MULTI-OMICS, THE LAST PARADIGM SHIFT IN MICROBIOLOGY

	Technological innovation	Paradigm shifts
1670	first microscope (Anthony van Leuwenhook)	access to the invisible world
1857	cultivation based approaches	Koch's explanation of the origin of the human diseases, concept of pathogenicity
1888	Winogradsky column	beginning of microbial ecology, the overwhelming majority of microbes are essential for ecosystem functioning
1983	molecular microbiology (PCR - FISH -Sanger seq.)	culture-independent microbiology
2010	multi-omics: metagenomics, metatrascriptomic, metabolomic (NGS, NMR, LC and GC-MS)	microorganisms occurs in complex assemblages in with species interaction are critical for dynamics and functional activities

UNTARGET CULTURE INDIPENDENT MICROBIOLOGY, THE DARK SIDE OF THE MOON

only minimal fraction (~2-10%) of the planet microbial diversity is cultivable

by studying the total microbial DNA in a sample, metagenomics revolutionized microbiology, shedding light on the total microbial diversity living on our planet

THE MILTI-OMICS ERA

NEW OMICS THECNOLOGIES REVOLITIONIZED MICROBIAL ECOLOGY

PROVIDING STRUCTURE AND FUNCTIONAL POTENTIAL OF WHOLE MICROBIAL COMMUNITIES IN THEIR NATURAL HABITATS

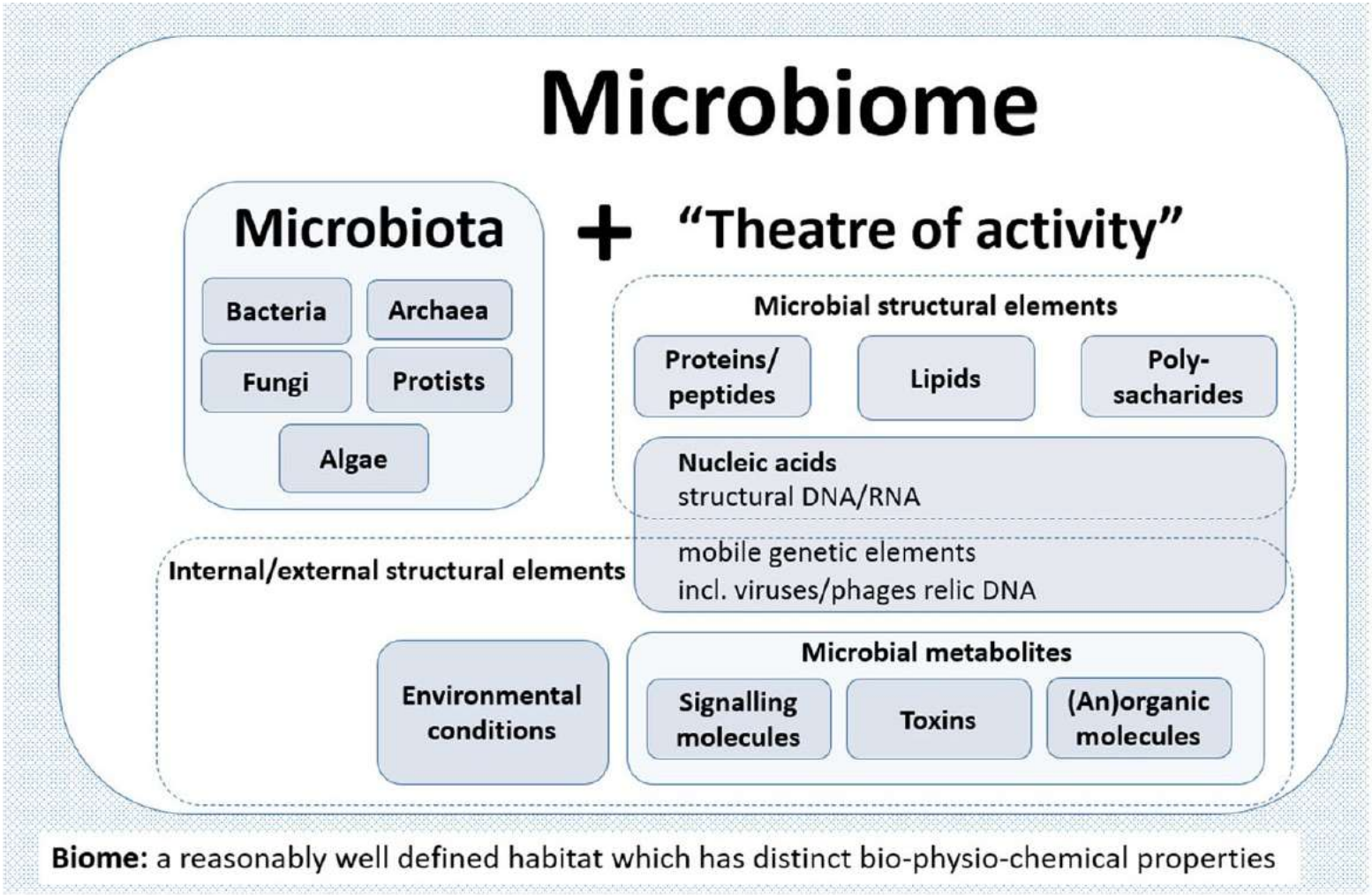
NEW ACTORS AND NEW DEFINITIONS

MICROBIOTA
all living microbes populating a given habitat
(bacteria, archea, fungi)

MICROBIOME (Whipps 1988)
characteristic microbial community in a well-defined habitat which has distinct physiochemical properties as their "theatre of activity"

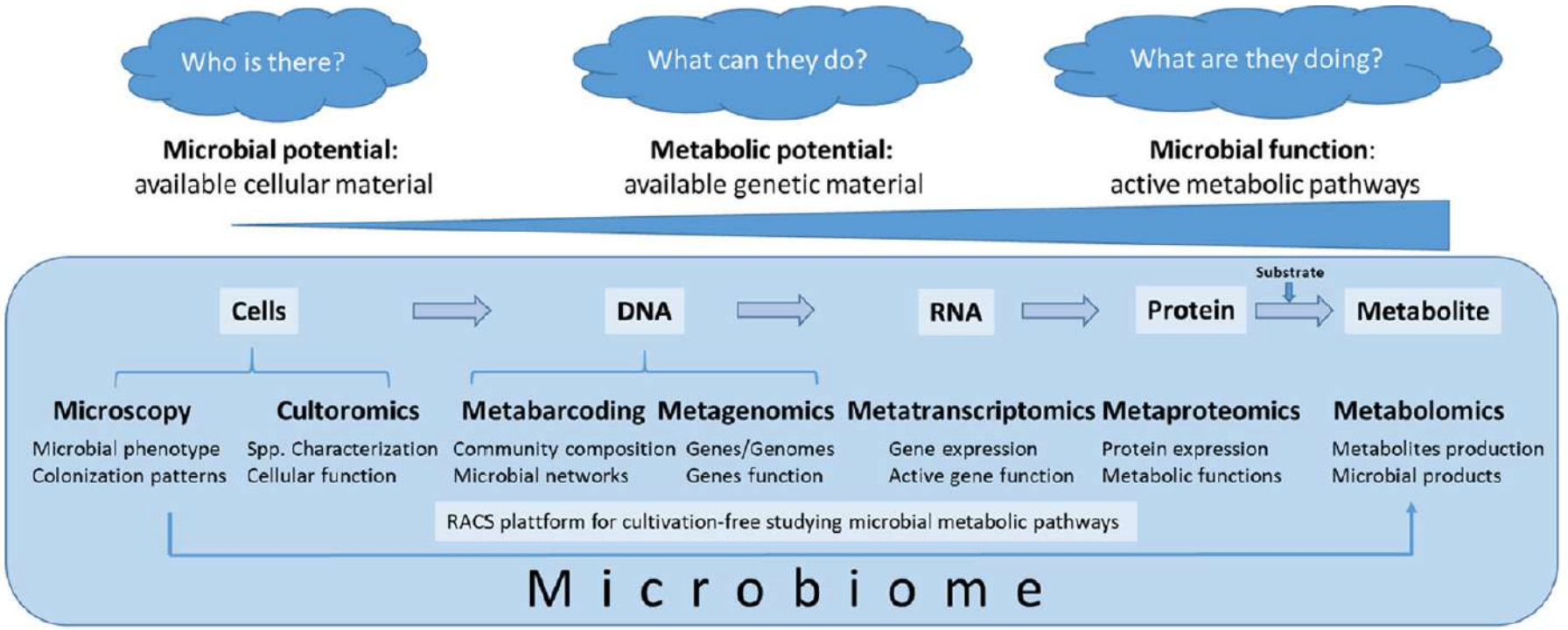
METAGENOME
collection of genomes and genes from the member of the microbiota

THEATRE OF ACTIVITY



MULTI-OMIC AND MICROBIOME ASSESMENT

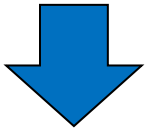
METAGENOMICS IS CENTRAL IN THE WORKFLOW FOR PRODUCING MICROBOME KNOWLEDGE



NON-OMICS METADATA, TO COMPLETE THE THEATRE OF ACTIVITY

ENVIRONMENTAL MICROBIOMES

HOST ASSOCIATED MICROBIOMES

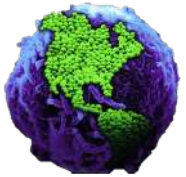


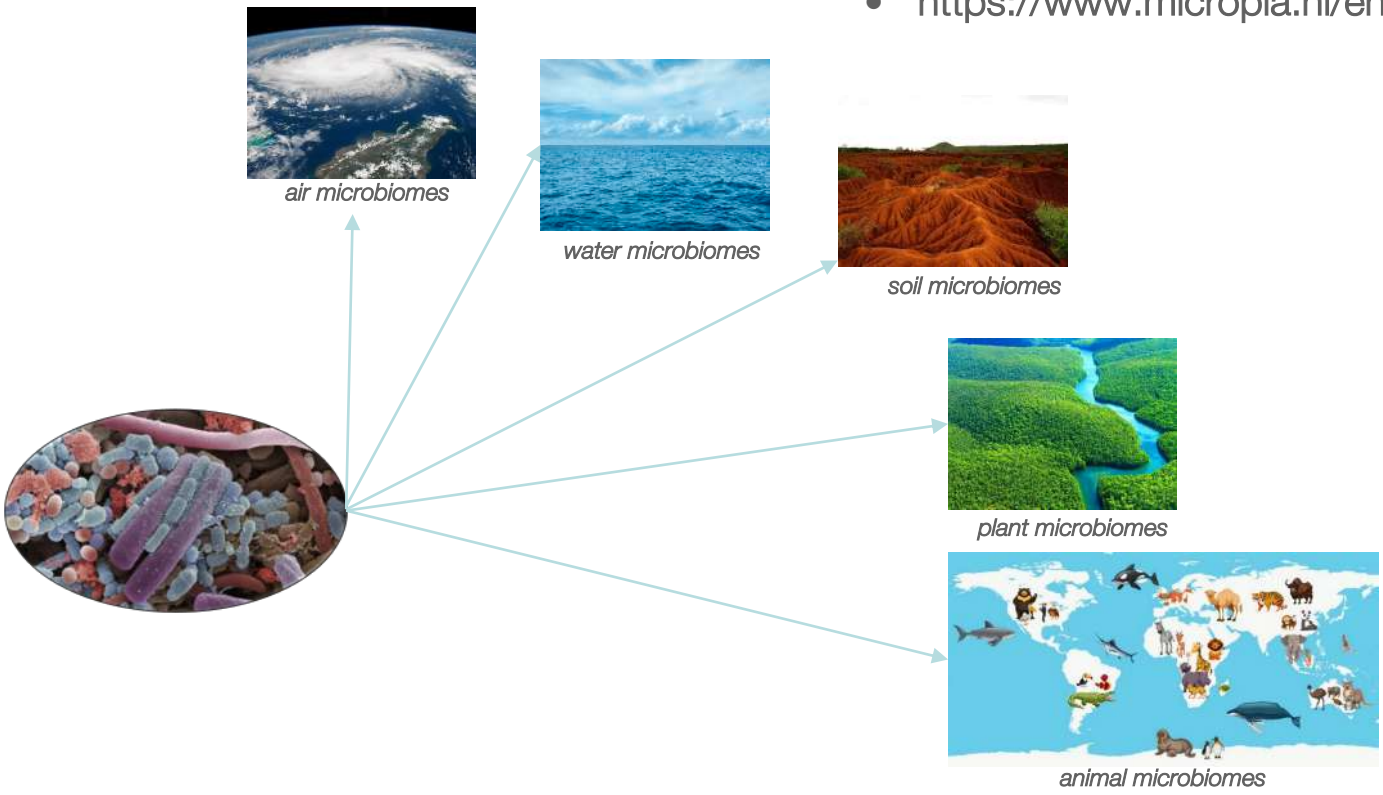
ECOLOGY AND GEOCHEMICAL VARIABLES

HOST PHYSIOLOGY AND PHYSIOCHEMICAL VARIABLES

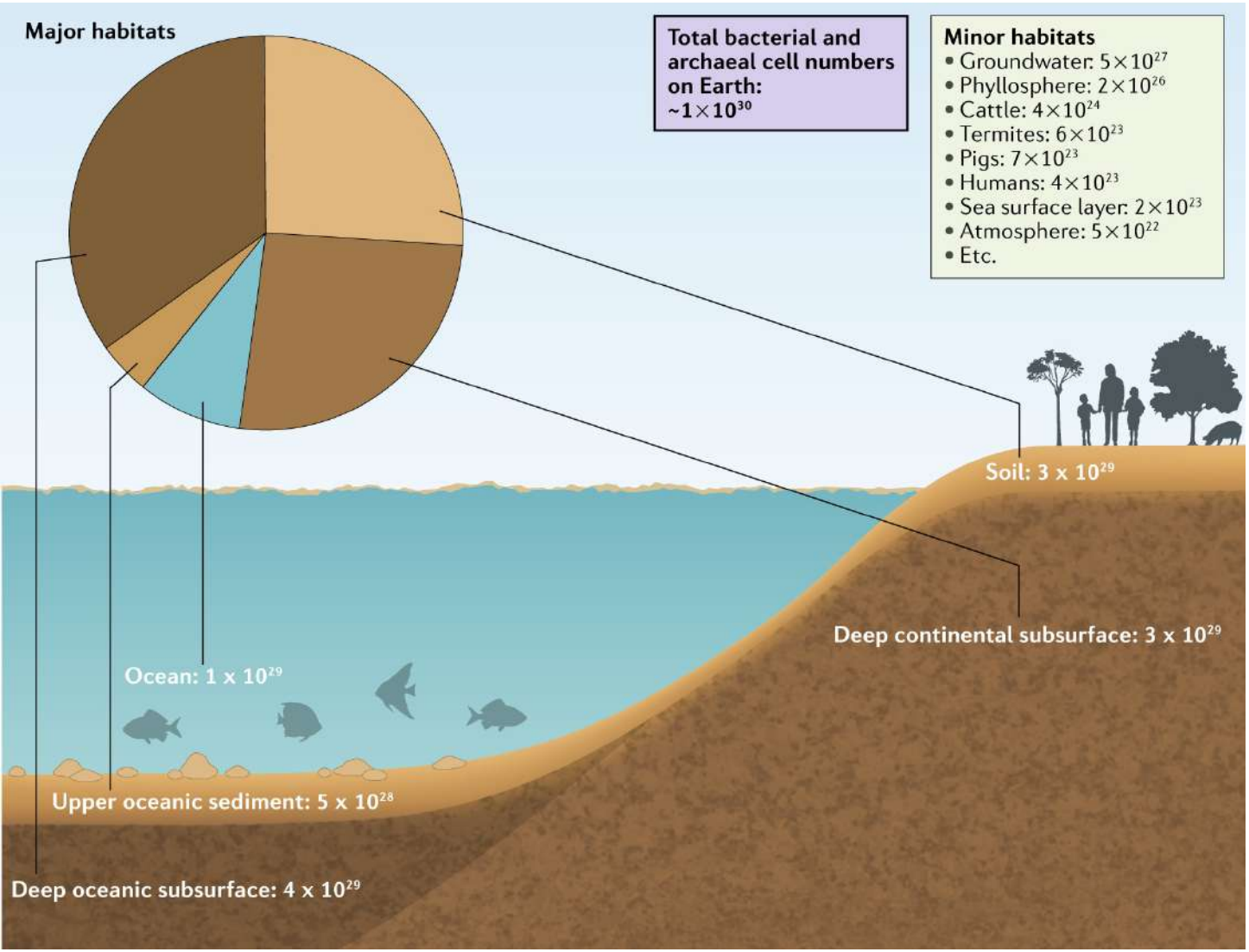
Microorganisms are all around us and live as microbiomes

- …there are 100 million times as many bacteria on Earth (13×10^{28}) as stars in the universe, and viruses are even more (13×10^{31})
  - https://www.micropia.nl/en/



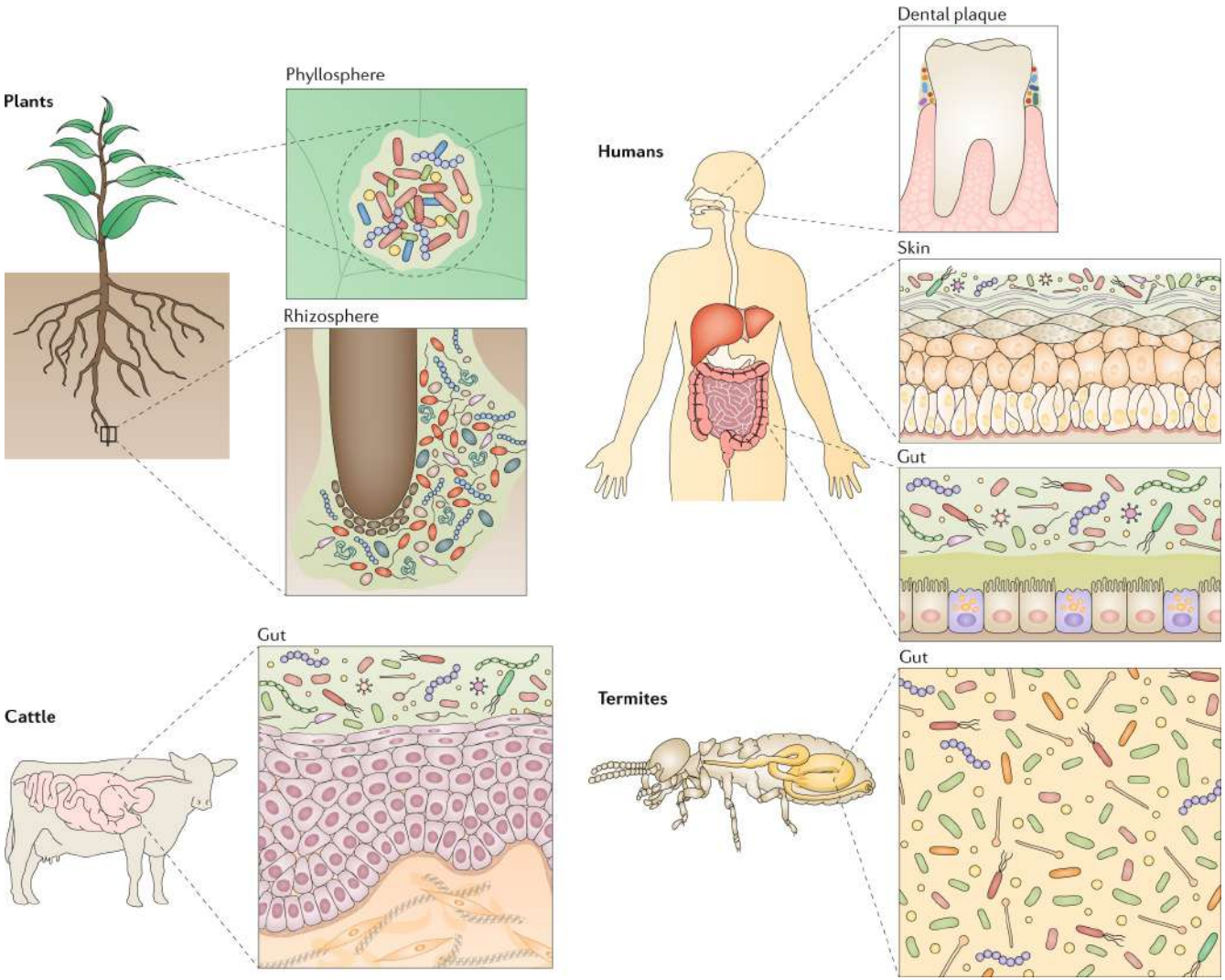


MOST BAVCTERIA AND ARCHEA ON EARTH EXISTS IN BIG 5 HABITATS



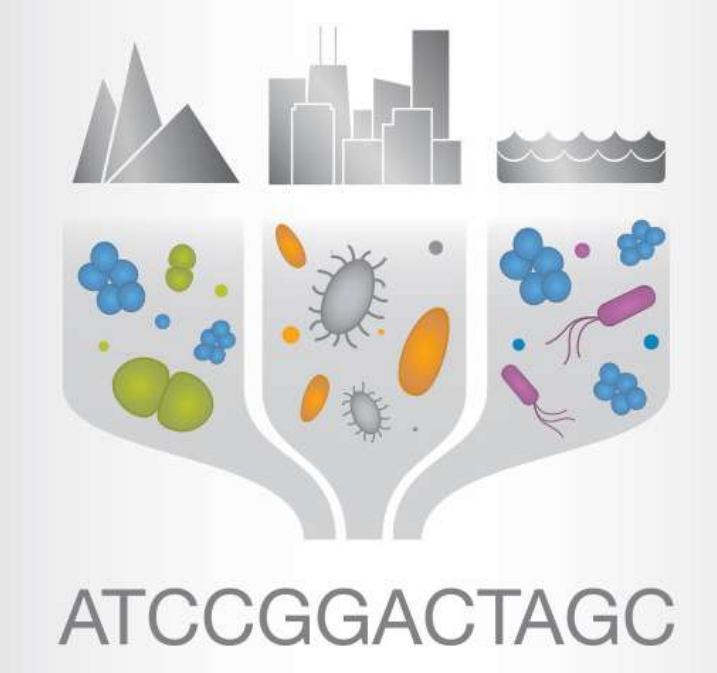
deep continental and oceanic **subsurfaces hold 60%** of all microbial cells in the biosphere

NUMERICALLY MINORITARY BUT BIOLOGICALLY STRATEGIC, THE HOLOBIONTS MICROBIOMES



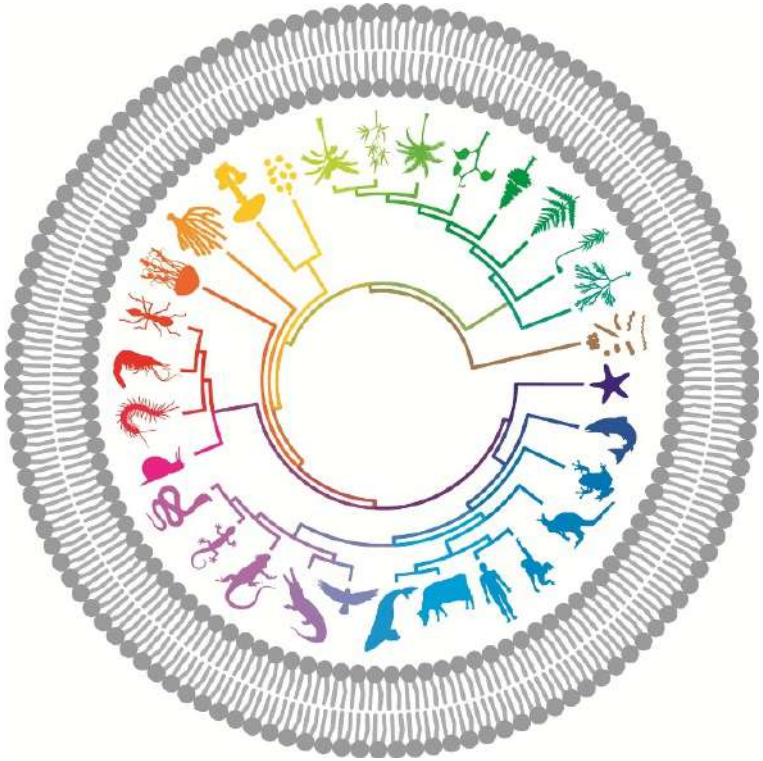
3x10^{24} plant and animal associated fast growing microbial cells (aerobic and anaerobic heterotrophs)

ENV. MICROBIOMES



environmental microbial communities are the basement of life on earth, being responsible of the biogeochemical cycles (N, P and S), C recycling and food transformation, and being the **source from which microbiomes component are selected**

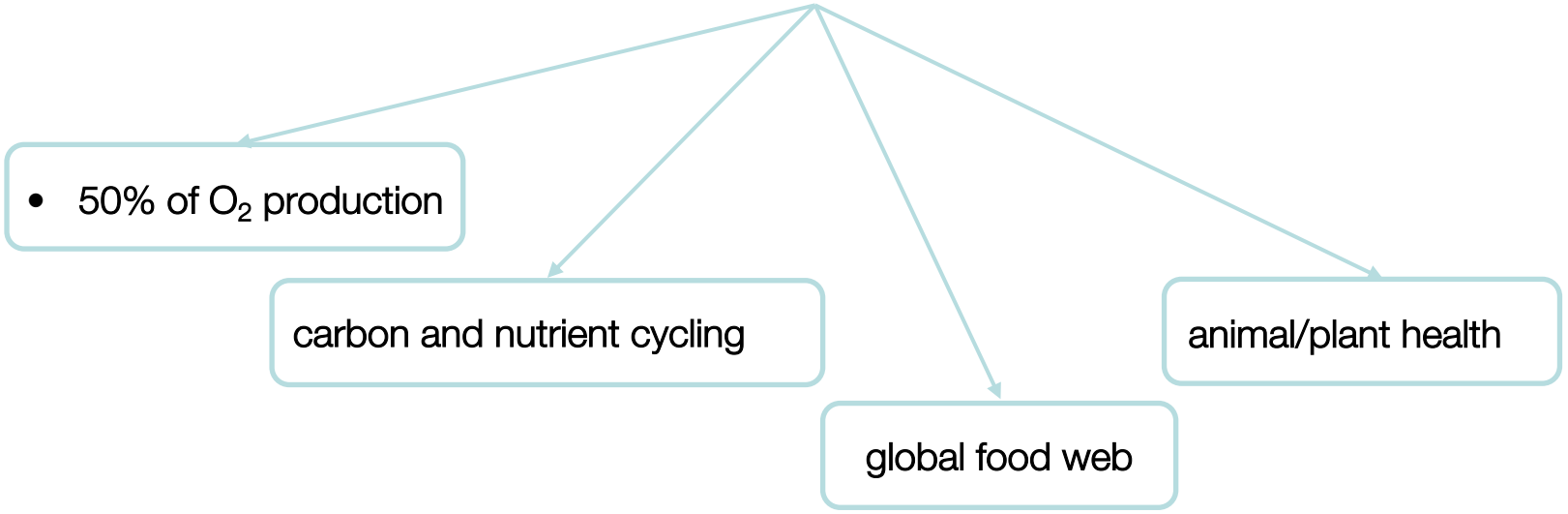
MICROBIOMES



all the macroorganisms populating our planet lives as **holobionts, defined as animals or plants together with associated microorganisms living on them**. The holobionts microbiomes contribute to the host phenotype by providing essential physiological functions

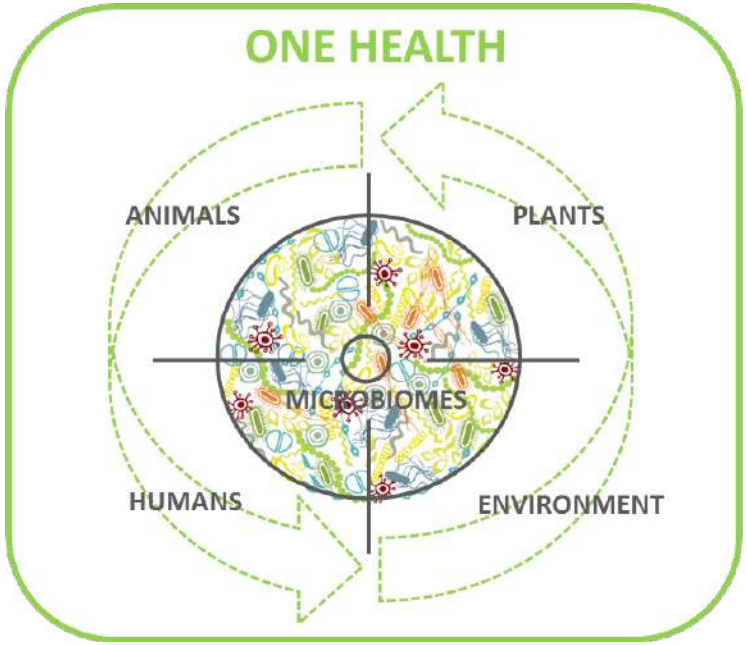
Microbiomes centrality

the microbiome world constitutes the life support system for the biosphere



NGS BOOSTED METAGENOMICS

high throughput **NGS sequencing** technologies and **dedicated bioinformatics pipelines** combined in modern metagenomics, transforming microbiology



allowing to study the planet microorganisms - beyond the limits of culturing - **NGS-based metagenomics offers important biotechnological promises**, pinpoint the **centrality of microorganism** to planetary, animal and human **health**

$1x10^{-20}$ % OF THE TOTAL DNA OF EARTH HAVE BEEN SEQUENCED

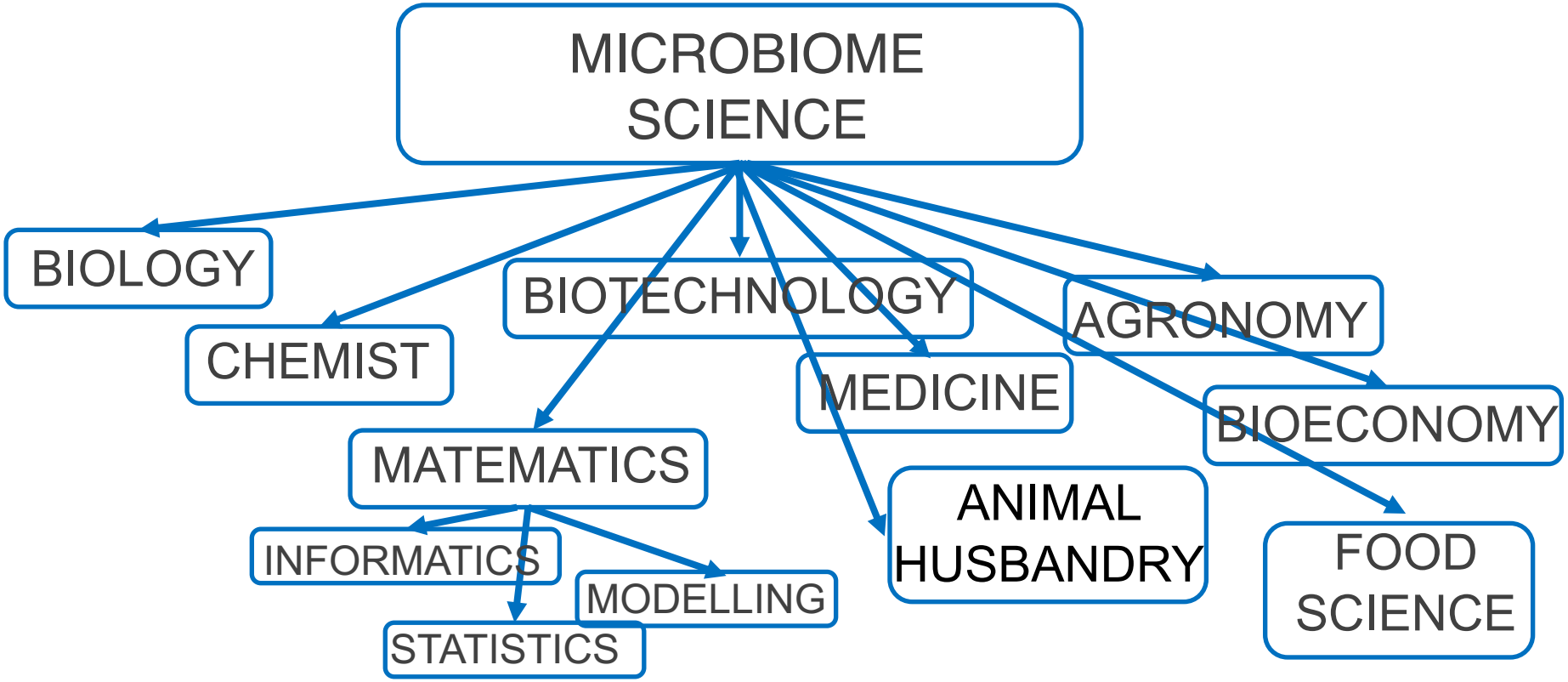
the Earth Microbiome Project is a systematic attempt to characterize global microbial taxonomic and functional diversity for the benefit of the planet and humankind

a better understanding of microbiome-dependent ecosystem services can provide microbiome-based solutions for our society and planet

MICROBIOME BIOTECHNOLOGY

LEVAREGING SPECIFIC ECOLOGICAL CONCEPTS FROM NATURAL MICROBIOMES FOR THE IMPLEMENTATION OF CONCRETE MICROBIOME BASED ACTIONS FOR IMPROVED PLANET HEALTH

MICROBIOME SCIENCE IS HIGHLY MUTIDISCIPLINARY



KNOWLEDGE **TECHNOLOGY** **APPLICATION**

INTRODUCTION TO METAGENOMICS

metagenomics is the untargeted sequencing of the genetic material present in a given sample

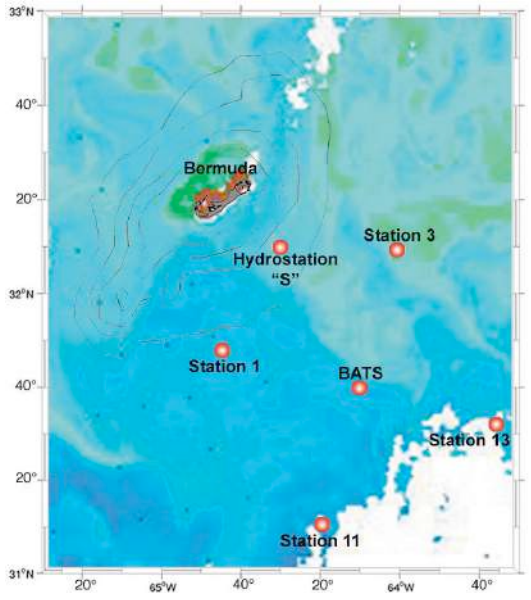
MICROBIAL METAGENOMICS

microbial metagenomics is the study of the total microbial DNA (virus, bacteria and fungi) present in a given sample

Environmental Genome Shotgun Sequencing of the Sargasso Sea

J. Craig Venter,[1]* Karin Remington,[1] John F. Heidelberg,[3] Aaron L. Halpern,[2] Doug Rusch,[2] Jonathan A. Eisen,[3] Dongying Wu,[3] Ian Paulsen,[3] Karen E. Nelson,[3] William Nelson,[3] Derrick E. Fouts,[3] Samuel Levy,[2] Anthony H. Knap,[6] Michael W. Lomas,[6] Ken Nealson,[5] Owen White,[3] Jeremy Peterson,[3] Jeff Hoffman,[1] Rachel Parsons,[6] Holly Baden-Tillson,[1] Cynthia Pfannkoch,[1] Yu-Hui Rogers,[4] Hamilton O. Smith[1]

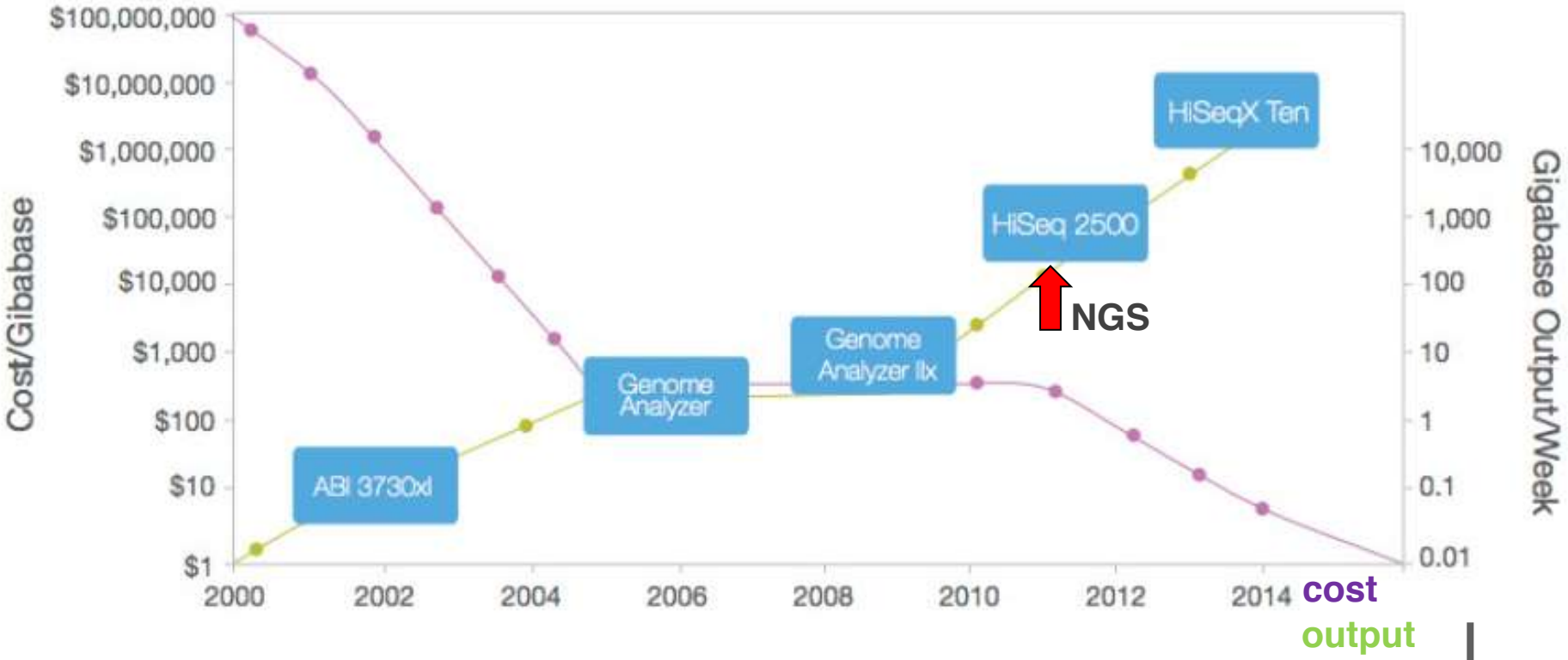
We have applied "whole-genome shotgun sequencing" to microbial populations collected en masse on tangential flow and impact filters from seawater samples collected from the Sargasso Sea near Bermuda. A total of 1.045 billion base pairs of nonredundant sequence was generated, annotated, and analyzed to elucidate the gene content, diversity, and relative abundance of the organisms within these environmental samples. These data are estimated to derive from at least 1800 genomic species based on sequence relatedness, including 148 previously unknown bacterial phylotypes. We have identified over 1.2 million previously unknown genes represented in these samples, including more than 782 new rhodopsin-like photoreceptors. Variation in species present and stoichiometry suggests substantial oceanic microbial diversity.



