



University of Naples "Federico II"

Marine Microbial Diversity

Methods in Marine Microbial Diversity 2



*General questions in
studying microbial diversity*

Milos shallow water hydrothermal vent biofilm



General questions in studying microbial diversity

Who's there



What are they doing



Who's doing what?



How are they doing it?



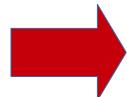
How much?



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.

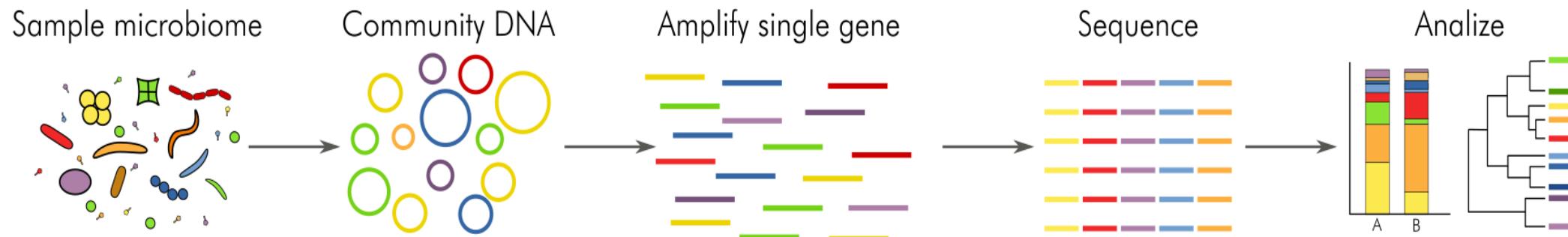


amplicon-sequencing 16S rRNA

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





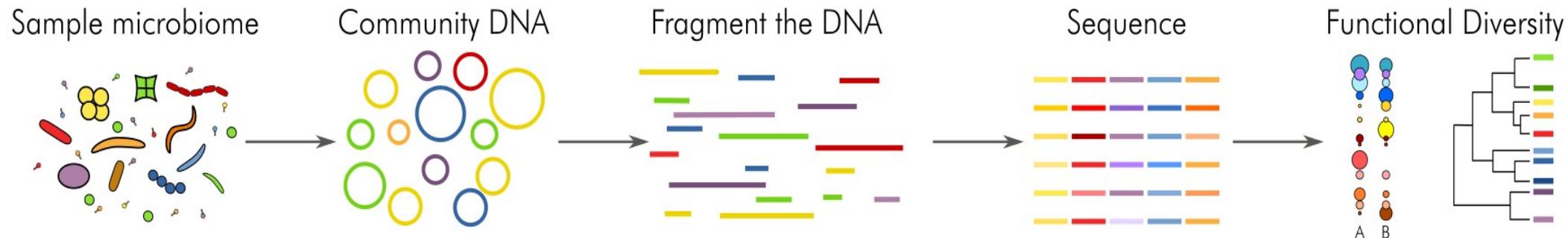
Metagenomic

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

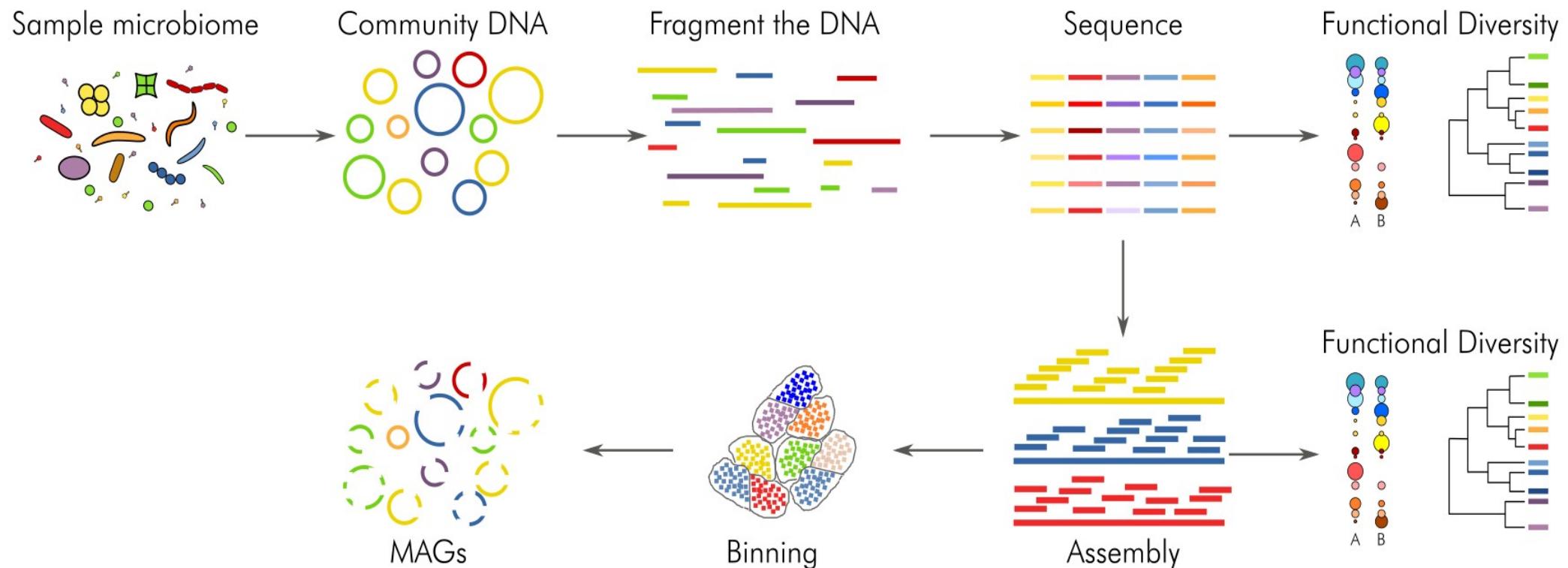




Metagenomic – MAGs

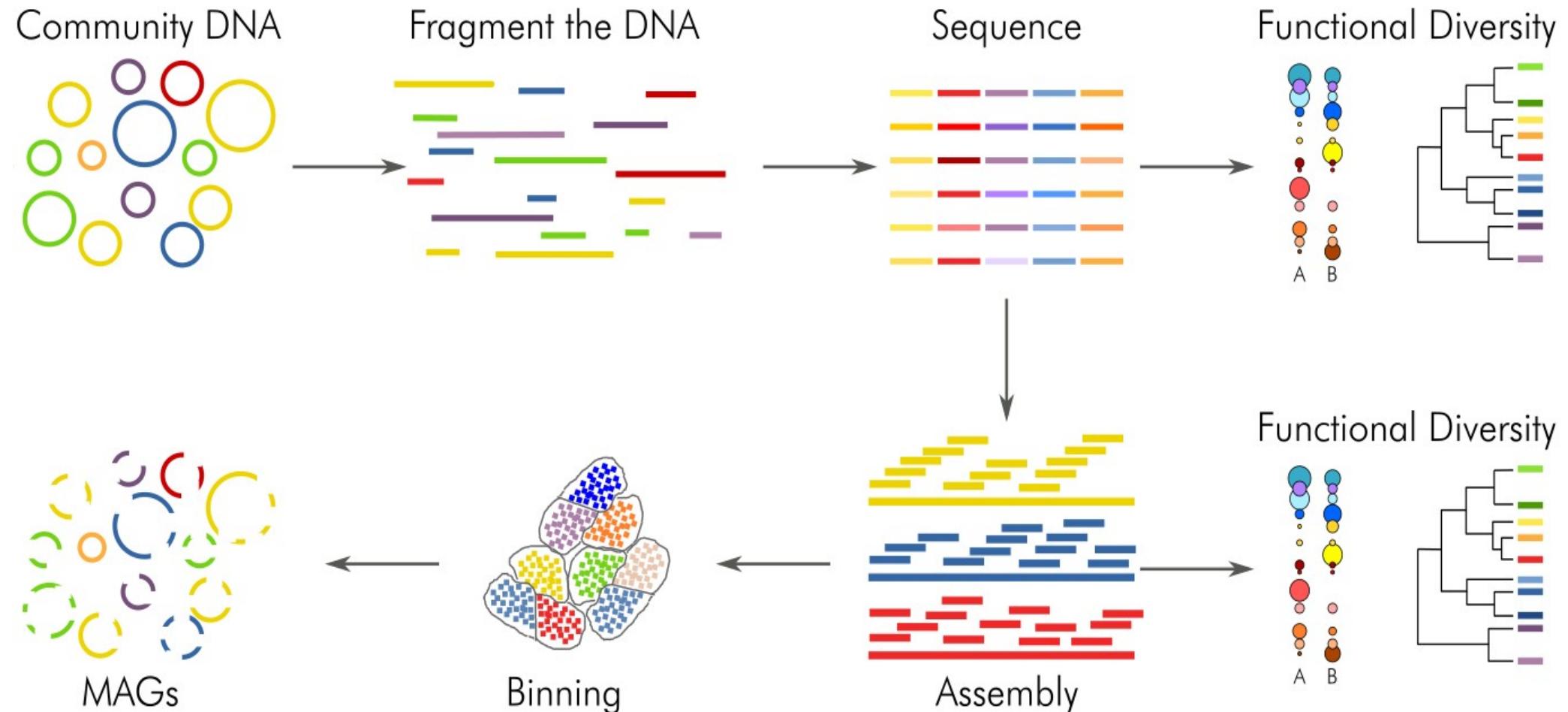
One of the main downstream possibility given by metagenomic is the reconstruction of complete genomes from the environmental sample

