

lucie.bittner@upmc.fr



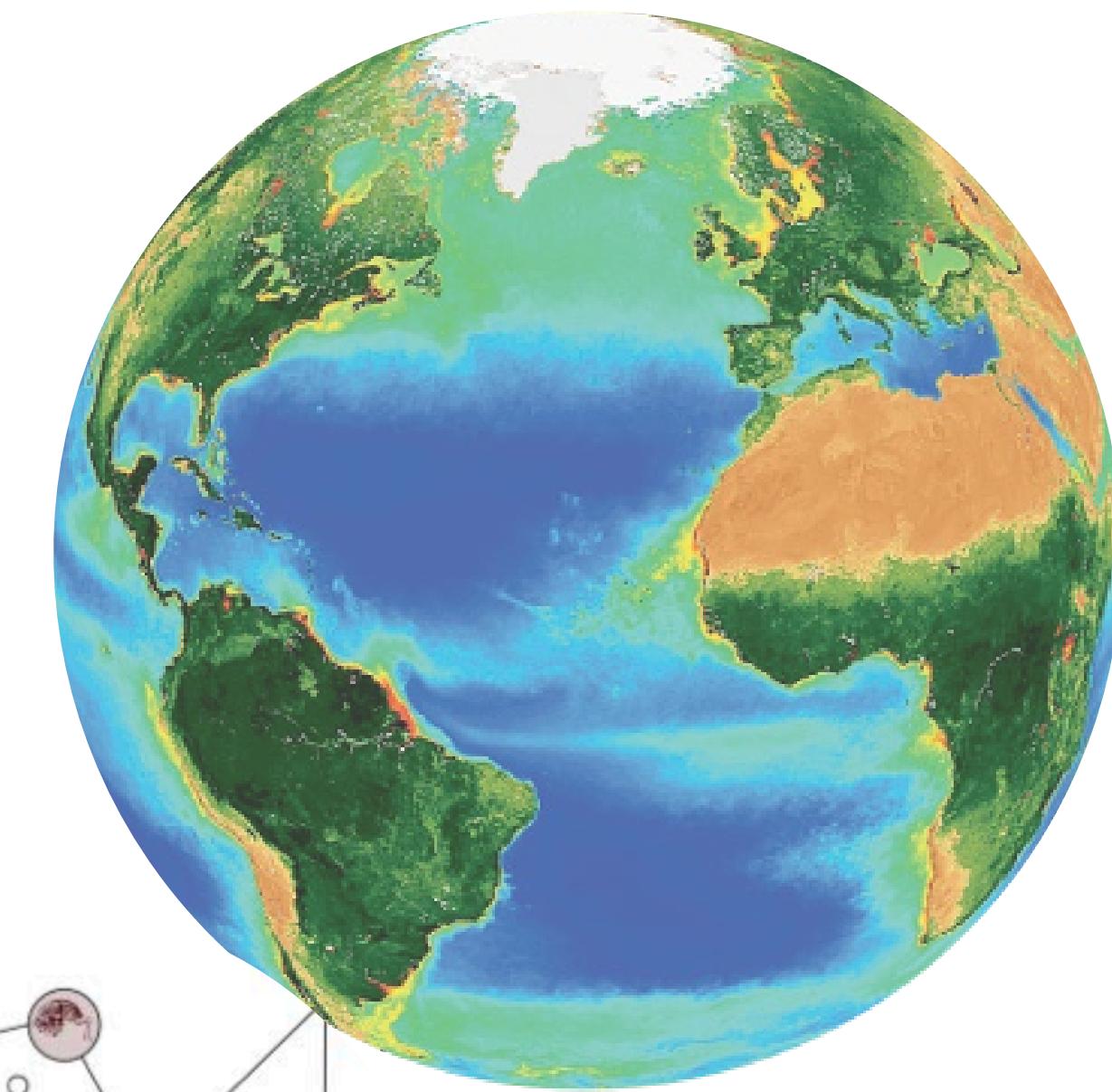
ISYEB
Institut de Systématique
Evolution Biodiversité