Venter *et al.*, Science *2004*

SEQUENCING COST AND DATA OUTPUT

in 2012 **next generation sequencing (NGS)** technology incredibly boosted the sequencing power, while reducing the sequencing costs. Thus opening the way to metagenomic applications



GOALS FOR A METAGENOME STUDY

for a microbial community or a microbiome in a given environment, eg. soil, water, air or holbionts body the goals are:

PROFILE THE TAXONOMIC COMPOSITION → **PROFILE THE FUNCTIONAL POTENTIAL**

EXPLOITATION FOR BIOTHECNOLOGY

INFLUENCE ON SYSTEM LEVEL PROCESSES ← **DEFINE THE MAIN ECOLOGICAL SERVICES**

BASEMENT DEFINITIONS

MICROBIOME: a microbial community associated with a given host or a given environment

METAGENOME: the overall genome of a **MICROBIOME**

METAGENOMICS: the study of the **METAGENOME** by shotgun NGS sequencing

SAMPLE: a sample for a given host or environment containing a **MICROBIOME**

COVARIATE: a variable characterizing a given host or environment hosting a **MICROBIOME**

TYPES OF MATAGENOME STUDIES

the object of a metagenome study is **the microbiome**, the overall genome of a microbial community in a given sample

MARKER GENE

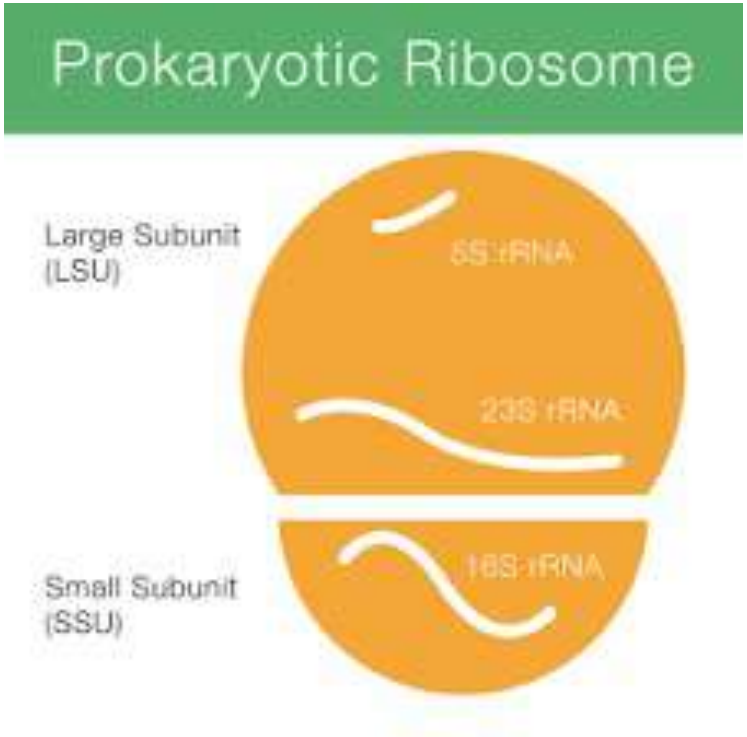
use primers targeting a specific region of a gene of in order to determine the microbial phylogenesis

WHOLE METAGENOME

shot-gun sequencing of the whole microbiome to determine phylogenesis and the functional repertoire

MARKER GENE ANALYSIS

the **16S rDNA gene** is the best phylogenetic clock for prokaryotes

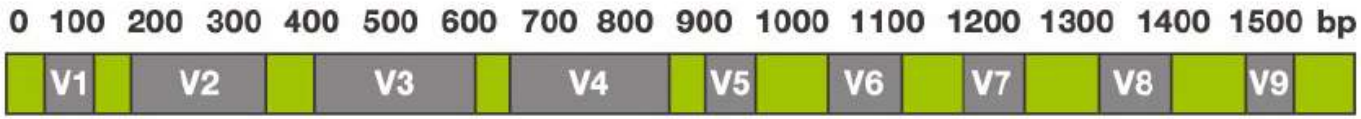


- ALL THE ORGANISMS POSSESS AT LEAST ONE COPY OF SUCH GENE
  universal marker

- CONSERVED GENE, SAME FUNCTION FOR ALL ORGANISMS
  same housekeeping function and same selective pressure for all organisms

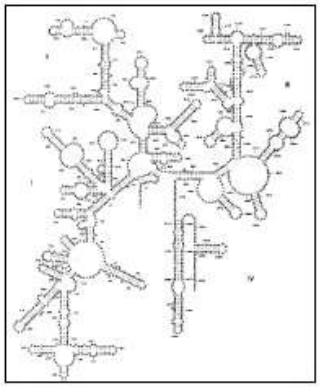
- NOT INVOLVED IN LATERAL GENE TRANSFER
  only vertical transmission

THE 16 rDNA GENE



CONSERVED REGIONS: unspecific applications

VARIABLE REGIONS: group or species-specific applications

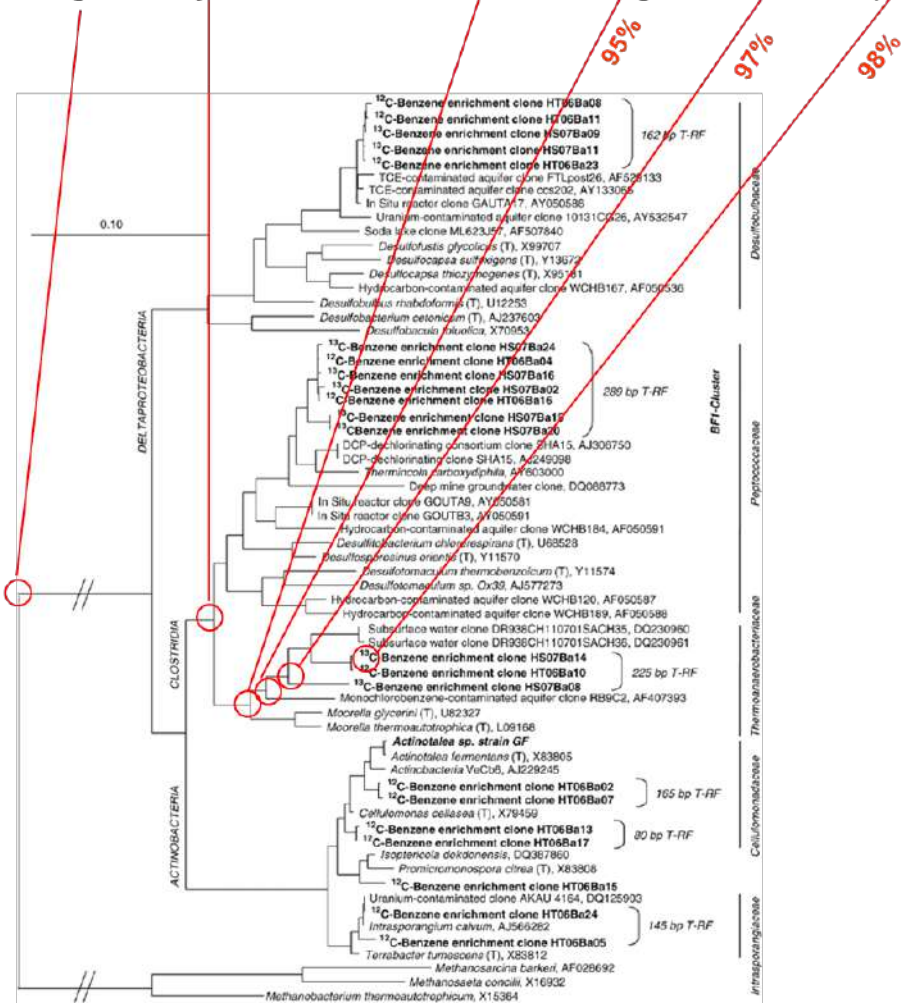


CONSERVED REGIONS → **AMPLIFY, CLONE AND SEQUENCE THE GENE FROM UNKNOWN MICOORGANISMS**

VARIABLE REGIONS → **IDENTIFICATION AND PHYLOGENESIS**

THE 16 rDNA GENE BASED PHYLOGENETIC TREE

Dominio-Regno-Phylum-Classe-Ordine-Famiglia-Genere-Specie



the % of homology of the 16S rDNA gene is used to weight the degree of similarities among microorganisms

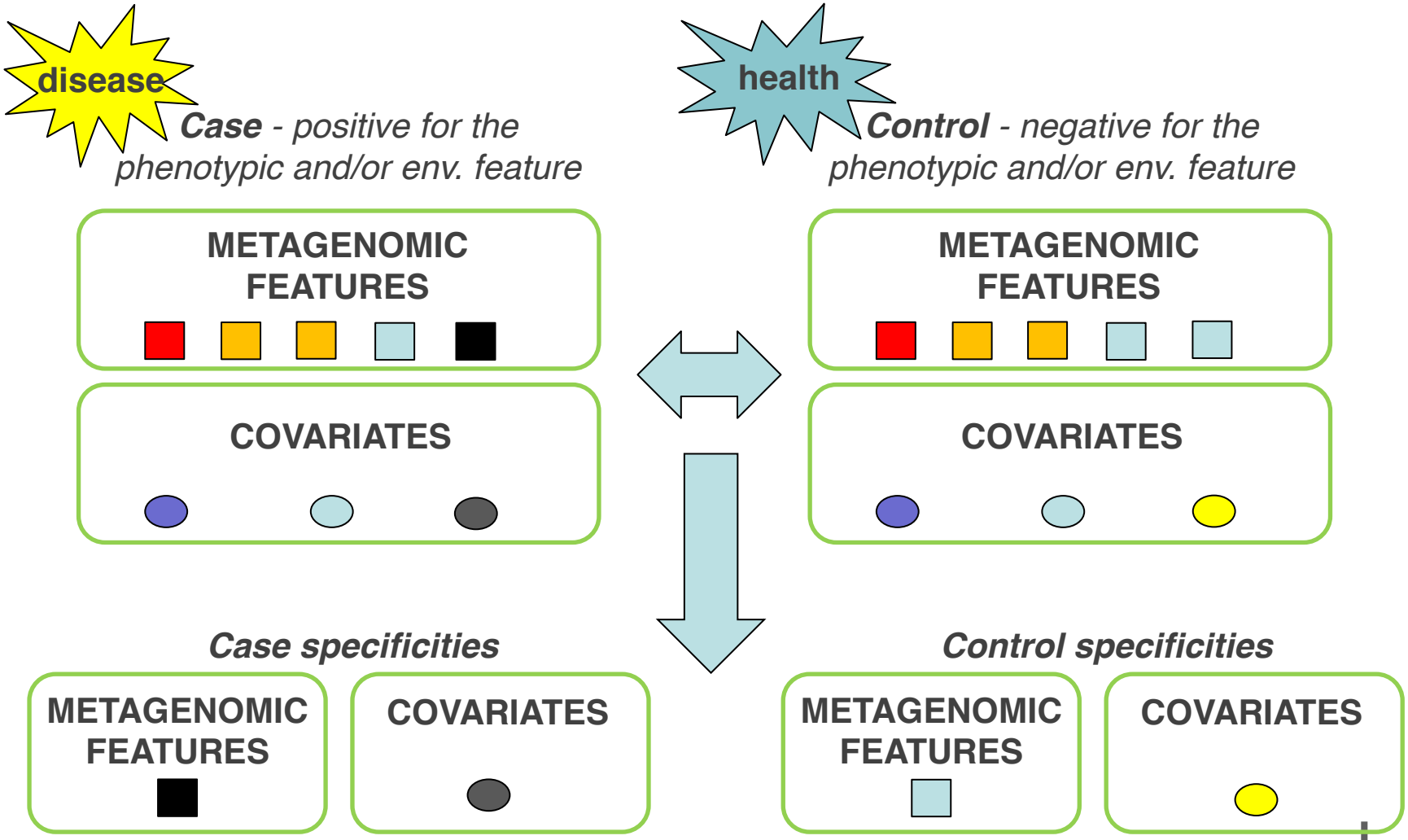
microorganisms are arranged in a phylogenetic tree on the bases of the percentage of homology of the 16S rDNA gene

TYPE OF QUESTIONS METAGENOMIC DEAL WITH

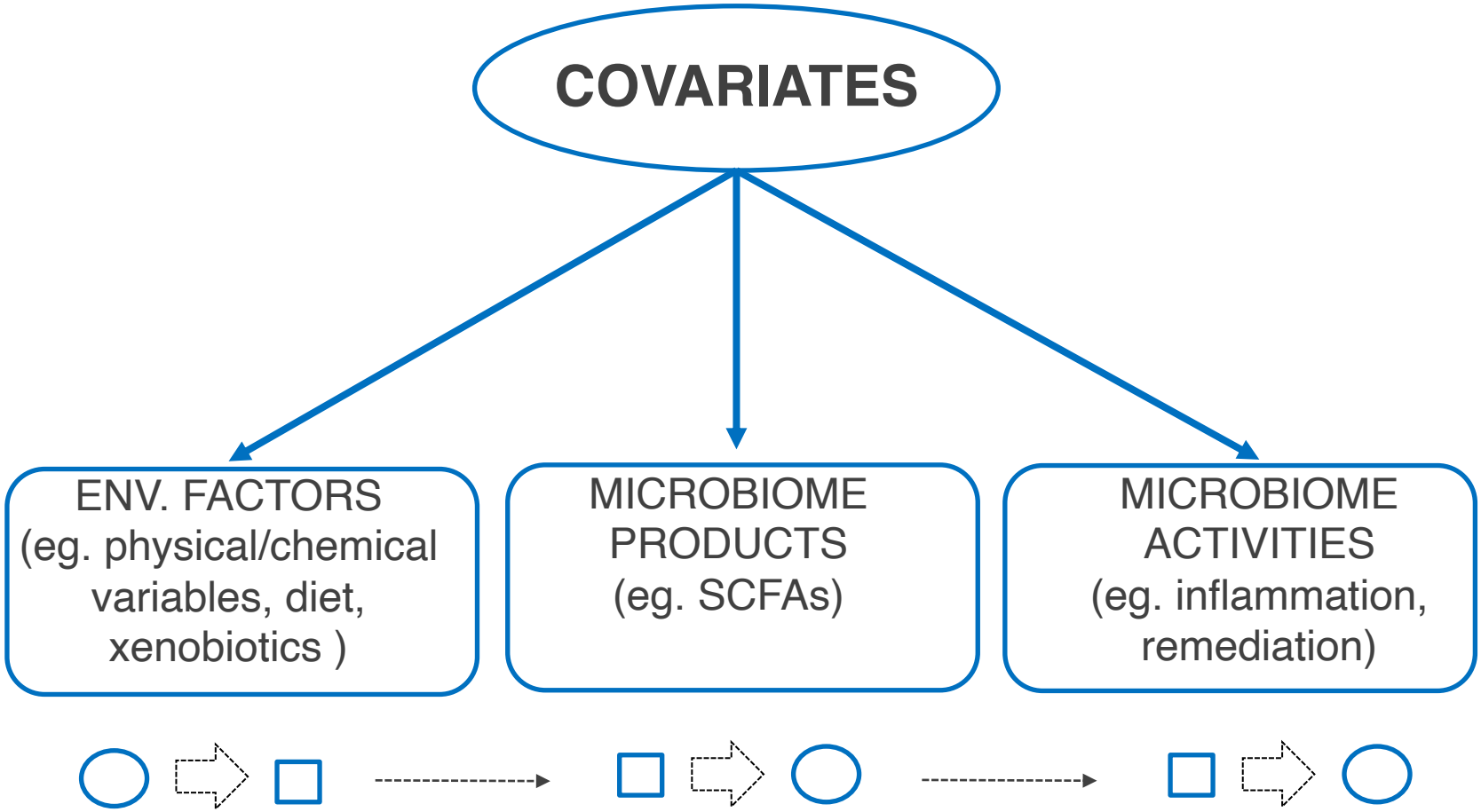
is the microbiome populating a given holobiont **associated** with a specific phenotypic and/or env. feature?

is the microbiome a given holobiont **a determinant** of a specific phenotypic and/or env. feature?

THE GREAT MAJORTY OF METAGENOME STUDIES ARE COMPARATIVE

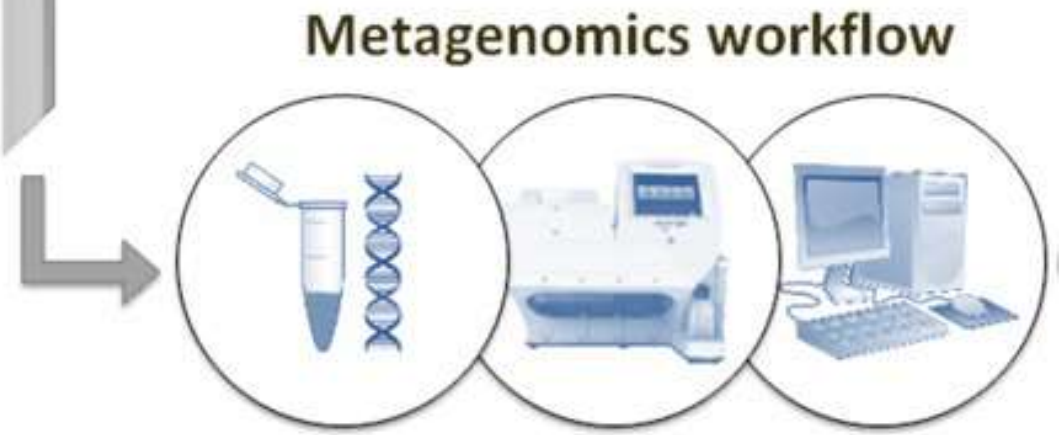


TYPES OF MICROBIOME COVARIATES



HOW TO CONDUCT A METAGENOME STUDY

- study design → *process to be explored*
- sampling and storage → *statistic robustness*
- DNA extraction → *adequate to the matrix*
- library preparation → *level of multiplexing*
- NGS sequencing → *sequencing power*
- bioinformatics → *optimal pipelines*
- biostatistics → *creative statistics*



NEXT GENERATION SEQUENCING TECHNOLOGY

SHORT-READS THECNOLOGY

illumina®

up to 3 Tb per cell with read lengths ranging between 100-300 bp

SHOT GUN METAGENOMICS

LONG-READS THECNOLOGY

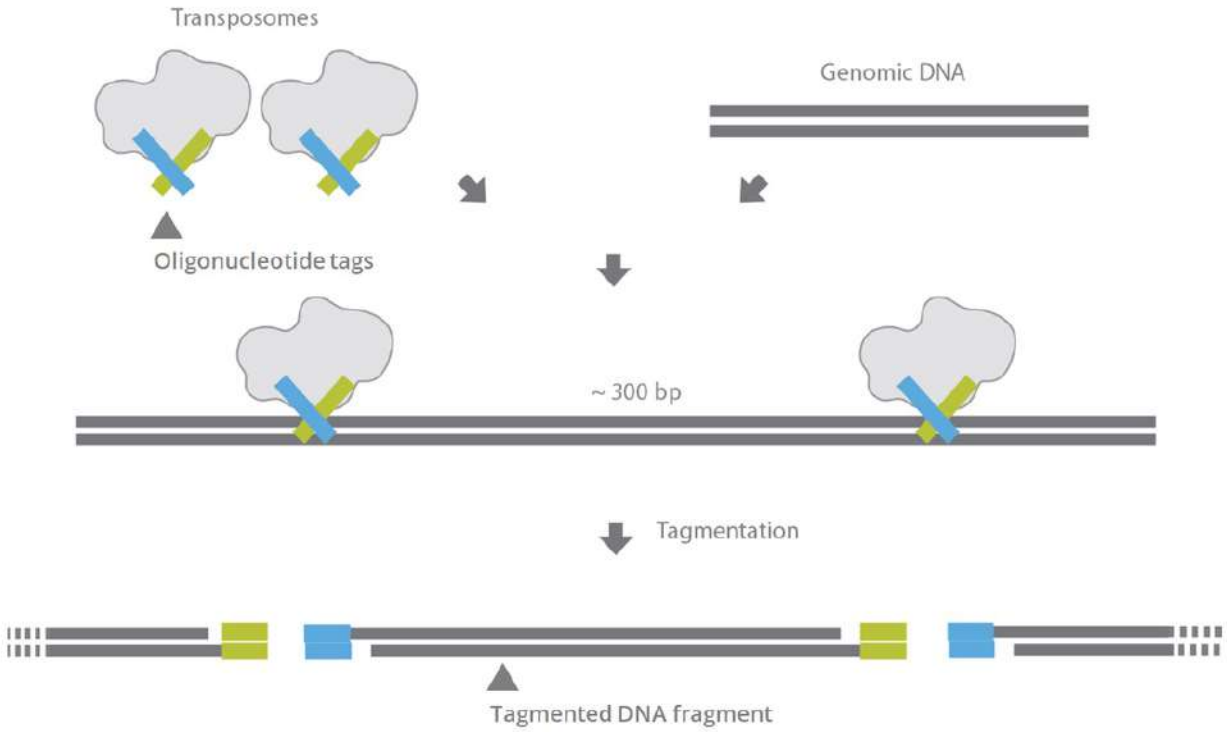
PACBIO®

20 Gb per cell with average read lengths up to 30 kb

SINGLE CELL SEQUENCING

SHOT GUN METAGENOMICS, DNA TAGMENTATION

tagmentation reaction of a metagenome from a given sample involves the transposon cleaving and tagging of the double stranded DNA with a universal overhang

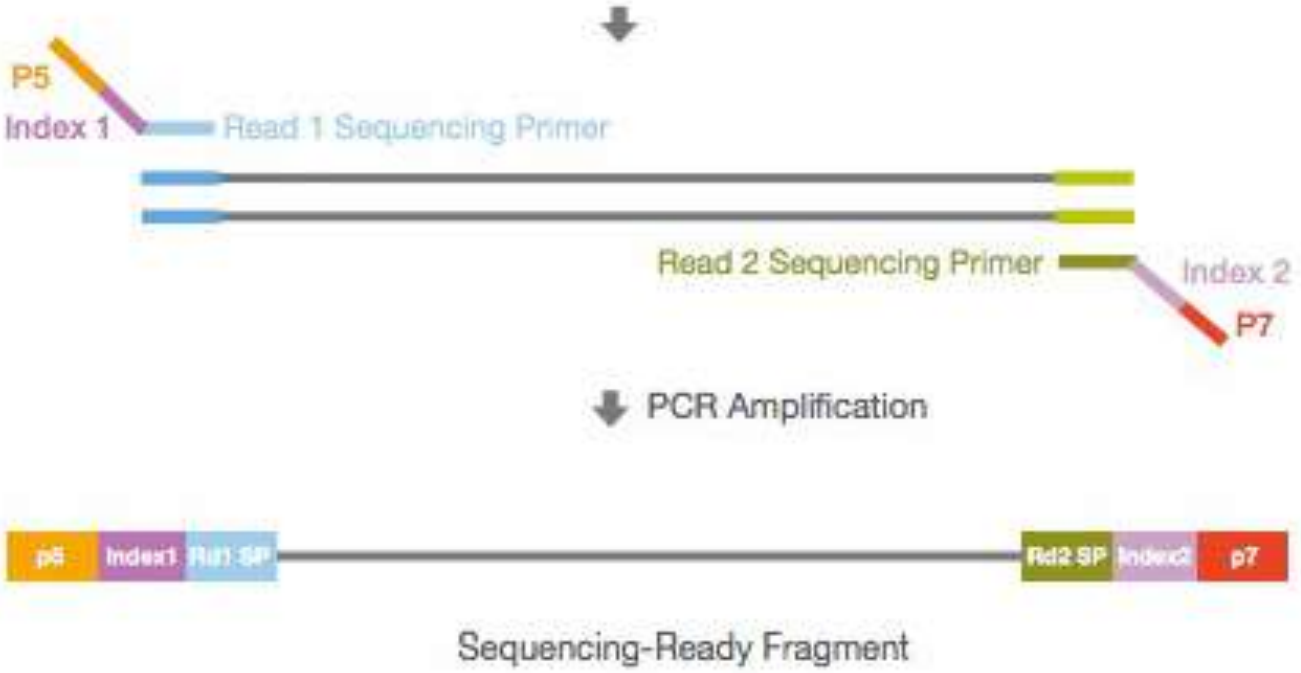


10 ng-1 μg starting DNA

n. 300 pb fragments

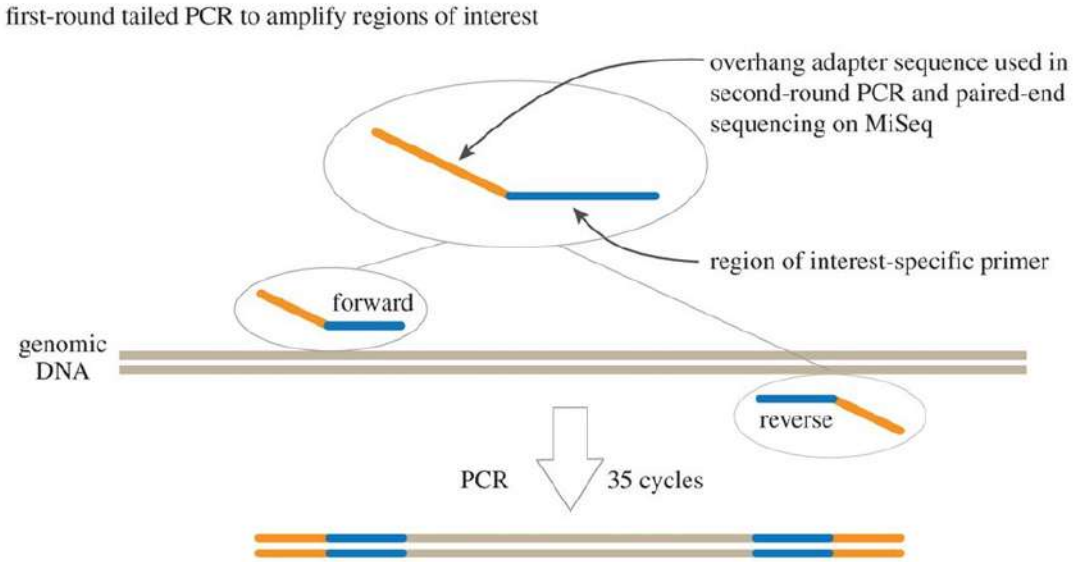
FRAGMENTS CLONING AND BARCODING BY INDEX PCR

Index PCR allows to **clone and tag** each 300 pb DNA fragment from a given sample with R1 and R2 sequencing primers, a unique combination of two barcode index and the P5 and P7 regions for bridge PCR on the flowcell



16S SEQUENCING, MARKER GENE PCR

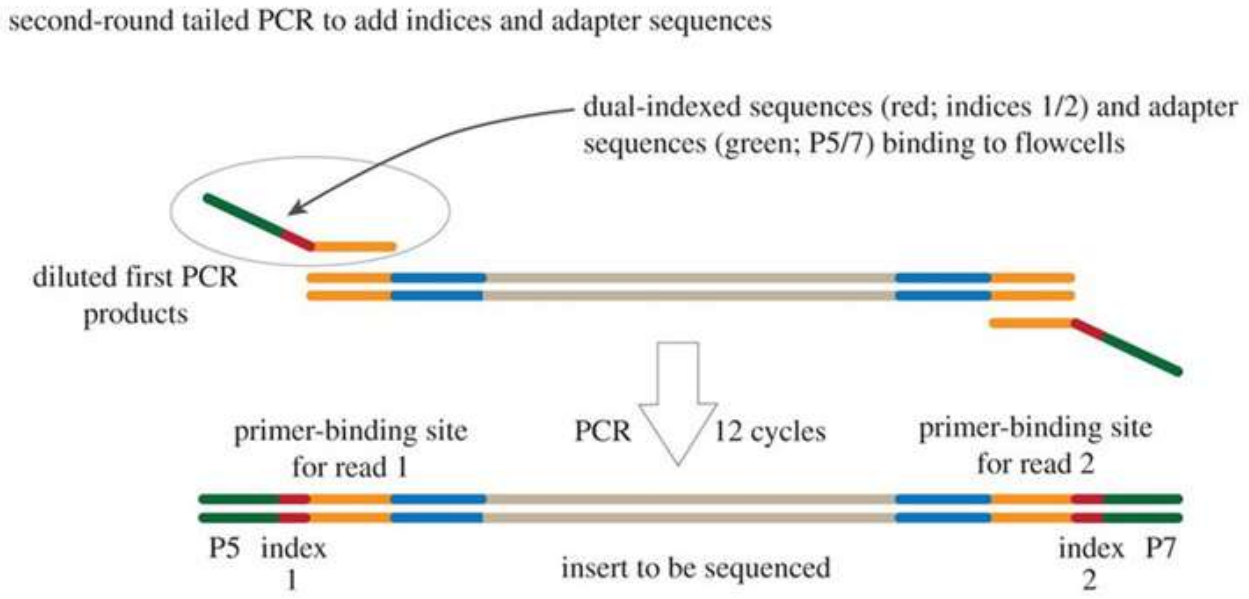
25 cycles of PCR for the amplification of the V3/V4 region (450 bp) of the 16rDNA gene from a metagenome in a given sample



L and R primers possessed an universal overhang adapter for the following index PCR

BARCODING BY INDEX PCR

Index PCR allows to tag each 450 pb V3/V4 DNA fragment from a given sample with R1 and R2 sequencing primers, a unique combination of two barcode index and the P5 and P7 regions for bridge PCR on the flowcell



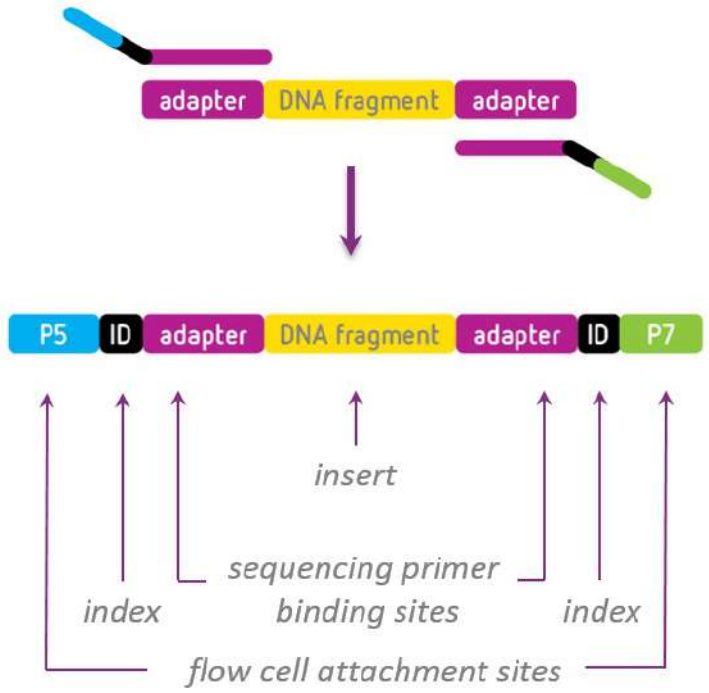
SAMPLE RADY FOR NGS

the final outcome of a sample preparation procedure for NGS is the insertion of n. 300–450 bp DNA fragments from the metagenome to be sequenced (both marker gene and shot gun metagenomics) between:

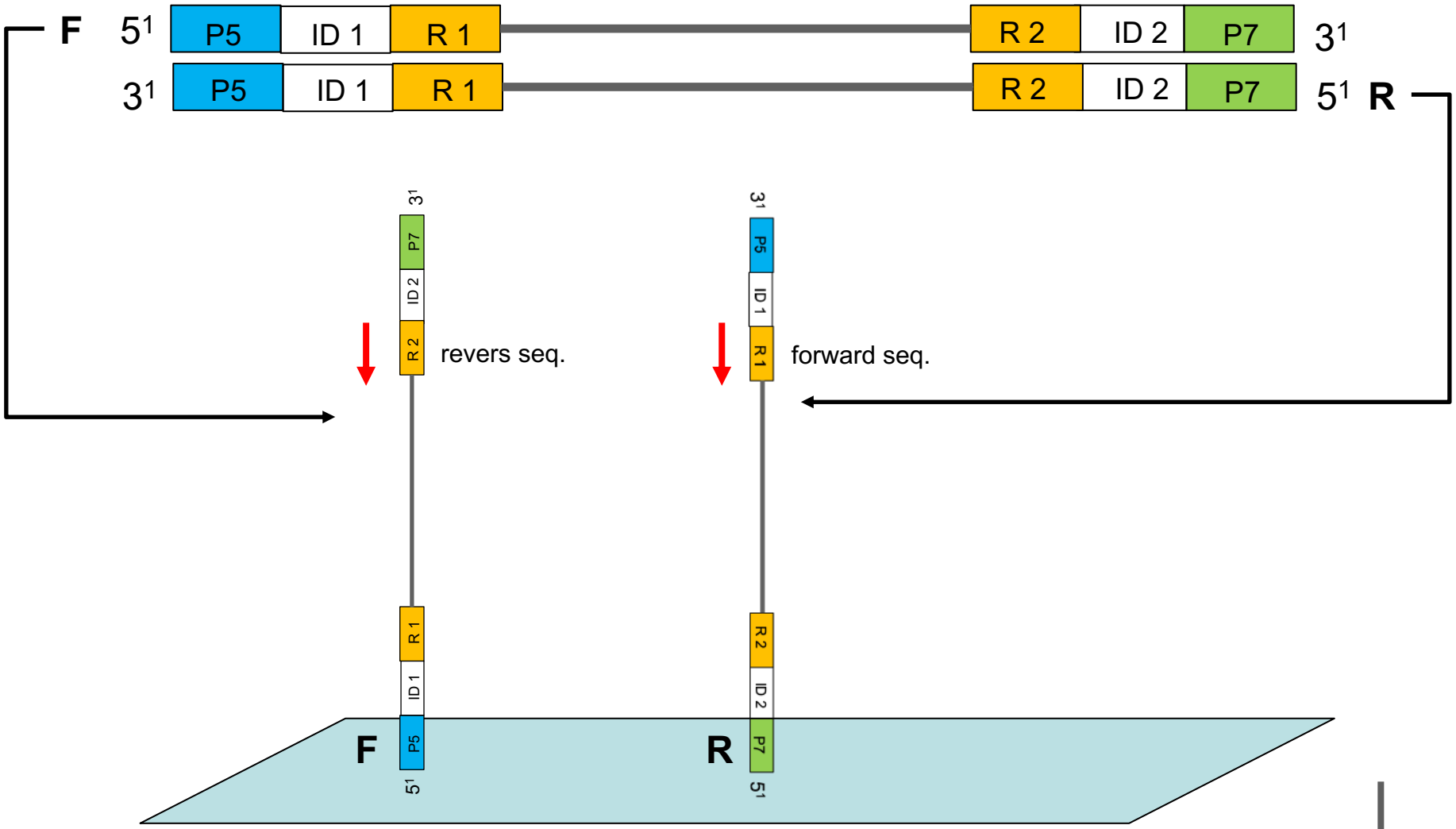
- L and R sequencing primers binding sites;
- L and R samples specific barcodes index
- L and R flow attachment sites

fluorimetric quantification (Qubit)

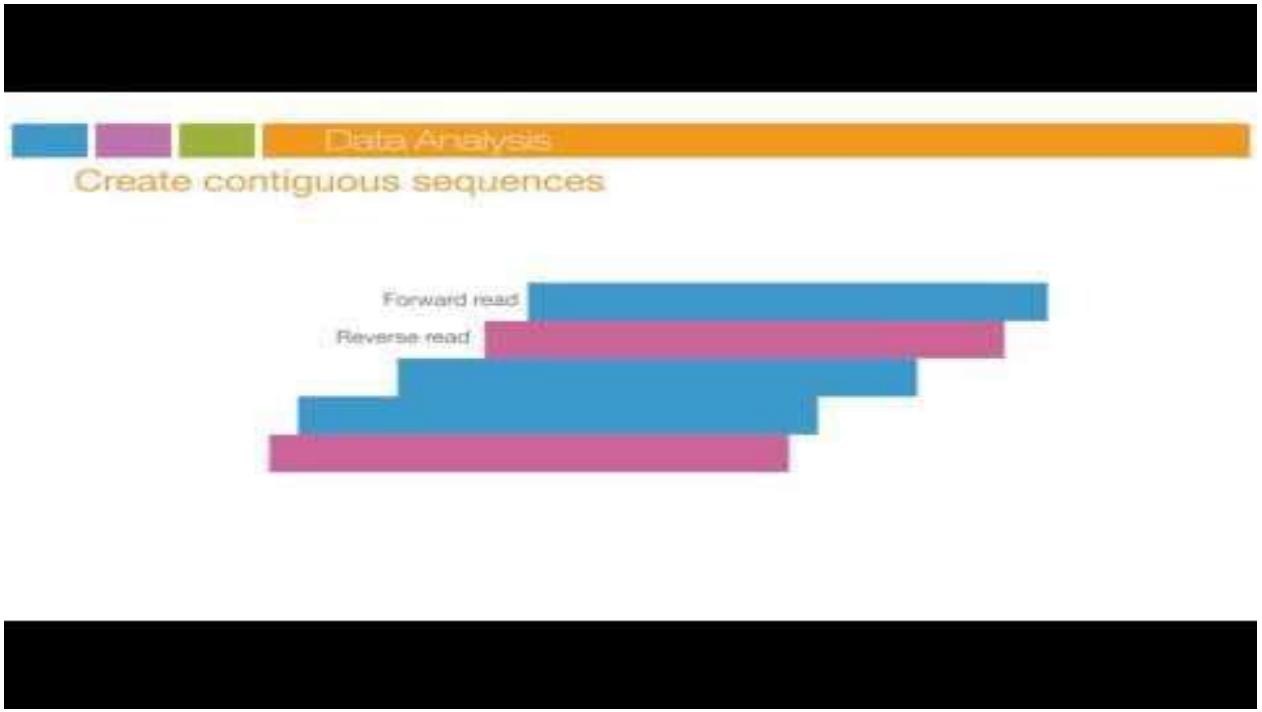
dilution at 4 nM



FLOW CELL, SURFACE SEQUENCING



ILLUMINA SEQUENCING THECNOLOGY



https://youtu.be/womKfikWlxM

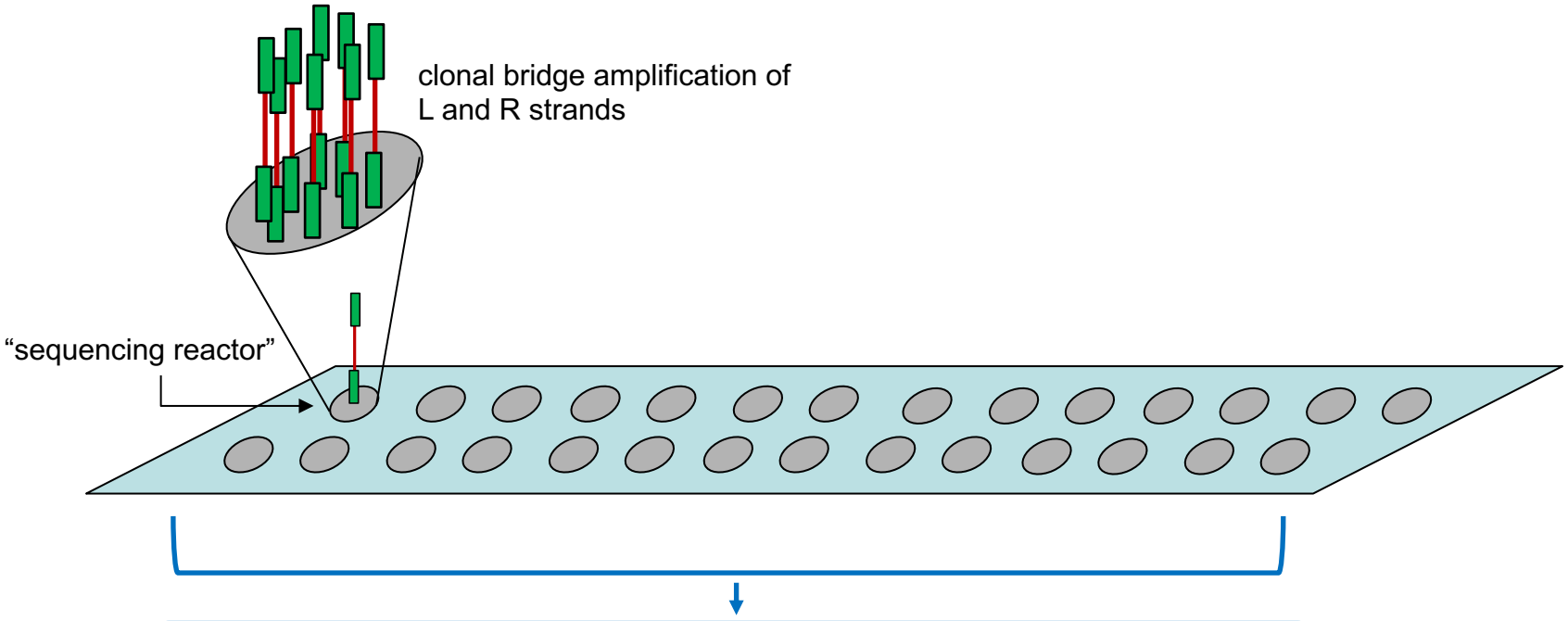
HOW CAN WE FIGURE OUT THE TOTAL SEQUENCING POWER?

