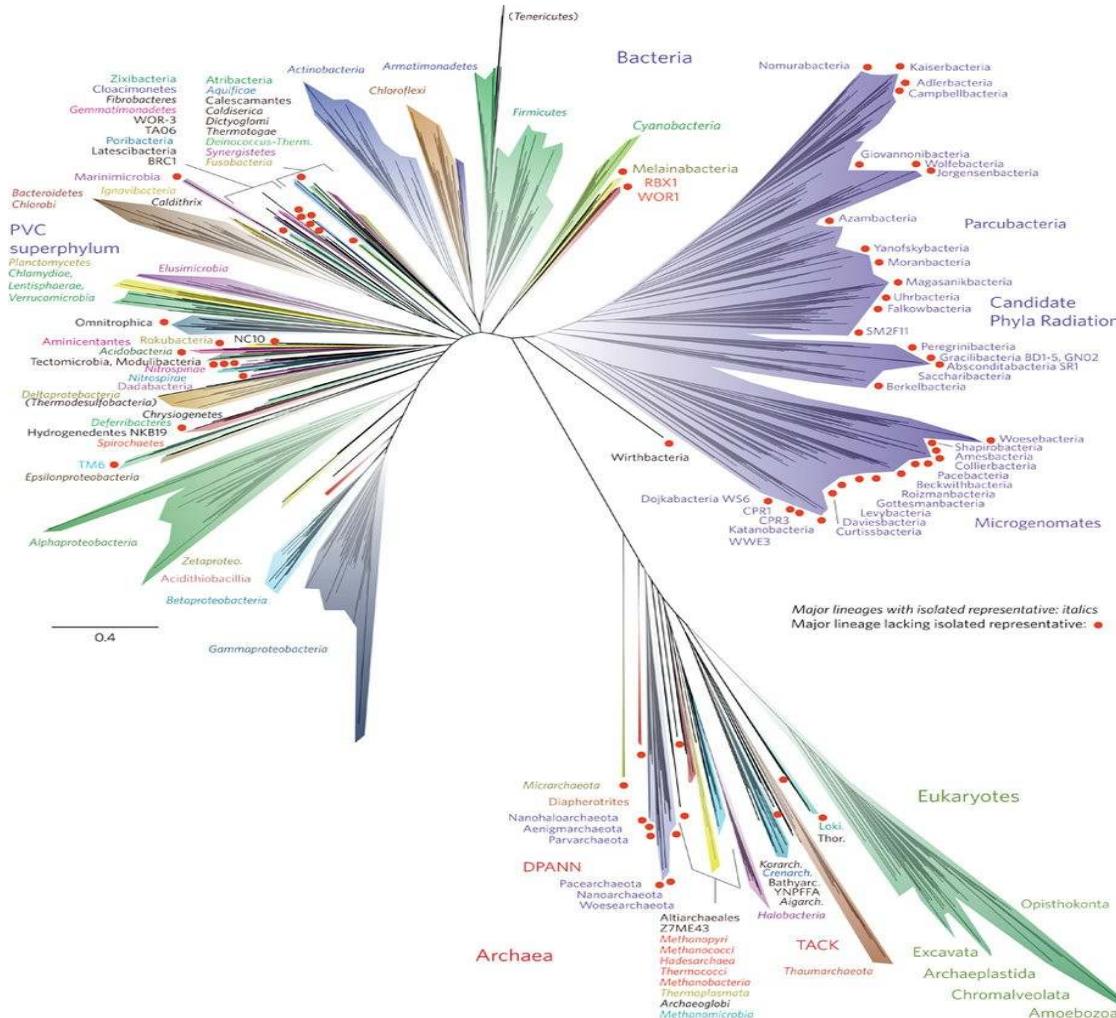
Metagenomic – MAGs

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



Metametabolomic

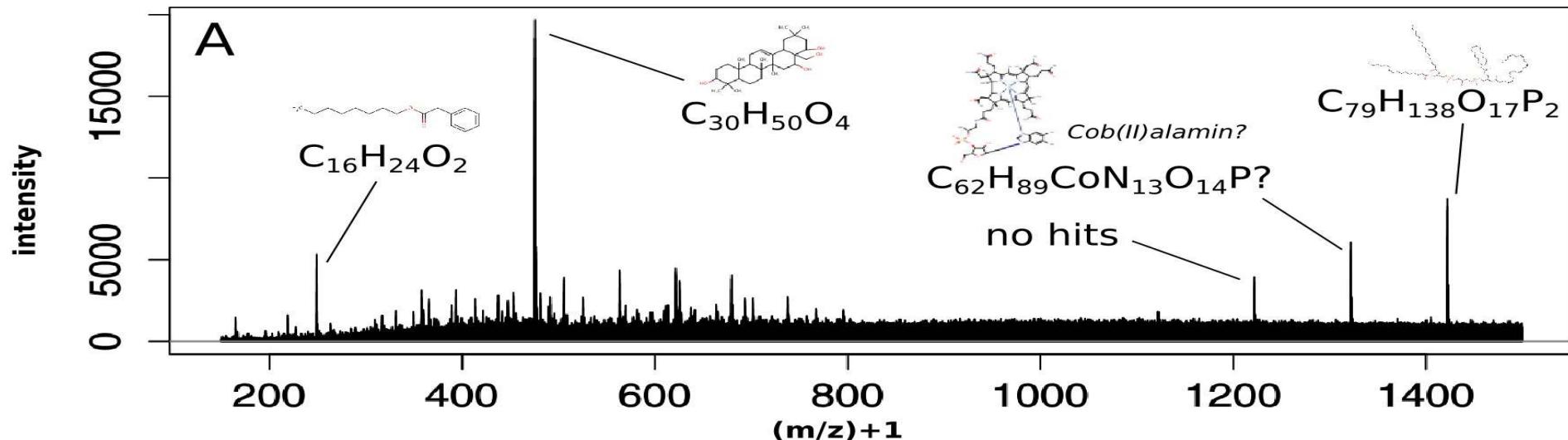
Metametabolomics refers to the use of different techniques (like MS or NMR) to identify the metabolites from an environmental sample

Metametabolomic give access to the metabolic products of a community