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**



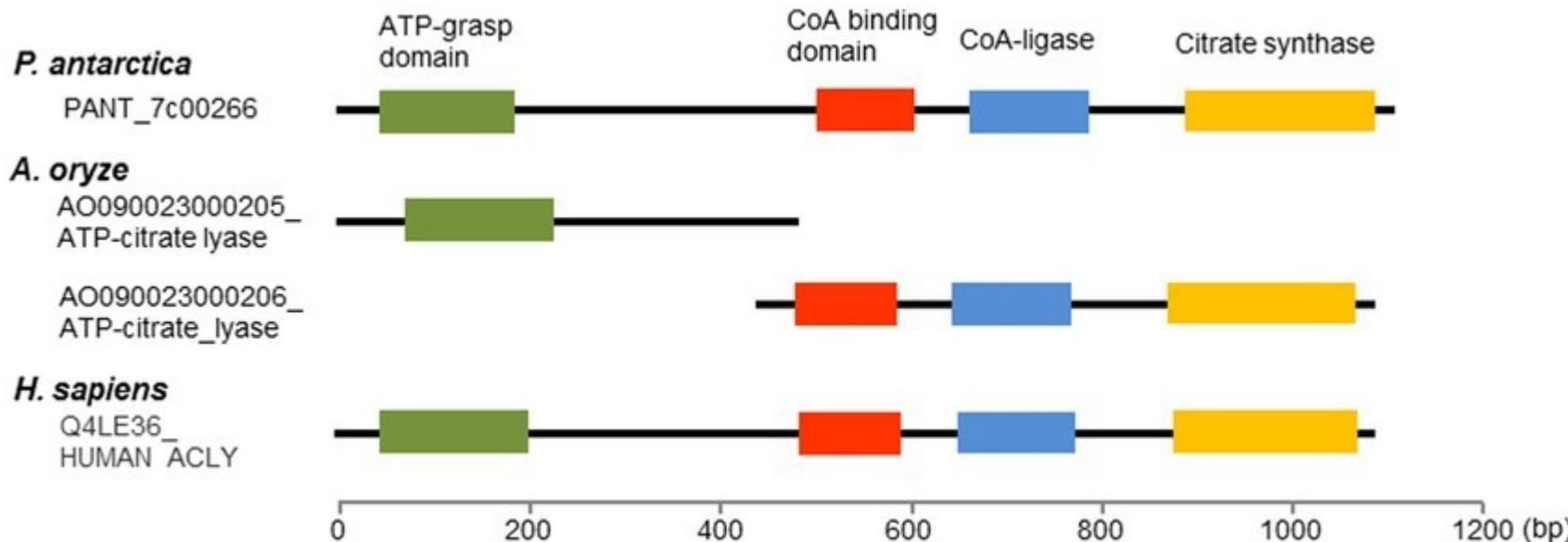
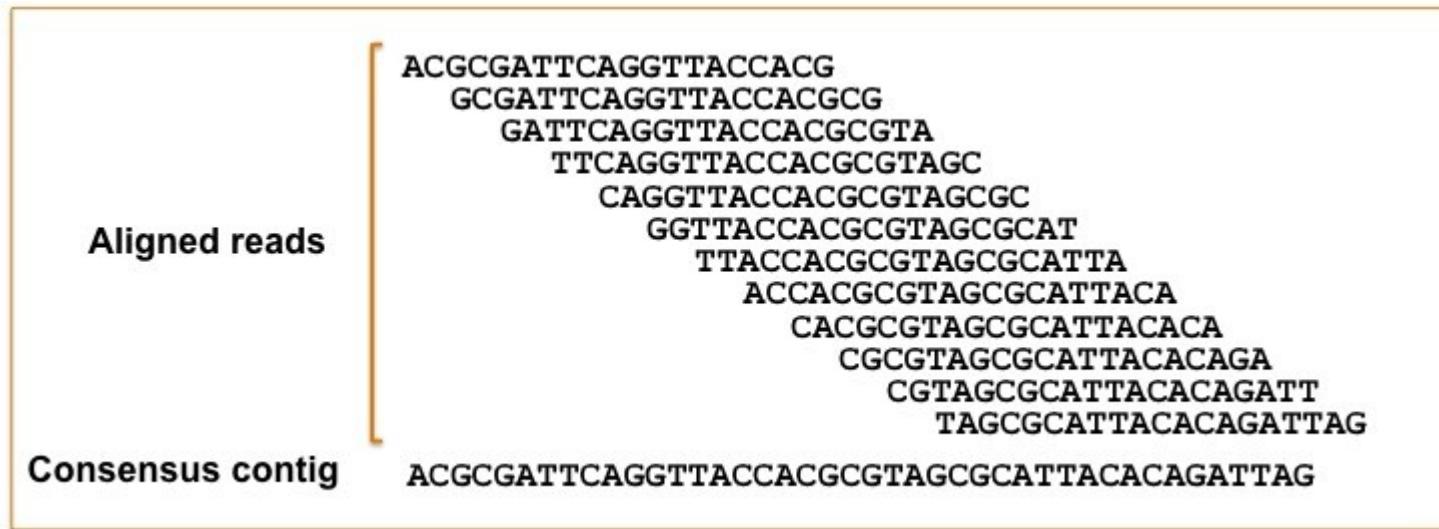