SINGLE READS
PAIRED-ENDS

SEQUENCING
REACTION

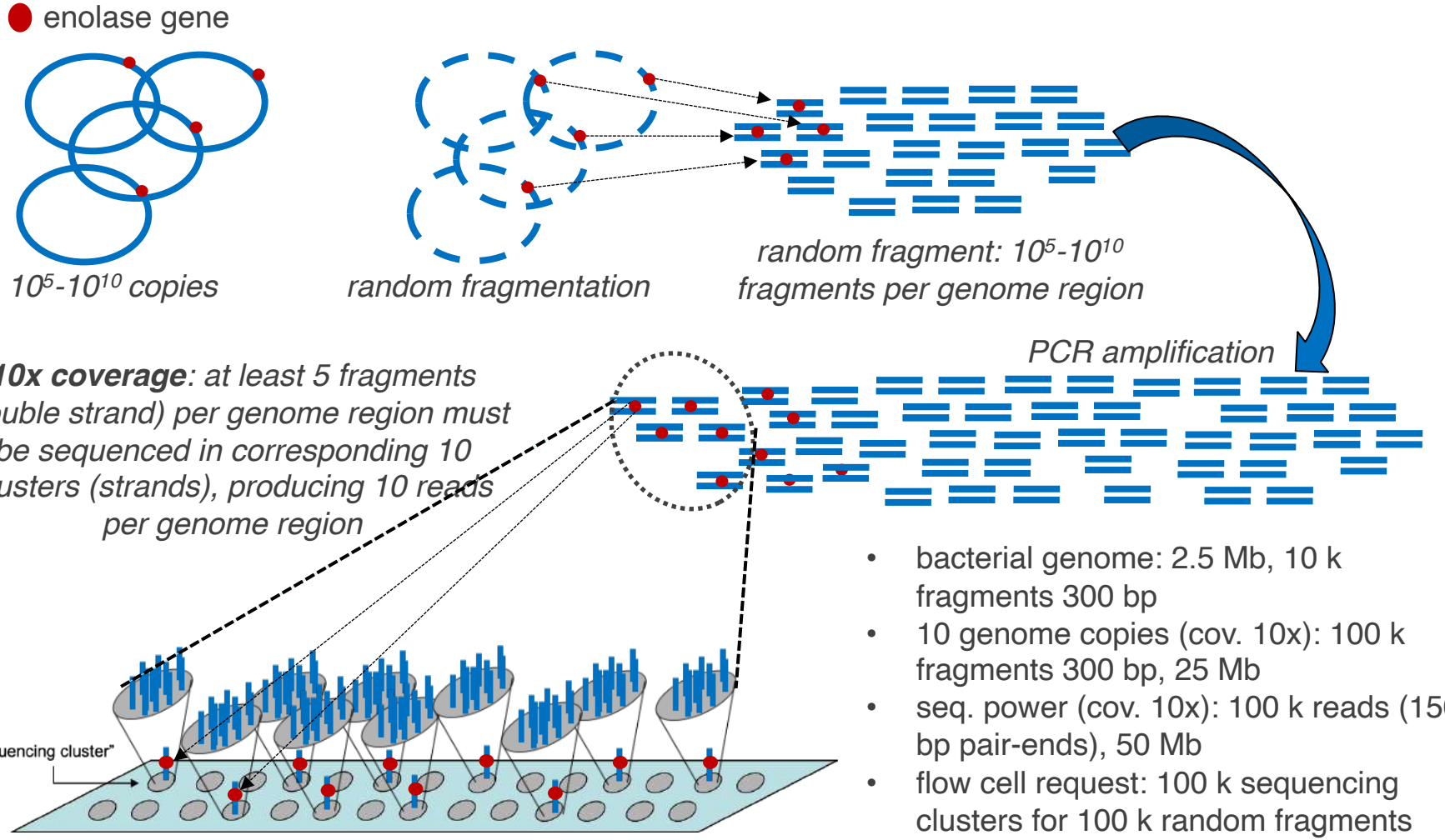
10X COVERAGE

FOR A GIVEN DNA FRAGMENT, AT LEAST 10 COPIES NEEDS TO BE SEQUENCED IN 10 REACTORS



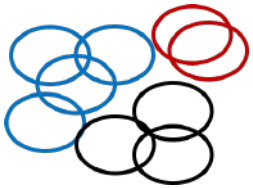
TOTAL SEQUENCING POWER DEPEND ON READS LENGTH AND ON THE TOTAL NAMBER OF SEQENCING REACTORS TO BE AVAILABLE PER FLOW CELL

GENOME SEQUENCING DEPTH, ILLUMINA PAPELINE

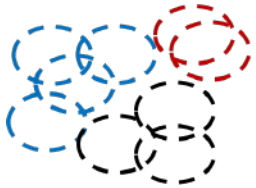


- bacterial genome: 2.5 Mb, 10 k fragments 300 bp
- 10 genome copies (cov. 10x): 100 k fragments 300 bp, 25 Mb
- seq. power (cov. 10x): 100 k reads (150 bp pair-ends), 50 Mb
- flow cell request: 100 k sequencing clusters for 100 k random fragments each 300 bp

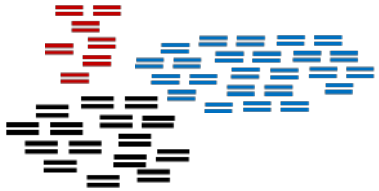
METAGENOME SEQUENCING DEPTH, ILLUMINA PAPELINE



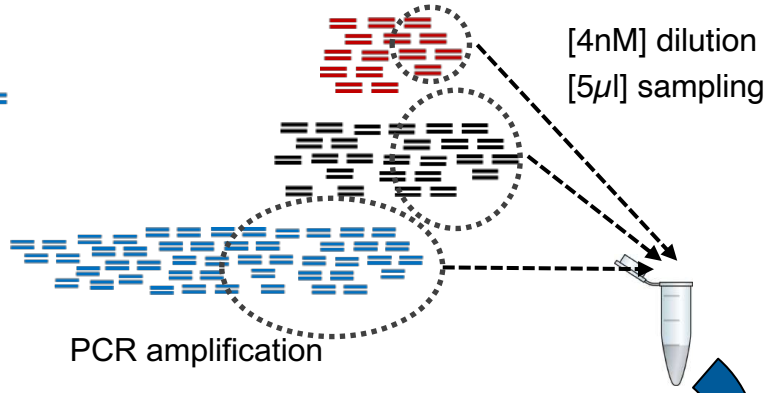
metagenome, up to 200 genomes types at different proportional abundance (50-0.01%)

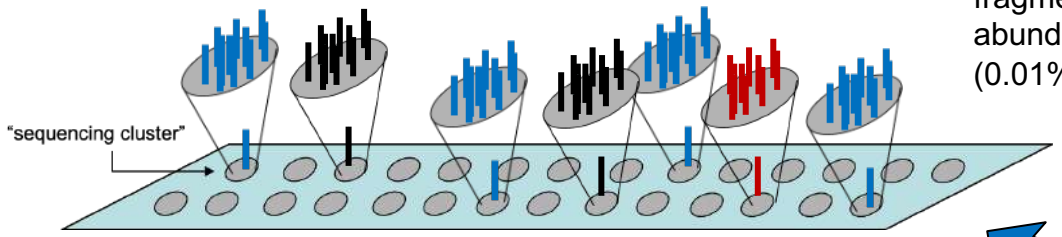


random fragmentation



random fragments, n. of fragments from a given genome type reflect its proportional abundance





at least 100 k 300pb fragments for the less abundant genome (0.01%) – 10x cov.

- 50-400 gnome types (species)
- rel. abb. ranging from 50 to 0.01%
- low abb. species: 100 K 300bp fragments for 10x minimal cov.
- high abb species: >>>>> 100 k 300bp fragments
- seq. power: 5 to 100 M reads (pair-ends 150 bp), 1-25 Gb total sequencing
- flow cell request: 5 to 100 M reads sequencing clusters for 5 to 100 M random metagenome fragments

GOOD PRACTICES PER SAMPLE

- single **bacterial genome** = around 50 Mb; 250 K reads (pair-ends 100-125 bp)
- 16S rRNA **marker gene metagenome** = 4 to 18 Mb; 10-30 K reads (400 bp, single reads or 2x300/2x250 bp paired-ends)
- shot gun **metagenome** sequencing = between 1 to 25 Gb; 5 to 100 M reads (pair-ends 100-125 bp)
- **human genome** = 50 Gb, 250 M reads (pair-ends 100-125 bp)

the **level of multiplexing** is selected on the bases of the desired per-sample sequencing depth

ILLUMINA PLATFORM, THE PREDOMINANT CHOICE FOR SHOT GUN METAGENOMICS

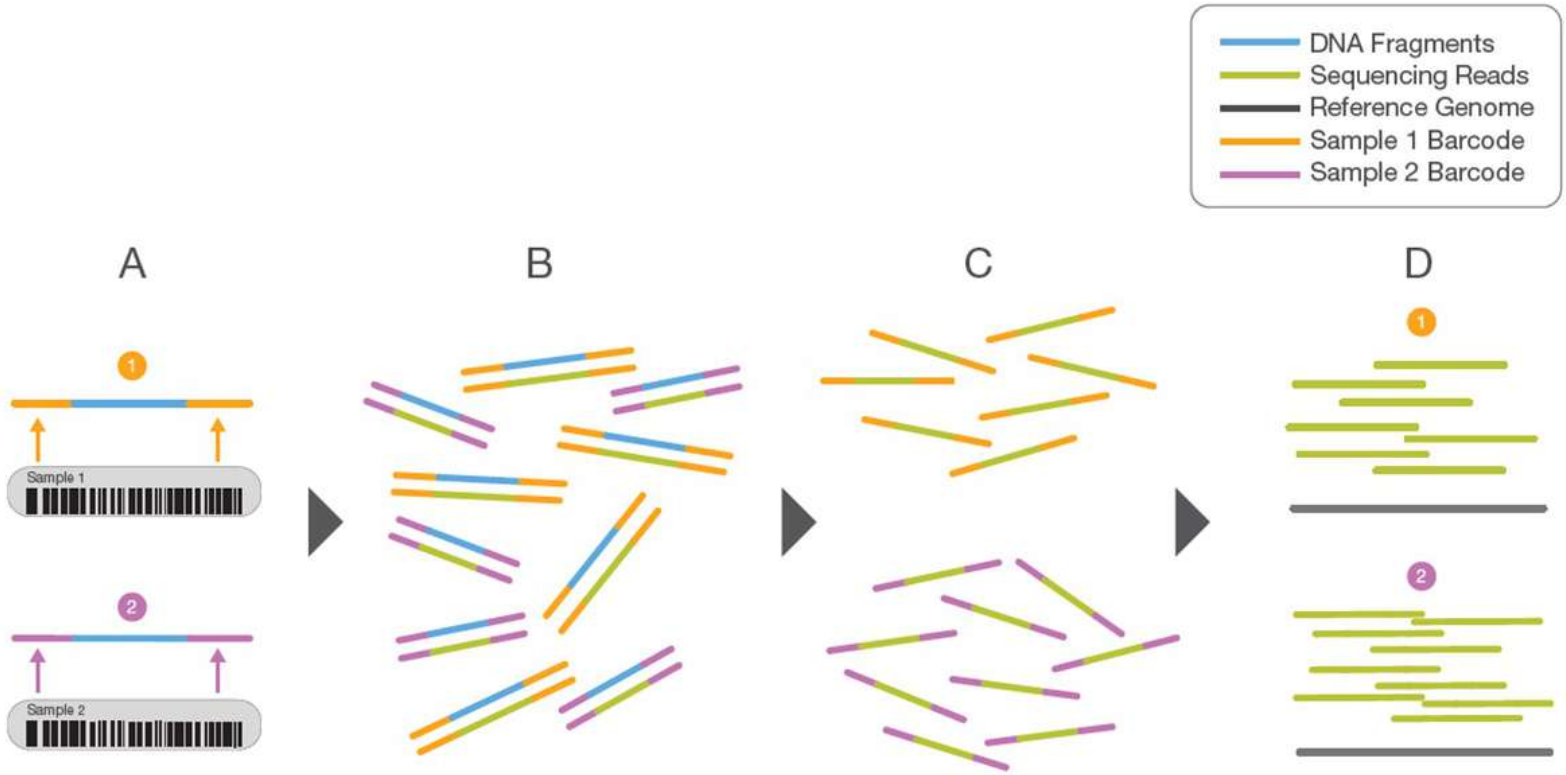
	Benchtop Sequencers		Production-Scale Sequencers	
				
	iSeq 100 System	MiniSeq System	MiSeq Series ⊕	NextSeq Series ⊕
Run Time	9–17.5 hours	4–24 hours	4–55 hours	12–30 hours
Maximum Output	1.2 Gb	7.5 Gb	15 Gb	120 Gb
Maximum Reads Per Run	4 million	25 million	25 million †	400 million
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 300 bp	2 × 150 bp

ILLUMINA PLATFORM

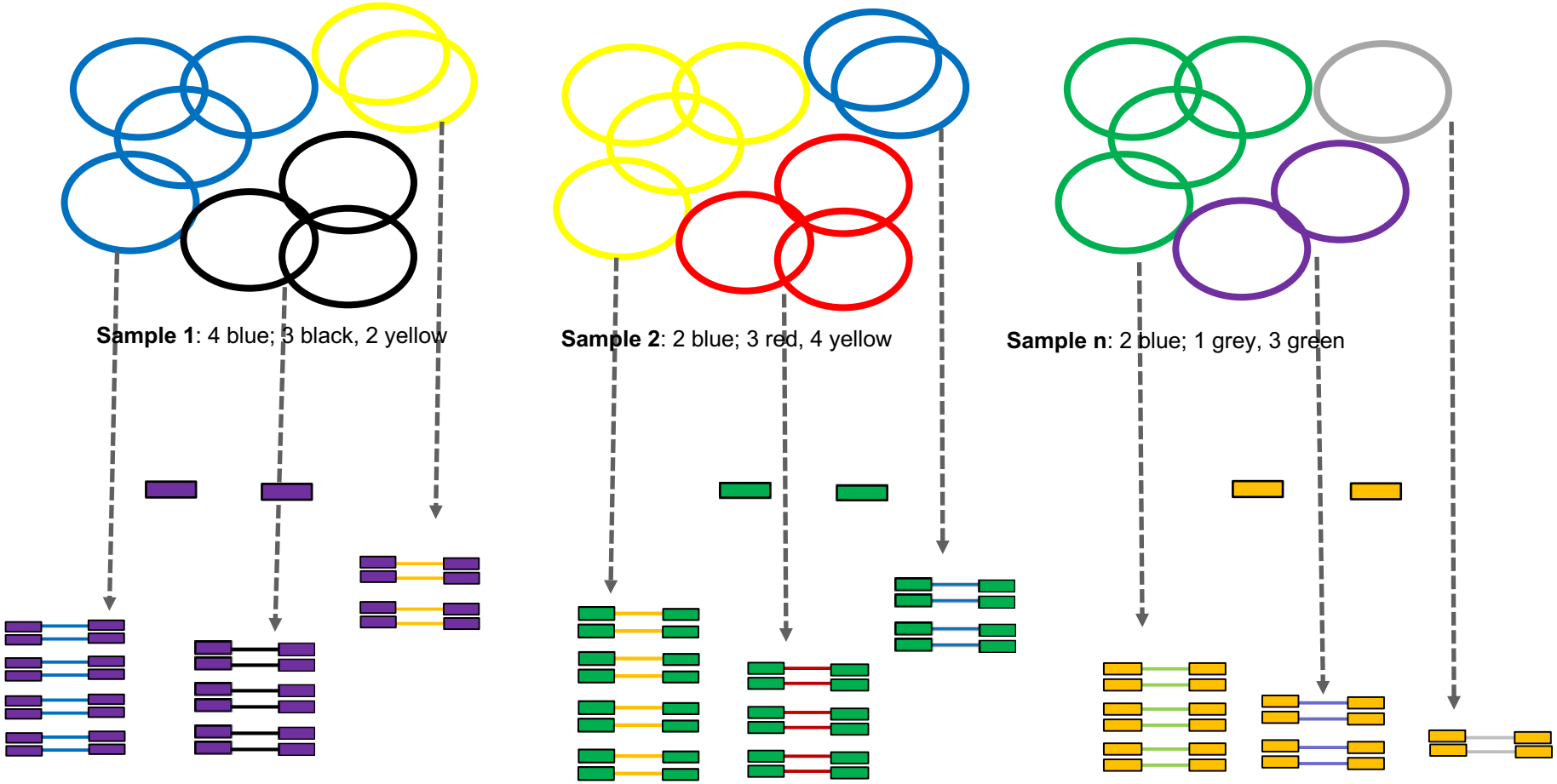
	Benchtop Sequencers	Production-Scale Sequencers		
	 NextSeq Series	 HiSeq 4000 System	 HiSeq X Series‡	 NovaSeq 6000 System
Run Time	12–30 hours	< 1–3.5 days	< 3 days	~13–25 hours (dual S1 flow cells) ~16–36 hours (dual S2 flow cells) ~44 hours (dual S4 flow cells)
Maximum Output	120 Gb	1500 Gb	1800 Gb	6000 Gb
Maximum Reads Per Run	400 million	5 billion	6 billion	20 billion
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp

NGS MEANS MULTIPLEXING WITH BARCONDING

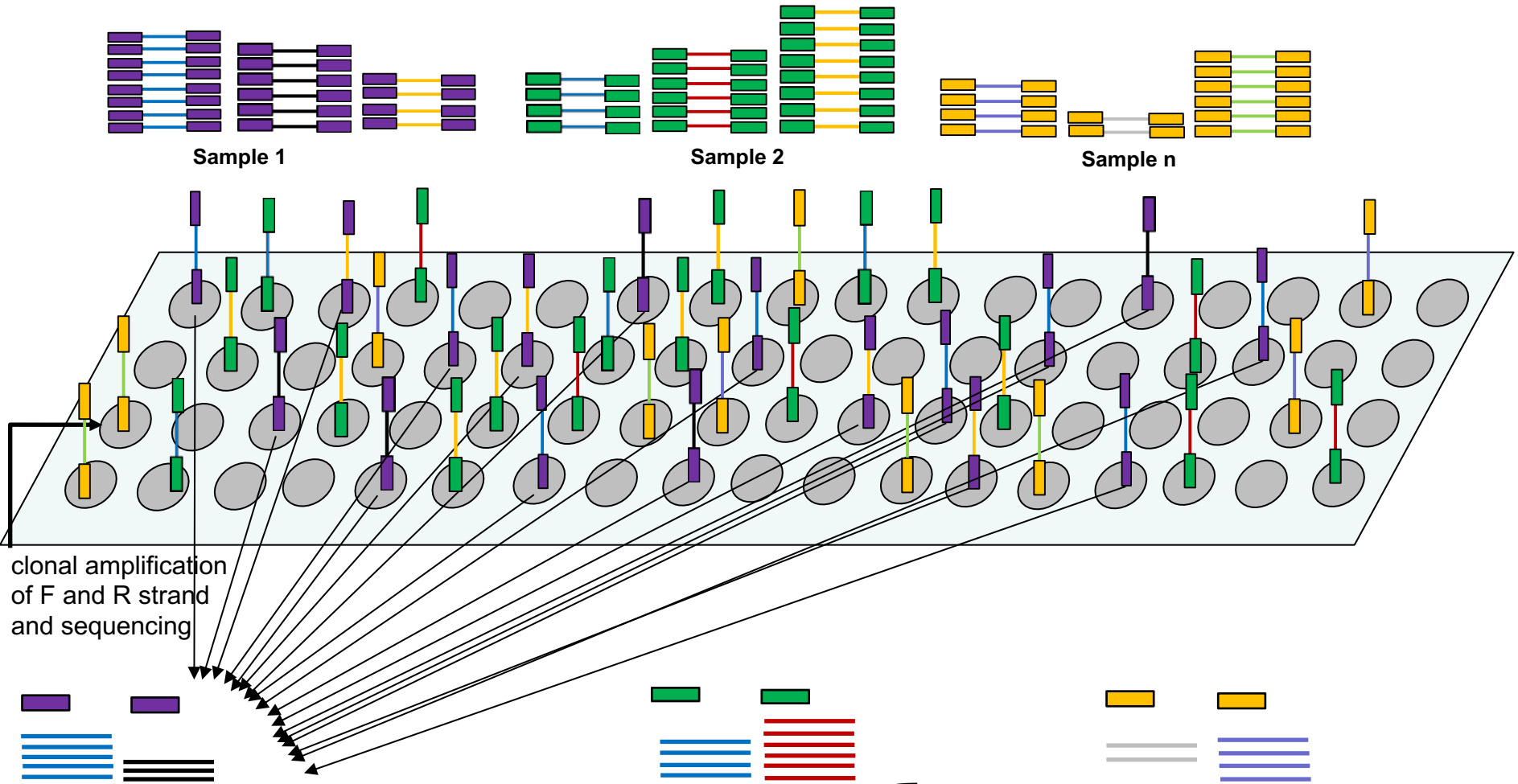
considering that optimal sequencing require 10x coverage of the target DNA, the extreme sequencing power of the Illumina platforms allow multiplexing, sequence on the same flow cell up to 384 different **barcoded samples**



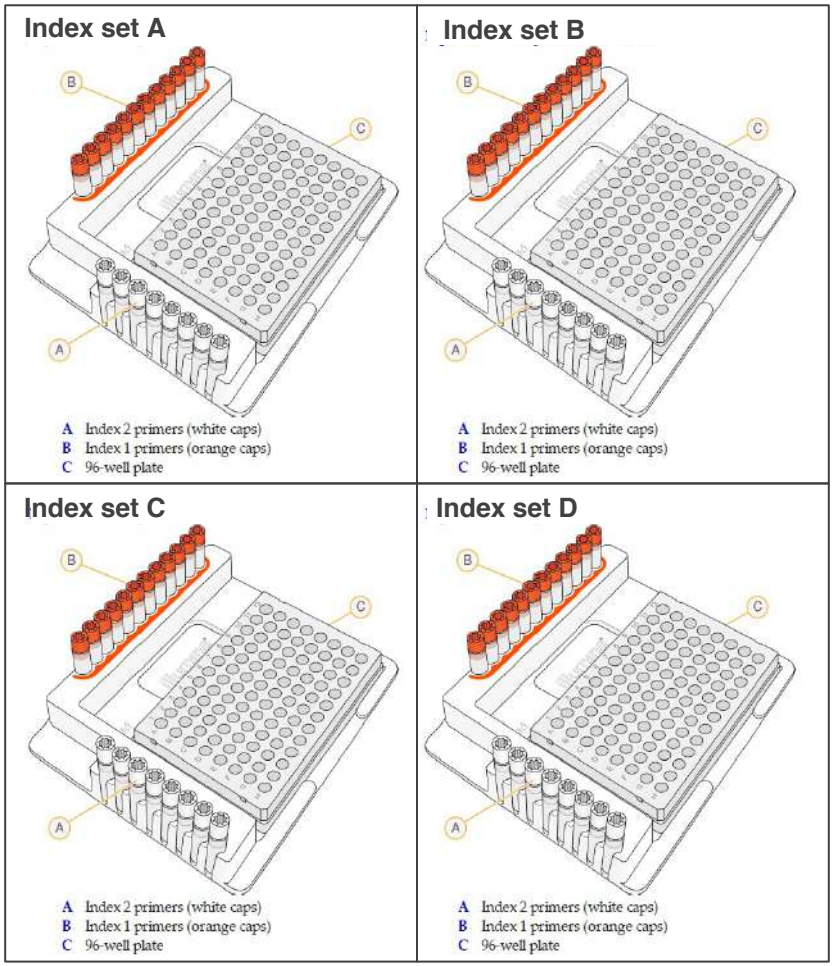
METAGENOMES BARCODING FOR MULTIPLEXING



FLOW CELL MULTIPLEXING, DE-MULTIPLEXING AND QUANTIFICATION, GENERAL CONCEPT



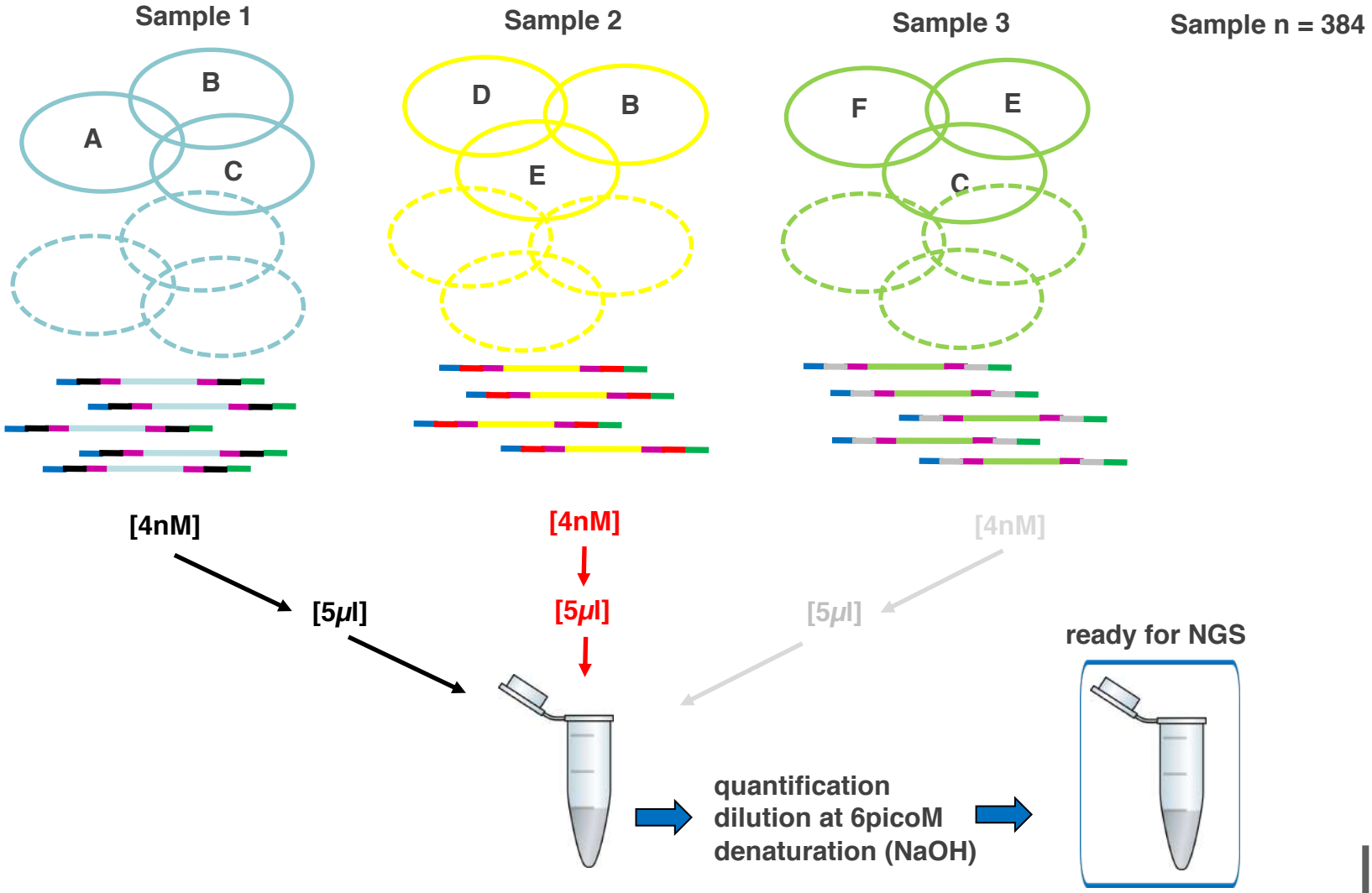
METAGENOME SEQUENCING LIBRARY PREPARATION



each set (A,B,C,D) include specific array of **12 red index-1 L** and **8 blank index-2 R** primers each endowed with a specific index ID, combining in 96 different index PCR for barcoding 96 different matagenomic samples

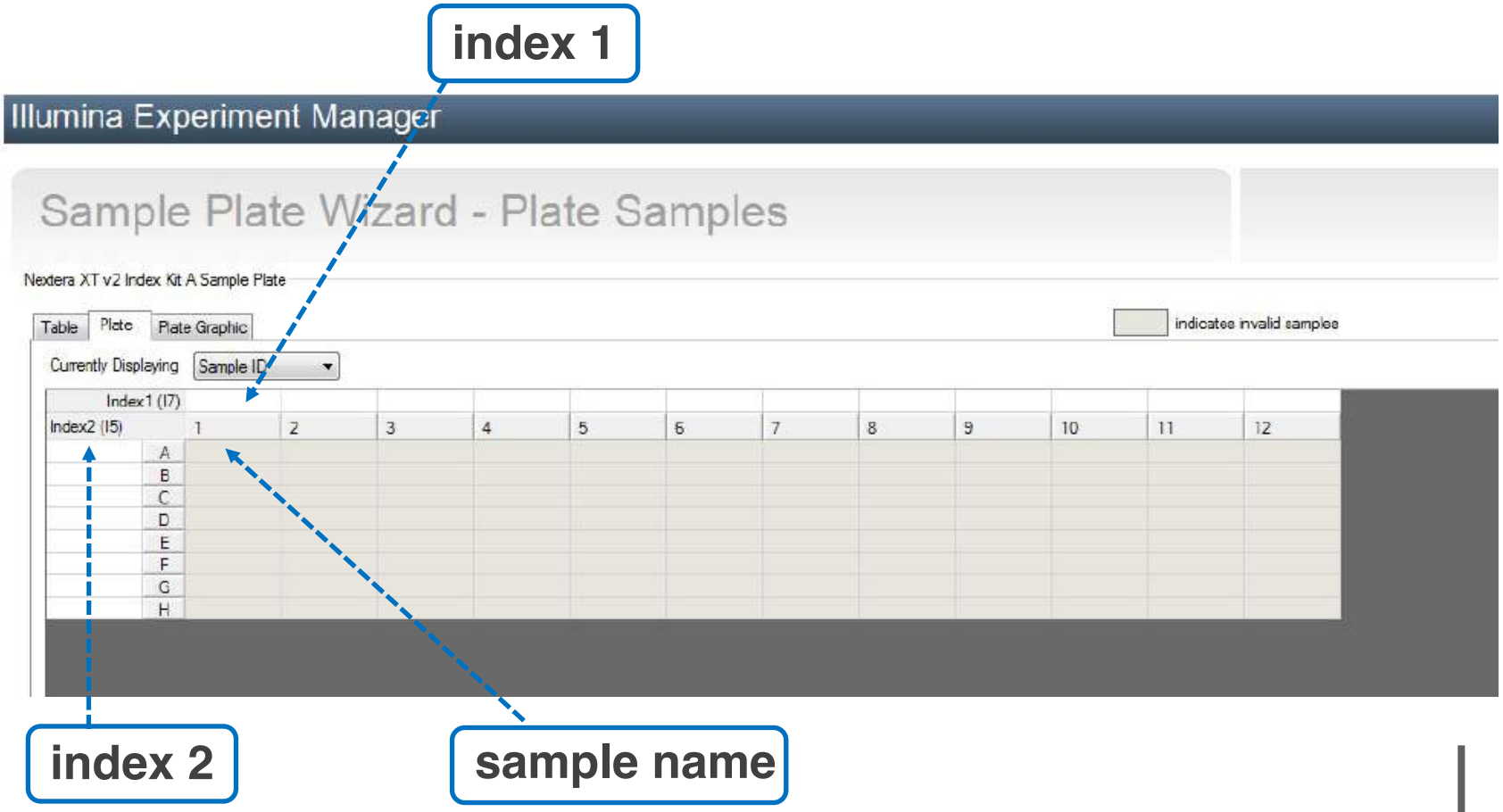
a total of 384 different index PCR for barcoding 384 different metagenomics samples

POOLING


Sample 1
Sample 2
Sample 3
Sample n = 384
A
B
C
D
B
E
F
E
C
[4nM]
[5µl]
[4nM]
[5µl]
[4nM]
[5µl]
ready for NGS
quantification
dilution at 6picoM
denaturation (NaOH)

LOADING A RUN

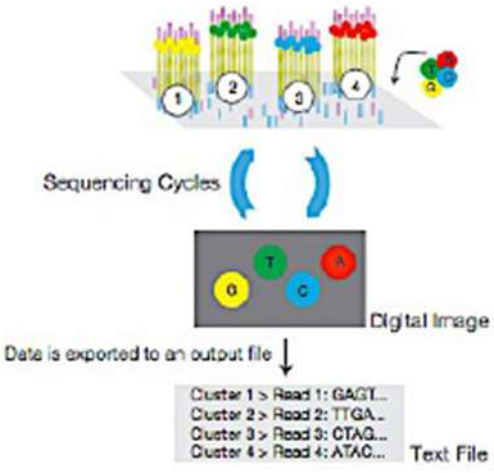
each multiplexed sample is tagged by a peculiar combination of indexes 1 and 2 and a sample name



SEQUENCING OUTPUT FROM THE ILLUMINA PLATFORM

the output of an Illumina run are **demultiplexed reads,** reads are grouped in 2 Fastq files (pair-end 1, pair-end 2) according to the sample of origin

- all pair-end 1 reads, sequenced in the first round, are separated according to coordinates in the flow cell
- all pair-end 2 reads, sequenced in the second round, are separated according to coordinates in the flow cell
- each sequencing cluster - with specific coordinates – has thus a specific set of pair-ends 1 and 2 sequences, according to the specific combination of indexes 1 and 2, respectively
- each sequencing cluster is assigned to a given samples according to index 1 and 2 combination
- assignment of two separate Fastq files, pair-end 1 and pair-end 2, corresponding to each sequencing cluster to matching sample



2 Fastq PER SAMPLE

2 Fastq per sample,
paired-end 1: all reads pair-end 1
paired-end 2: all reads pair-end 2

Fastq_paired-end1

Fastq_paired-end2

Fastq FILE FORMAT

Read

```
@M02887:45:000000000-APDB2:1:1101:14992:1857 1:N:0:6
CCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCCACGCCGCGTGAGTGAAGAAGTATTTCGGTTTGTAAAGCTCTTTCAGCA
GGGAAGAAAATGACGGTACCTCCCTAAGAAGCCCCCGCTATCTACCTGCCACCAGCCGCGGTTTTACGTAGGGGGCAAGCGTTATCCCGAATTTCTGGGTGTAA
AGTGTGCGTAGGTGGTATGGCAAGTCACTAGTGAAAACCCAC
+
AAAAAA>11D1>1AABABFFGGGCFAGH211GFHHHFF/A//GCGGHHGE1F1BG/B>EEFGGEGE/>FG12BFGHE2>2BGBE?/?EHFFG211BF1F>221BB<C/B/
000<F1>1</C//1<1///0?<01101<F.–<–.<=0<000../0;.:CE?@–.-..09;/;;.E.––9;9A-AB//–9-;/;BBBFBB@////9/;:/––-//-9//:9-//-9/////9/;/:;/B?––
```

Header

```
@M02887:45:000000000-APDB2:1:1101:17599:1882 1:N:0:6
```

Sequence

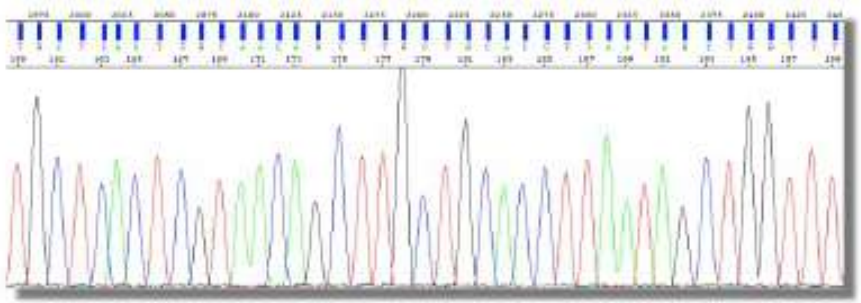
```
CCTACGGGGGGCAGCAGTGGGGAATCTTGCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGCGTGATGAAGTATTTCGGTTTGTAAAGCTCTTTCAGC
AGGGAAGATTCTTACCGTTCCTTCCTAAGAAGCCCCCGCTAACTTCCTTCCCTCCGCCGCCGTTCTTCGTAGGGGGCCTGCGTTTTCCTGCTTTCCTGGGTGTTA
CGGTTCCGTTGGTGGTTTGGCAAGTCCGCTGTGCAAACCCTGG
```

Quality value phred33

```
+
>AA>A11>10/>/??>>BF0AEA<B1FG211FGHHDCA//</<<?CHCF<1?11?//<>@C@CCG@.-CA./;0;CF000;F/-9.;;FBBF009BF0C0;09=.C?––-///:/:9/-:—///
9/9/;////9:9--9—/9//////9-/9-;-9-9--9—//-;-;A—9——;A-:/B////:;//:///-/-9-//—:-/-/:—-9—9//:;9/——///:999-—
```

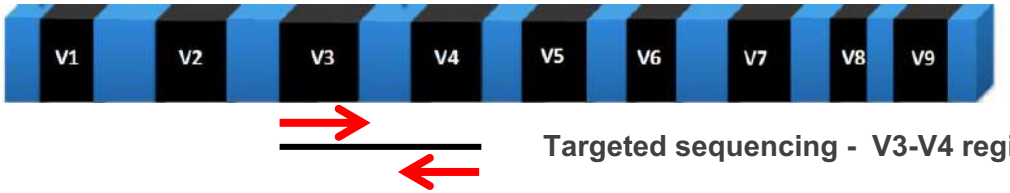
!"#$%&'()*+,-./0123456789:;<=>ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~

Low **High**



SEQUENCE PROCESSING, MARKER GENE APPROACH, THE 16S CASE

WORKFLOW ANALYSIS OF 16S DATA



Targeted sequencing - V3-V4 region of the 16S rDNA

GENEARAL PIPELINE

READS SELECTION	→ Quality trimming
OTUs CLUSTERING	→ 97% identity
FILTERING FOR CHIMERIC SEQUENCES	→ ChimeraSlayer
TAXONOMIC CLASSIFICATION	→ RDP classifier

READS SELECTION

PANDAseq: paired-end assembler for illumina sequences. For each sample:

- joint pair ends
- reads quality filter (10% error threshold)
- reads length filtering (460 bp ± 100 bp)

PANDAseq-output
A single Fastq file for each sample containing the assembled pair ends

SPLIT LIBRARIES

Split libraries:

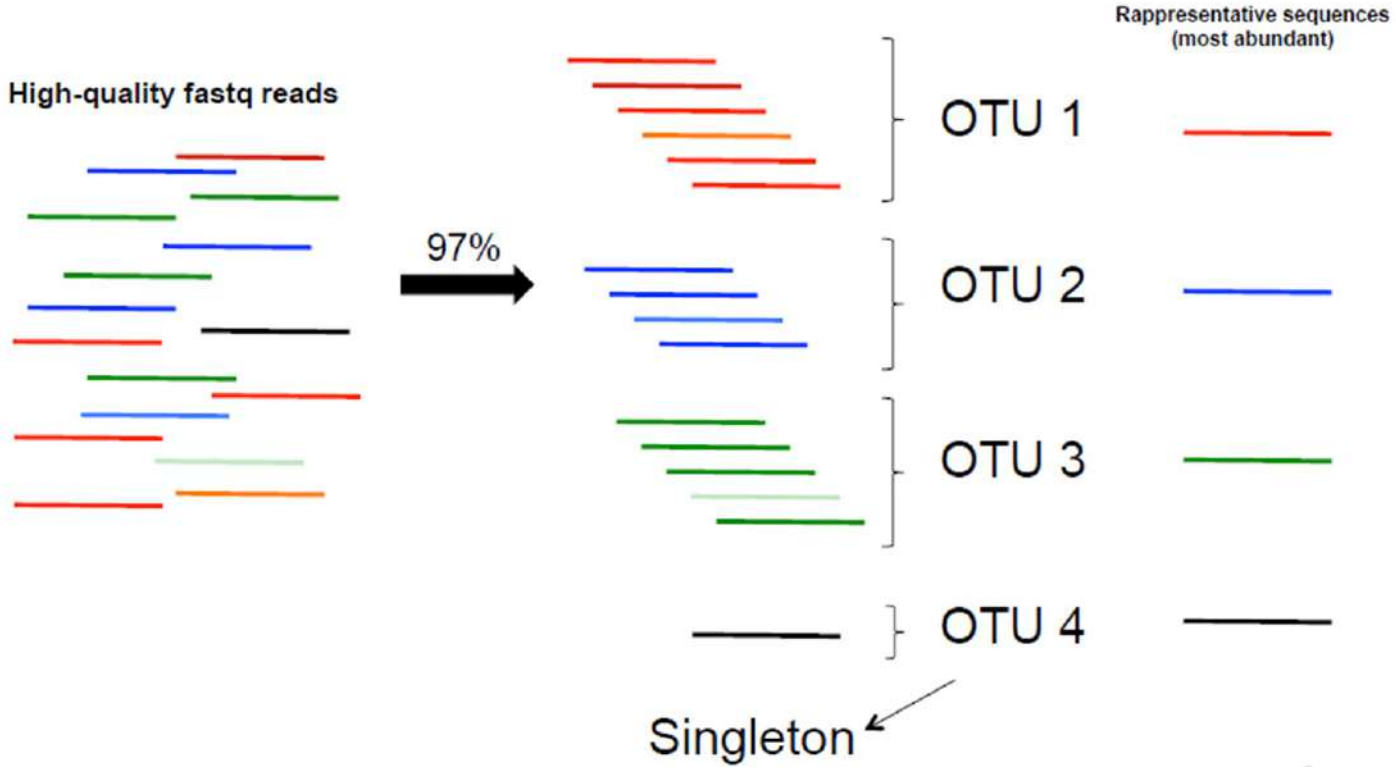
- rename reads according to the sample of origin
- transform Fastaq to Fasta file

for each sample it is created a single Fasta file with all the reads corresponding to the two joined pair ends tagged according to the sample name

FROM NOW SAMPLE TAGGED READS CAN BE PUT TOGETHER IN THE "SAME BASKET"

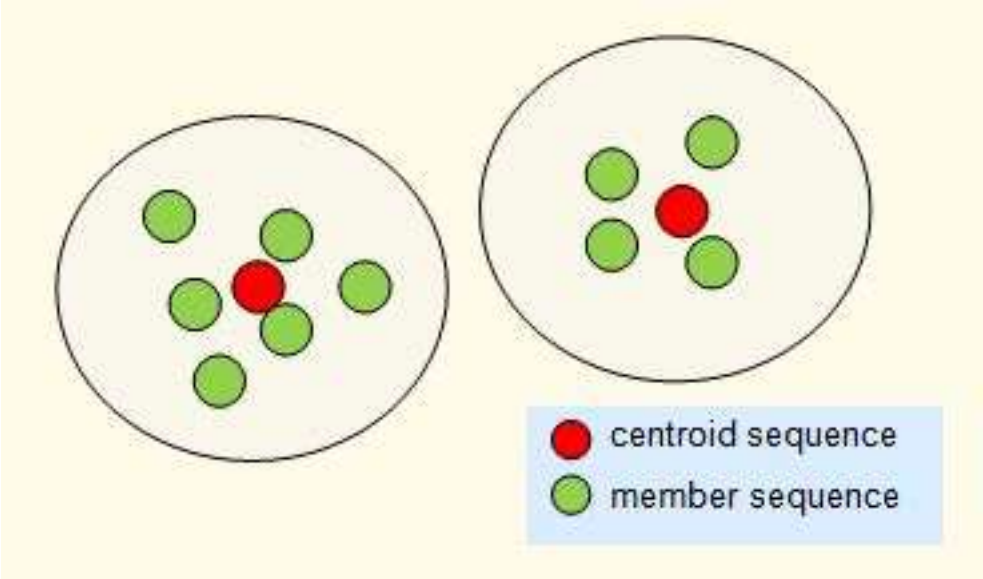
QIIME 1 - THE HISTORC APPROACH - CREATION OF THE OPERATIONAL TAXONOMIC UNITS (OTUs)

sequences from all the samples are clustered into groups sharing a % of similarity thresholds (OTUs). Various threshold of sequence identity are used to represent different taxonomic levels: 97% genus, 95% families ….