institut
universitaire
de France

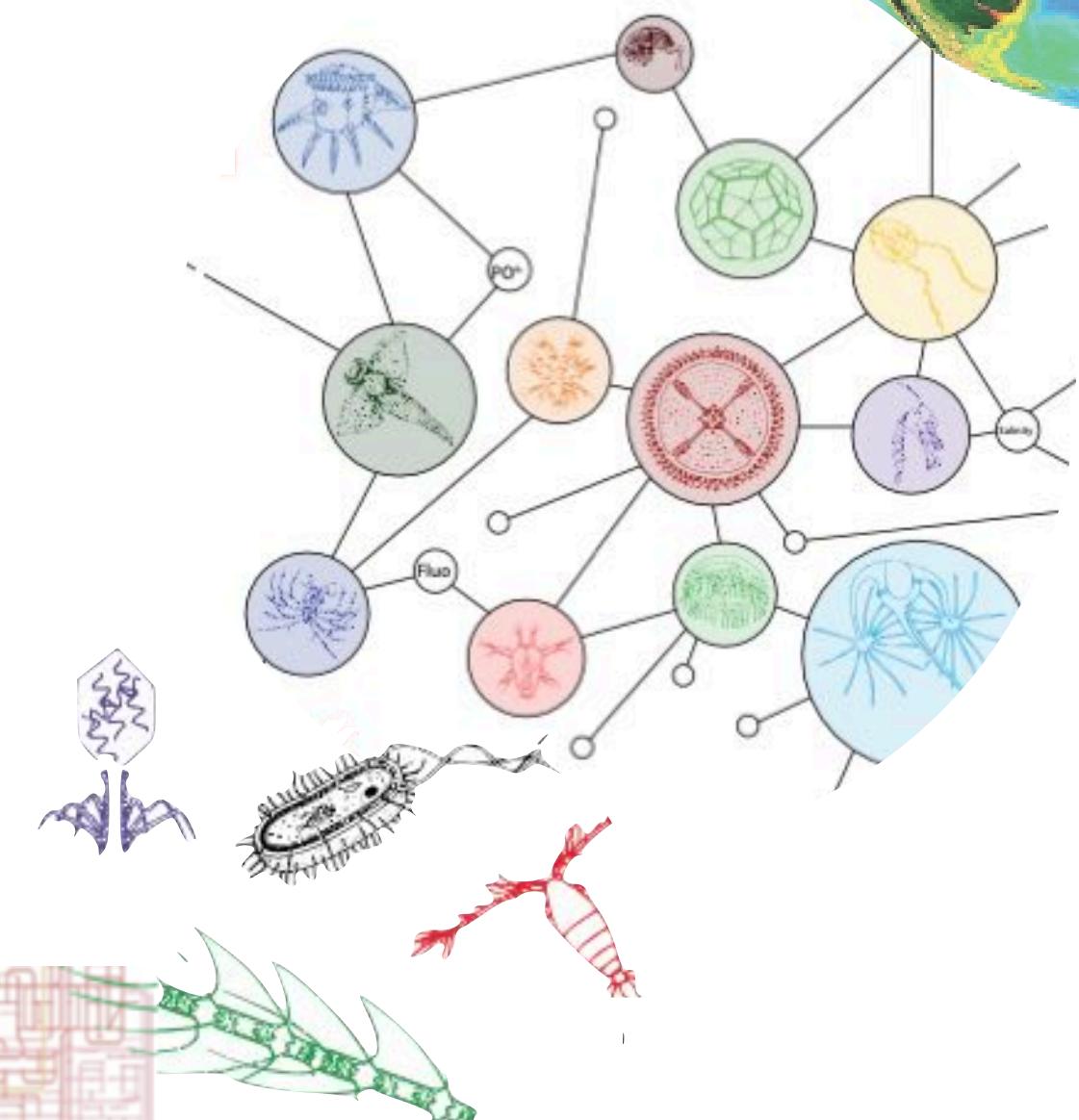


Atelier de
BioInformatique



Introduction to marine (meta)omics

Bridging the gap between meta-omics,
biodiversity perceptions and
earth system science