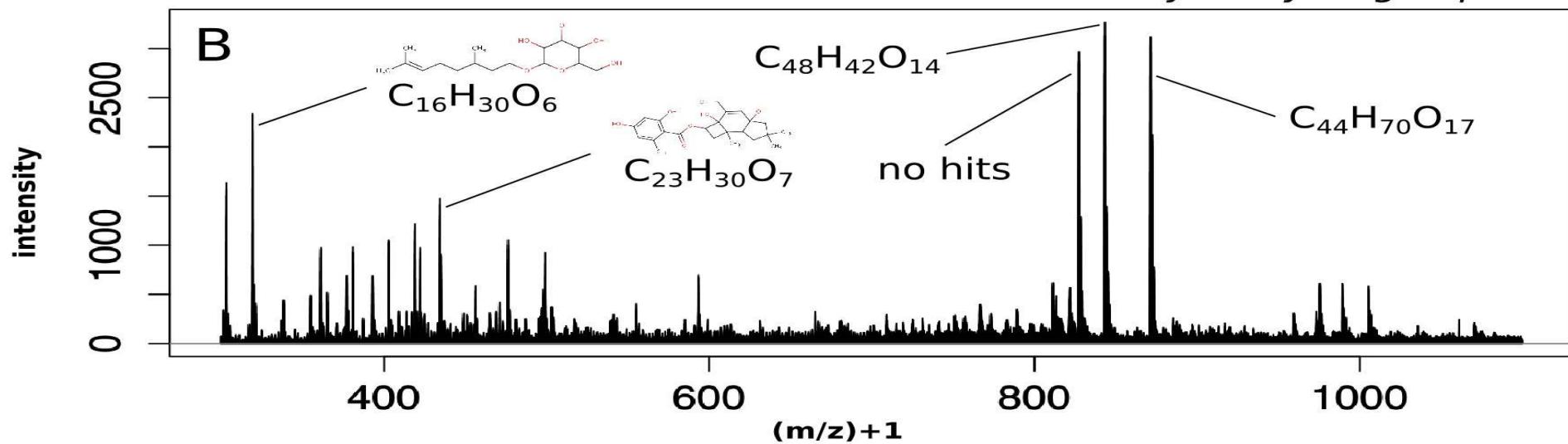
It is extremely challenging, as a single techniques cannot encompass the molecular diversity of metabolites, it is difficult to identify extracellular metabolites part of dissolved organic matter from intracellular metabolites and a single mass can refer to multiple compounds