OTUs ARE OPERATIONAL AND NOT BIOLOGICAL ENTITIES

cluster sequence into OTUs consolidate similar sequences (97% ID) into single feature, merging sequence variants including those introduced by sequence errors into a single OTUs



this method **misses subtitle and real biological sequence variation** that would be consolidated in a single OTU

OTU PICKING

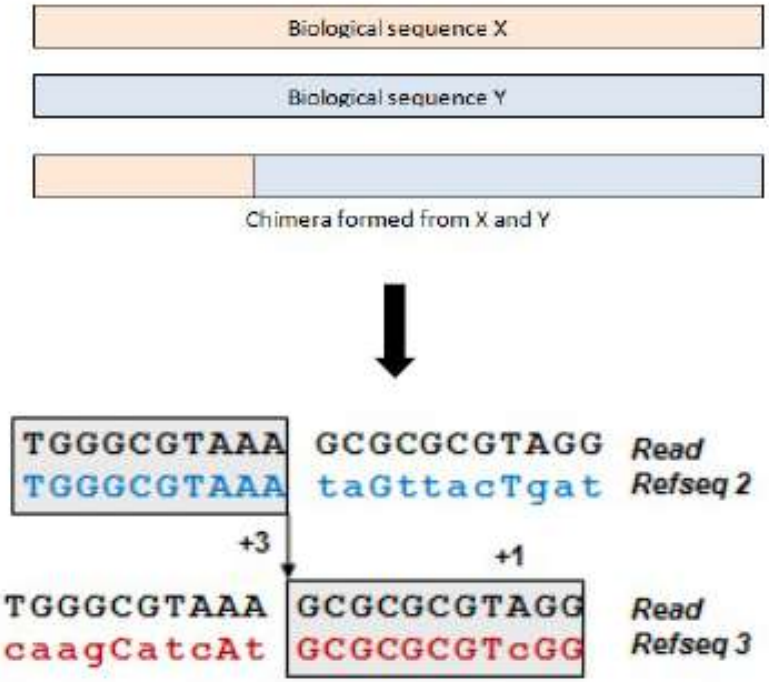
UCLAST – UQIIME (more recent)

de novo OTUs picking, sequences are clustered into OTUs without any external reference

clustering at 97%

filtering for chimera sequences by Chimera Slayer, remove singleton

Chimeric sequences



Chimeras are common in amplicon sequencing where closely related sequences are amplified. The majority of chimeras are believed to arise from incomplete extension in PCR. During subsequent cycles, a partially extended strand can bind to a template derived from a different but similar sequence. This then acts as a primer that is extended to form a chimeric sequence

QIIME 2, CREATION OF THE AMPLICON SEQUENCE VARIANTS (ASVs)

algorithms such as “DADA2” use error profiles to resolve sequence data into **exact sequence features called ASVs**. The resulting output from this method is a table of all DNA sequences from the whole sample set and counts of these different sequences per sample

DADA2

correction of Illumina errors

↓

filtering chimeric sequences

↓

filtering for low quality reads

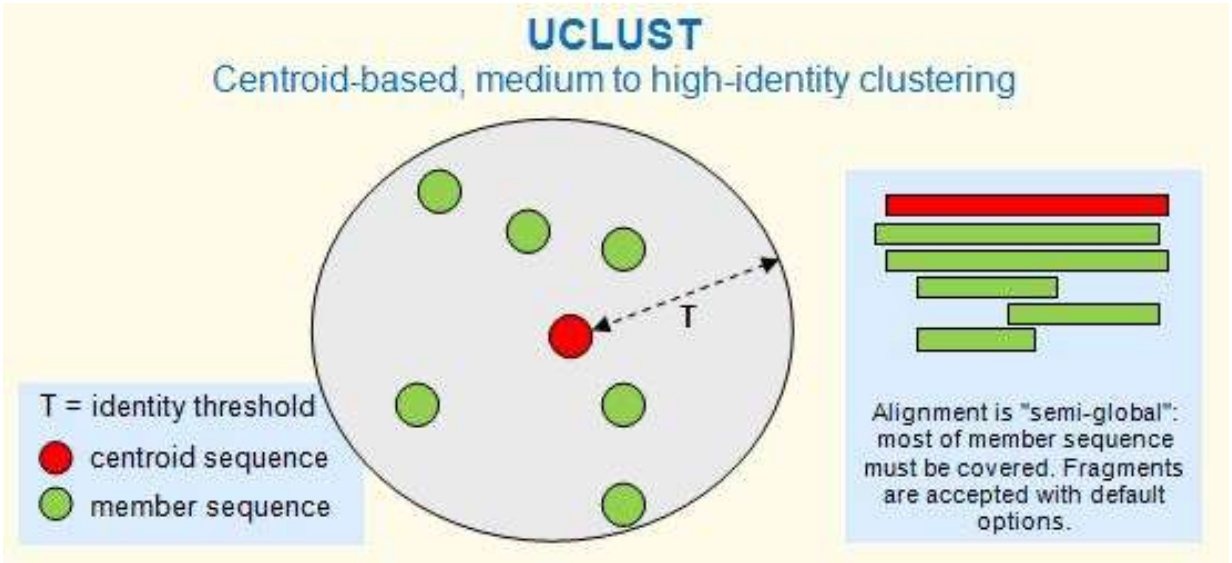
↓

creation of high resolution ASVs

QIIME 2, CREATION OF OTUs FROM THE ASVs

THE ALGORITHMS VSERACH – AN EVOLUTION OF UCLAST – IS USED FOR CLUSTERING ASVs IN 97% OTUs

ASVs can be clustered at each level of identity (eg. 95% for family; 90% for the order)



QIIME 2, TAXONOMIC ASSIGNEMENT OF ASVs or OTUs

the algorithms VSEARCH is also used for the for the ASVs or OTUs taxonomic assignment

A PARSIMONIOUS APPROACH

FOR EACH ASVs or OTUs the most represented seq is selected:

- ✓ Identification of the annotated sequences sharing 80% of homology
- ✓ selection of the best matching 100 sequences
- ✓ annotation according to the taxonomic identity matching the 51% of the best matching sequances

DATABASES FOR TAXONOMY ASSIGEMENT, OTUs and ASVs

selection of a **representative sequence** for each OTUs or ASVs

taxonomic assignment
(alignment against a reference database by VSEARCH)

green genes
(fast and optimized for the human GM)

GREENGENES
The 16S rRNA Gene Database and Tools

SILVA
(more complete, mandatory for env. samples but slow)

silva ngs

for each OTUs or ASVS a **taxonomic label and an ID** code in provided, the ones lacking a taxonomic designation are collapsed into unclassified bacteria. OUTs or ASVs assigned to mitochondria and chloroplasts can be filtered out

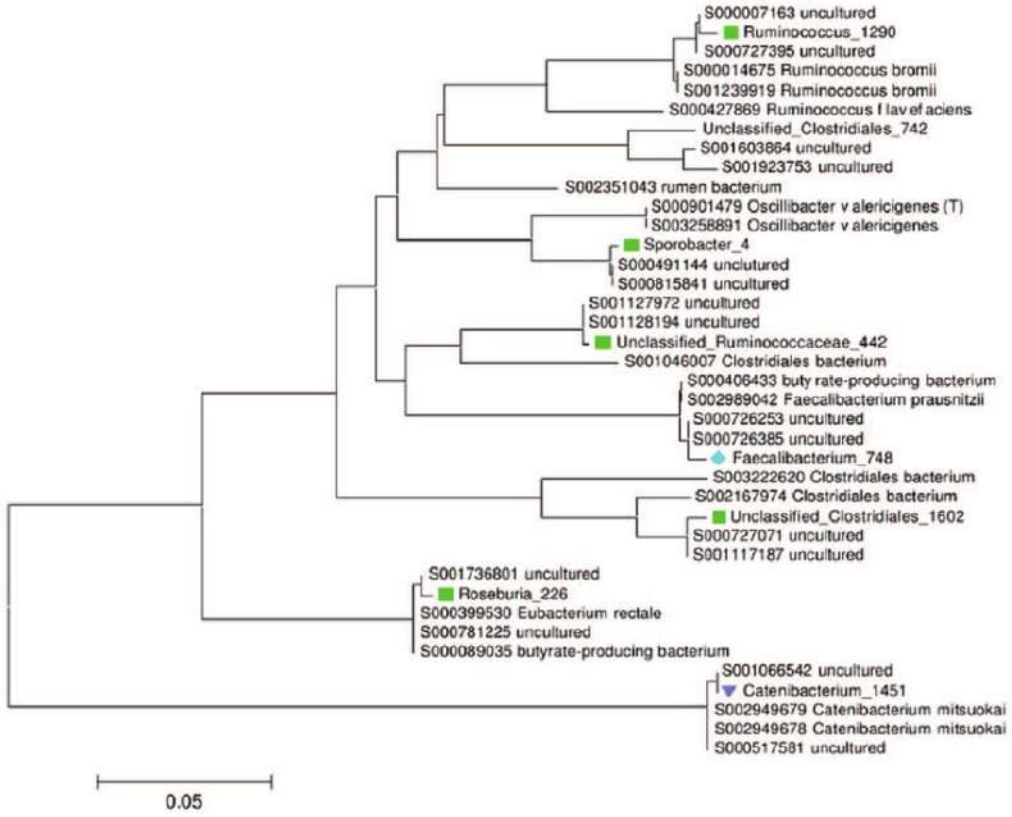
OTUs and ASVs ASSIGEMENT PREDICTIVE METHOD

assignment can be traditionally performed basing on homology, new predictive methods have been developed, allowing the assignment by applying predictive methods - **the SKLearn algorithm**

- ✓ by Artificial Intelligence (AI), tags - as repeated motives – characteristic for each taxa are identified;
- ✓ the taxonomic identification depend on the presence of the tags characteristic for each taxa

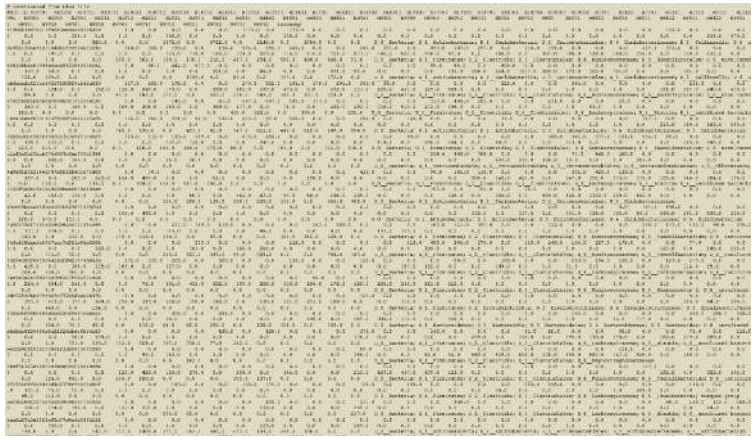
CONSTRUCTION OF A PHYLOGENETIC TREE

a **phylogenetic tree** of all the represented OTUs or sub-OTUs in the whole samples set is created



each high quality Illumina read is now tagged with the corresponding samples name and an OTUs/AVSs ID corresponding to taxonomy

CREATION OF A PROJECT OUT/sub-OTU TABLE



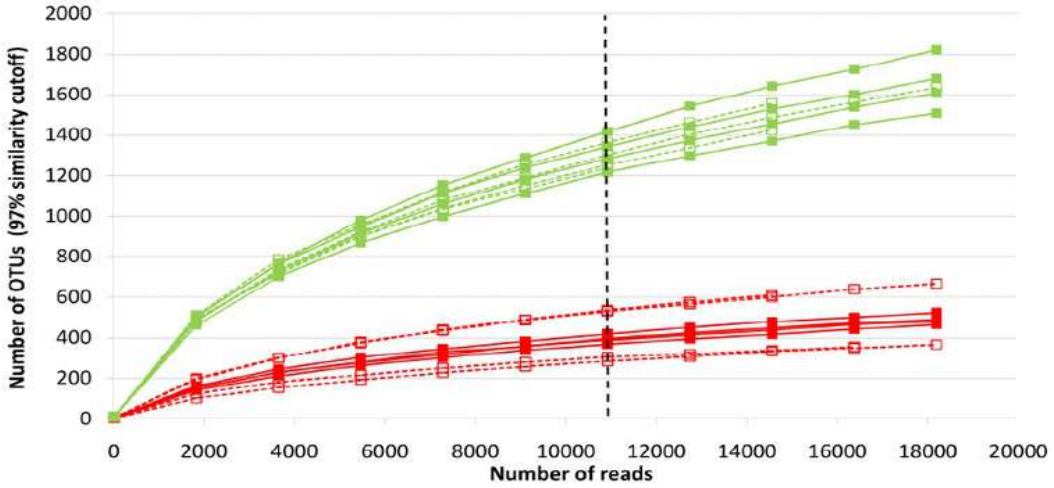
project OUT/sub-OTU table

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	…….	Taxonomy
OTU ID	reads count	……						*….Bifidobacterium*
OTU ID	……							
OTU ID								
OTU ID								
OTU ID								
OTU ID								
OTU ID								
OTU ID								
OTU ID								
……..								
	total reads count							

NORMALIZATION AND REREFACTION

Normalization
normalization of the OTUs table according to the lowest acceptable number of reads per sample = X. For each sample, X reads are stochastically selected and assigned to an OTU ID to product a **normalized OUT table**.

Multiple rarefaction
stochastic reads selection at multiple rarefaction levels and count of on the number of OTUs. If the rarefaction at X reads achieve **the plateau**, the majority of the reads diversity is kept

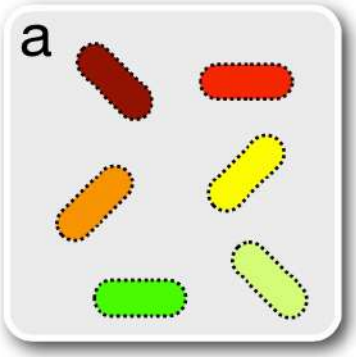


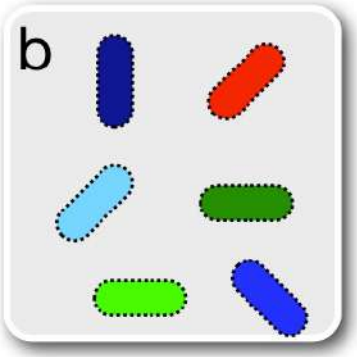
rarefaction curve

Diversity in compositional studies

How to describe and compare diversity?

- α diversity: How many taxa are in a sample?
  - e.g., 6 taxa (colors) in **a** and 6 in **b**
  - *e.g., Are polluted environments less diverse than pristine?*
- β diversity: How many taxa are shared between samples?
  - e.g., 2 shared taxa (colors) between **a** and **b**
  - *e.g., Does the gut microbiota differ between people with and without irritable bowel disease?*

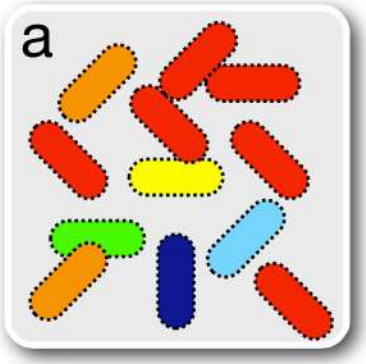


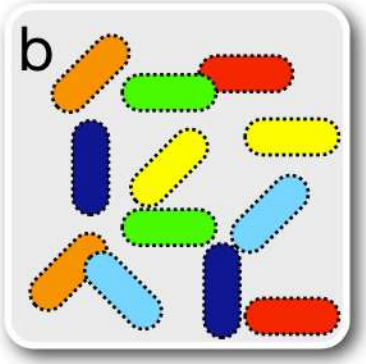


Diversity II

Qualitative versus quantitative measures

- Qualitative: Considers presence/absence only
  - α: how many in each? 6 taxa in both **a** and **b**
  - β: how many shared? all, the samples are identical
- Quantitative: Also considers relative abundance
  - α: accounts for evenness; **b** is more diverse than **a**
  - β: samples are considered more similar if the same taxa are numerically dominant vs. rare; **a** and **b** are no longer identical due to differences in abundance

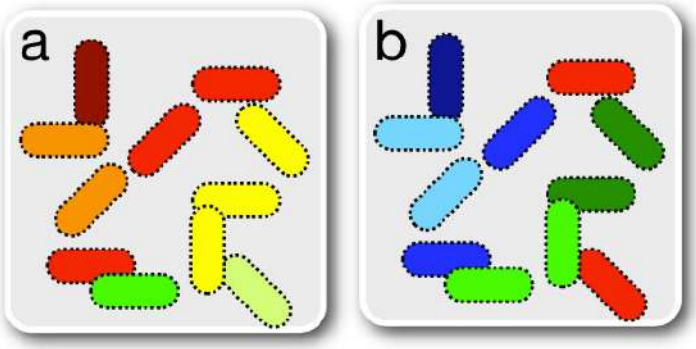




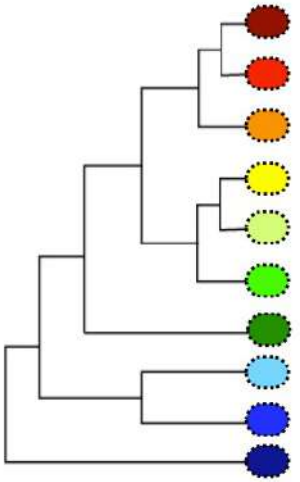
Phylogenetic Diversity

What is a *phylogenetic* diversity measure?

- α diversity (*within* samples):
  - given a phylogeny, how much divergence is in a sample?
  - e.g., **b** is more diverse than **a**; more divergent lineages

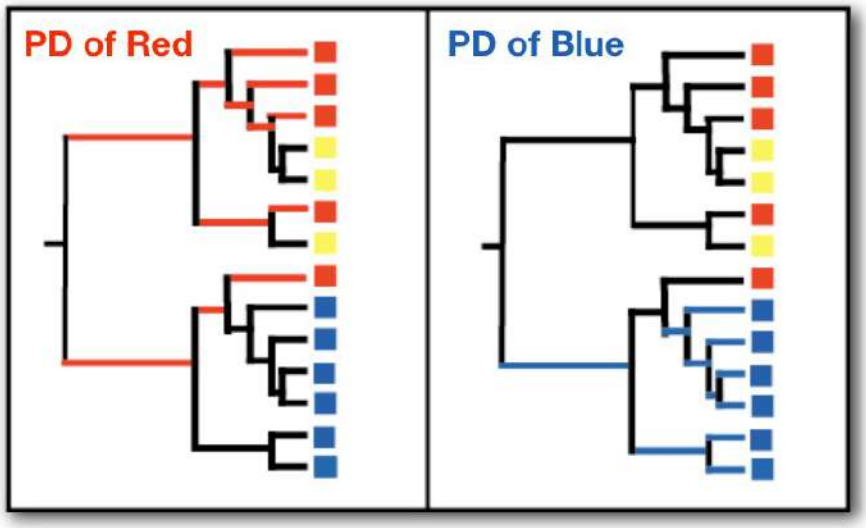


- β diversity (*between* samples):
  - given a phylogeny, how much divergence is shared between samples?
  - e.g., lineages in **b** are distantly related to lineages in **a**



Alpha Diversity
Phylogenetic Diversity (Faith's PD)

Phylogenetic Diversity (PD)



- Sum of branches leading to sequences in a sample; a qualitative measure of α diversity

- Sample with lineages spanning the most branch length in tree contains the most phylogenetically (and perhaps functionally) diverse community

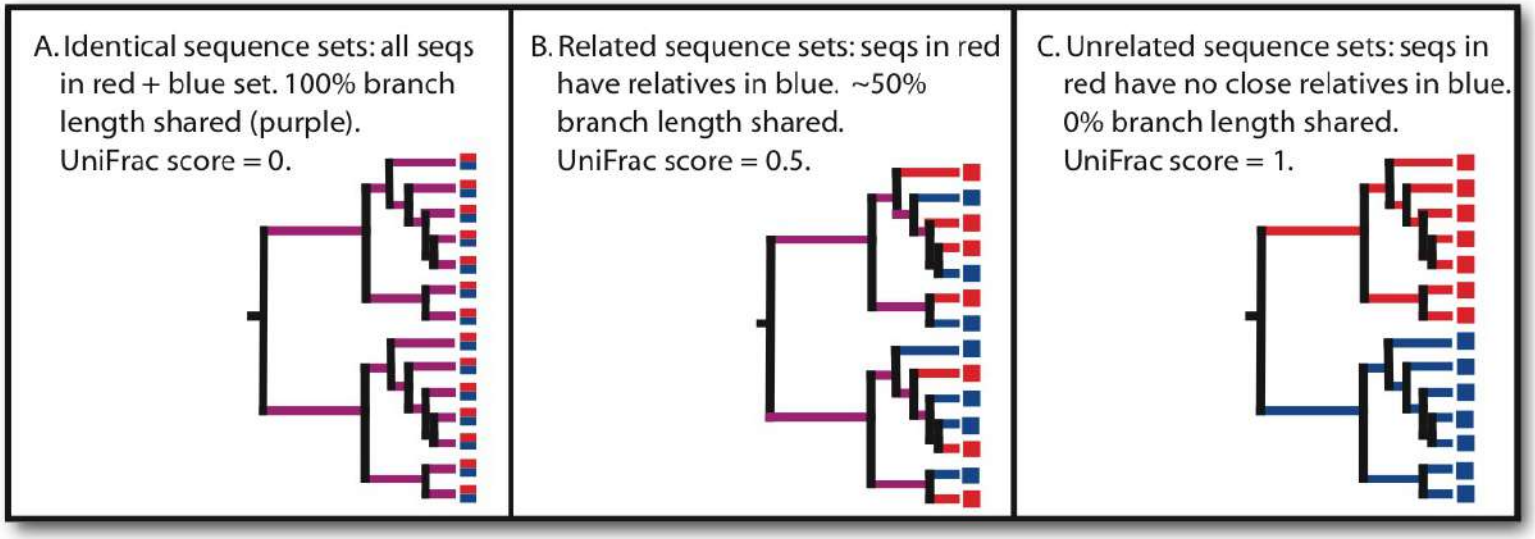
ALPHA DIVERSITY INDEXES

- n of observed OTUs
- Faith's phylogenetic diversity
- Shannon Index
- Chao Index
- Simpson Diversity Index, the only one accounting for evenness and this indicated for relative abundance values

Beta-Diversity: UniFrac

Unique Fraction (UniFrac) metric

- A branch length-based, qualitative phylogenetic β diversity measure
- Distance = fraction of the total branch length that is unique to any sample



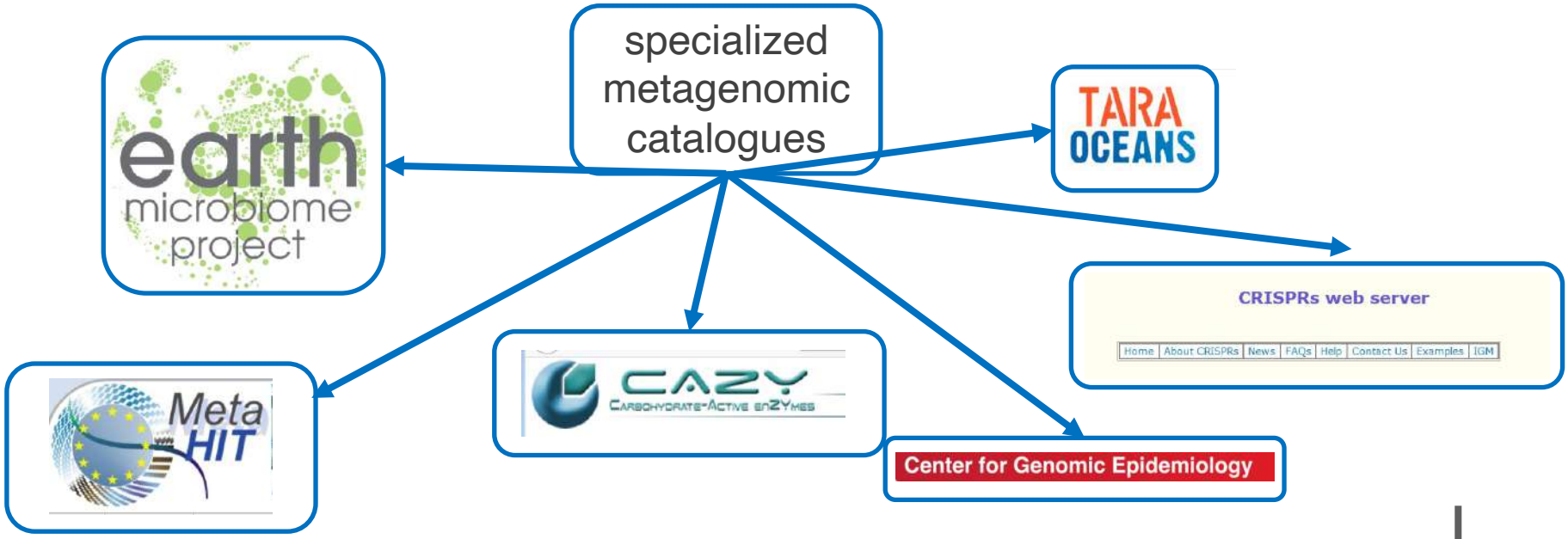
VADEMECUM
for a Bacterial Ecology BioStat Analysis

1. Structure
2. Alpha – Diversity
3. Beta – Diversity

METAGENOMICS, READ BASED PROFILING

read mapping takes the unassembled DNA sequence reads and compares them against a reference database to assign taxonomy and annotation genes

for well characterized environments are already available **curated genome databased**, from poor characterized environments the use of large comprehensive databases needs to be considered (eg. NCBI):



ASSEMBLY FREE METAGENOMICS, MAIN LIMITATIONS

the main limitation is that previously **uncharacterized microbes** is difficult to profile

the diversity of reference genomes available for some sample types - as the human gut – is now extensive enough to make assembly free taxonomic profiling efficient

analysis of more diverse environments – soil and oceans – is hampered by lack of reference genomes

the accuracy of assembly free metagenomics will improve as **more reference genomes** and high quality metagenomics assemblies become available

ASSEMBLY FREE METAGENOMICS, PAPELINE

Filtering the metagenomics reads

- quality
- length > 60 bp
- filtering out the host sequences

↓

READS MAPPING AGAINST REFERENCE DATABASES

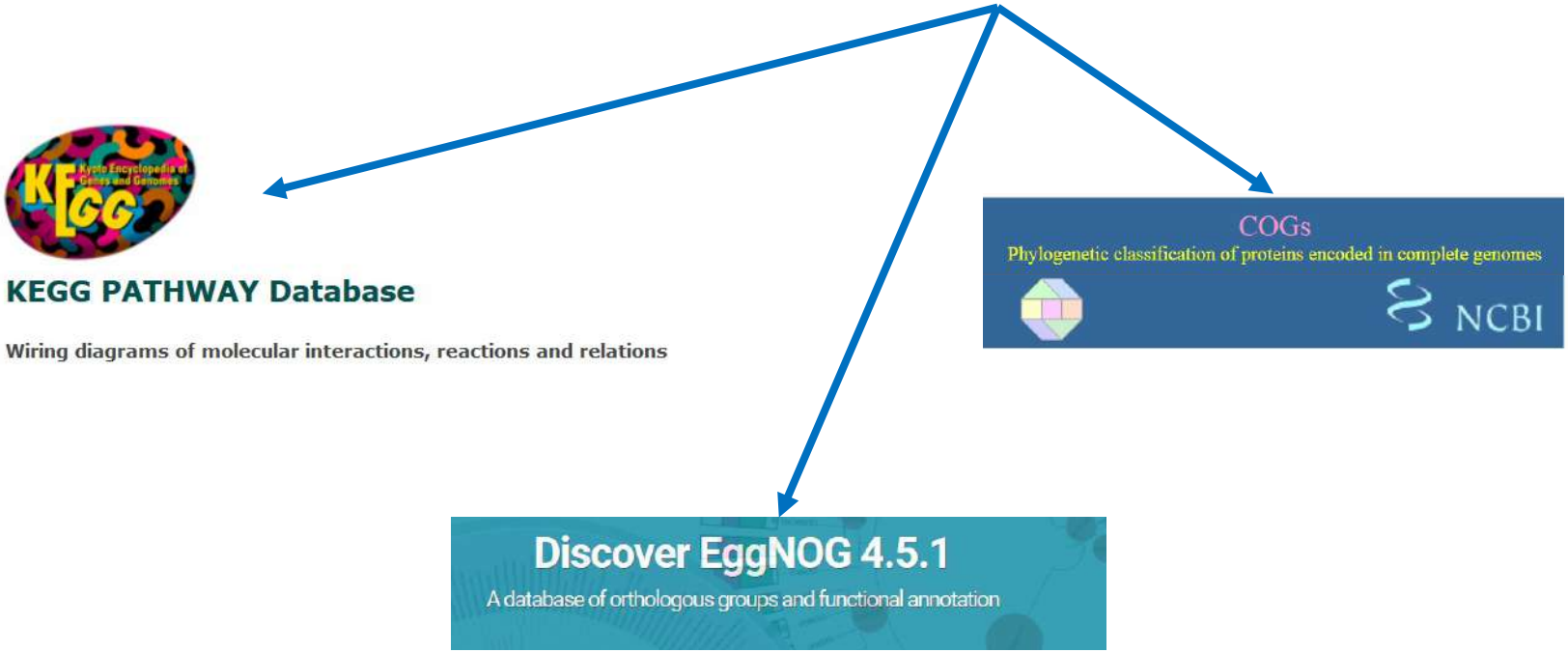
Taxonomic composition

Functional composition

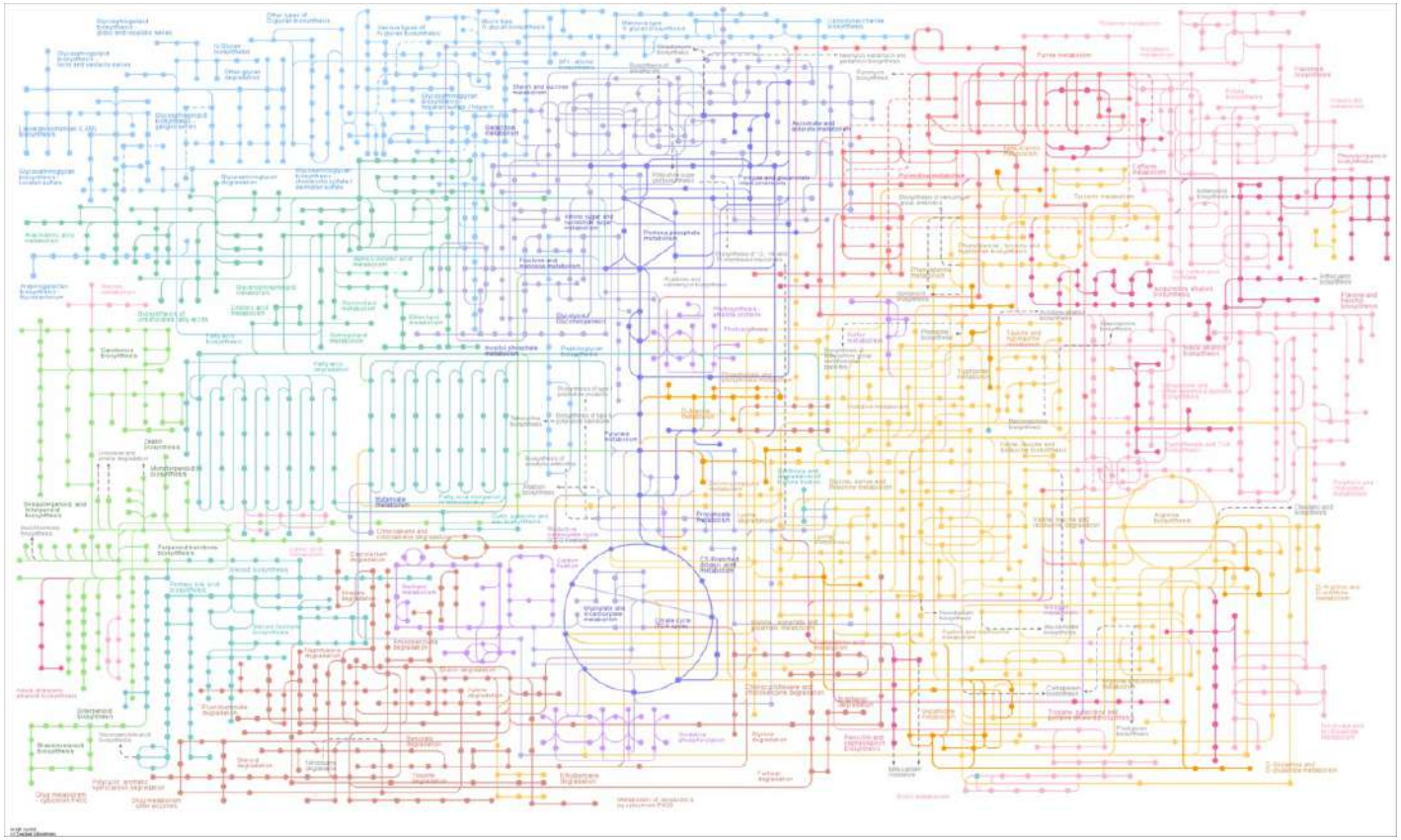
MetaCV, HISTORICAL ALGORITHM FOR FUNCTIONAL ASSIGNEMENT

MetaCV

MetaCV is a composition and phylogeny-based algorithm to classify very short metagenomic reads (75-100 bp) into specific taxonomic and functional groups.

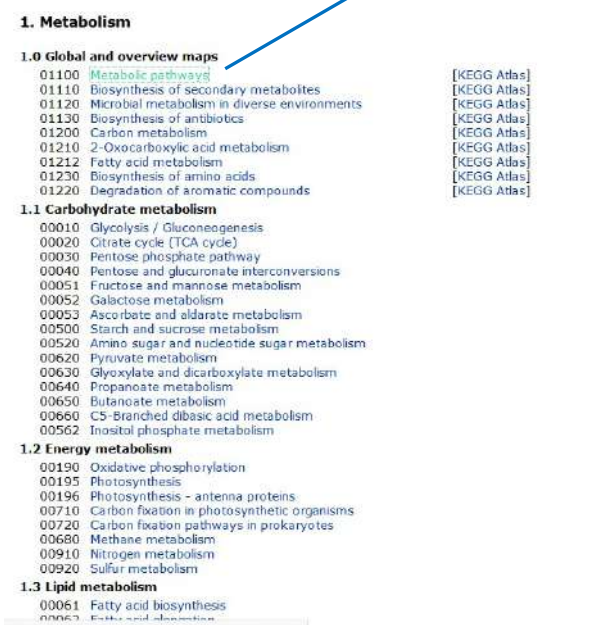


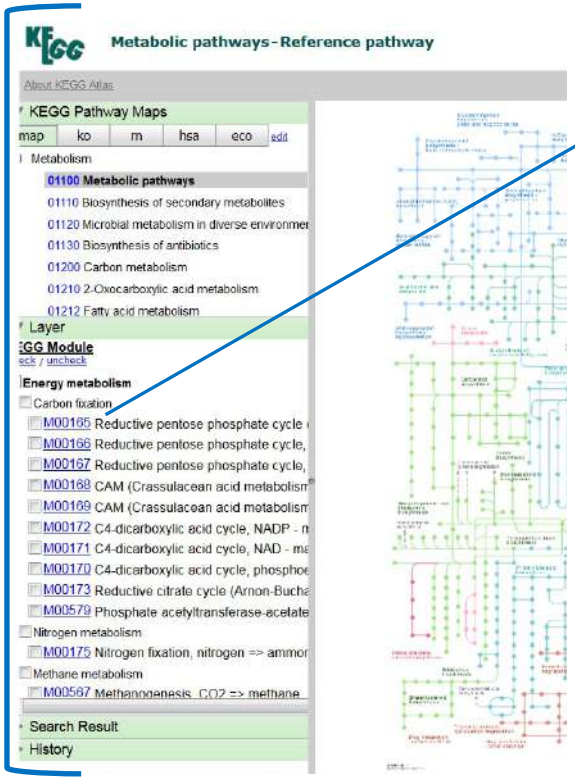
KEGG PATHWAY DATABASE, THE ONLY ONE HIERARCHIAL

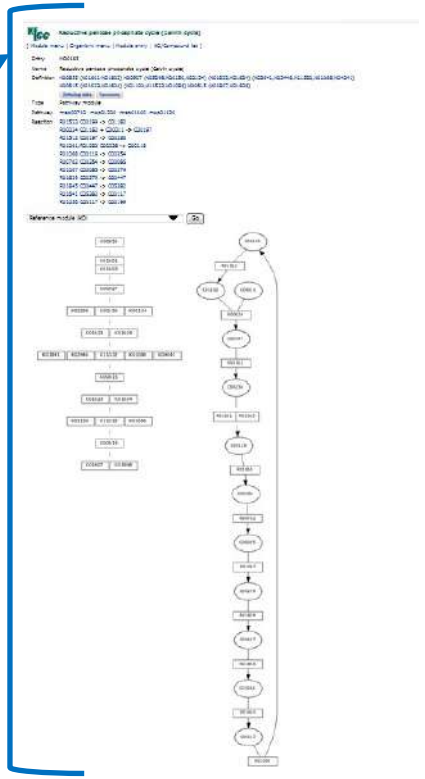


single protein families are aggregated into higher level metabolic pathways and functional modules

KEGG, HIERARCHIAL MAP OF LIFE PROCESSES AND PRODUCTS







bioBakery, THE MOST USED ALGORITHM FOR FUNCTIONAL AND TAXONOMIC ASSIGNEMENT

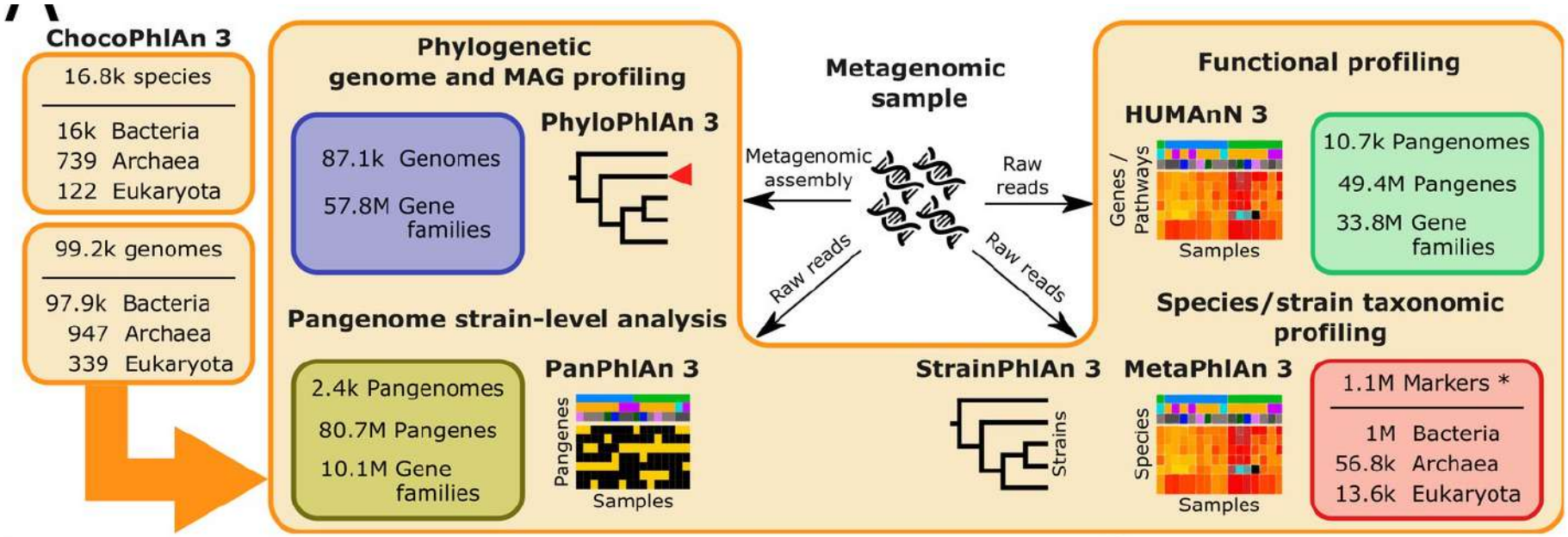
eLife TOOLS AND RESOURCES

Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3

Francesco Beghini[1†], Lauren J McIver[2†], Aitor Blanco-Miguez[1], Leonard Dubois[1], Francesco Asnicar[1], Sagun Maharjan[2,3], Ana Mailyan[2,3], Paolo Manghi[1], Matthias Scholz[4], Andrew Maltez Thomas[1], Mireia Valles-Colomer[1], George Weingart[2,3], Yancong Zhang[2,3], Moreno Zolfo[1], Curtis Huttenhower[2,3*], Eric A Franzosa[2,3*], Nicola Segata[1,5*]