Metagenomic – MAGs





Metagenomic – MAGs

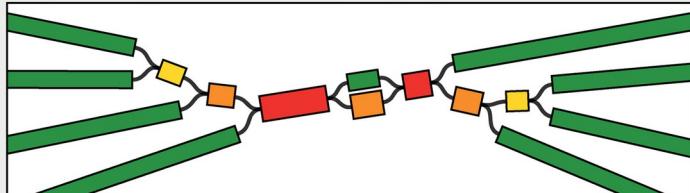




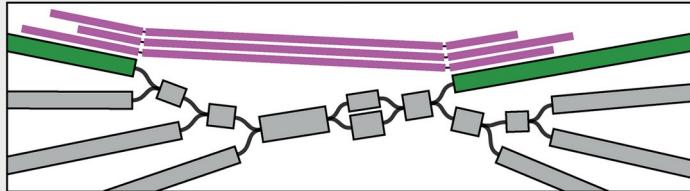
Metagenomic – MAGs

D. Long read bridging

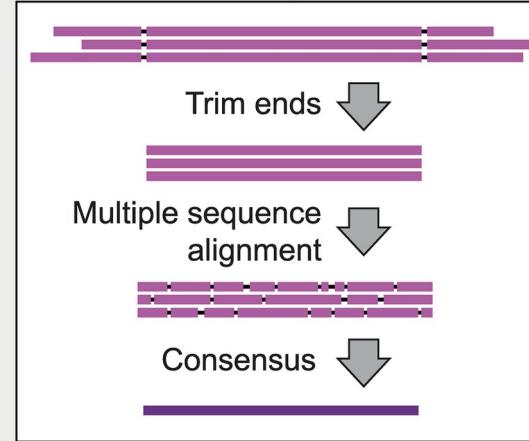
Repeat region in unbridged graph



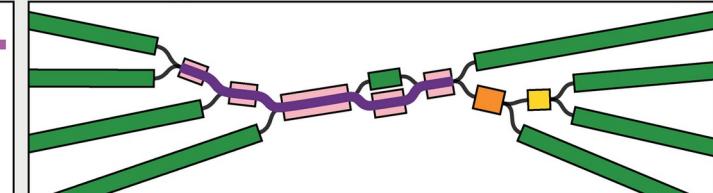
Semi-global long read alignment



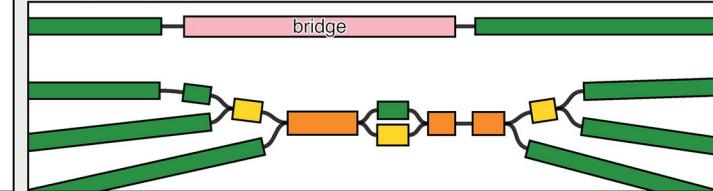
Consensus read sequence



Path finding

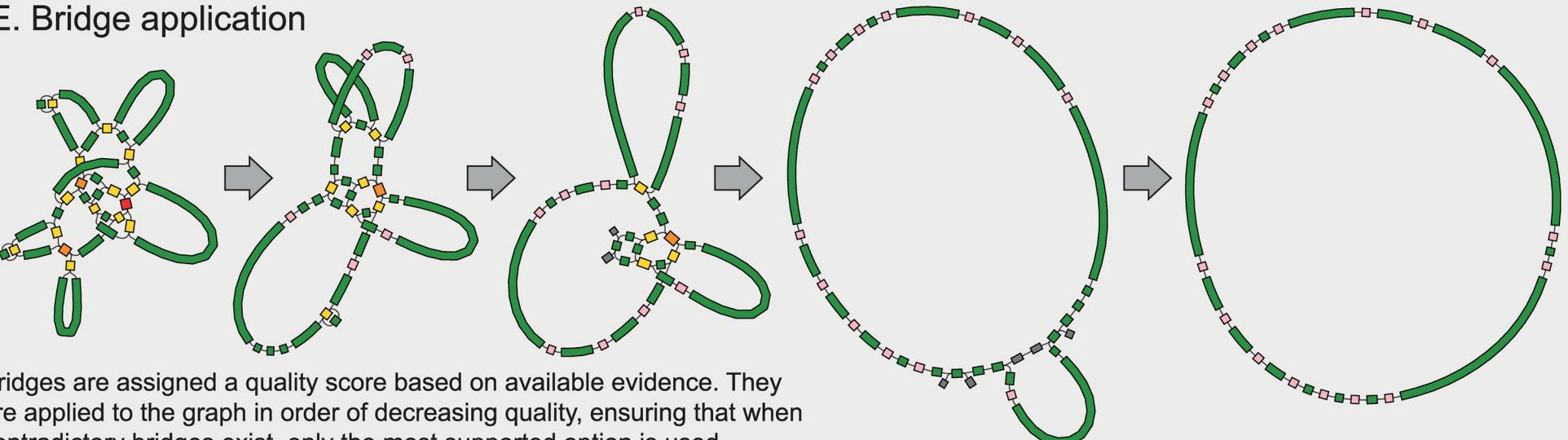