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

Themovibrio ammonificans



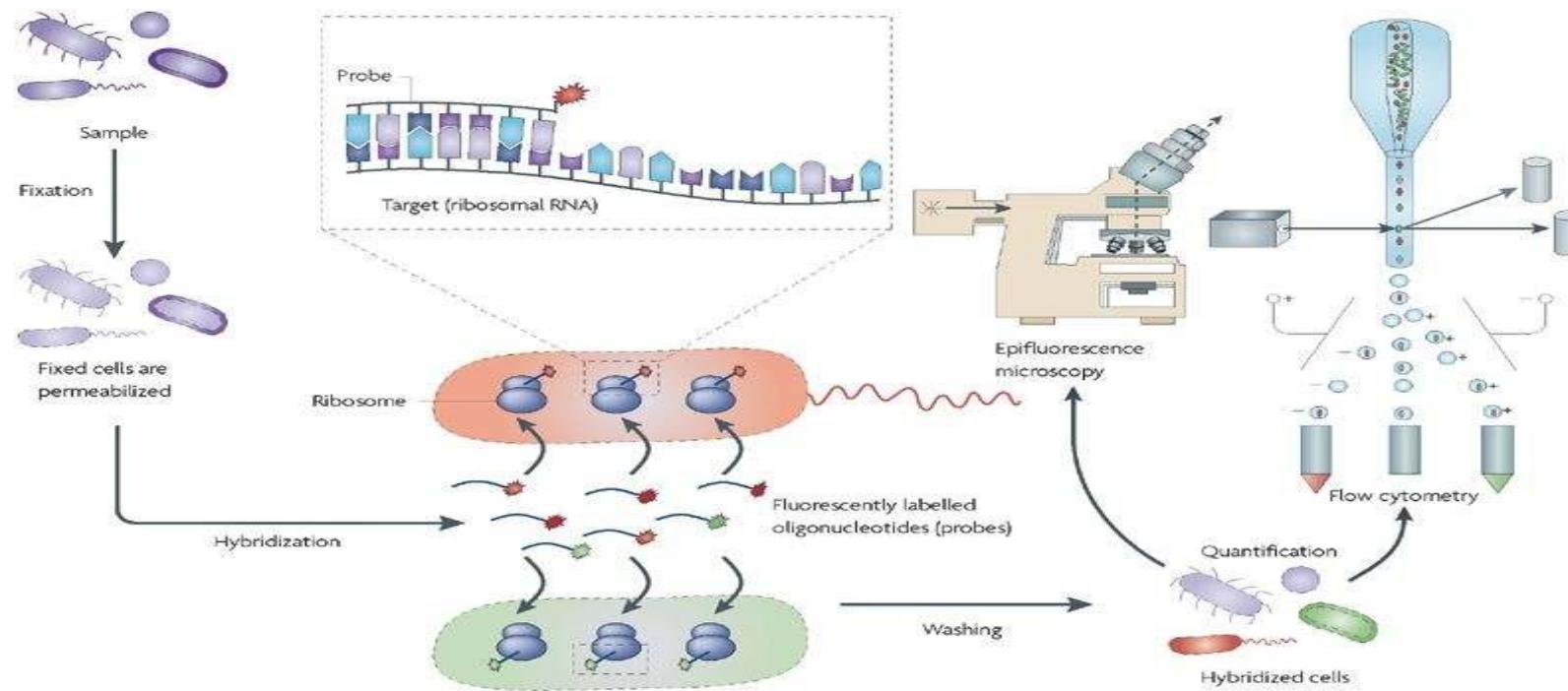
Phorcysia hydrogeniphila