[1]Department CIBIO, University of Trento, Trento, Italy; [2]Harvard T.H. Chan School of Public Health, Boston, United States; [3]The Broad Institute of MIT and Harvard, Cambridge, United States; [4]Department of Food Quality and Nutrition, Research and Innovation Center, Edmund Mach Foundation, San Michele all'Adige, Italy; [5]IEO, European Institute of Oncology IRCCS, Milan, Italy

bioBakery3 Integrates different tools for taxonomic and functional assignment



Human3, PAPERLINE FOR FUNCTIONAL ASSIGNEMENT

Bowtie 2, an alignment tool optimized for short reads

MATAGENOMIC READS

1 TAXONOMIC MAPPING (MetaPhlan)

2 GENERATION OF A DATABASE OF CIRATED GENOMES OF THE SPECIES IDENTIFIED (ChocoPhlan)

3 READS MAPPING AGANIS THE GENERATED DATABASE

diamond

DIAMOND

reads translation to AA sequence and alignment against a protein database

4 UNMAPPED READS ARE MAPPED AGAINS UNIREFF90 (a protein database from Uniprot with clusters at 90%)

OUTPUT FILE

```
#sample4:       C007_S5_metacv_out.res  6455854
#sample5:       C010_S29_metacv_out.res 4979914
#sample6:       C011_S26_metacv_out.res 3611638
#sample7:       C012_S30_metacv_out.res 4363524
#sample8:       C017_S25_metacv_out.res 6717174
#sample9:       C018_S24_metacv_out.res 5838704
#sample10:      C021_S2_metacv_out.res  4536184
#sample11:      C024_S28_metacv_out.res 4927190
#sample12:      C042_S6_metacv_out.res  5456614
#sample13:      C048_S4_metacv_out.res  6359168
#sample14:      C052_S31_metacv_out.res 4112028
#sample15:      C054_S1_metacv_out.res  3372110
#sample16:      K100_S14_metacv_out.res 3458300
#sample17:      K105_S17_metacv_out.res 6334566
#sample18:      K106_S15_metacv_out.res 4522916
#sample19:      K108_S19_metacv_out.res 5740208
#sample20:      K113_S10_metacv_out.res 2560170
#sample21:      K119_S20_metacv_out.res 5777134
#sample22:      K124_S18_metacv_out.res 4075222
#sample23:      K125_S12_metacv_out.res 7896674
#sample24:      K300_S8_metacv_out.res  5670976
#sample25:      K301_S16_metacv_out.res 3624672
#sample26:      K303_S13_metacv_out.res 4431994
#sample27:      K304_S9_metacv_out.res  7756536
#sample28:      K306_S11_metacv_out.res 7142380
#sample29:      S010_S1_metacv_out.res  7004482
#sample30:      S020_S2_metacv_out.res  5750220
#sample31:      S030_S3_metacv_out.res  7325858
#sample32:      S050_S4_metacv_out.res  7004368
#sample33:      S080_S5_metacv_out.res  7060442
#sample34:      S110_S7_metacv_out.res  5154174
#sample35:      S120_S8_metacv_out.res  8495976
#sample36:      S130_S9_metacv_out.res  4766878
#sample37:      S140_S10_metacv_out.res 7327644
#sample38:      S150_S11_metacv_out.res 6248770
#sample39:      S180_S12_metacv_out.res 7355476
#sample40:      S190_S13_metacv_out.res 3377250
#sample41:      S200_S14_metacv_out.res 3691300
#sample42:      S210_S15_metacv_out.res 7271778
#sample43:      S220_S16_metacv_out.res 6262062
#sample44:      S240_S17_metacv_out.res 6327276
#sample45:      S260_S18_metacv_out.res 4891384
#sample46:      S280_S19_metacv_out.res 3514292
#sample47:      S290_S20_metacv_out.res 6255352
#sample48:      S300_S21_metacv_out.res 3857810
#sample49:      S320_S22_metacv_out.res 4062988
#sample50:      S330_S23_metacv_out.res 6734654
#sample51:      T12_S27_metacv_out.res  3528590
```

output-reads

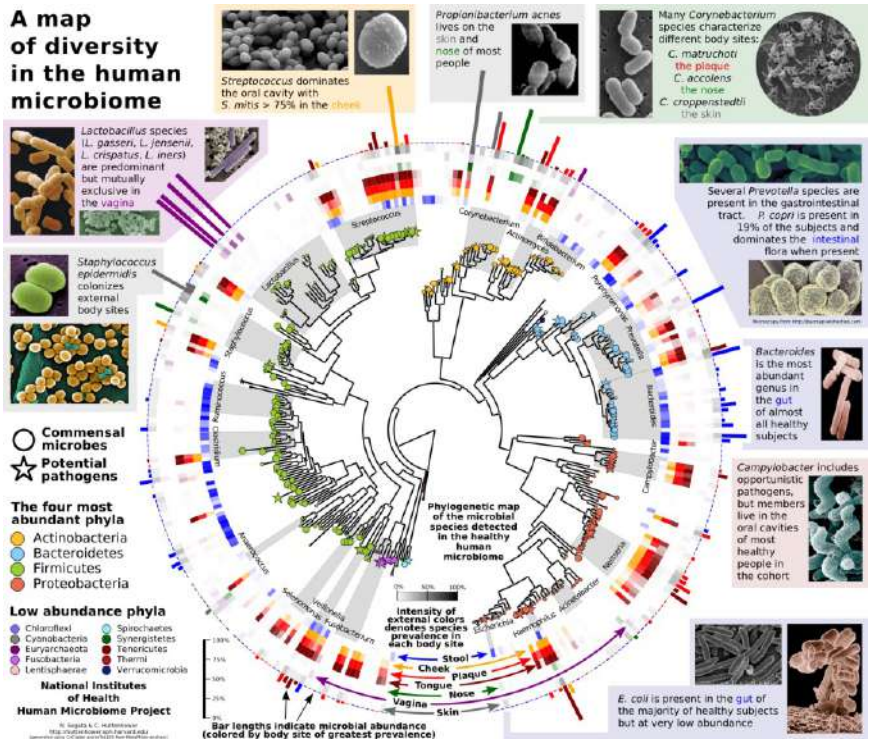
output-functions

```
#table[1] num of reads per KEGG level 1
level1  sample1 sample2 sample3 sample4 sample5 sample6 sample7 sample8 sample9 sample10        sample11        sample12        sample13        sample14        sample15
        sample28        sample29        sample30        sample31        sample32        sample33        sample34        sample35        sample36        sample37        sample38
mple50          sample51        sample52        sample53        sample54        sample55        sample56        sample57        sample58        sample59        sample60        sa
L1_1    80803   114491  110033  68147   101335  75876   147114  77189   72134   95263   137179  102729  68440   178031  148959  128326  120702  77551   71222   71485   67915
9983  95494   159029  125839  89933   119952  82210   84347   157450  82796   95357   87673   92563   125383  104393  113044  101515  Metabolism
L1_2    1917    3039    2454    1689    2054    1578    3570    2042    1644    2534    2989    2587    1558    3559    3804    3645    3096    1981    1905    1981    1614
38    2234    4376    2680    2365    2963    2383    1943    4120    2104    2374    2244    2674    3514    2756    3171    2870    Organismal Systems
L1_3    19803   33633   39419   18819   25743   18144   41060   19127   20734   25982   41461   28985   18192   67719   34828   41317   33526   21400   17996   19379   14621
970   30754   50373   50490   23515   39437   22869   21861   45563   22154   25652   23743   24562   35568   28387   26203   24868   Environmental Information Processing
L1_4    4088    7757    5903    3364    5571    4124    6959    4330    4184    4872    6359    5583    3532    11267   8233    9926    6307    4788    5077    6712    3906
38    5864    8639    5280    4523    8581    3674    4474    8080    4552    5363    6192    5057    8131    6539    6046    6061    Cellular Processes
L1_5    29943   42173   50411   25631   38925   30415   51465   28751   27850   32907   51900   38164   27747   92133   58467   50140   44013   31077   27344   32011   26013
907   39145   61082   58018   30763   54007   26661   31980   53692   31169   34800   32466   33941   47161   39033   43811   38011   Unclassified
L1_6    2532    3390    3602    2166    3492    2608    4466    2582    2430    3159    4689    3228    2314    5826    4227    3698    3705    2878    2606    2501    2430
97   2946   4180   3263   2941   3853   2670   2736   4254   2716   2943   2837   2836   3643   3137   3548   2934   Human Diseases
L1_7    49522   65769   63439   39621   61058   42037   85202   46755   44472   57291   83498   63506   42279   92560   79394   73134   71250   46026   46424   46700   40215
962   52958   92430   59448   55591   67326   52301   52043   88624   50271   53851   52213   53449   71056   59583   63220   56813   Genetic Information Processing

#table[2] num of reads per KEGG level 2
level2  sample1 sample2 sample3 sample4 sample5 sample6 sample7 sample8 sample9 sample10        sample11        sample12        sample13        sample14        sample15
        sample28        sample29        sample30        sample31        sample32        sample33        sample34        sample35        sample36        sample37        sample38
mple50          sample51        sample52        sample53        sample54        sample55        sample56        sample57        sample58        sample59        sample60        sa
L2_01   21948   31900   30020   18779   25388   20097   39040   20504   20895   26972   37070   26481   17452   51186   40633   35849   32241   22331   18443   20090   17794
488   27579   40984   37898   26145   32831   23443   22738   43028   23013   26438   24428   26018   35214   29229   32567   28720   Carbohydrate Metabolism
L2_02   4898    8491    7463    4432    6712    4746    11493   4485    4979    6871    9647    7127    3711    13209   11404   9809    9540    4828    4074    4703    3965
431   6375   13276   9687   6244   8450   5999   5954   13110   5261   6472   5568   6736   9154   7453   8057   7152   Lipid Metabolism
L2_03   28325   37659   36170   23454   35197   26518   48073   27151   24391   32595   45488   34396   24365   51286   47205   41487   40320   26853   25623   24563   24570
243   32491   51487   39334   30798   39331   27524   28014   50692   28458   32616   29934   29934   40676   34604   36183   32831   Amino Acid Metabolism
L2_04   9012    13945   12708   7715    12437   9107    18238   8655    7397    10342   15383   12203   7663    21926   19377   15527   11529   8592    7978    8227    8016
304   11562   19834   14559   9540    14663   8619    9628    19682   9726    11463   10243   10775   15107   12601   13211   11673   Metabolism of Cofactors and Vitamins
L2_05   3173    3996    4895    2788    3543    2821    4896    2871    2941    3446    4877    4064    2696    7462    4899    4257    4357    2936    2616    2348    2436
37   3818   5699   6550   3809   4496   3424   3282   5083   2942   3315   3145   3263   4240   3705   3885   3465   Xenobiotics Biodegradation and Metabolism
L2_06   1673    3370    2394    1618    2270    1664    4853    1798    2040    2402    2972    2311    1281    4696    4337    3978    3254    1932    1565    1821    1248
03   2354   3653   3264   1960   2905   1990   2320   5354   2236   2640   2566   2681   3973   2933   3648   3013   Biosynthesis of Other Secondary Metabolite
L2_07   16930   18964   18717   13235   17623   14622   22695   15076   13463   16779   20502   17840   14873   27415   24294   20659   19433   15035   14161   13098   13297
509   17205   22688   21493   16386   20381   14668   15252   22809   15471   17055   15872   15696   21116   18339   20602   17875   Energy Metabolism
```

ALGORITHM FOR TAXONOMY ASSIGNEMENT

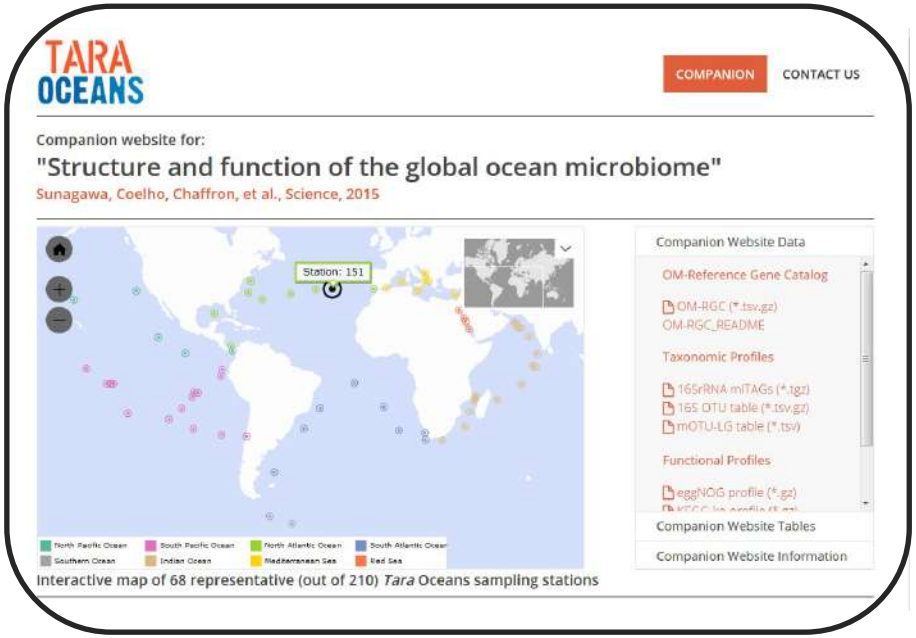
MetaPhlAn3 is a computational tool for profiling the composition of microbial communities from metagenomic shotgun sequencing data.



relies on unique **clade-specific marker genes** identified from 3,000 reference genomes,

OCEAN

ocean microbiomes contains the **largest microbial reservoir** from the planet. At least 35.000 prokaryotes species and 40 M non redundant novel genes



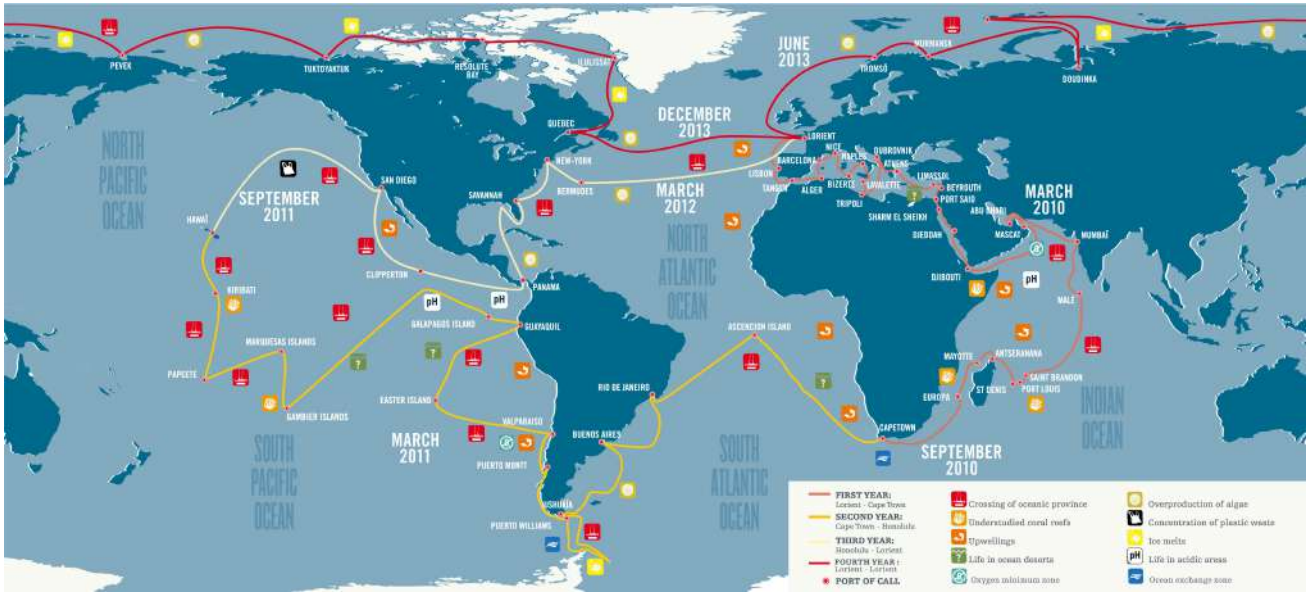
interactive webpage showing microbiome composition across oceans sites

K strategists:
oligotrophic microorganisms whose metabolism is adapted to low nutrient [ ]

R strategists:
copiotropic microorganism whose exhibit phases of rapid growth in nutrient rich condition but are outcompeted by k strategists in nutrient poor conditions

TARA OCEANS

launched in September 2009, the schooner's 8th and 9th expeditions (Tara Oceans and Tara Oceans Polar Circle) has been a **three year voyage around the world**, with fifty stopovers. Its purpose has been to investigate planktonic and coral ecosystems in the perspective of climate changes. 150 international scientists have taken part.



TARA EXPEDITIONS

Tara is a unique ship for scientific discovery and adventure, **still sailing and collecting samples around the world,** see https://oceans.taraexpeditions.org/en/m/about-tara/

MICROBIAL ECOLOGY IN THE OCENAS

Sunagawa et al., Science 2015

Special Issue ***TARA* OCEANS**

INTRODUCTION TO SPECIAL ISSUE

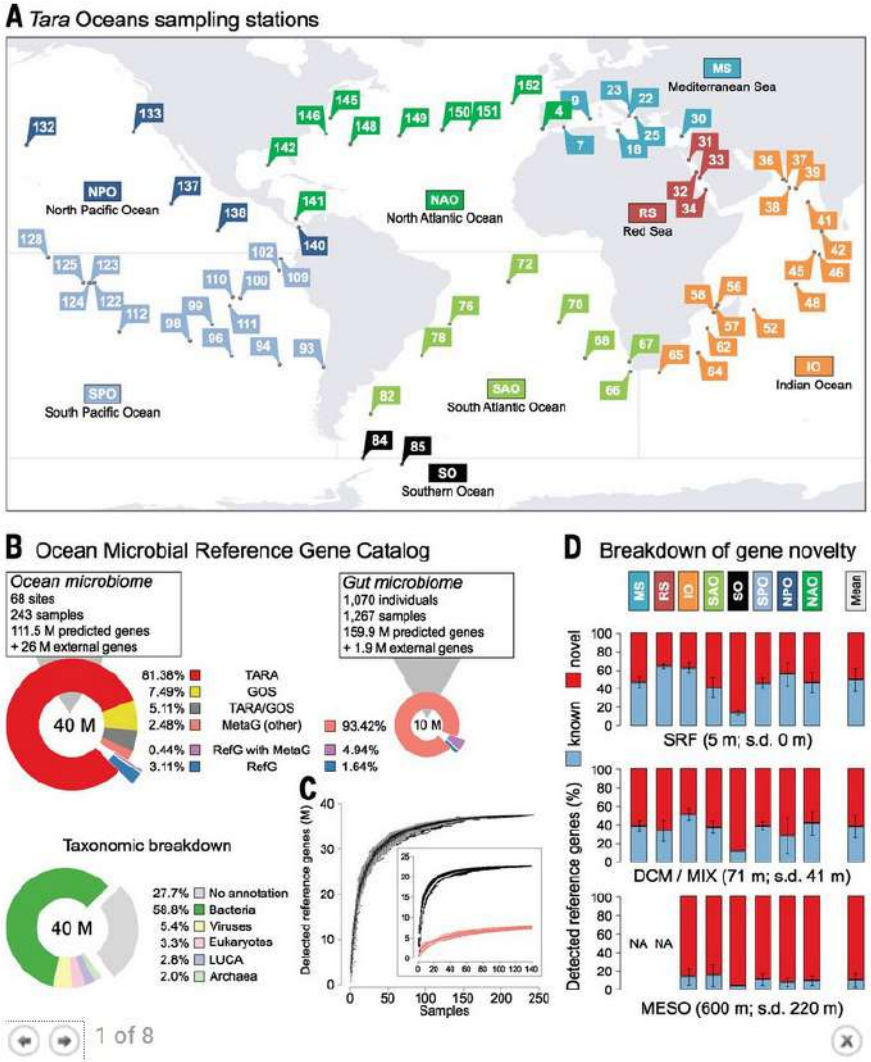
Tara Oceans studies plankton at planetary scale

BY P. BORK, C. BOWLER, C. DE VARGAS, G. GORSKY, E. KARSENTI, P. WINCKER

SCIENCE | 22 MAY 2015 : 873 |

Full Text PDF

analysis of 243 ocean microbiome samples, collected at **68 locations representing all main oceanic regions** (except for the Arctic) from **three depth layers**, which were subjected to metagenomic Illumina sequencing. By integrating these data with those from publicly available ocean metagenomes and reference genomes, we assembled and annotated a reference gene catalog, which we use in combination with phylogenetic marker genes to derive **global patterns of functional and taxonomic microbial community structures**



DIVERSITY

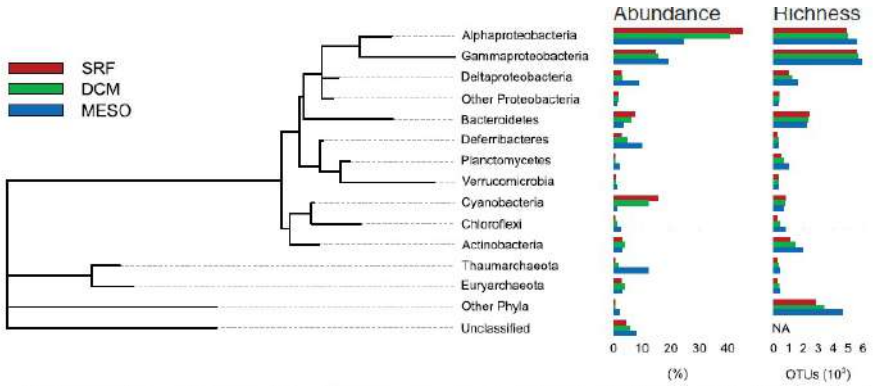


Fig. 2. Taxonomic breakdown of *Tara* Oceans samples. A phylum-level (class-level for Proteobacteria) breakdown of relative abundances is shown for all prokaryotic samples from three depth layers along with the number of detected taxa at the OTU level. SRF, surface water layer; DCM, deep chlorophyll maximum layer; MESO, mesopelagic zone.

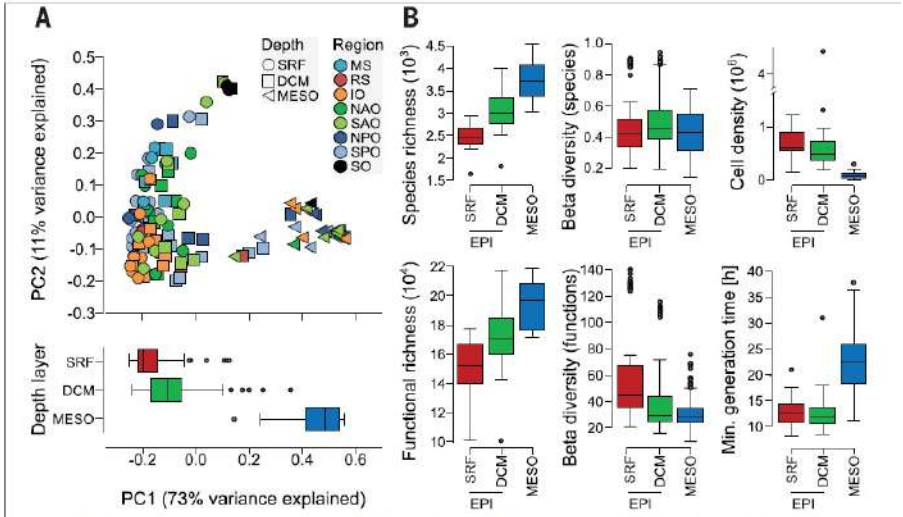


Fig. 3. Depth stratification of the ocean microbiome. (**A**) Principal coordinate (PC) analysis performed on community composition dissimilarities (Bray-Curtis) of 139 prokaryotic samples based on 16S $_{m}$tag relative abundances shows that samples are significantly separated by their depth layer of origin, i.e., surface (SRF), deep chlorophyll maximum (DCM), or mesopelagic (MESO). Boxplots of the first PC illustrate differences between depth layers. Differences between samples from SRF and DCM were significant, but small compared to those with mesopelagic samples. Abbreviations for ocean regions are the same as in Fig. 1. (**B**) For a matched sample set from 20 stations where SRF, DCM, and MESO were sampled, calculations of within-sample species richness (top left) and between-sample diversities (top-center; Bray-Curtis) and cell densities per milliliter (top right) suggest an increase in species richness and a decrease in cell density with depth (pairwise Mann-Whitney U-test: $P < 0.001$), whereas no significant trend was found for between-sample dissimilarity. For gene functional groups (bottom left and center), richness increased with depth, whereas between-sample dissimilarity decreased. Minimum potential generation time of microbial communities (bottom right) is predicted to be higher in the mesopelagic compared to the epipelagic (EPI).

depth is a major determinant of ocean microbiome stratification

ENVIRONMENTAL DRIVERS OF COMPOSITIONAL STRUCTURE

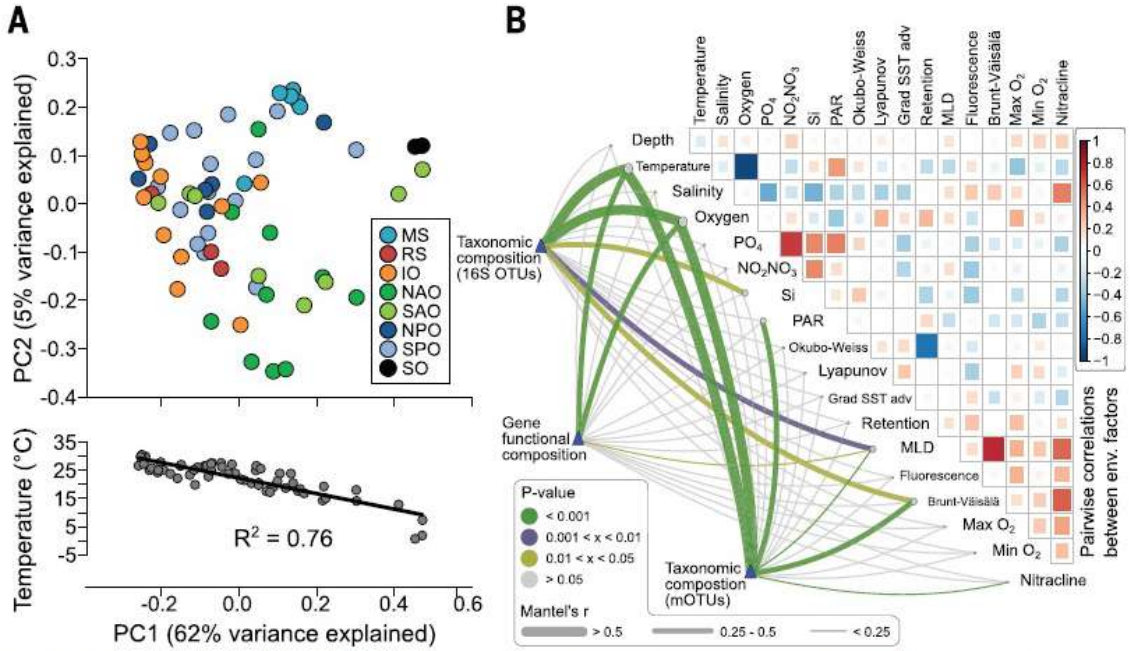


Fig. 5. Environmental drivers of surface microbial community composition. (**A**) Principal coordinate (PC) analysis of surface samples shows that samples are not clearly grouped by their regional origin (top), but rather separated by the local temperatures as shown by the strong correlation (R^2: 0.76) between the first PC and temperature (bottom). (**B**) Pairwise comparisons of environmental factors are shown, with a color gradient denoting Spearman's correlation coefficients. Taxonomic [based on two independent methods: $_{m}$tags (*12*) and mOTUs (*13*)] and functional (based on biochemical KEGG modules) community composition was related to each environmental factor by partial (geographic distance–corrected) Mantel tests. Edge width corresponds to the Mantel's *r* statistic for the corresponding distance correlations, and edge color denotes the statistical significance based on 9,999 permutations.

temperature and oxygen are dominant drivers of variation in surface water microbiomes

FUNCTIONAL STRUCTURING

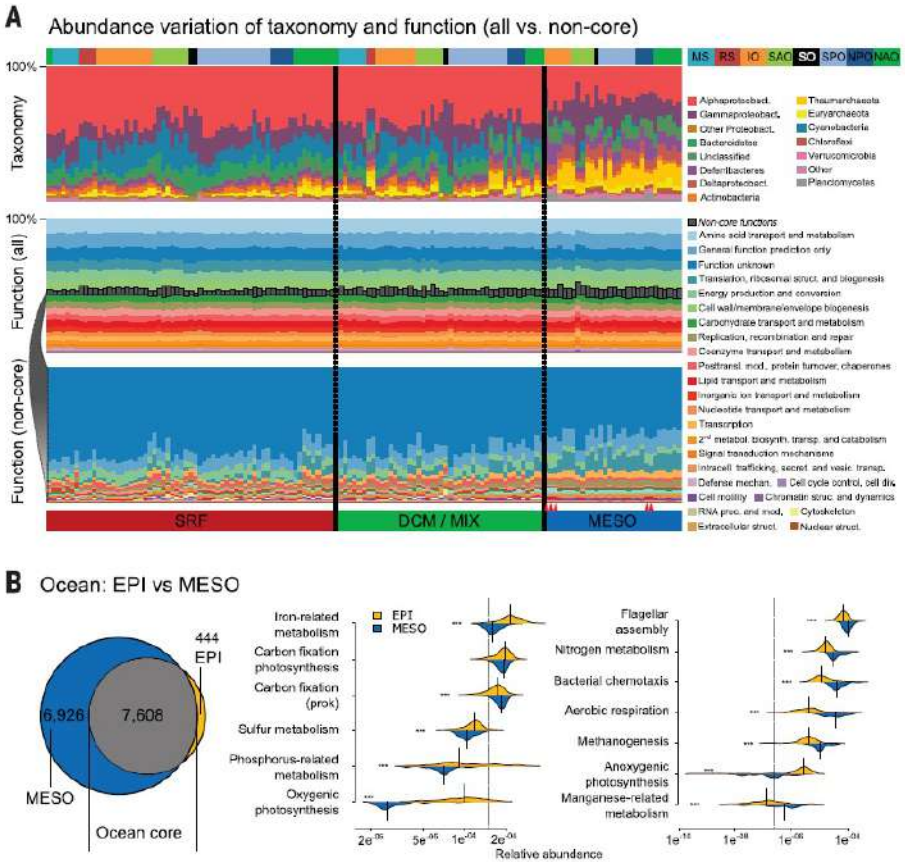
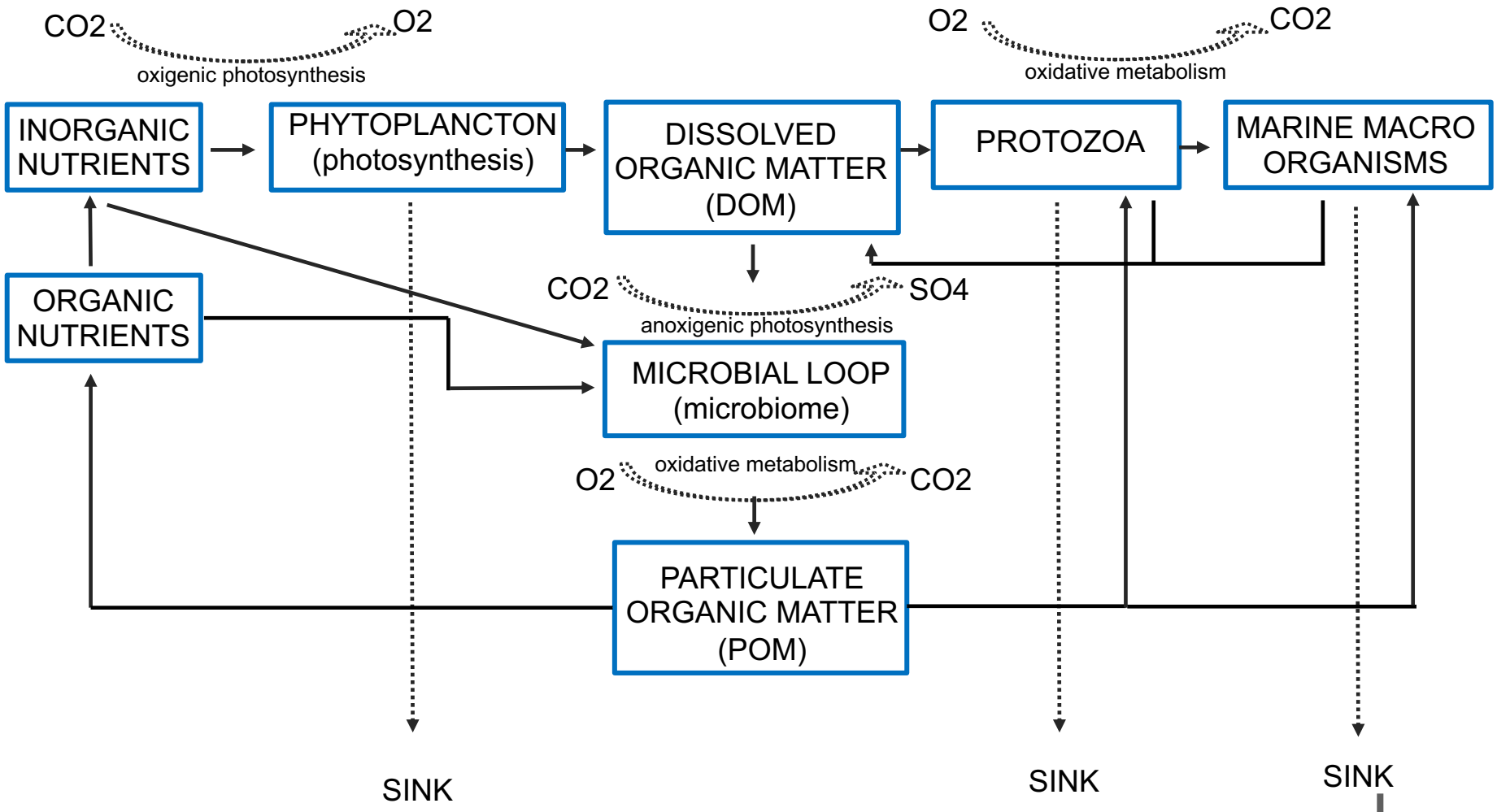


Fig. 8. Functional structuring of the ocean microbiome. (**A**) Phylum-level (class-level for Proteobacteria) taxonomic variability is higher (top, median relative SD = 65%) relative to the functional composition (OG functional categories) of ocean microbial samples (center, median relative SD = 7%). Removal of functions that are ubiquitous in the ocean environment reveals the variable, noncore fraction (bottom, median relative SD = 47%), which amounts on average to 4% of the total gene abundance. Red triangles on *x* axis highlight mesopelagic samples collected in oxygen minimum zones of the Indian Ocean and Eastern Pacific, which show increased levels of lipid metabolism in noncore functions. (**B**) Venn diagram (left) showing that core OGs in the epipelagic layer of the ocean are almost completely contained in mesopelagic core OGs (left). The bean charts (right) display differential abundances of marker genes (based on KO annotations) for selected functional processes in the ocean. Asterisks denote Mann-Whitney U test results (***$P < 0.001$).

the mesophilic waters are the richest in functional diversity

Importance and centrality of the microbial loop in epipelagic waters



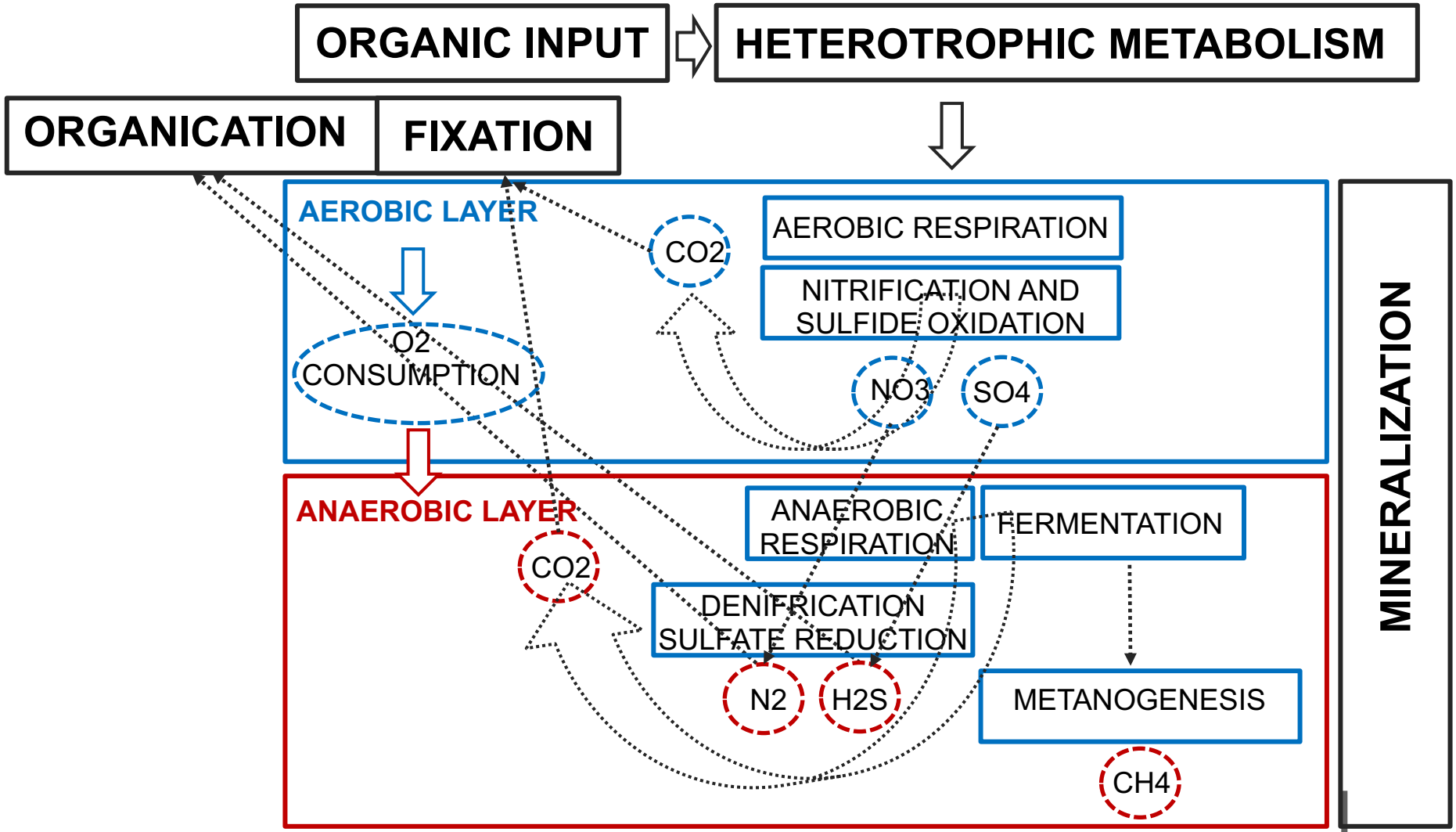
Epipelagic microbiome and microbial loop

INCREASE THE COMPLEXITY OF AVAILABLE NUTRIENTS PRODUCING POM AND MICROBIAL CELLS, SUPPORTING THE ECOSYSTEM DIVERSITY

CIRCOL NUTRIENTS

COMPETE WITH PHYTOPLANCTON FOR NUTRINETS , CONTRASTING CYNOBACTERIAL BLOOM IN THROPHIC ECOSYSTEMS

Marine microbiomes are central for the biology of marine sediments



Balanced microbiome activities are mostly important in trophic sediments

MICROBIOMES ARE STRATEGIC FOR NUTRIENT CIRCULARIZATION SUSTAINING THE WHOLE ECOSYSTEM DIVERSITY

DESTOING THE BALANCE BETWEEN DEGRADING AND ASSIMILATING COMPONENTS COMPROMIZE NUTRIENT CIRCULARIZATION AND MACROSCALE ECOSYSTEM DIVERSITY

Host associated marine microbiomes are integral for the host and ocean health

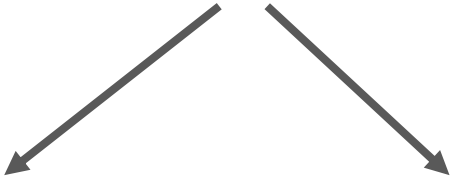
BASEMENT OF TROPHIC CHAINS

BENTHONIC AND PELAGIC SPECIES

All macro-organisms populating our planet exist as holobionts

Holobionts are defined as animals or plants together with associated microorganisms living on them

HOLOBIONTS EXIST WITHIN A RANGE OF SYMBIOSIS



OBLIGATORY SYMBIOSIS

e.g. herbivores, termites, corals, sponges, legumes …

WEAK SYMBIOSIS

e.g. carnivores, omnivores, non-human primates and **human beings** …

Fitness contribution of microbiomes to their host species

- NITROGEN FIXATION IN LEGUMINOSES
- CELLULOSE DEGRADATION IN RUMINANTS
- PHOTOSYNTHESIS BY MICROALGAE IN CORALS
- OXIDATION OF ORGANIC COMPOUNDS IN SPONGES

OBLIGATORY

- IMPROVEMENT AND REGULATION OF HOST NUTRITION AND METABOLISM
- MODULATION OF THE IMMUNE FUNCTION IN MAMMALIAN OMNIVORES
- PROTECTION FROM PATHOGEN ATTACK IN MAMMALIAN OMNIVORES

WEAK

At the evolutionary scale, microbiomes co-evolve with host species in the holgenome frame

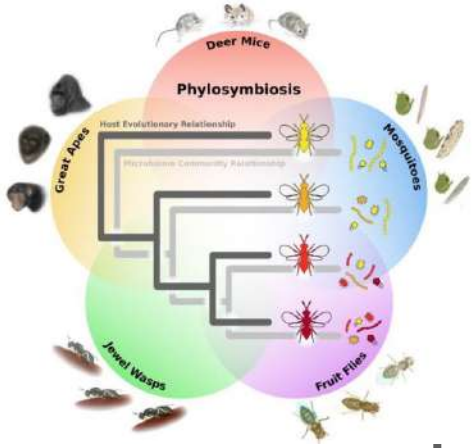
Even under weak symbiosis the holobiont gut microbiome (GM) provides **functional traits integral to the host physiology**

(e.g. nutrition, protection and immune regulation for the human GM)

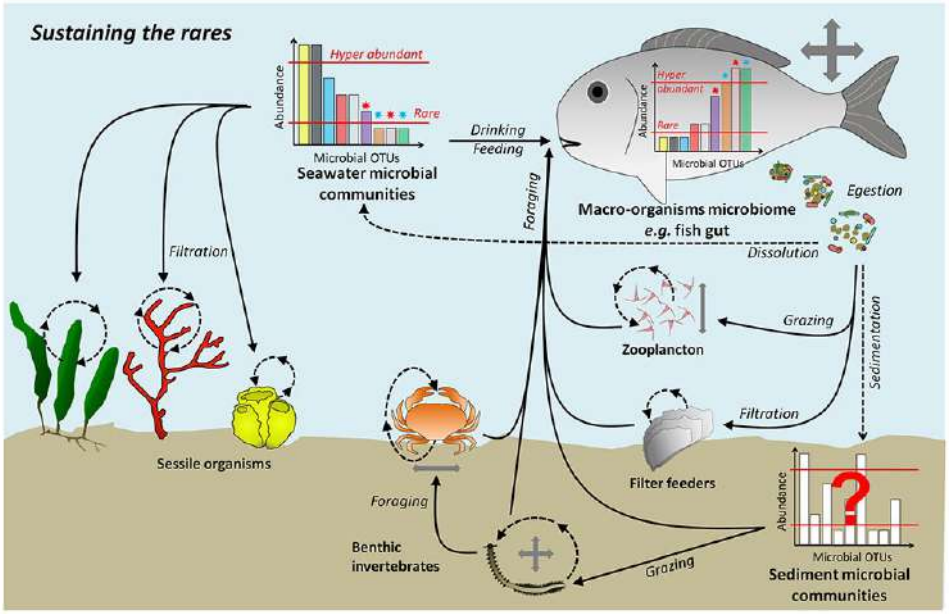
IMPACT ON THE HOST FITNESS ⇨ **CO-EVOLUTION** → **HOLOGENOME**: Host Genome ⟲ Metagenome