Bridged graph



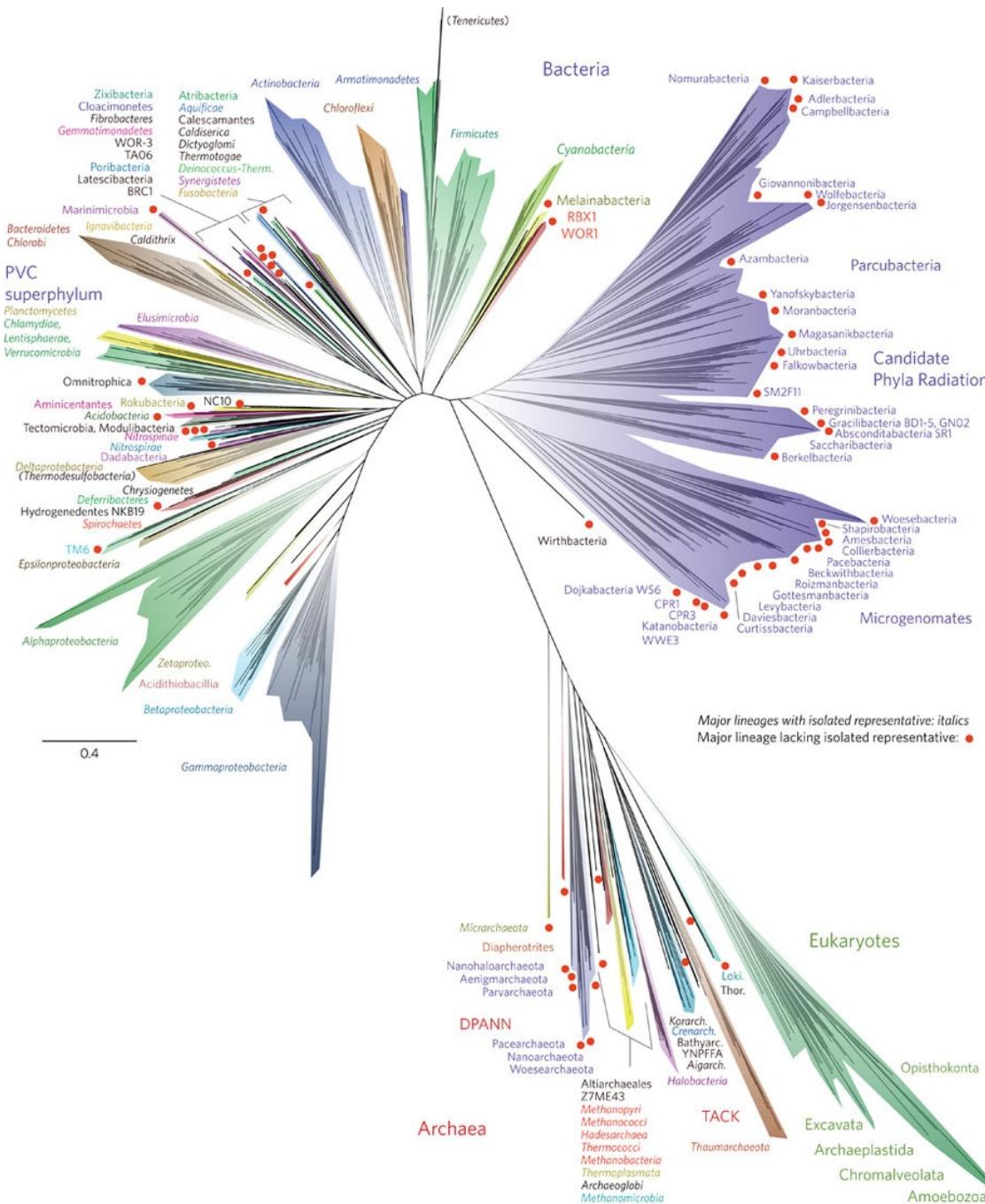
Bridges made using long reads can resolve larger repeats than short-read bridges. They are made from long reads which align to two or more single-copy contigs. The bridge sequence comes from the graph path between the two contigs, not the long reads, providing greater accuracy. When multiple possible bridge paths exist, the best path is chosen based on agreement with the long-read consensus sequence.