lucie.bittner@upmc.fr



ISYEB

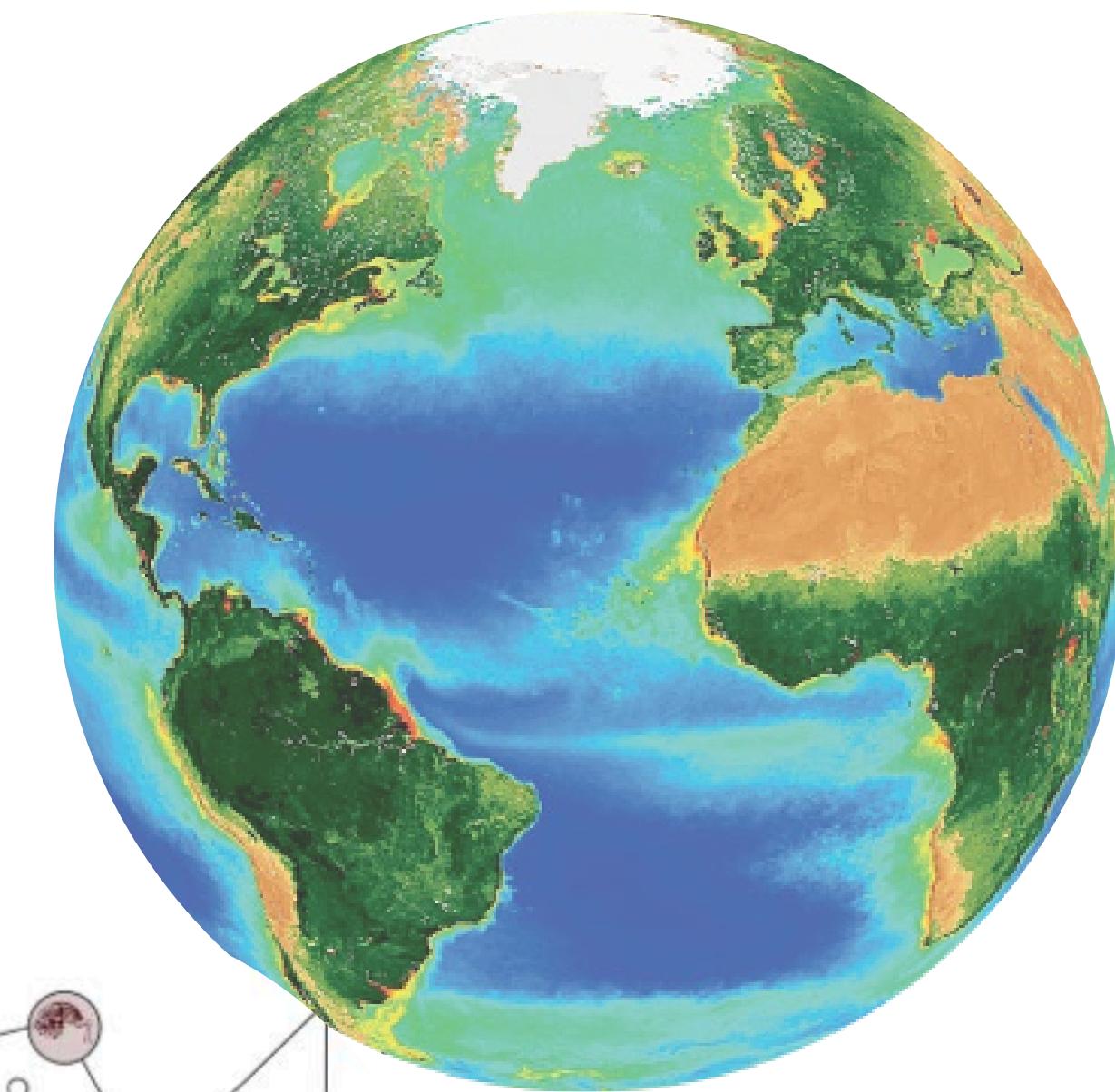
Institut de Systématique
Evolution Biodiversité