CO-EVOLUTION ⇩ **PHYLOSYMBIOSIS**

PHYLOSYMBIOSIS
GM recapitulates host phylogeny



MARINE ORGANISMS ARE INTEGRAL TO THE OCEAN MICROBIOMES

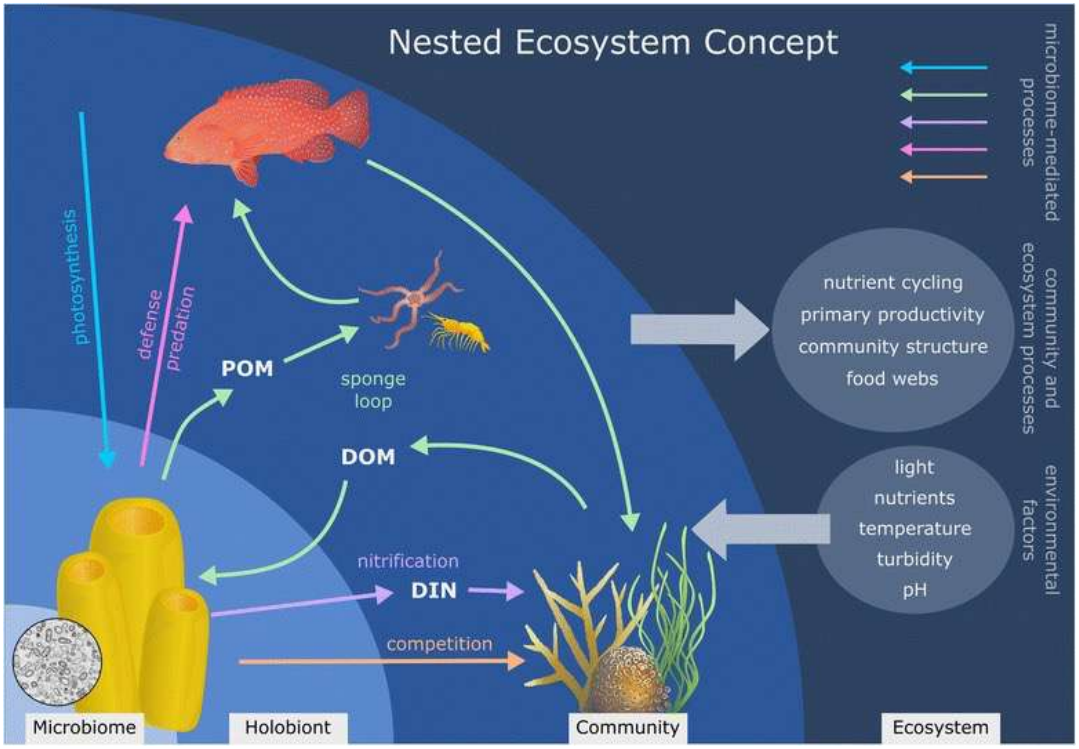


macroorganisms microbiomes harbor a high functional potential and are integral components of functional gene dynamics in aquatic bacterial communities

dissemination vectors

maintenance of microbial diversity at various scales in the marine environment

• The holobionts microbiome at the ecosystem scale



- the health of the holobionts basement of the marine trophic chains is instrumental for the health of the whole ecosystem

Anthropic factors, microbiomes and planet health

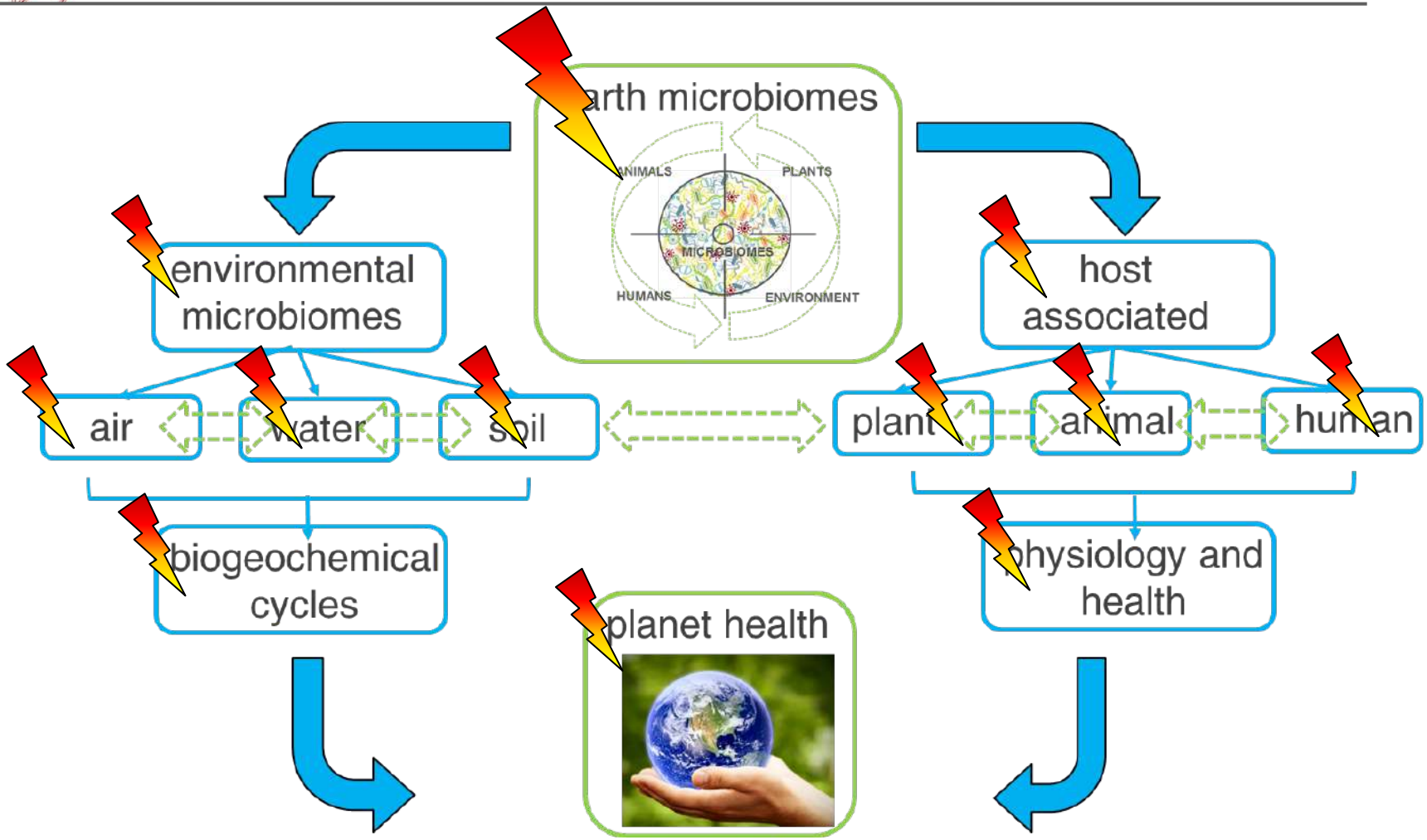
the central dogma of the microbial ecology

“Everything is everywhere: but the environment selects”
Martinus Wilhelm Beijerinck, early in the twentieth century

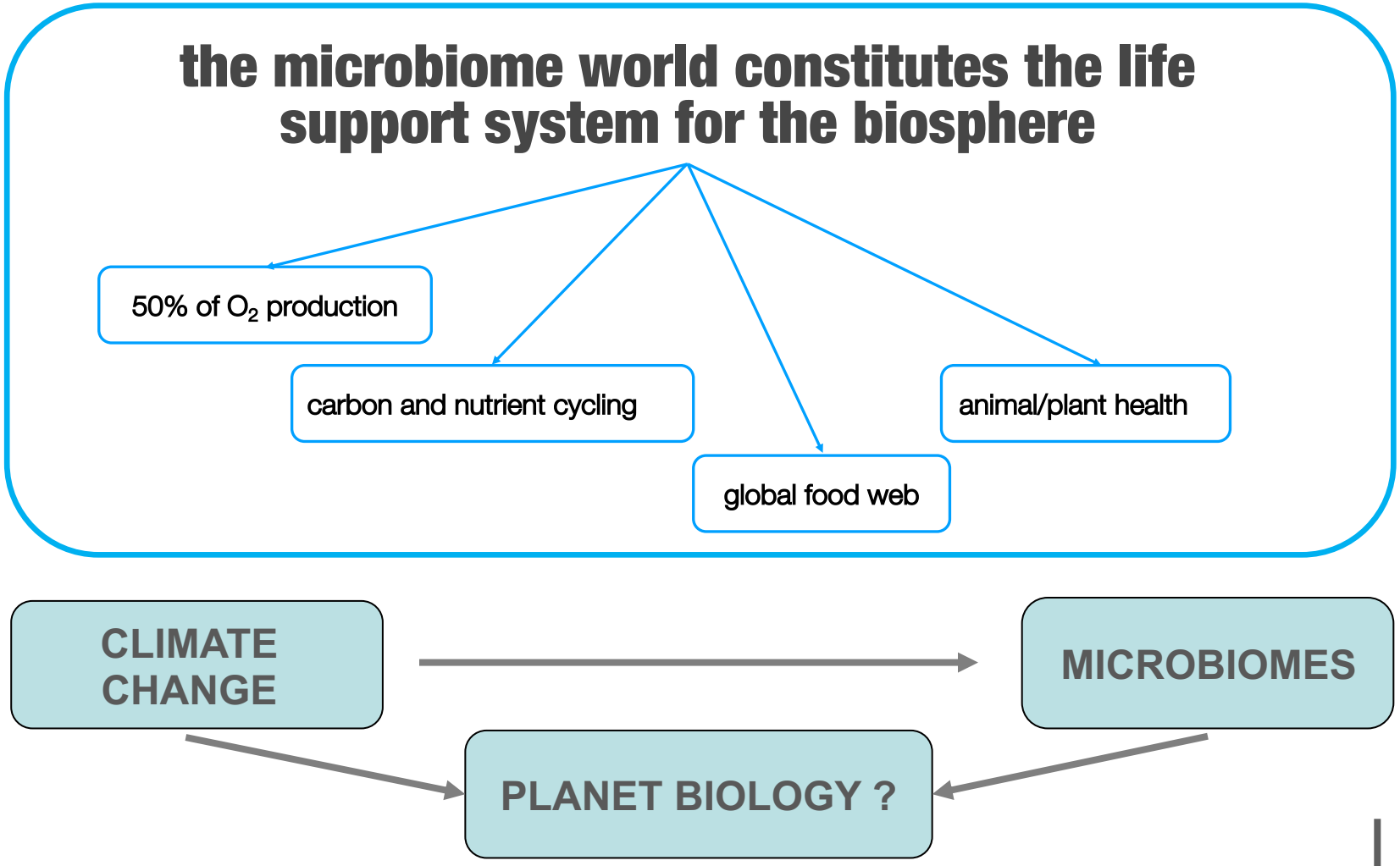
ubiquitous distribution and ecological determinism in microbial biogeography

Anthropic factors and climate change shape the microbiomes and microbiomes connection chains of the planet

MICROBIOMES AND ANTHROPIC STRESSORS

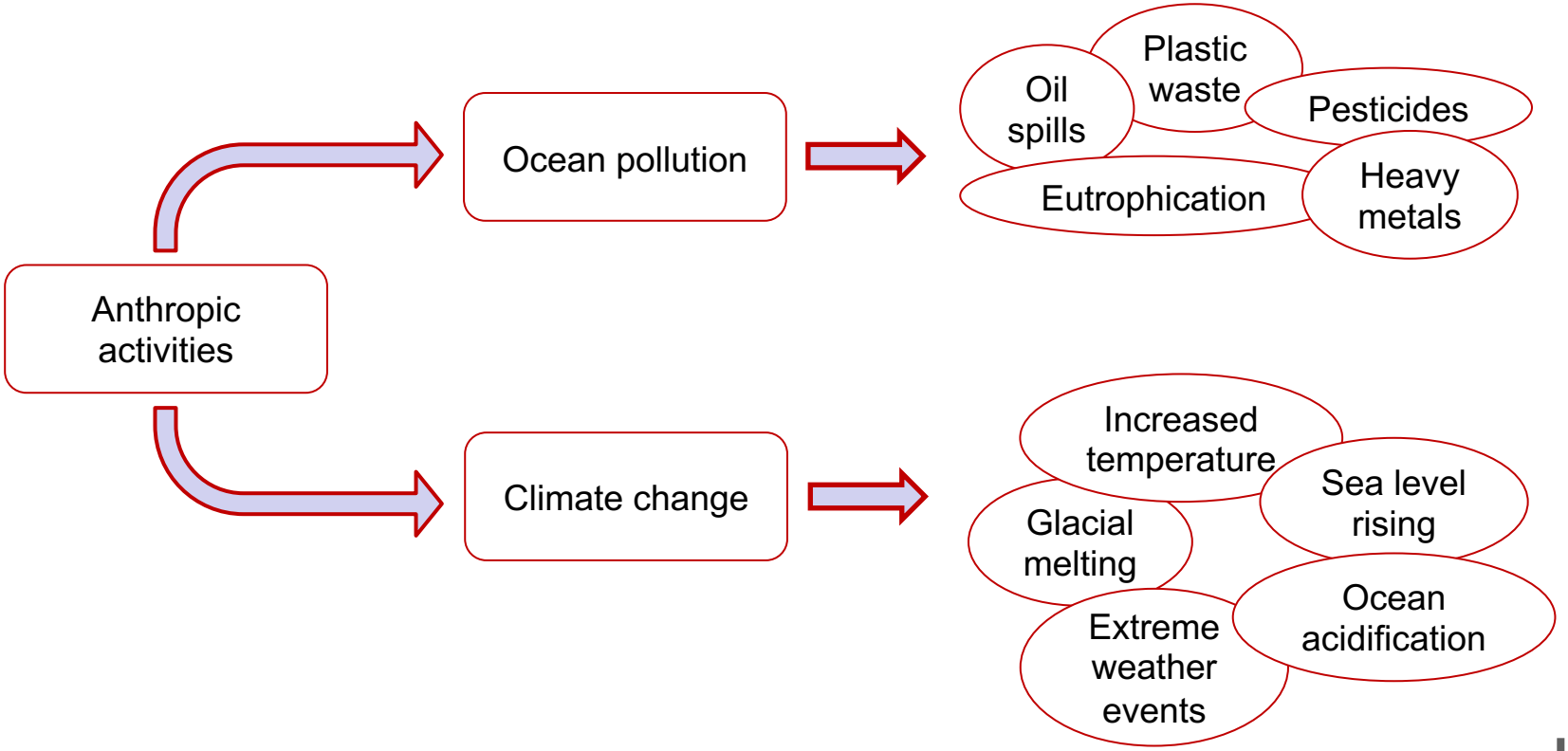


Microbiomes have a central role in the biology of climate change

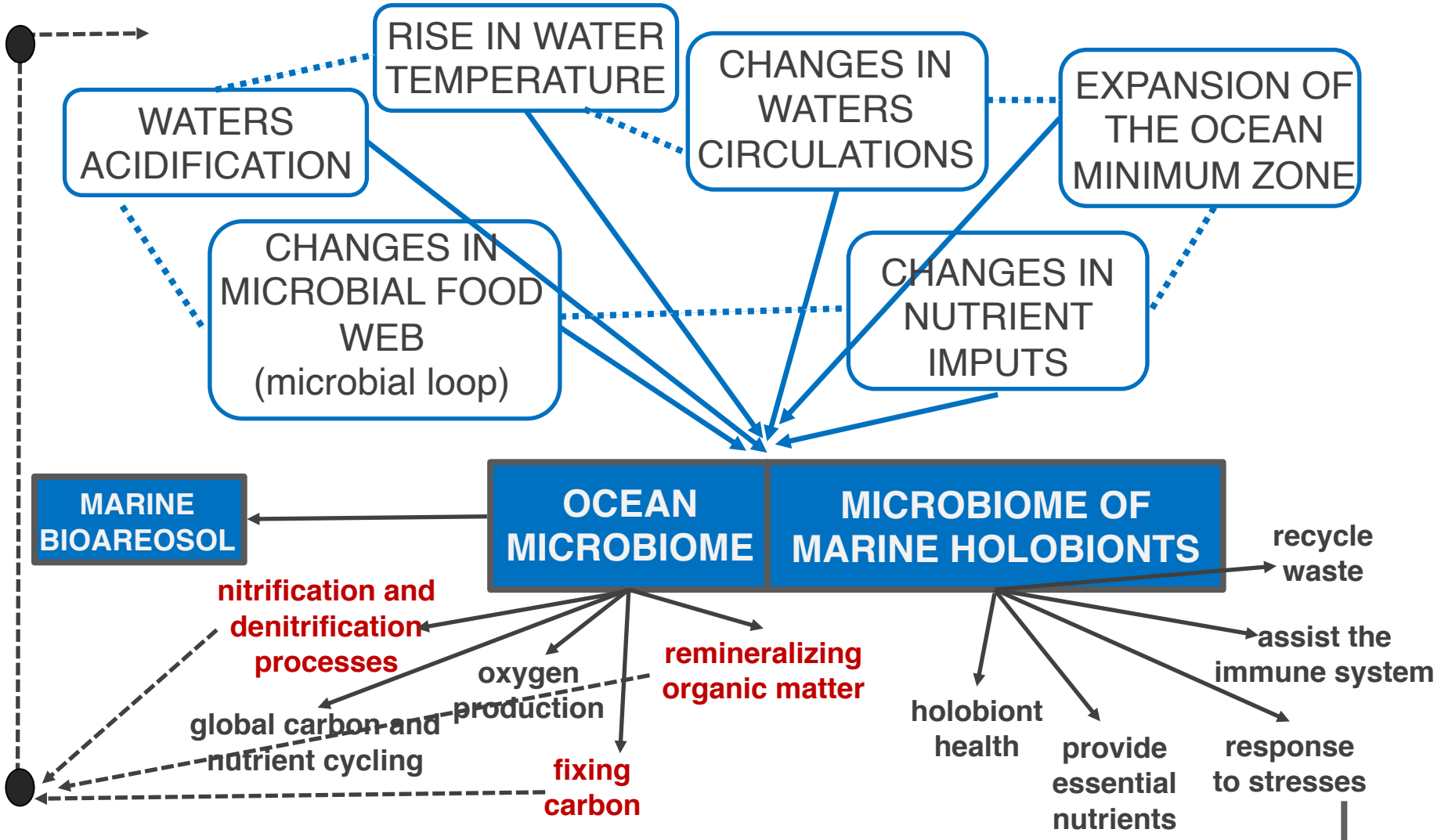


Microbiome disturbances

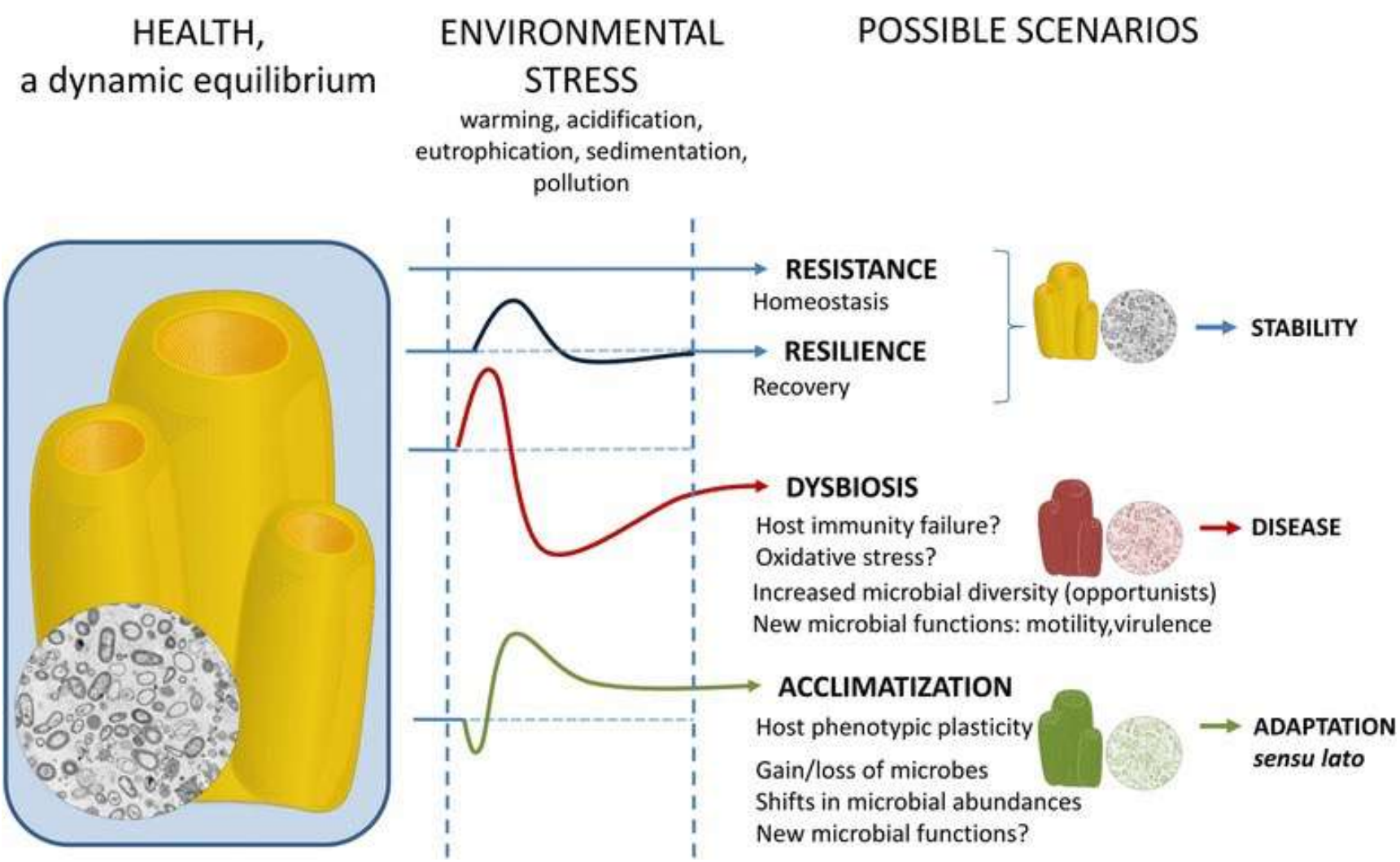
The fast and restless **increase** of World's human population is rising crescent concerns about the **threads** that **human activity** can pose to **seas and oceans ecosystems**.



CLIMATE CHNAGE IS PUSHING OCEAN MICROBIOME IN CONCITIONS OUTSIDE THE RECENT HYSTORICAL RANGE

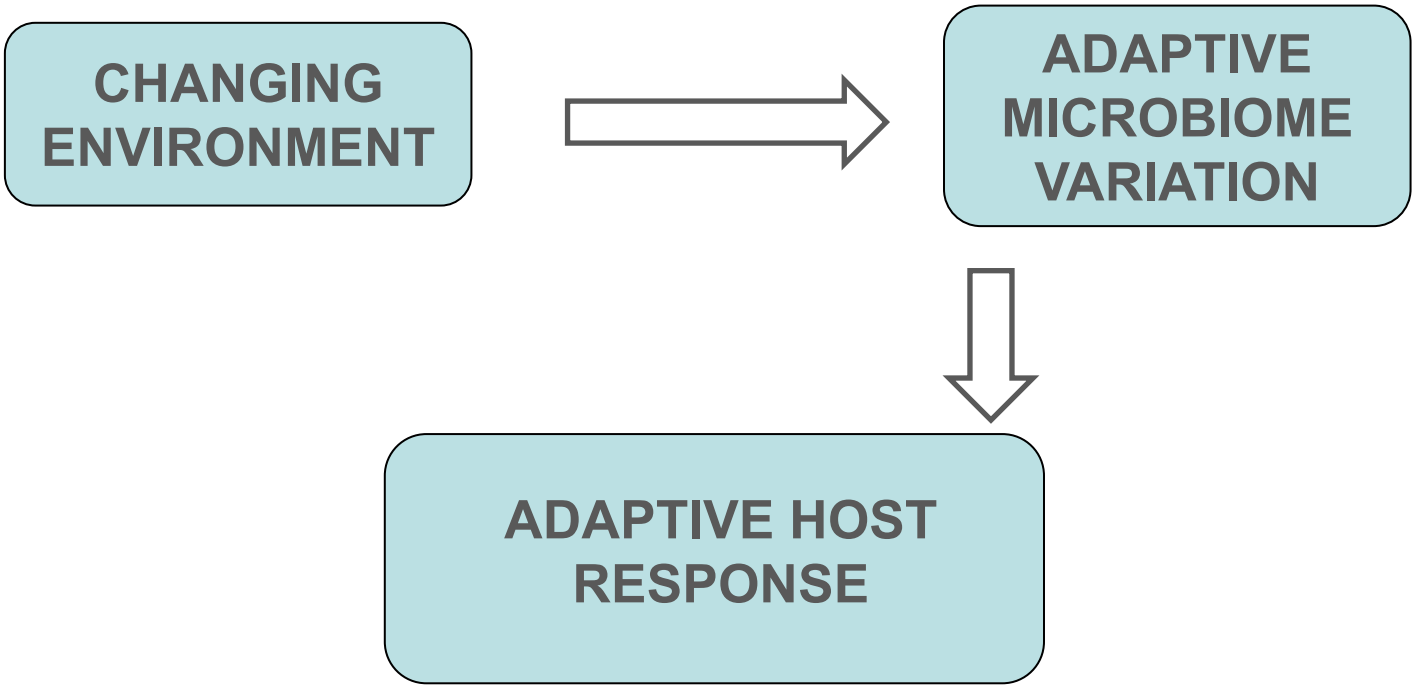


• Holobionts microbiome response to environmental stresses

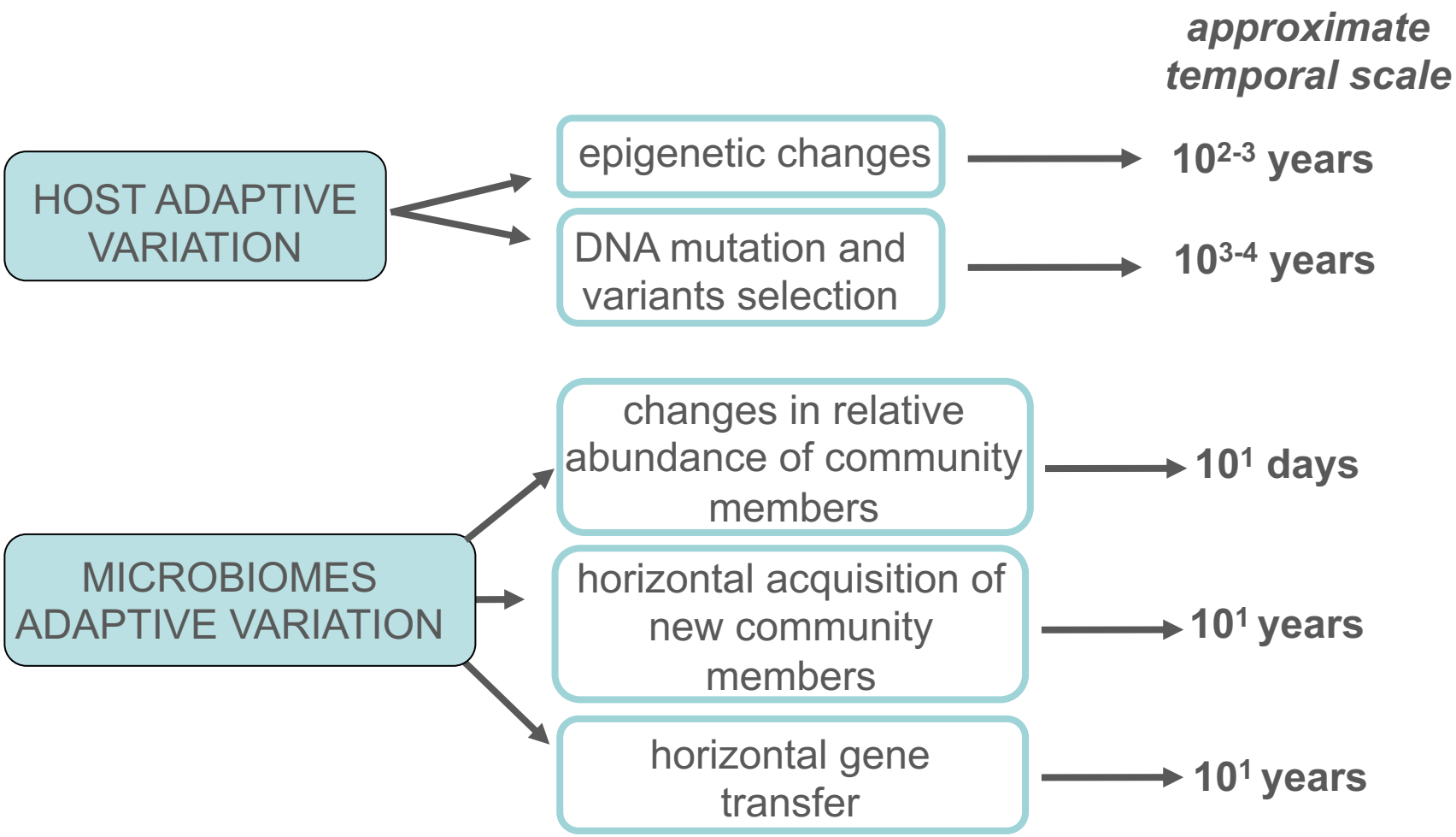


The holobionts microbiome as a key provider of phenotypic plasticity

HOLOBIONTS MICROBIOMES RESPOND RAPIDLY TO ENVIRONMENTAL CHANGES, SUPPORTING FAST ADAPTIVE RESPONSE OF THE HOST

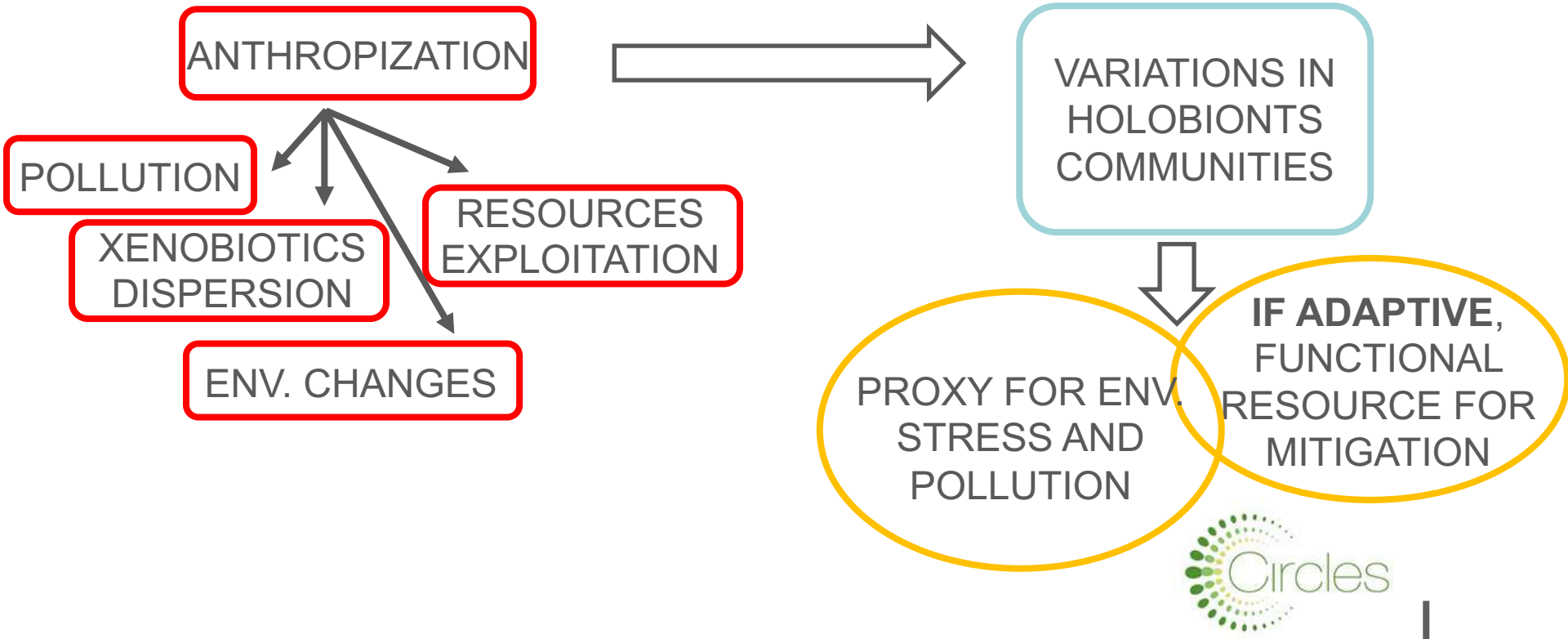


Microbiomes adapt faster than the host genome



Holobionts microbiomes for sustainability

Adaptive variations of holobionts microbiomes can be exploited as a proxy for environmental stress and pollution, as well as functional resource for its mitigation



Circles

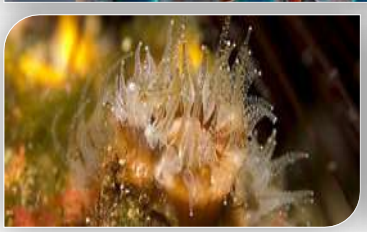
Coral reefs, the basement of the marine trophic chains

(Torda et al., Nat Clim Change 2017)

Marine ecosystems

Climate change (ocean acidification)

Adaptive responses

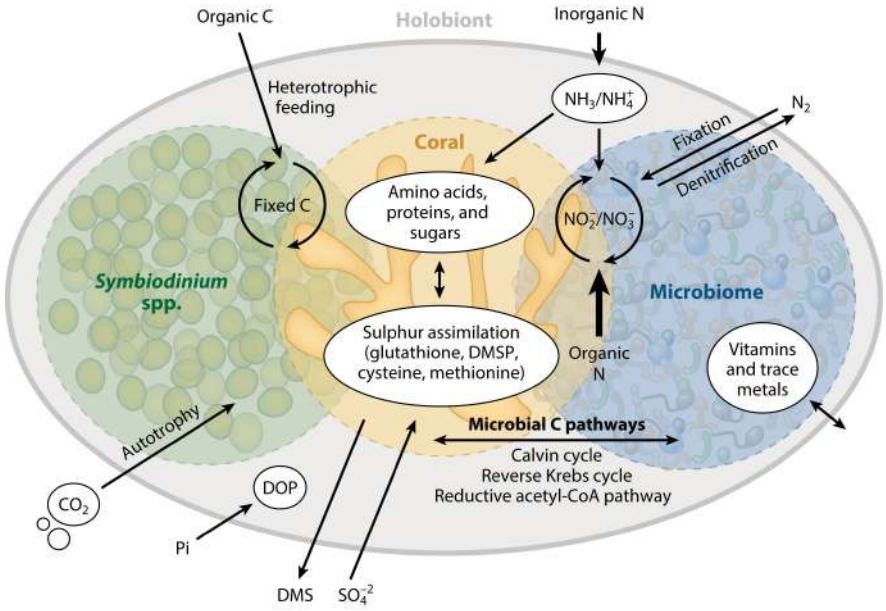


Balanophyllia europaea

Astroides calycularis

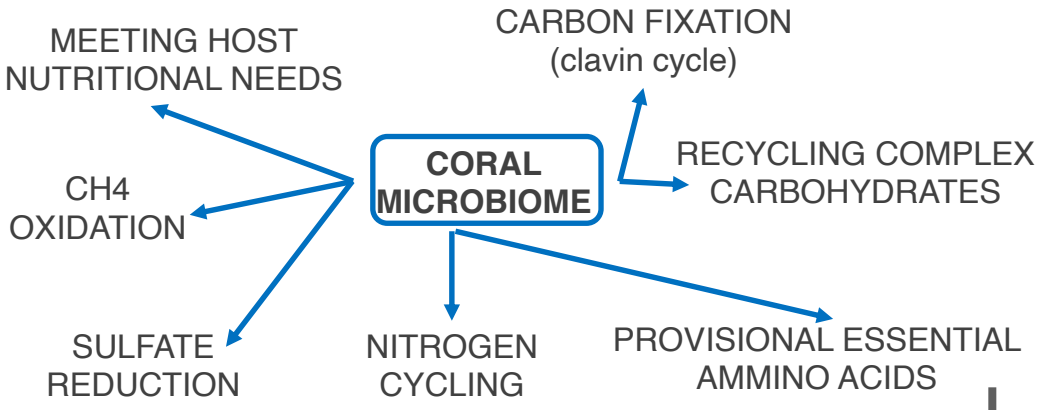
MICROBIOME

CORAL REEF MICROBIOMES AND CLIMATE CHANGE



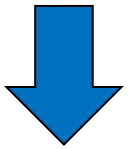
Bourne DG, et al. 2016.
Annu. Rev. Microbiol. 70:317–40

CORAL REEF REALY ON DIVERSE CONSORTIUM OF FREE-LIVING AND HOST ASSOCIATED MICROORGANISMS FOR THE CAPTURE, RETENTION AND RECYCLING OF NUTRIENTS AND TRACE ELEMENTS, ALLOWING TO THRIVE IN THE MARINE EQUIVALENT OF A DESERT



CLIMATE CHANGE AND DYSBIOSIS

CLIMATE CHANGE CAN DESTABILIZE MICROBIOMES LEADING TO DYSBIOSES THAT CAN DEVELOP INTO ALTERNATIVE STABLE STATES



USAGE OF MICROBIOMES AS BIOMARKERS FOR ECOSYSTEM HELATH FOR APPLICATION IN MARINE CONSERVATION AND RESTORATION

MICROBIOME CONTRIBUTION TO CORAL ACCLIMATATION

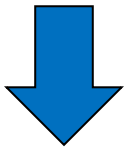
MICROBIOME CAN HELP INCREASE THE RESILIENCE OF REEF-BUILDING CORALS TO FUTURE CLIMATE CONDITIONS

EXTENSIVE TAXONOMIC AND METABOLIC DIVERSITY AND THE SHORT GENERATION TIME PROVIDES CONSIDERABLE POTENTIAL TO CONTRIBUTE TO THE HOLOBIONT ADAPTIVE RESPONSE

ADAPTIVE MICROBIOME MODIFICATIONS CAN BE VERTICALLY TRANSMITTED, INCREASING POPULATION FITNESS

MICROBIOME MANIPULATION FOR CORAL RESTORATION

NATURAL MICROBIOME OF STRESS RESISTANT CORALS CAN BE USED TO DEVELOP SYSTHETIC COMMUNITIES TO BE INOCULATED INTO DISEASED CORALS, ALLWING TO TRANSFER HELATH BENEFIT AND STRESS RESISTANCE



DEVELOPMENT OF CORAL STOPKC WITH AND ENHANCED MICROBIOME-MEDIATED STRESS RESILIENCE BY ASSISTED EVOLUTION

Contents lists available at ScienceDirect

Science of the Total Environment

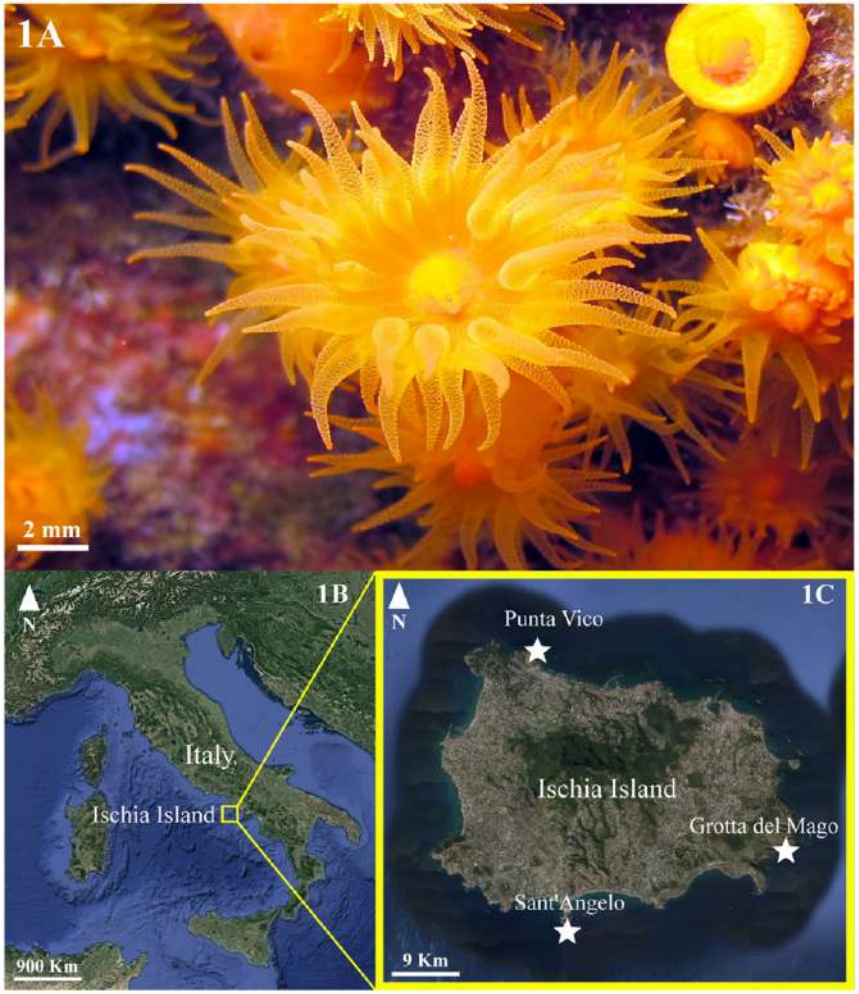
journal homepage: www.elsevier.com/locate/scitotenv