E. Bridge application



Bridges are assigned a quality score based on available evidence. They are applied to the graph in order of decreasing quality, ensuring that when contradictory bridges exist, only the most supported option is used.

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

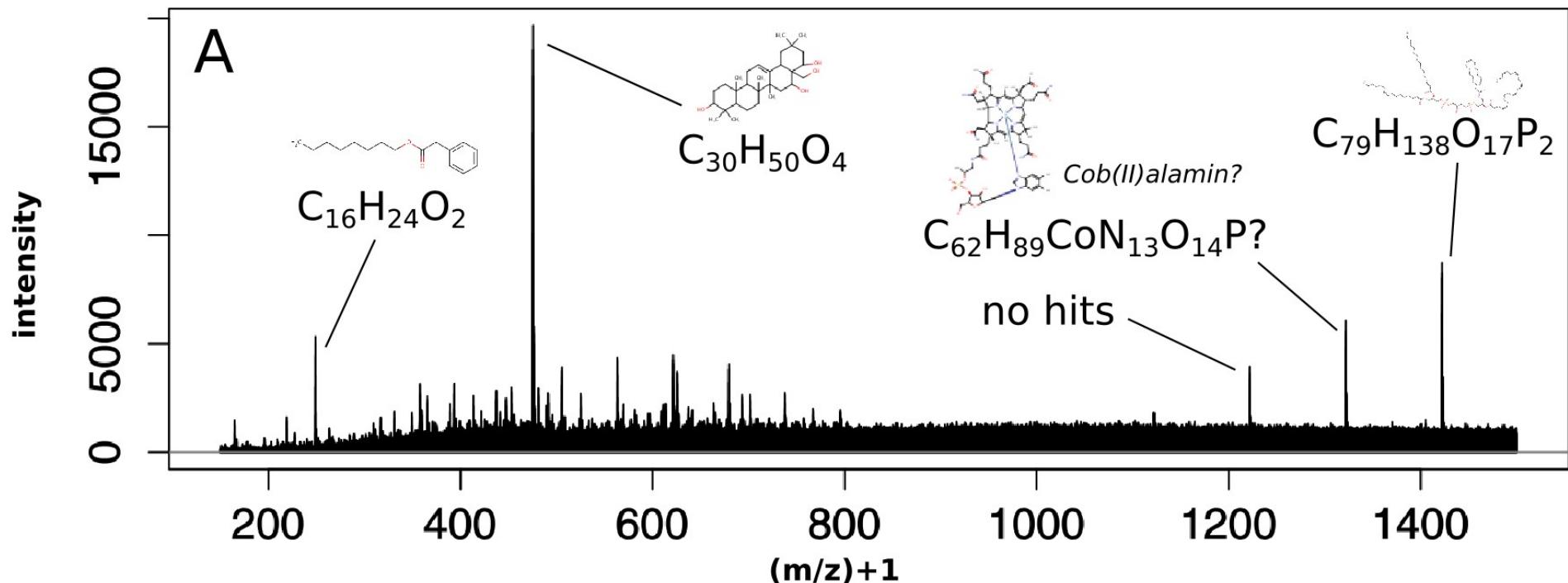
Metametabolomics refers to the use of different techniques (like MS or NMR) to identify the metabolites from an environmental sample