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

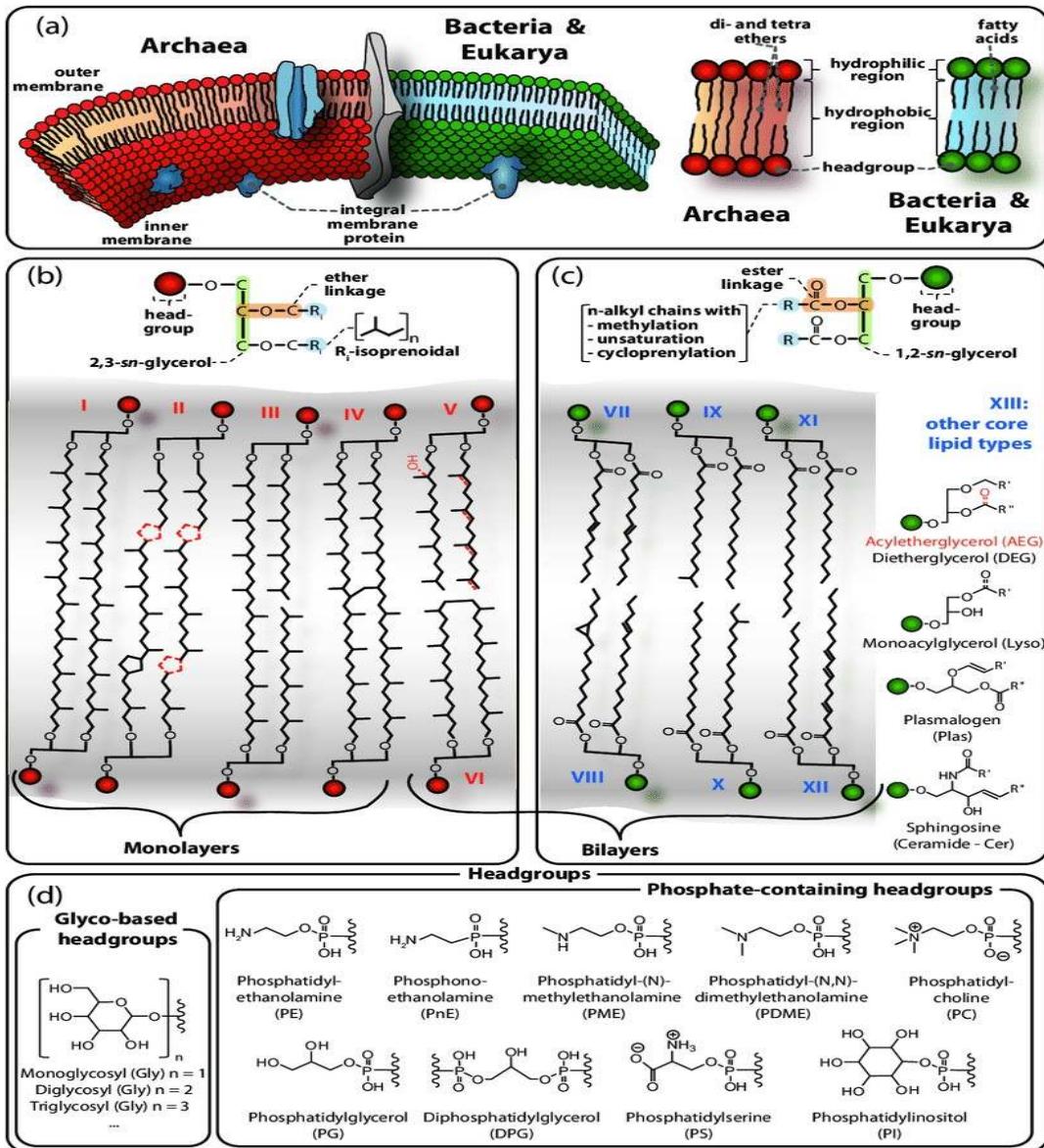




Lipids Biomarkers

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".

Thought exercise