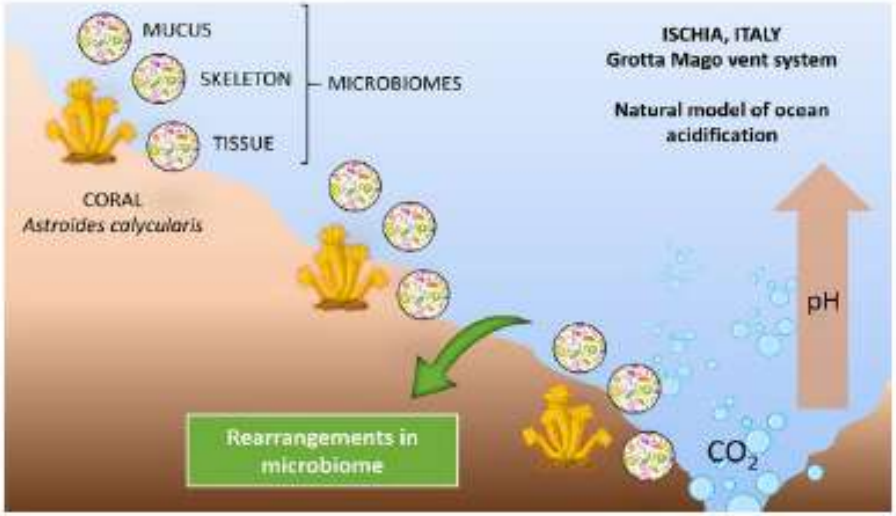
Patterns in microbiome composition differ with ocean acidification in anatomic compartments of the Mediterranean coral *Astroides calycularis* living at CO_2 vents

Elena Biagi [a,1], Erik Caroselli [b,c,1], Monica Barone [a], Martina Pezzimenti [b], Nuria Teixido [d,e], Matteo Soverini [a], Simone Rampelli [a], Silvia Turroni [a], Maria Cristina Gambi [e], Patrizia Brigidi [a], Stefano Goffredo [b,c,*], Marco Candela [a,c,**]

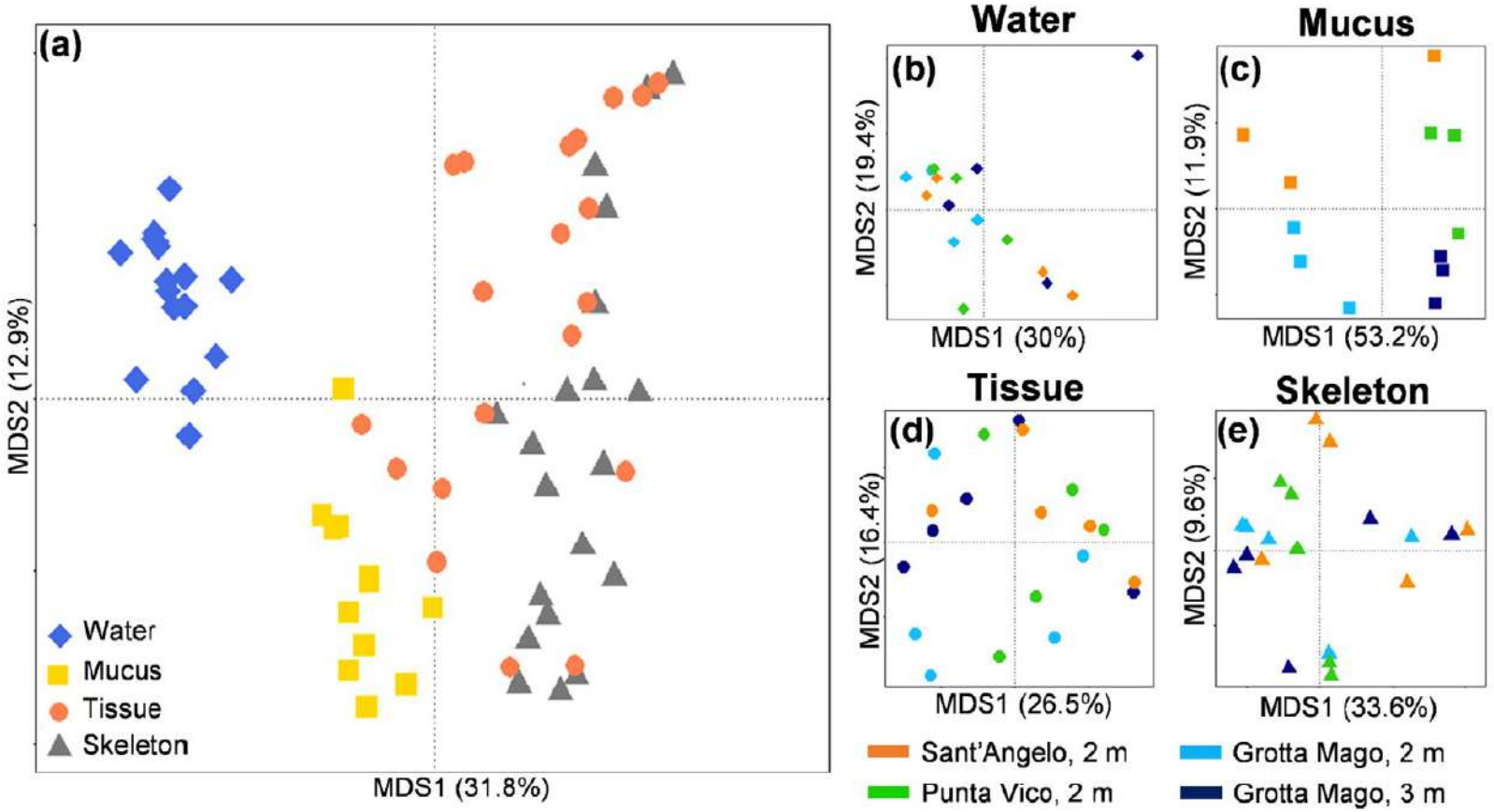
[a] Unit of Holobiont Microbiome and Microbiome Engineering (HolobioME), Department of Pharmacy and Biotechnology, University of Bologna, via Belmeloro 6, 40126 Bologna, Italy
[b] Marine Science Group, Department of Biological, Geological and Environmental Sciences, University of Bologna, via Selmi 3, 40126 Bologna, Italy
[c] Fano Marine Center, The Inter-Institute Center for Research on Marine Biodiversity, Resources and Biotechnologies, viale Adriatico 1/N, 61032 Fano, Pesaro Urbino, Italy
[d] Sorbonne Université, CNRS, Laboratoire d'Océanographie de Villefranche, 181 chemin du Lazaret, F-06230 Villefranche-sur-Mer, France
[e] Villa Dohrn-Benthic Ecology Center, Department of Integrative Marine Ecology, Stazione Zoologica Anton Dohrn, 80077 Ischia (Naples), Italy

Comparison of the microbiome of non-symbiotic solitary coral Astroides calycularis that naturally lives at a volcanic CO2 vent in Ischia Island (Naples, Italy), with that of corals living in non-acidified sites at the same island.

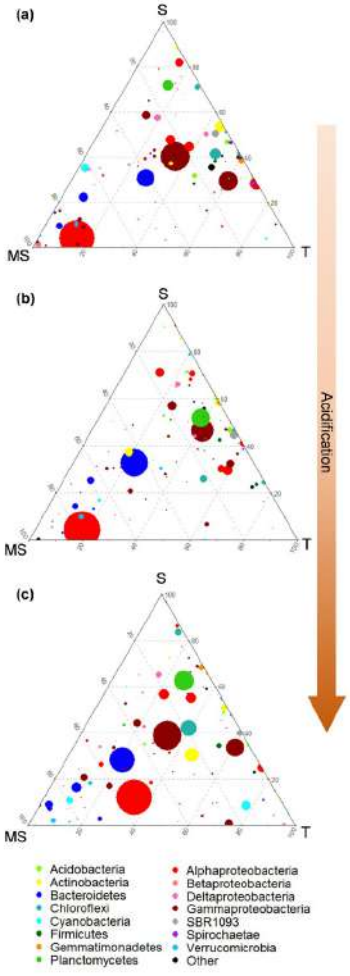


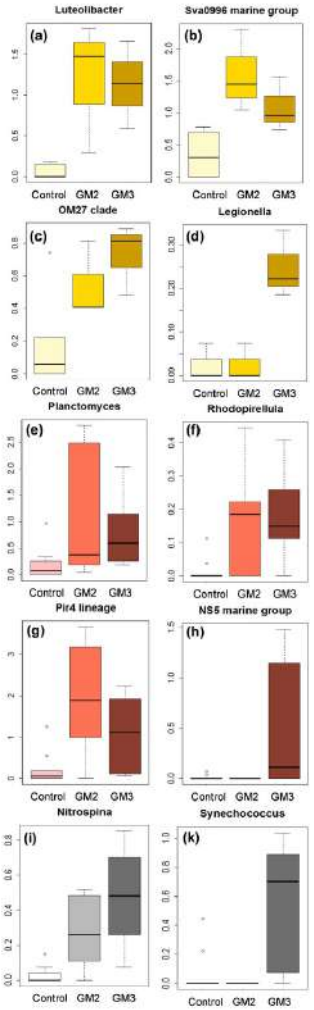


Microbiomes associated with the different coral anatomic compartments were different from each other and from the microbial communities of the surrounding seawater.



The mucus associated microbiome differed the most between the control and acidified sites.





- Coral microbiomes contribute to host acclimatization to environmental change.
- Natural CO2 gradients are a model of global change-induced ocean acidification.
- Non-symbiotic coral Astroides calycularis survives in a natural acidified site.

- Calycularis mucus microbiome is the most affected by low pH conditions.
- Low pH conditions induce changes in microbiome supporting nitrogen cycling.

Model V

Giorgia Palladino[1,4], Erik Caroselli[2,4], Teresa Tavella[1], Federica D'Amico[3], Fiorella Prada[2,4], Arianna Mancuso[2,4], Silvia Franzellitti[4,5], Simone Rampelli[1], Marco Candela[1,4], Stefano Goffredo[2,4], Elena Biagi[6]

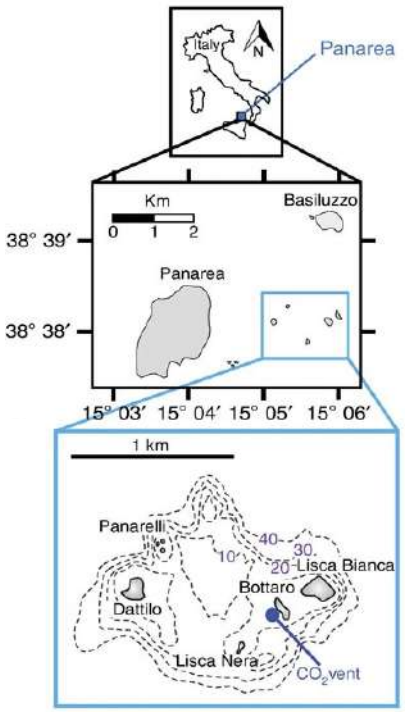
Metagenomic shifts in mucus, tissue and skeleton of the coral *Balanophyllia europaea* living along a natural CO_2 gradient

Under review on *ISME Communication*

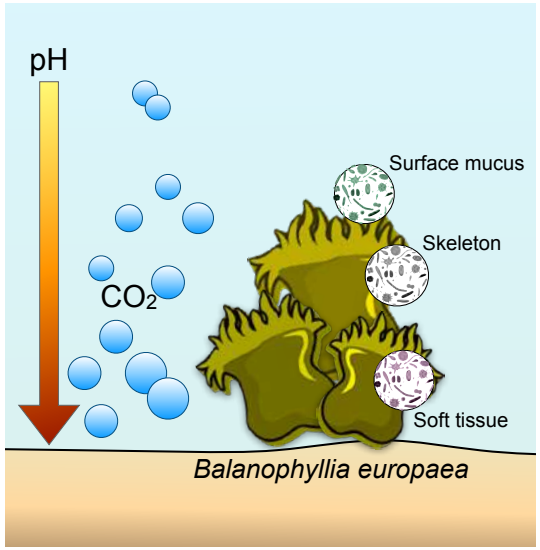
[1]Unit of Microbiome Science and Biotechnology, Department of Pharmacy and Biotechnology, University of Bologna, via Belmeloro 6, 40126 Bologna, Italy
[2]Marine Science Group, Department of Biological, Geological and Environmental Sciences, University of Bologna, via Selmi 3, 40126 Bologna, Italy
[3]Department of Medical and Surgical Sciences, University of Bologna, Via Massarenti 9, 40138, Bologna, Italy
[4]Fano Marine Center, the Inter-Institute Center for Research on Marine Biodiversity, Resources and Biotechnologies, viale Adriatico 1/N, 61032 Fano, Pesaro Urbino, Italy
[5]Animal and Environmental Physiology Laboratory, Department of Biological, Geological and Environmental Sciences, University of Bologna, via Sant'Alberto 163, 48123 Ravenna, Italy
[6]Department of Civil, Chemical, Environmental, and Materials Engineering, University of Bologna, Viale del Risorgimento 2, 40136 Bologna, Italy

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Experimental design



Model for coral microbiome acclimatization to different ocean acidification conditions in a natural study system

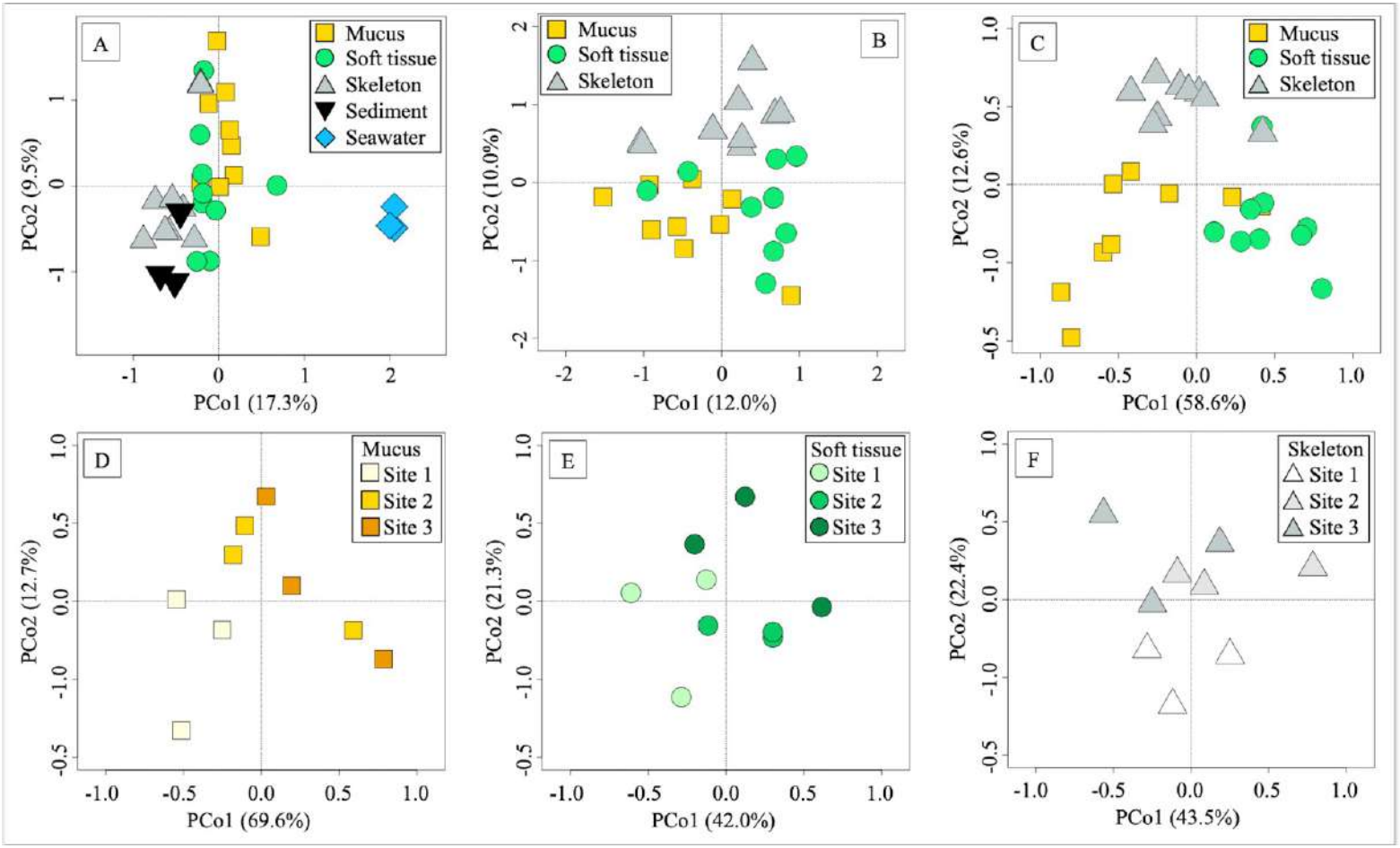


Study design

- 3 sampling sites (decreasing pH)
- 3 coral specimens/site
- 3 coral anatomic compartments
- 16S rRNA sequencing → phylogenetic composition
- shotgun sequencing → functional variations

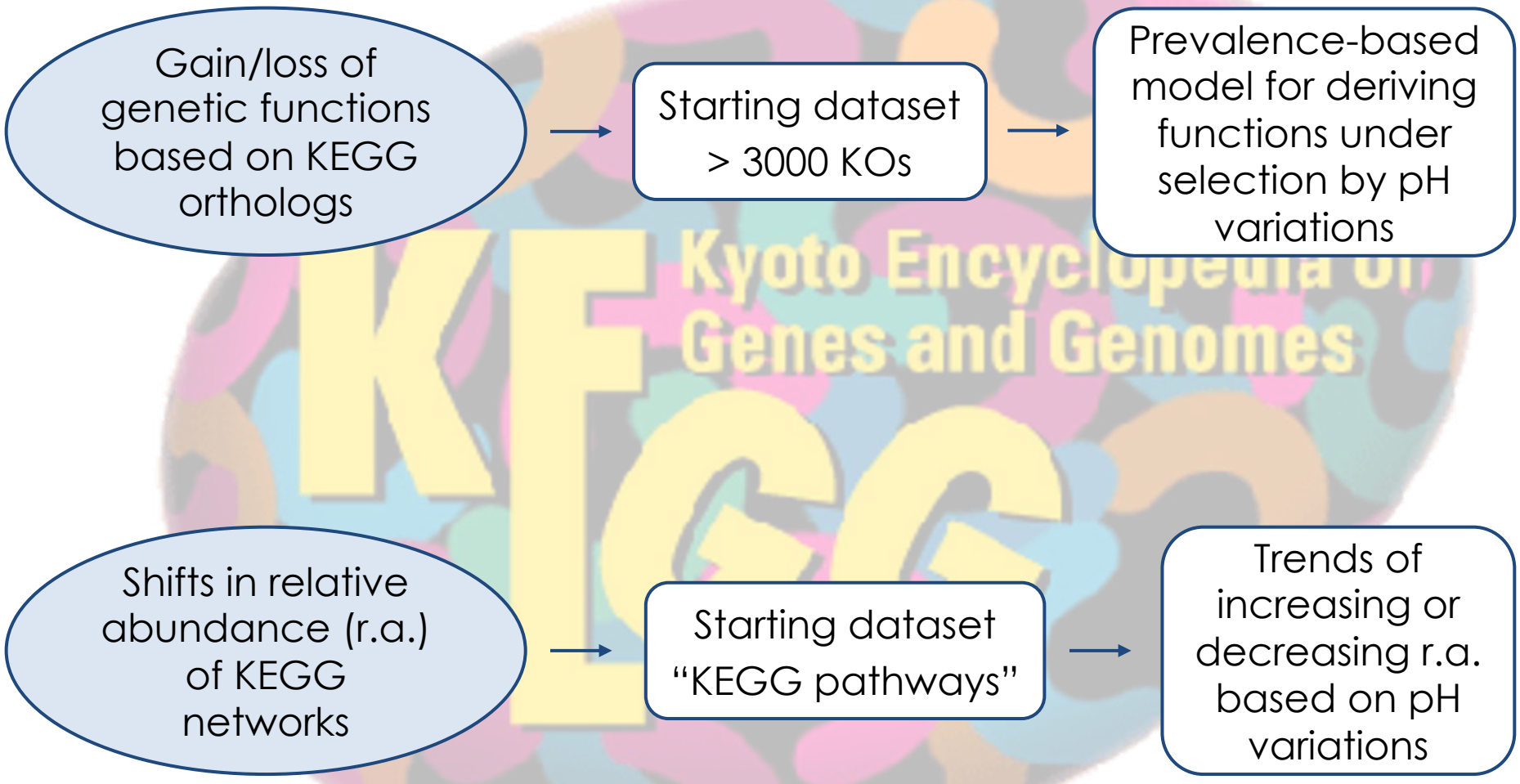
Microbiome compositional structure of B. europaea and the surrounding environment.

Principal Coordinate Analyses (PCoAs) of the Bray-Curtis distances calculated on microbiome profiles at genus taxonomic level, obtained from 16S rRNA sequencing (A and B) and phylogenetic assignation of metagenomic reads (C-F)



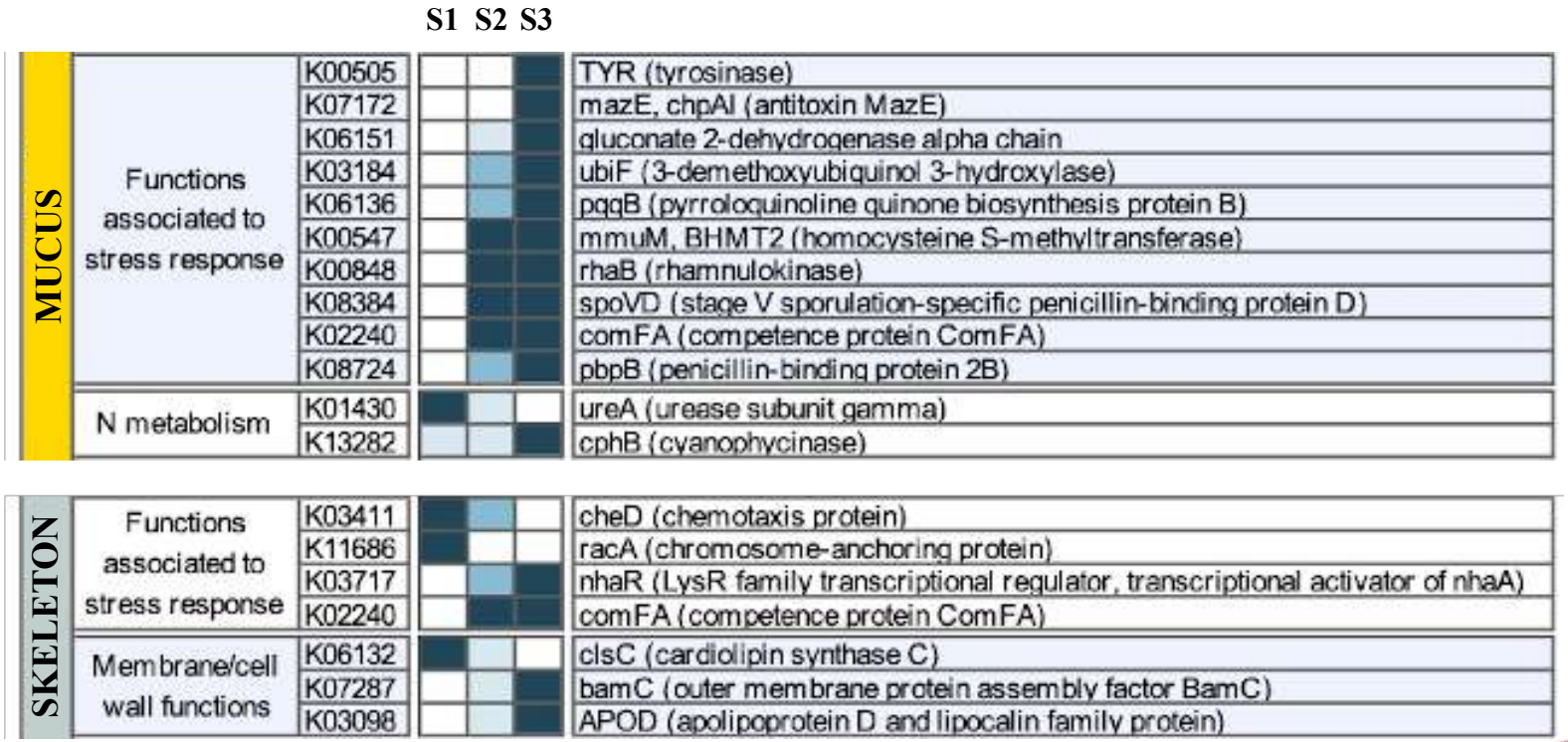
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

A double approach for metagenomic data exploration



A double approach for metagenomic data exploration

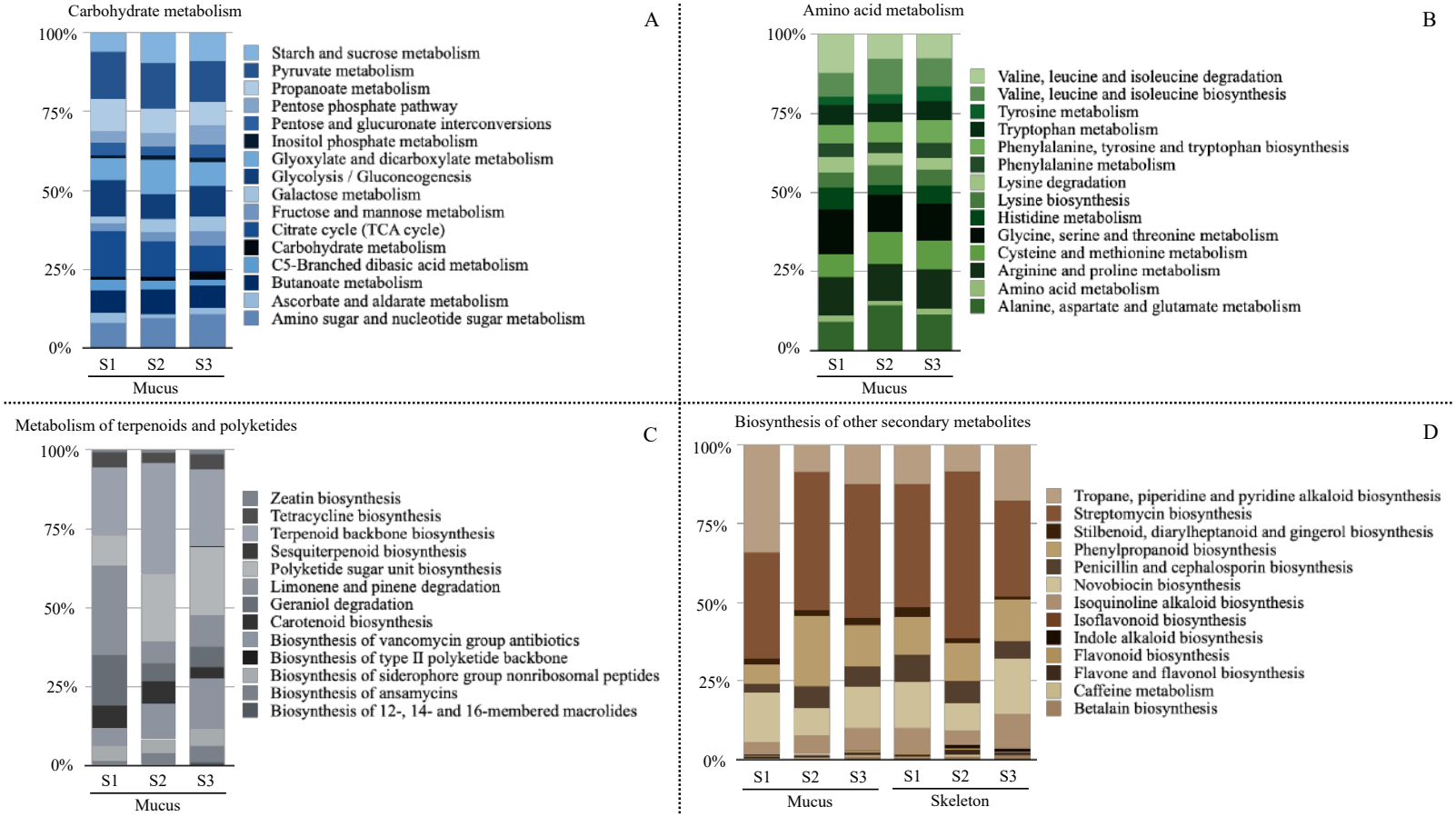
Gain/loss of genetic functions based on KEGG orthologs



			S1	S2	S3	
MUCUS	Functions associated to stress response	K00505				TYR (tyrosinase)
		K07172				mazE, chpAI (antitoxin MazE)
		K06151				gluconate 2-dehydrogenase alpha chain
		K03184				ubiF (3-demethoxyubiquinol 3-hydroxylase)
		K06136				pqqB (pyrroloquinoline quinone biosynthesis protein B)
		K00547				mmuM, BHMT2 (homocysteine S-methyltransferase)
		K00848				rhaB (rhamnulokinase)
		K08384				spoVD (stage V sporulation-specific penicillin-binding protein D)
		K02240				comFA (competence protein ComFA)
		K08724				pbpB (penicillin-binding protein 2B)
	N metabolism	K01430				ureA (urease subunit gamma)
		K13282				cphB (cyanophycinase)
SKELETON	Functions associated to stress response	K03411				cheD (chemotaxis protein)
		K11686				racA (chromosome-anchoring protein)
		K03717				nhaR (LysR family transcriptional regulator, transcriptional activator of nhaA)
		K02240				comFA (competence protein ComFA)
	Membrane/cell wall functions	K06132				clsC (cardiolipin synthase C)
		K07287				bamC (outer membrane protein assembly factor BamC)
		K03098				APOD (apolipoprotein D and lipocalin family protein)

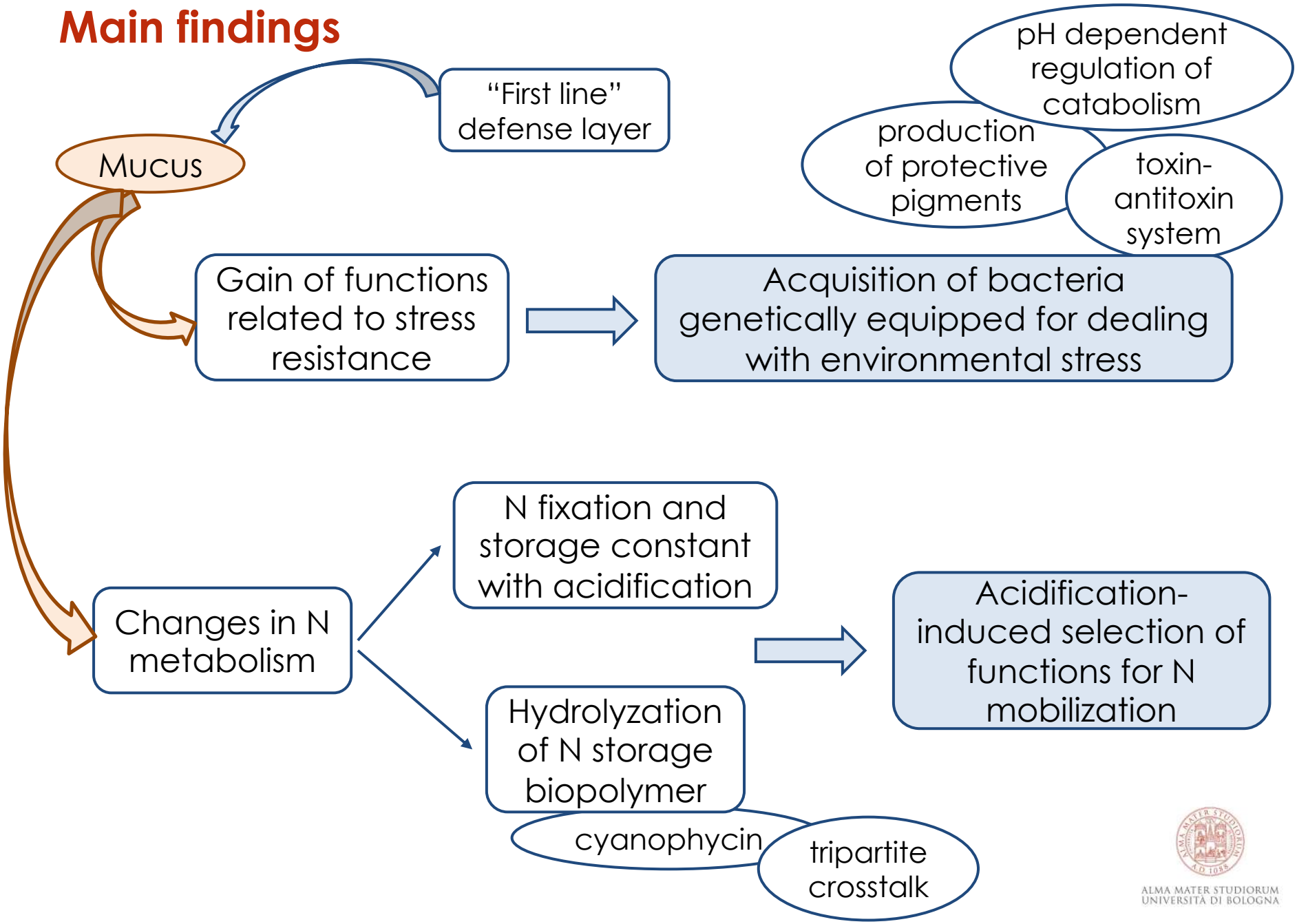
A double approach for metagenomic data exploration

Shifts in relative abundance (r.a.) of KEGG networks



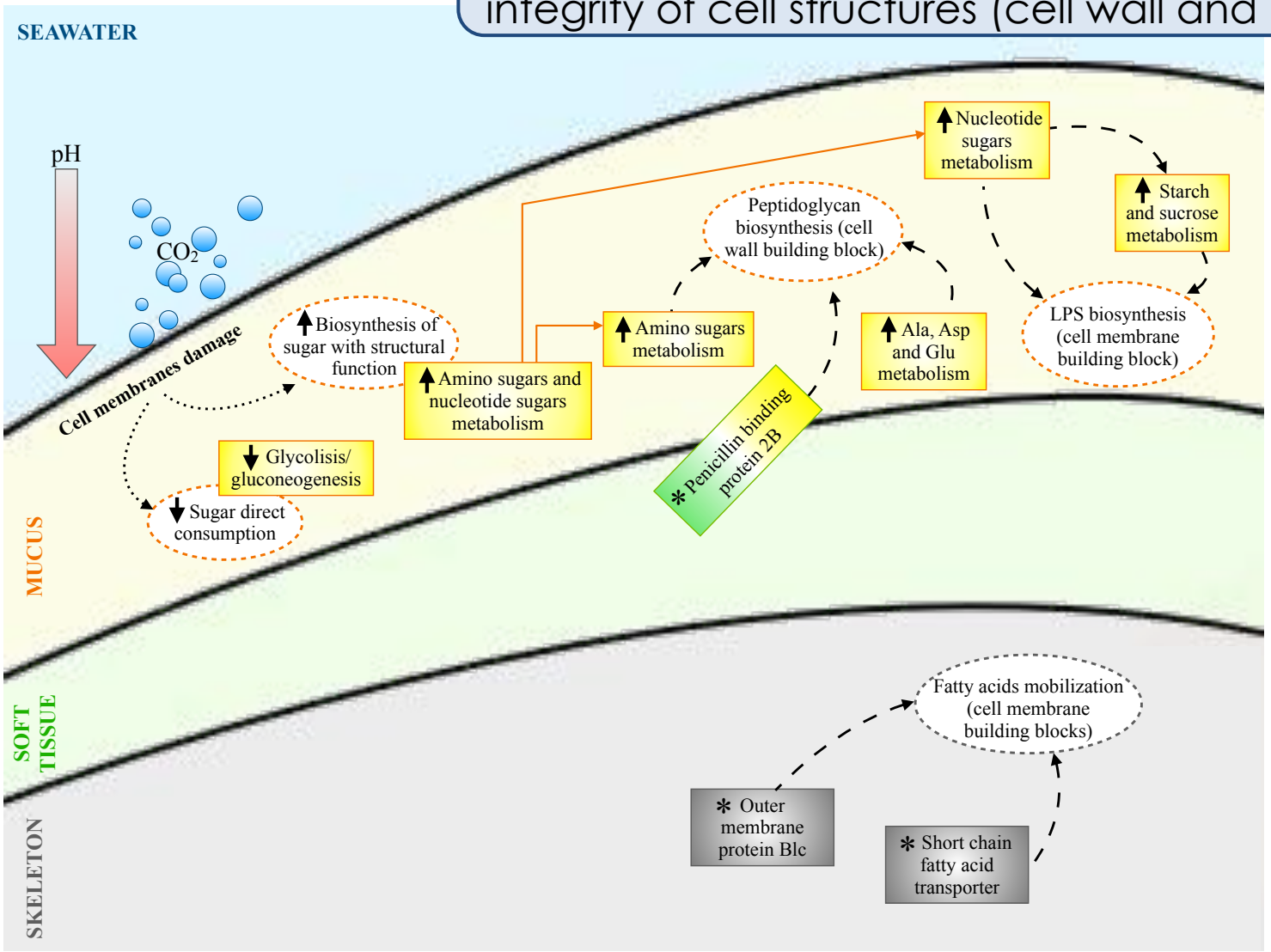
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Main findings

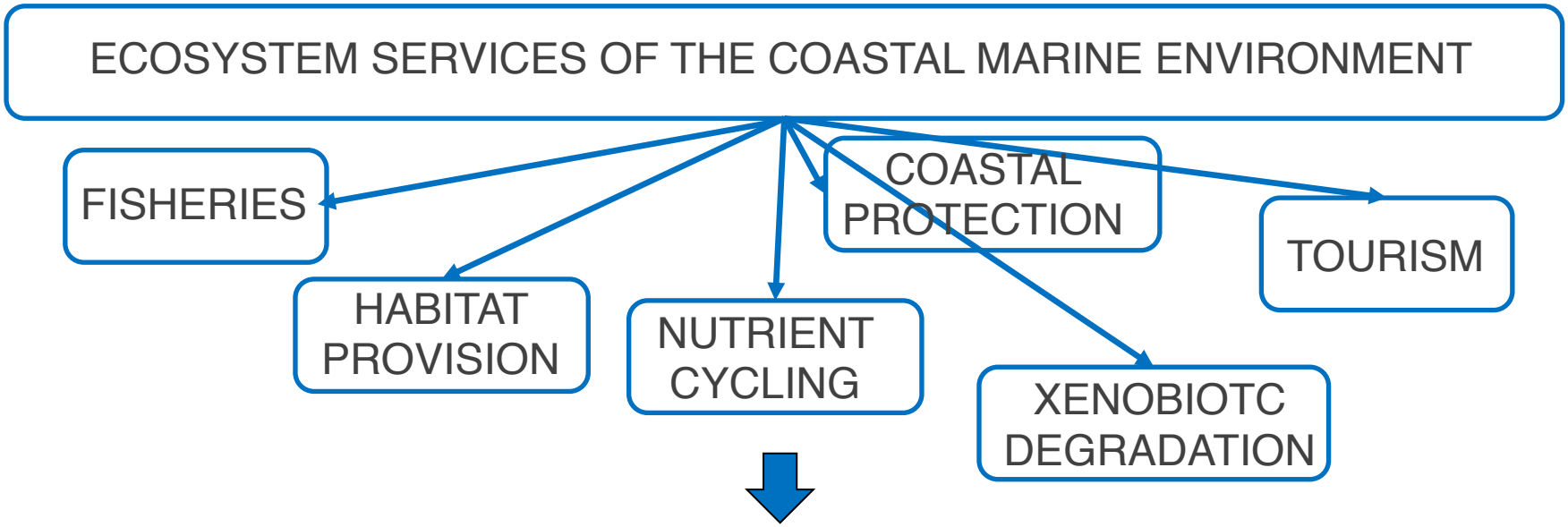


Main findings

Shift in carbohydrate metabolic pathways from energy production to the maintenance of the integrity of cell structures (cell wall and membrane)



COASTAL RESEAERCH IN THE MICROBIOME FRAME



MARINE MICROBIOMES (water and sediments)

MICROBIOMES OF THE HABITAT FORMING HOLOBIONTS

microalgae **mussels** **sea grass** **corals** **anemones** **sponge** **mangrove**

MICROBIOME DYNAMIC IN COASTAL HOLOBIONTS

DYNAMICS INVOLVES DIFFERENT SCALES AND DRIVERS

HOST PHYSIOLOGY	ENVIRONMENTAL CONDITIONS	GEOGRAPHY	EVOLUTION

SHORT TERM DYNAMICS → **LONG TERM DYNAMICS**

ESTABLISHMENT OF THE HOST-MICROBIOME INTERACTIONS

CHEMICAL SIGNALING, SECONDARY METABOLITES PRODUCED BY HOST SPECIES (independently or in response to env. stimuly)

ENVIRONMENTAL MICROBIOME POOL

MICROBE – MICROBE INTERACTION

DIFFERENT HOST LIES IN A CONTINUUM AMONG THESE PROCESSE AND CAN DYNAMICALLY CHANGE THE SELECTION STRATEGIES IN RESPONSE TO SPECIFIC NEEDS AND ENVIRONMENTAL CONDITIONS