Metametabolomic give access to the metabolic products of a community

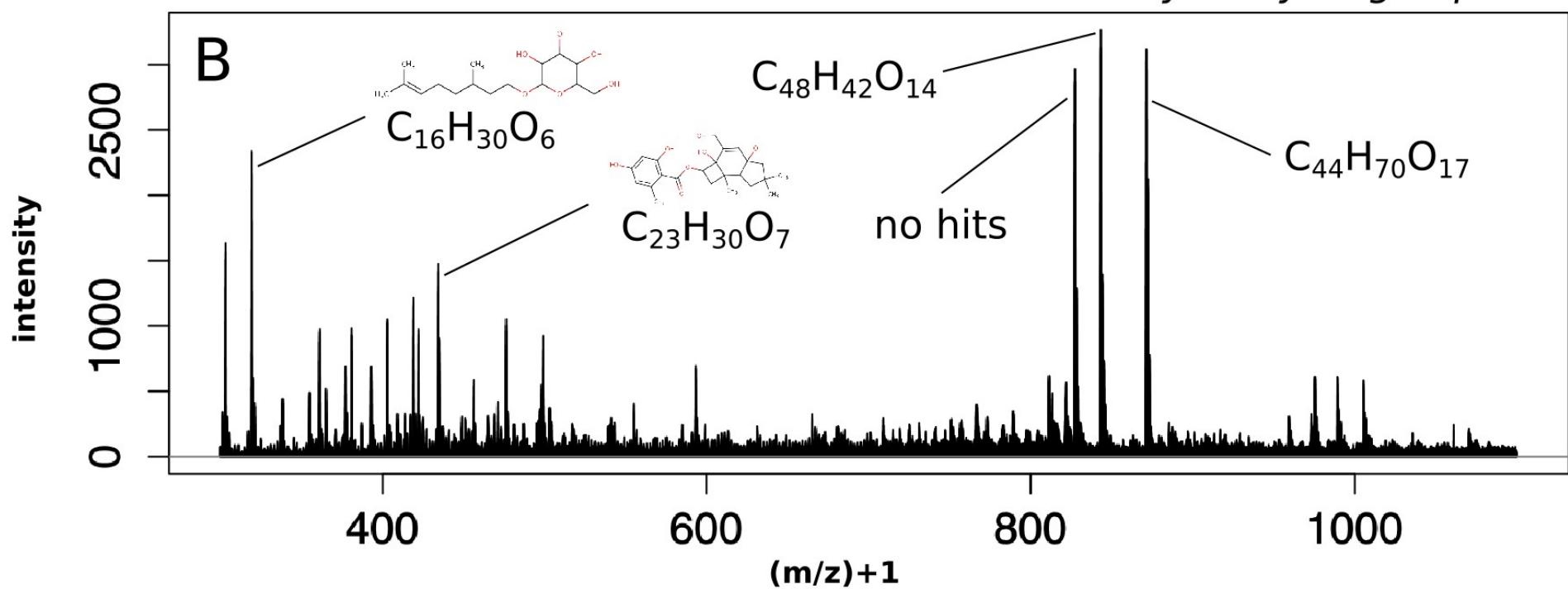
It is extremely challenging, as a single techniques cannot encompass the molecular diversity of metabolites, it is difficult to identify extracellular metabolites part of dissolved organic matter from intracellular metabolites and a single mass can refer to multiple compounds