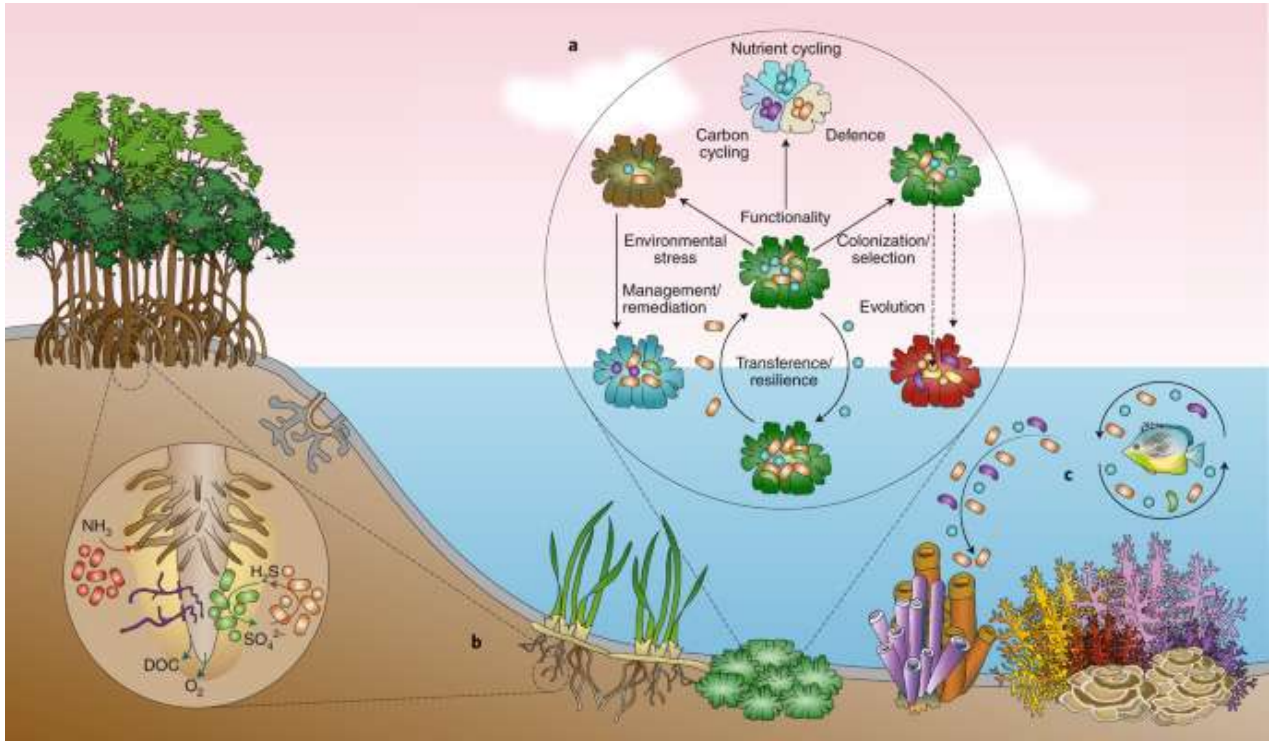
THE ENVIRONMENT ACT AS A SOURCE FOR HOLOBIONT MICROBIOME

the intrinsic environmental variability linked to seasonal changes, perturbation events or a combination of these, strongly influence microbiome diversity and functionality

environmental stressors can interact opposing, additive or synergistic ways to influence host microbiome and their interactions, leading to positive, negative or neutral impacts on them

INDIRECT ECOLOGICAL INTERACTIONS

ECOLOGICAL INTERACTIONS WITHIN AND AMONG HOLOBIONTS CAN BE INDIRECT, MICROBIOME RECRUITMENT BY ONE HOST MAY BE AFFECXTED BY EXUDATES OF OTHER HOST

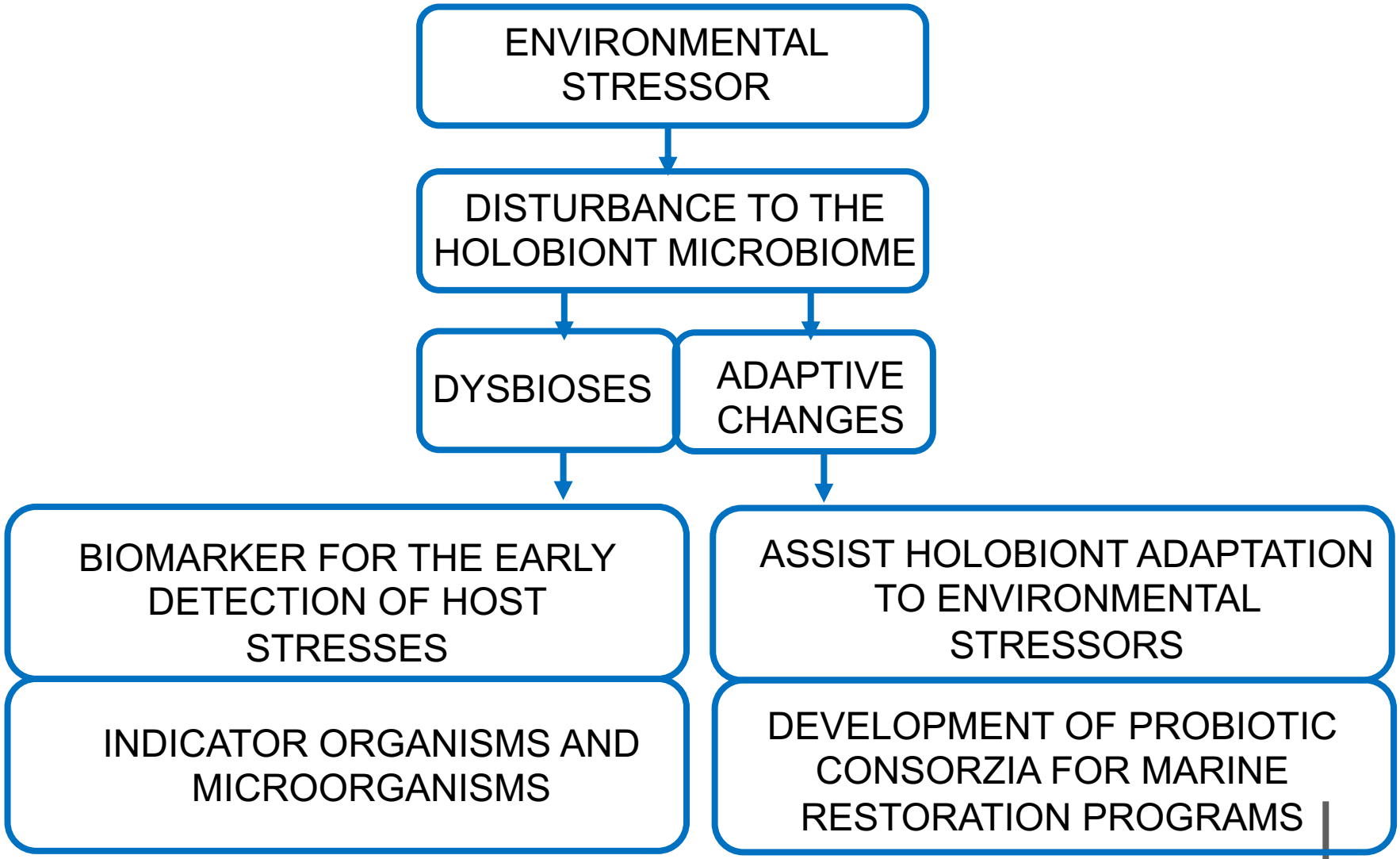


HOLOBIONT MICROBIOMES AND STRESS RESPONSE

AS HABITAT FORMING MARINE HOLOBIONTS POSSESS RELATIVELY SHORT GENERATION TIMES ARE INHERENTLY ABLE TO SELECT FOR MICROBIOME STRESS-DEPENDENT ADAPTIVE RESPONSES IMPROVING THE HOST FITNESS UNDER THE STRESS CHALLENGE

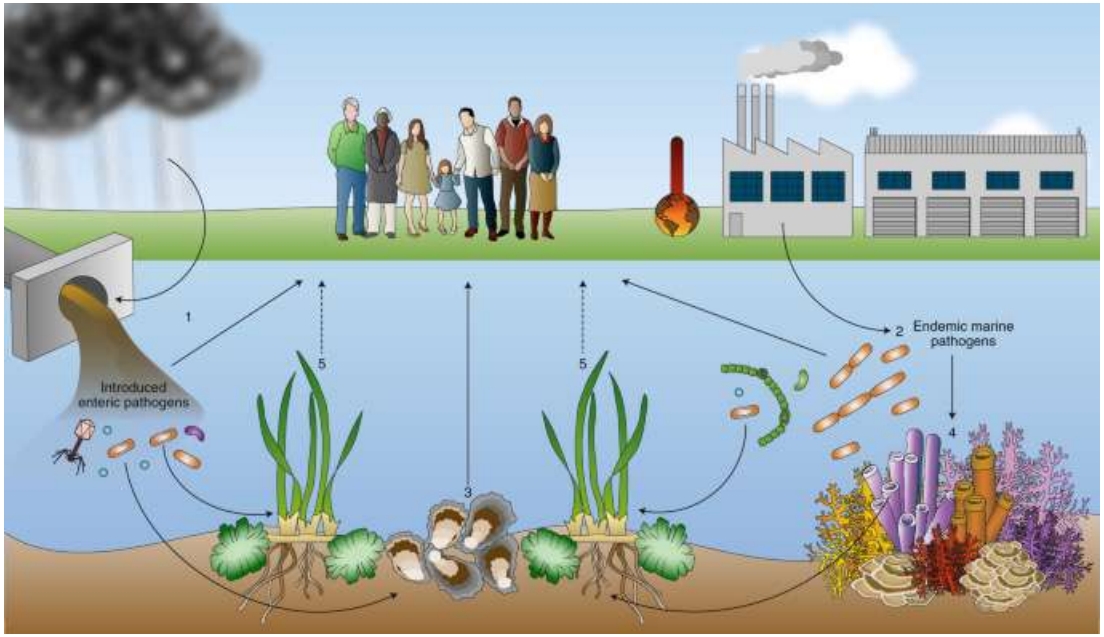
the microbiomes from habitat forming marine holobionts is likely to be instrumental in assisting adaptation of the host to climate change and atrophic stressors

MICROBIOME AND RESILIENCE OF THE HOLOBION HEALTH



Coastal holobiont microbiome at the interface with human health

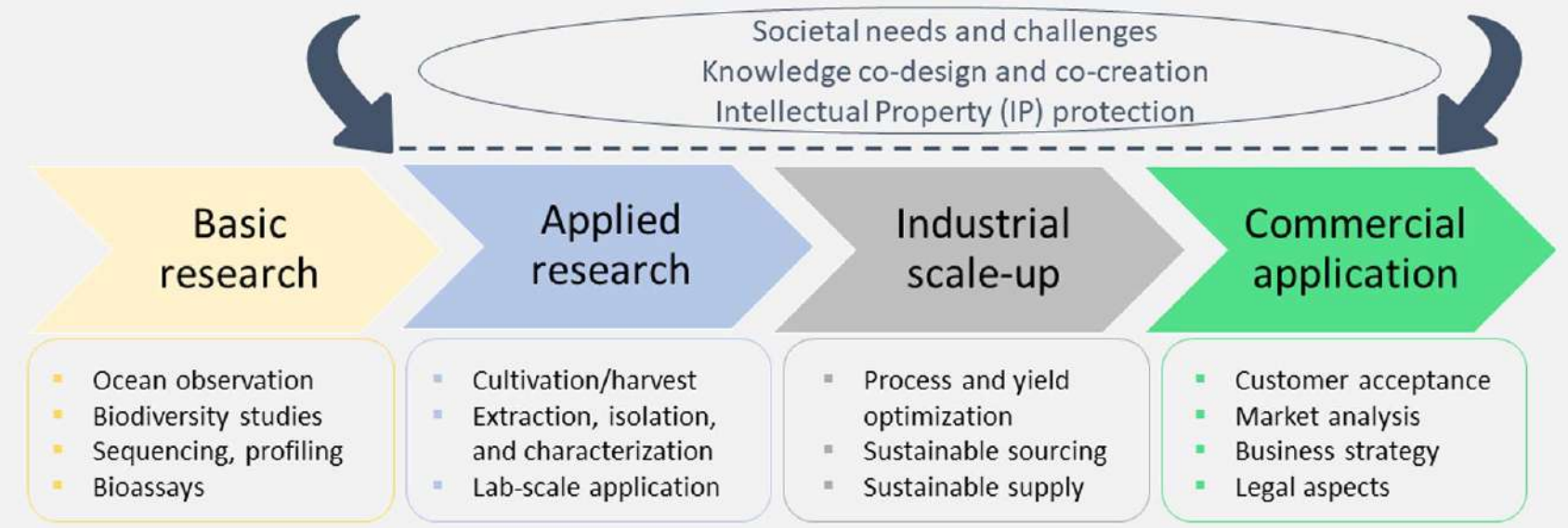
Human pathogens and enteric microbiomes are exogenously introduced to coastal habitats via swage and urban stormwater. **Release of of pollutants** provide a direct atrophic stress to the coastal habitat, with a cascade effect on marine microbiomes and holobionts



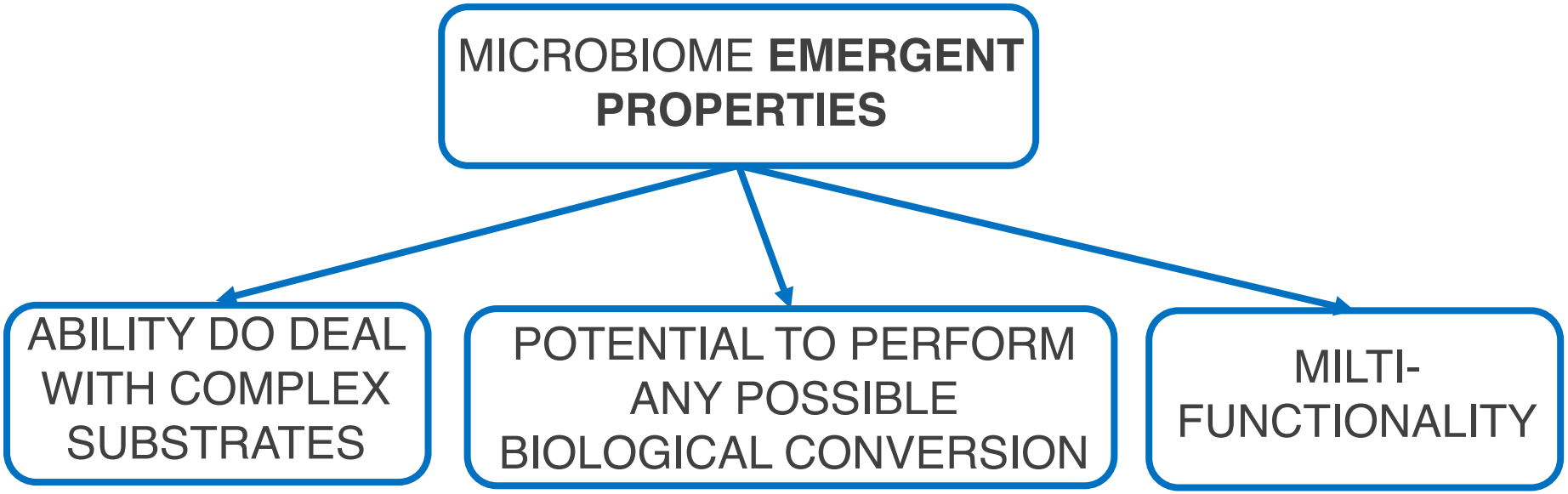
enteric pathogens becoming, transiently, incorporated into marine microbiomes following exposure to coastal pollution pose significant health risk and microbiome from coastal organisms potentially represent an hotspot and reservoir of human pathogen

there is now the evidence that some marine organisms (eg seargreaaes) may act as effective natural filtration systems removing pathogens from the coastal ecosystems, by the production of biocides

MARINE MICROBIOME BIODISCOVERY AND BIOTECH APPLICATIONS



EVOLVED TO DEAL WITH COMPLEX FUNCTIONS



understanding ultracomplex microbial communities and unraveling the **mechanisms providing the microbiome emergent properties**

THE MARINE MICROBIOME BIOTECNOLOGICAL POWER

MARINE HOLOBIONTS MICROBIOMES

SPONGES

SHELLFISH

CNIDARIANS

TUNICATES

BRYOZOANZ

FREE LIVING MARINE MICROBIOMES

WATER

SEDIMENTS

UNLIMITED BIOTRANFORMATION CAPACITY

UNIQUE CHEMICAL DIVERSITY OF BIOACTIVE METABOLITES

REMOVAL OR DEGRADATION OF CHAMICAL COMPOUNDS

BIOTRANSFORMATION AND VALORIZATION OF SIDE STREAMS

PROVISION OF INDUSTRIAL PRODUCTS

BIOREMEDIATION

AGRONOMY

FOOD

COSMETICS

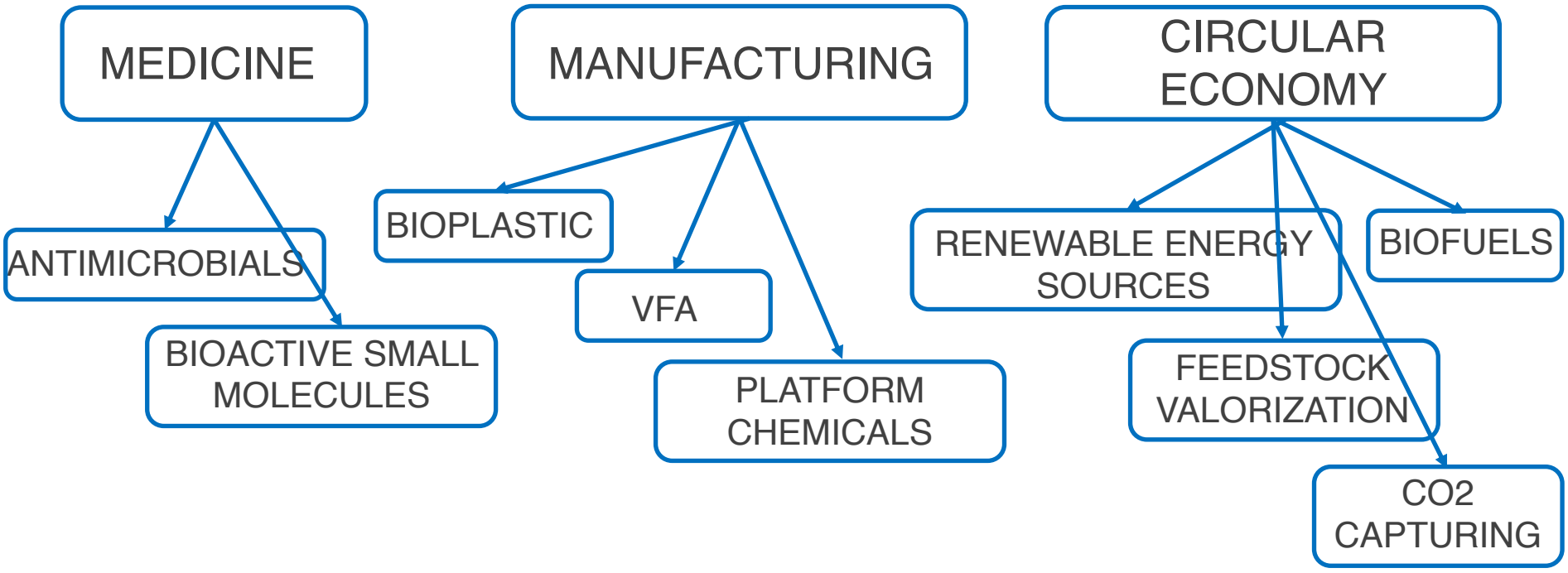
BIO-INSPIRED MATHERIALS

FEED

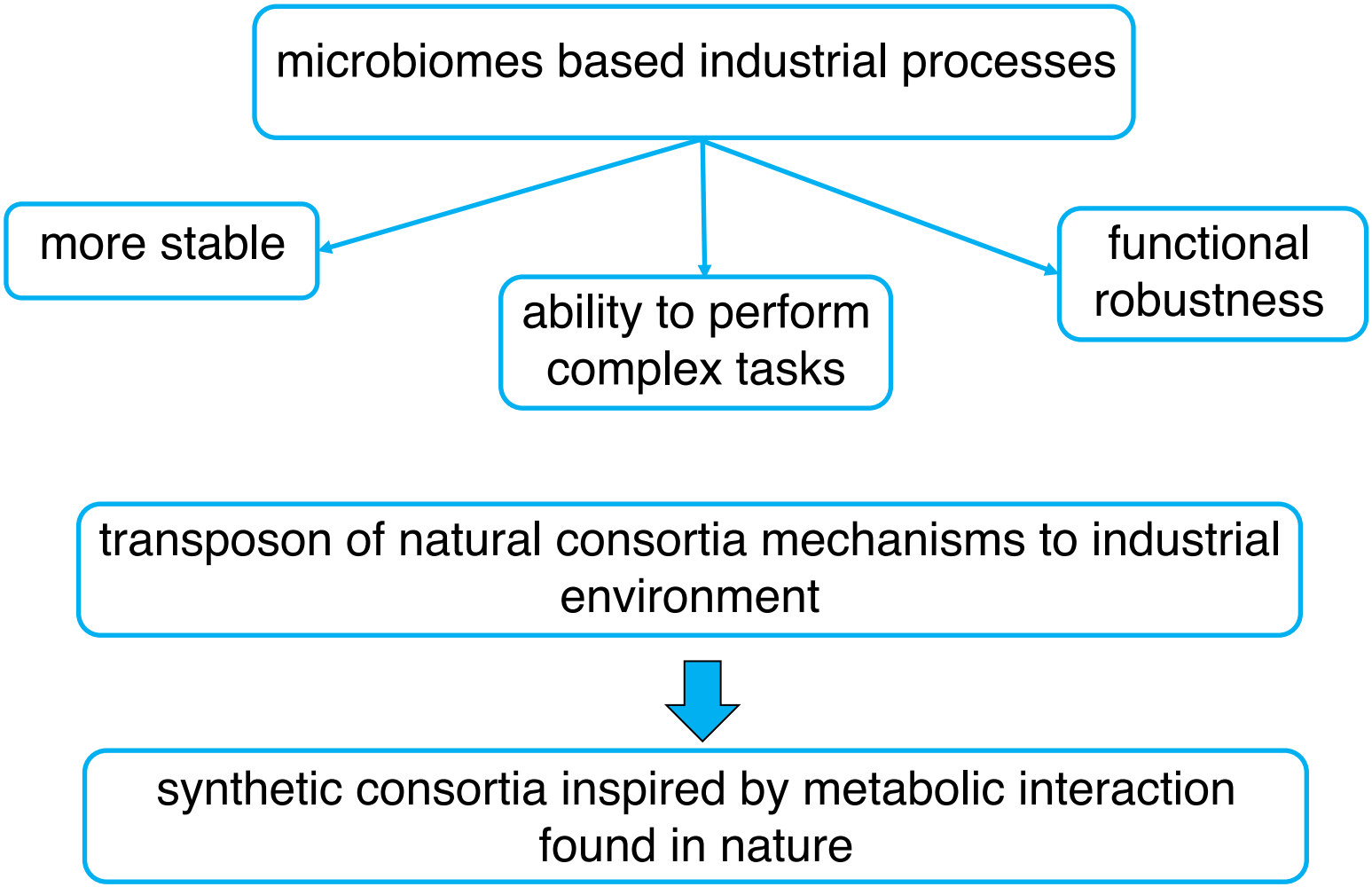
PHARMACEUTICALS

MICROBIAL COMMUNITIES

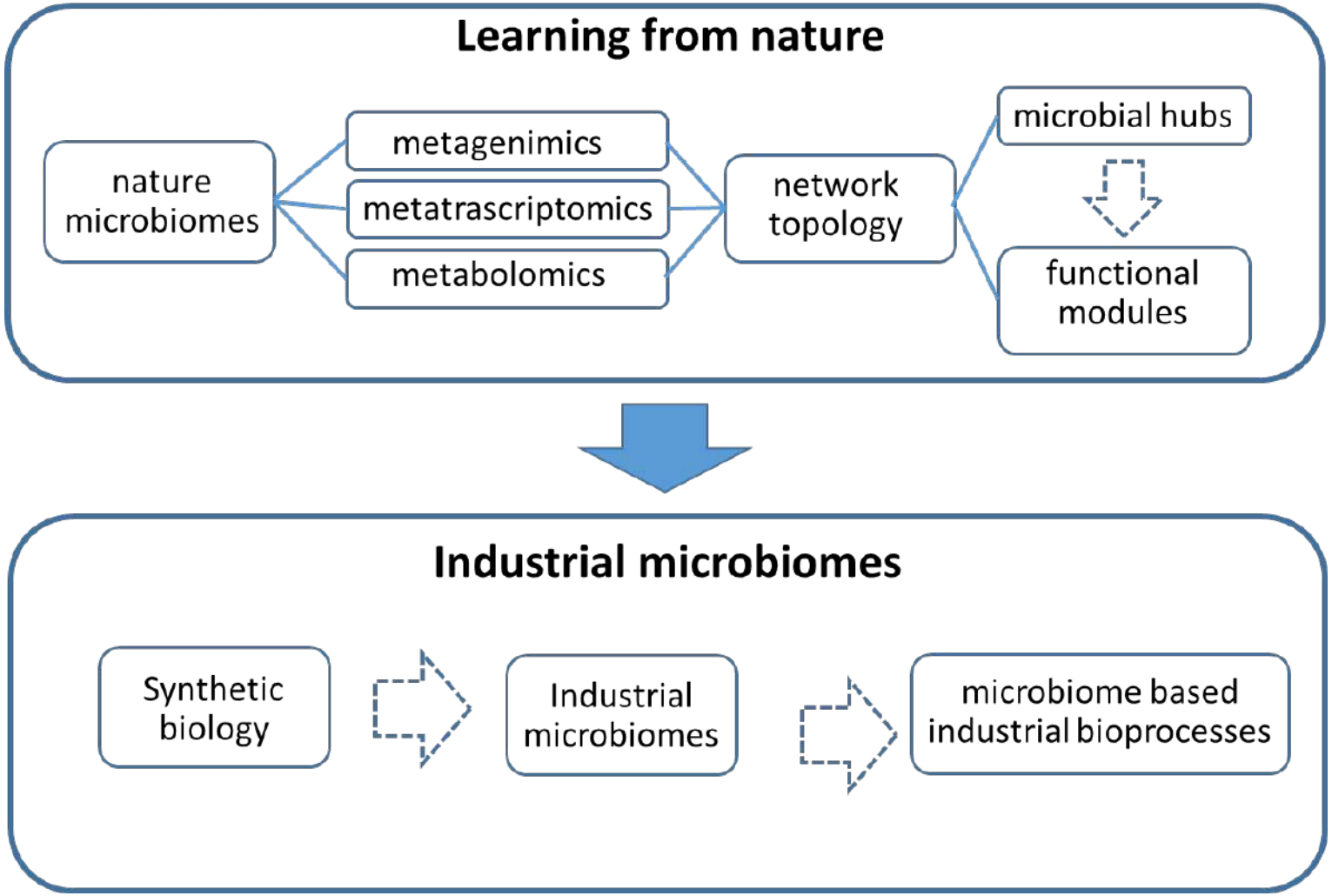
natural microbiomes have **a limitless transformative capacity** and provide an untapped source for **new, high performative biotechnological applications**



LEARNING FROM NATURE, MICROBIOMES BIOTECNOLOGY



Metagenomics, from nature to industrial microbiomes



FROM ECOLOGICAL AND FUNCTIONAL PRINCIPLES TO INDUSTRIAL MICROBIOMES

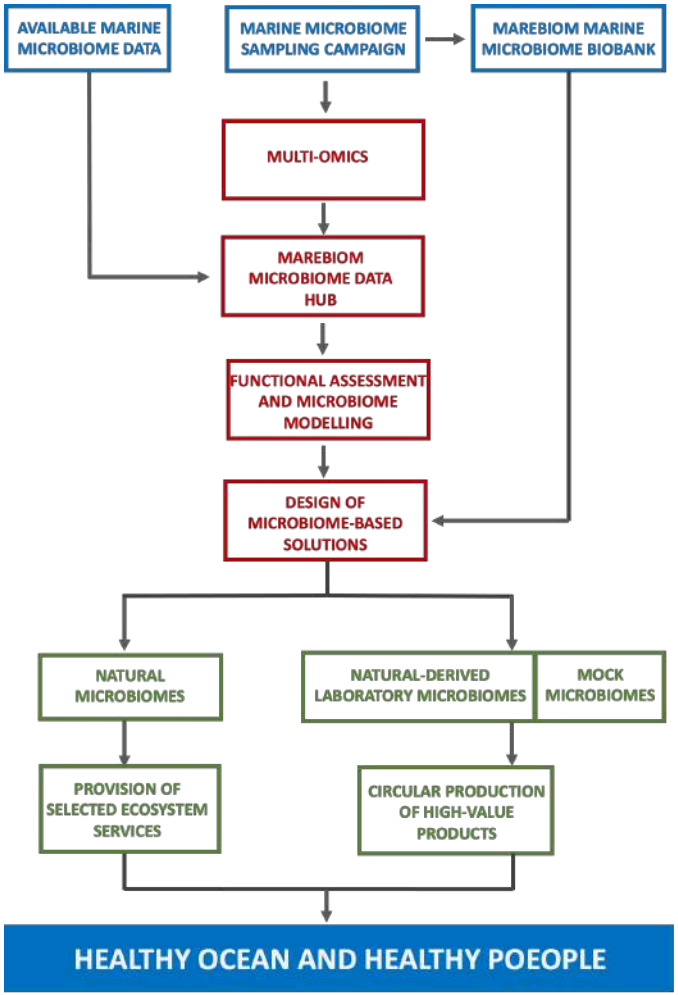
APPLY THE BEST AVAILABLE SCIENCE TO ELUCIDATE THE ECOLOGIC AND FUNCTIONAL MICROBIAL INTERACTIONS IN **NATURAL MICROBIOMES**

TRANSPOSE THE KNOWLEDGE TO THE RATIONAL DESIGN OF **INDUSTRIAL MICROBIOMES** TO BE EXPLOITED IN NOVEL MICROBIOME BASED BIOPROCESSES

Marine microbiomes for health and sustainability

Microbiomes provide an untapped source of solutions for the transition to more sustainable planet

Their complexity and multifunctionalities can be exploited for the sustainable production of industrial products, as well as for the provision of selected ecological services to protect the planet health.



Environmental exploration

Microbiome analysis

Microbiome-based applications

Compounds/ By-products

Market applications

Marine microbiomes

- Ecosystem studies
- Data analysis
- Technological development
- (Other habitat?)

Environmental microbiome

Host-associated microbiome

Microbiome characterization

Modelling and dynamics

Microbial signalling and communication

Natural and synthetic community development

New genetic pathways identification

Interaction with the host and the surrounding environment

Microbiome cultivation and preservation

Microbial fermentation and compounds extraction

Fish waste and side streams valorisation

Fish farmers involvement as feedstock providers and end-users for aquaculture feeding

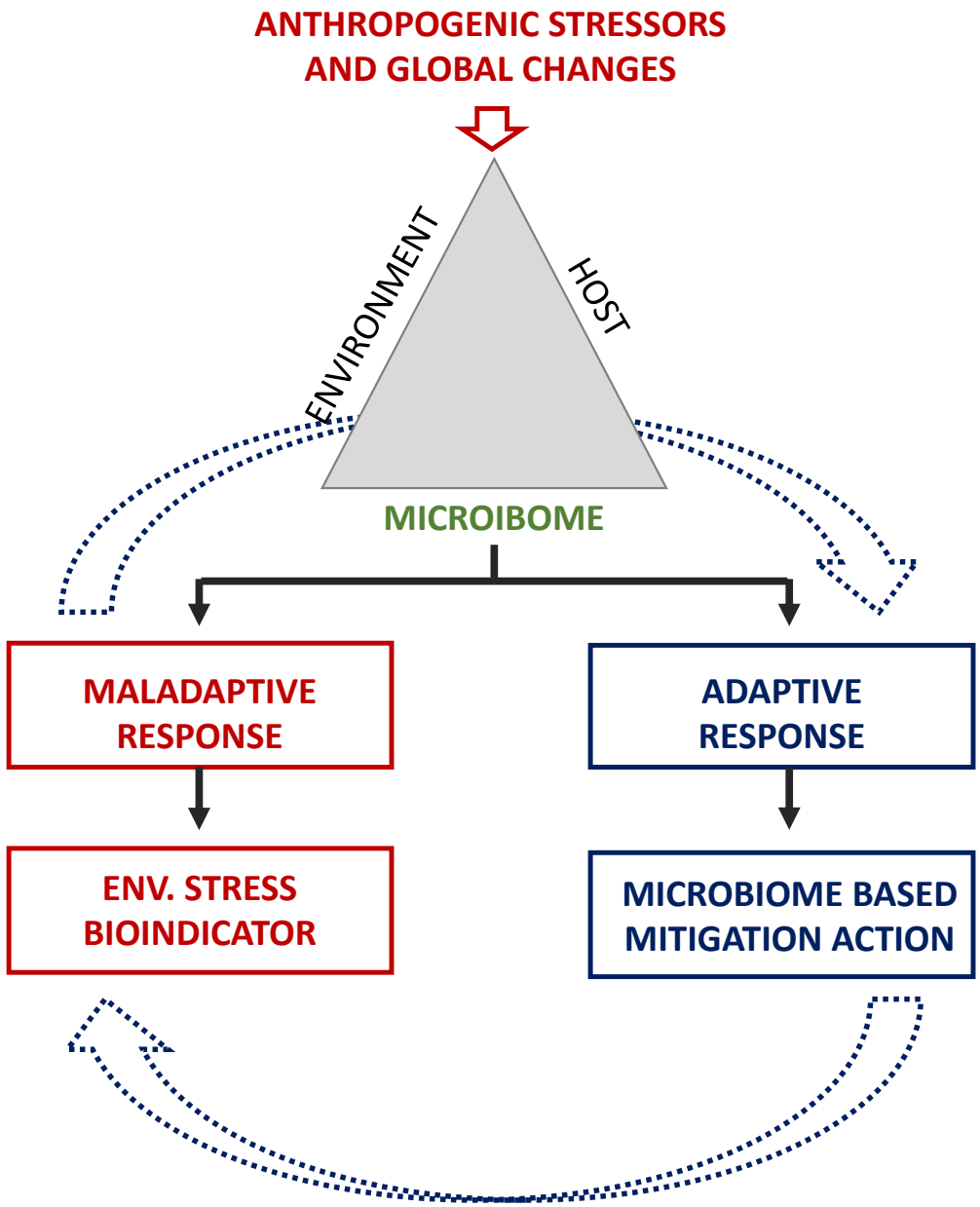
- **Peptides**
- **PUFAs**
- **Chitin and chitosan**
- **Enzymes**
- **Small molecules**
- **Others**

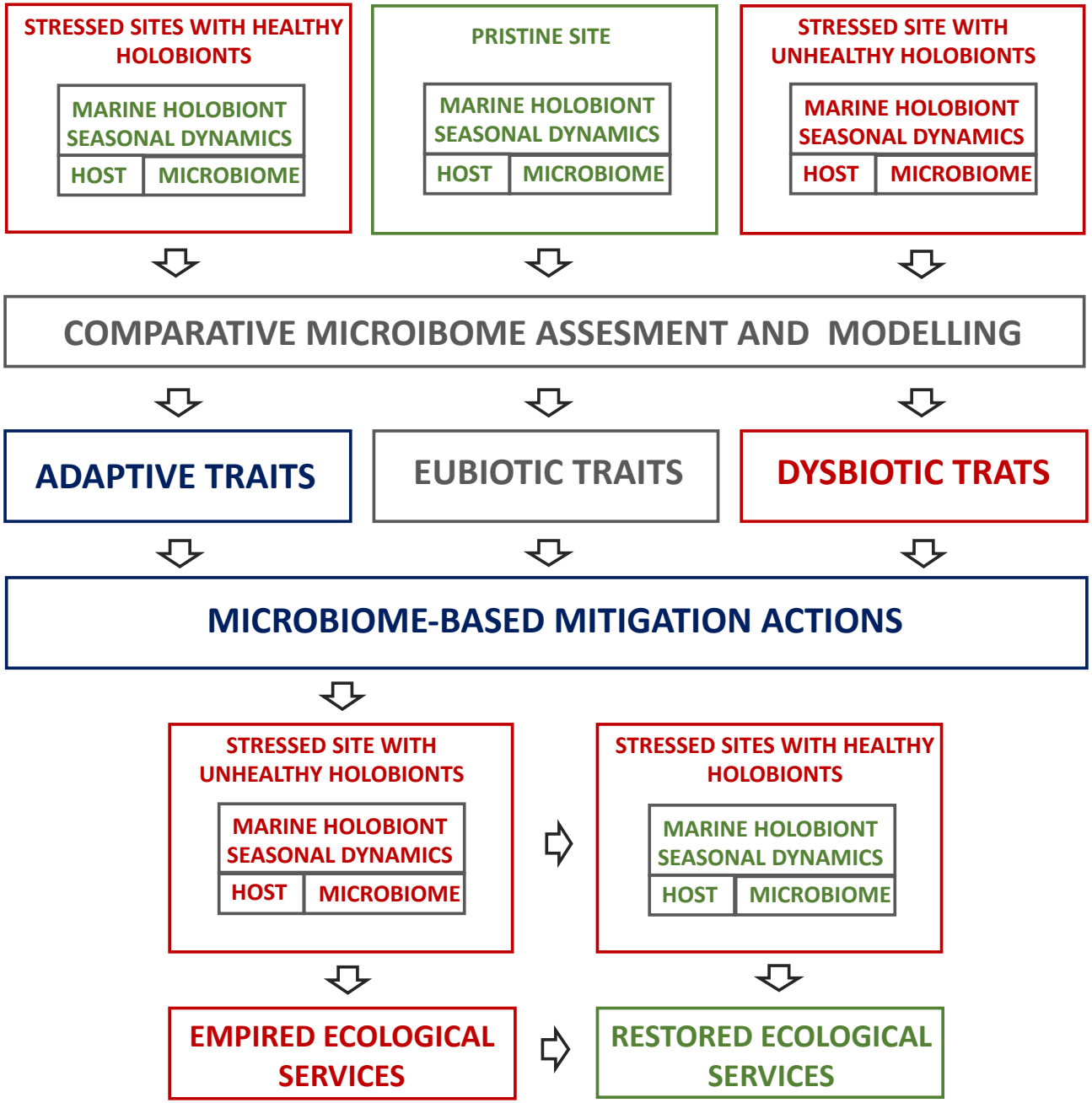
Circularity

Health

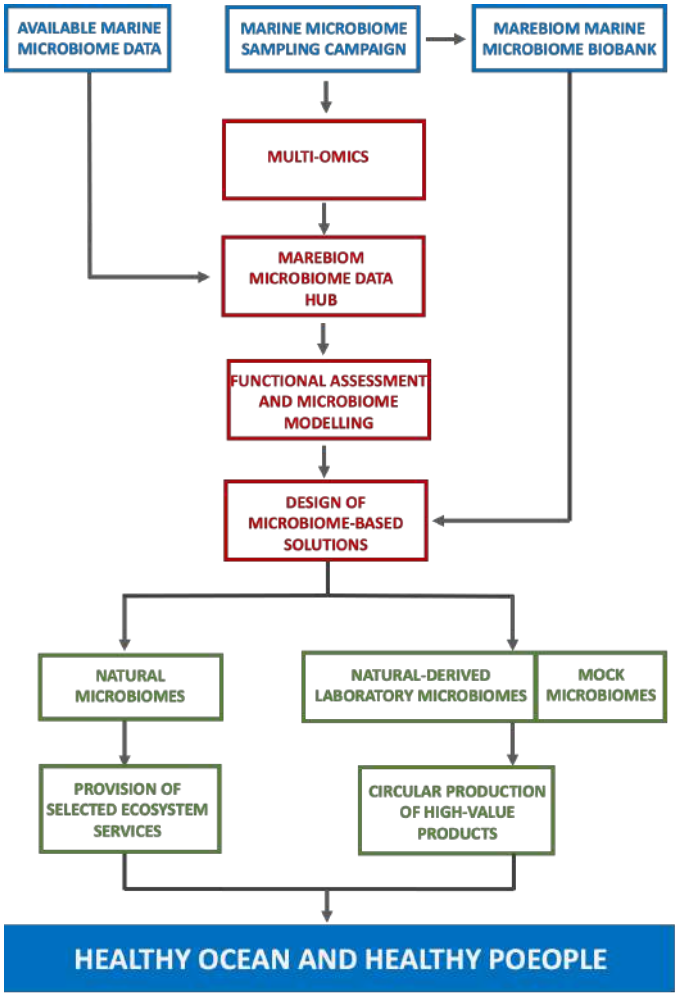
Cosmeceuticals

Food and Nutraceuticals


ANTHROPOGENIC STRESSORS
AND GLOBAL CHANGES
ENVIRONMENT
HOST
MICROIBOME
MALADAPTIVE
RESPONSE
ADAPTIVE
RESPONSE
ENV. STRESS
BIOINDICATOR
MICROBIOME BASED
MITIGATION ACTION


STRESSED SITES WITH HEALTHY HOLOBIONTS
MARINE HOLOBIONT SEASONAL DYNAMICS
HOST
MICROBIOME
PRISTINE SITE
MARINE HOLOBIONT SEASONAL DYNAMICS
HOST
MICROBIOME
STRESSED SITE WITH UNHEALTHY HOLOBIONTS
MARINE HOLOBIONT SEASONAL DYNAMICS
HOST
MICROBIOME
COMPARATIVE MICROIBOME ASSESMENT AND MODELLING
ADAPTIVE TRAITS
EUBIOTIC TRAITS
DYSBIOTIC TRATS
MICROBIOME-BASED MITIGATION ACTIONS
STRESSED SITE WITH UNHEALTHY HOLOBIONTS
MARINE HOLOBIONT SEASONAL DYNAMICS
HOST
MICROBIOME
STRESSED SITES WITH HEALTHY HOLOBIONTS
MARINE HOLOBIONT SEASONAL DYNAMICS
HOST
MICROBIOME
EMPIRED ECOLOGICAL SERVICES
RESTORED ECOLOGICAL SERVICES

The pipeline for the microbiome based biodiscovery action



Functional assessment and microbiome modelling

1. search - agnostically - for new secondary metabolites Biosynthetic Gene Clusters (SM-BGCs);

2. construct de novo the metagenome-assembled genomes (MAGs);

3. evaluate the selective pressures driving clades differentiation in marine species;

4. assess the MAGs for selected functions;

5. model microbiomes, deriving modules, hubs and pathways involved in selected functions and phenotypes;

6. search for microbiome functionalities for the green biosynthesis of bioactive compounds