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

Themovibrio ammonificans



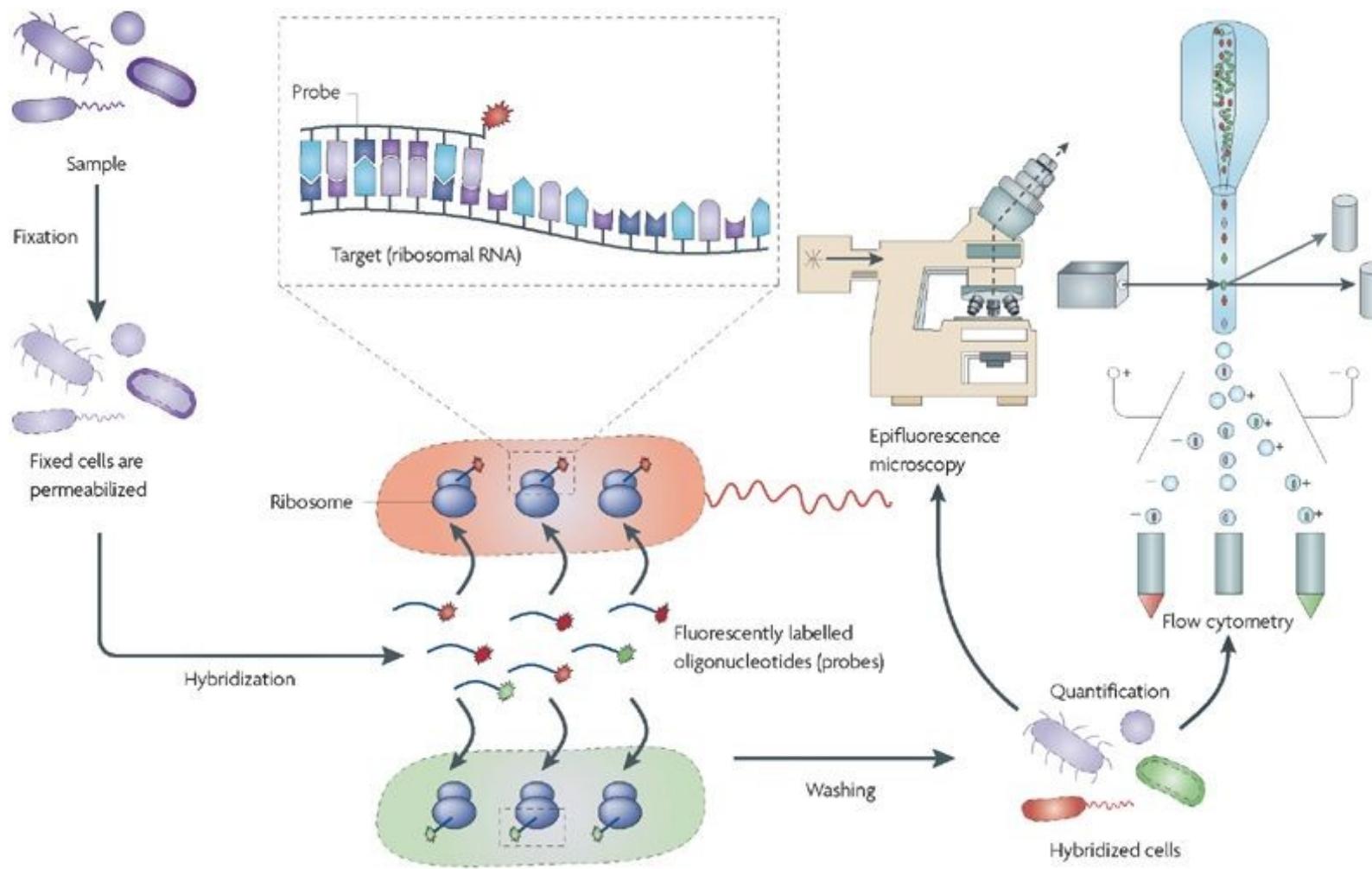
Phorcysia hydrogeniphila





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

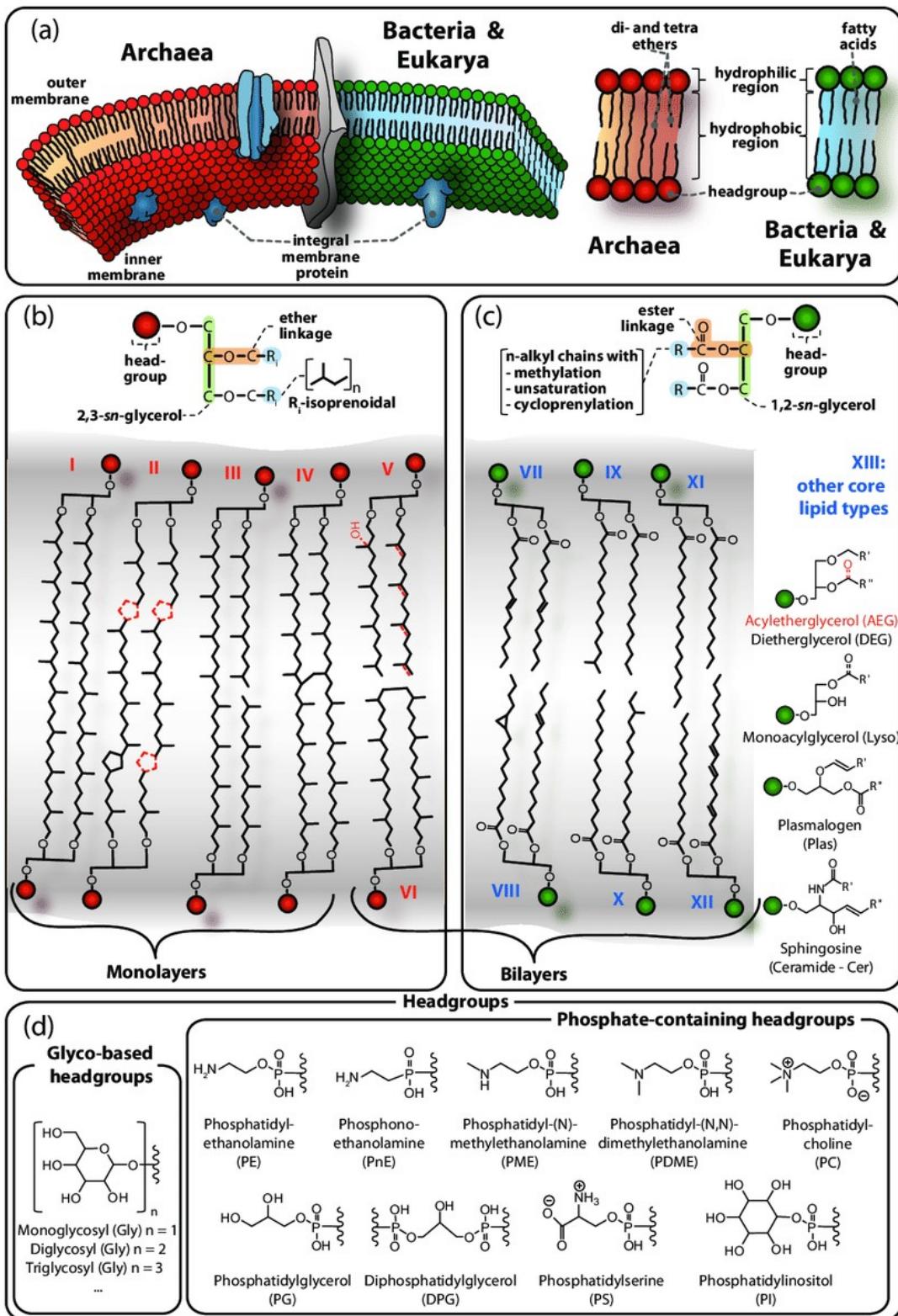




Lipids Biomarkers

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