institut
universitaire
de France

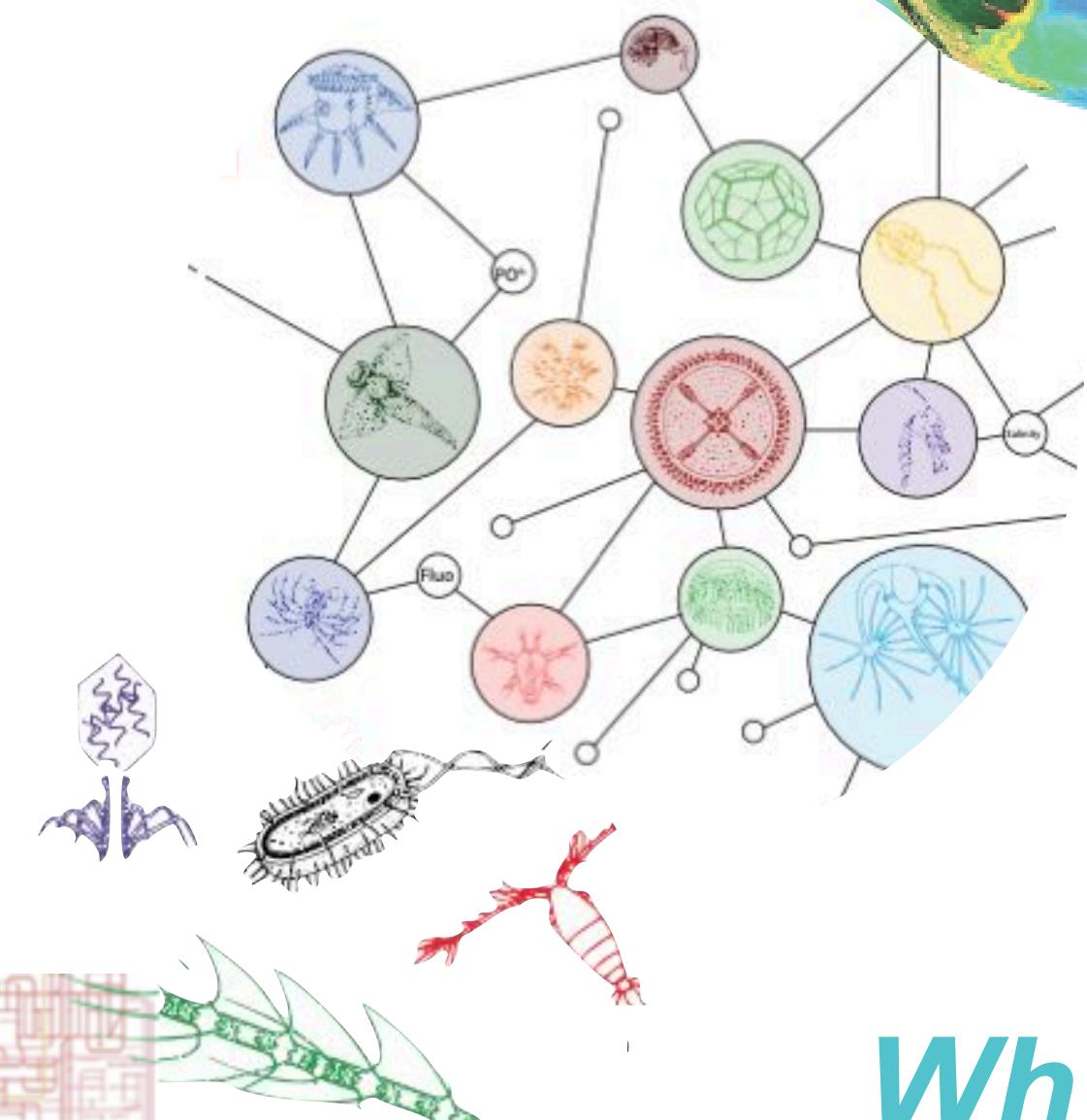
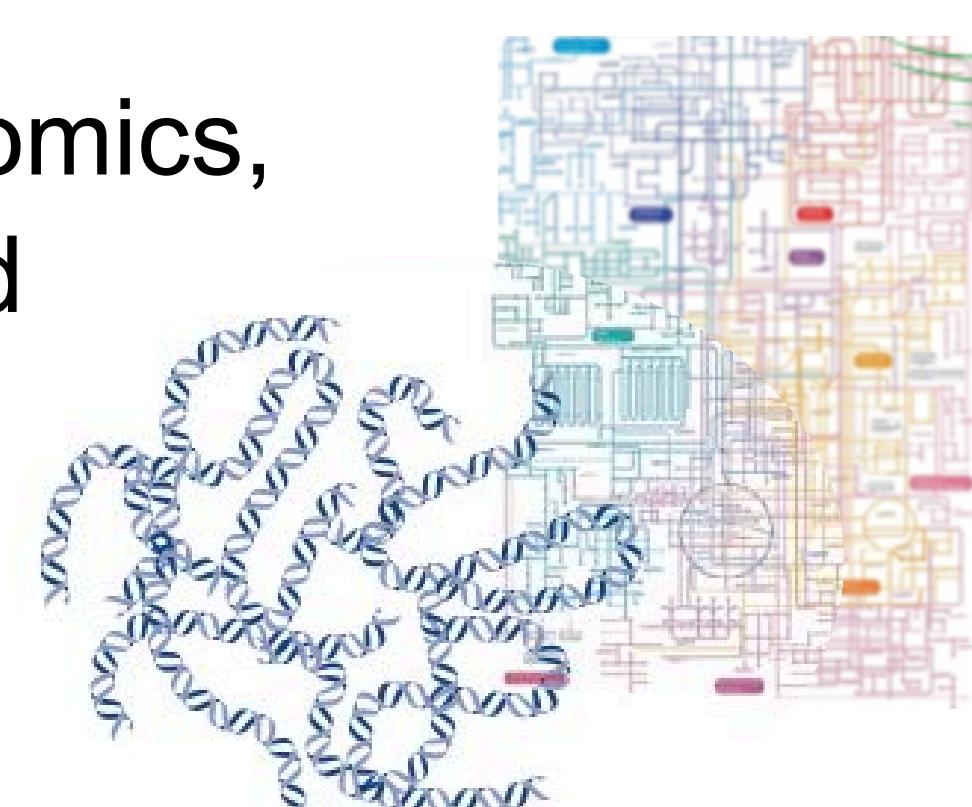


Atelier de
BioInformatique



Introduction to marine (meta)omics

Bridging the gap between meta-omics,
biodiversity perceptions and
earth system science



**Why do we
need ML ?**

DNA, RNA, 16S/18S, proteins,
(metabolites), ...

Environmental genomics

Microbiota, microbiomes

(meta)omics ?

Sequencing /
High-Throughput Sequencing (HTS)

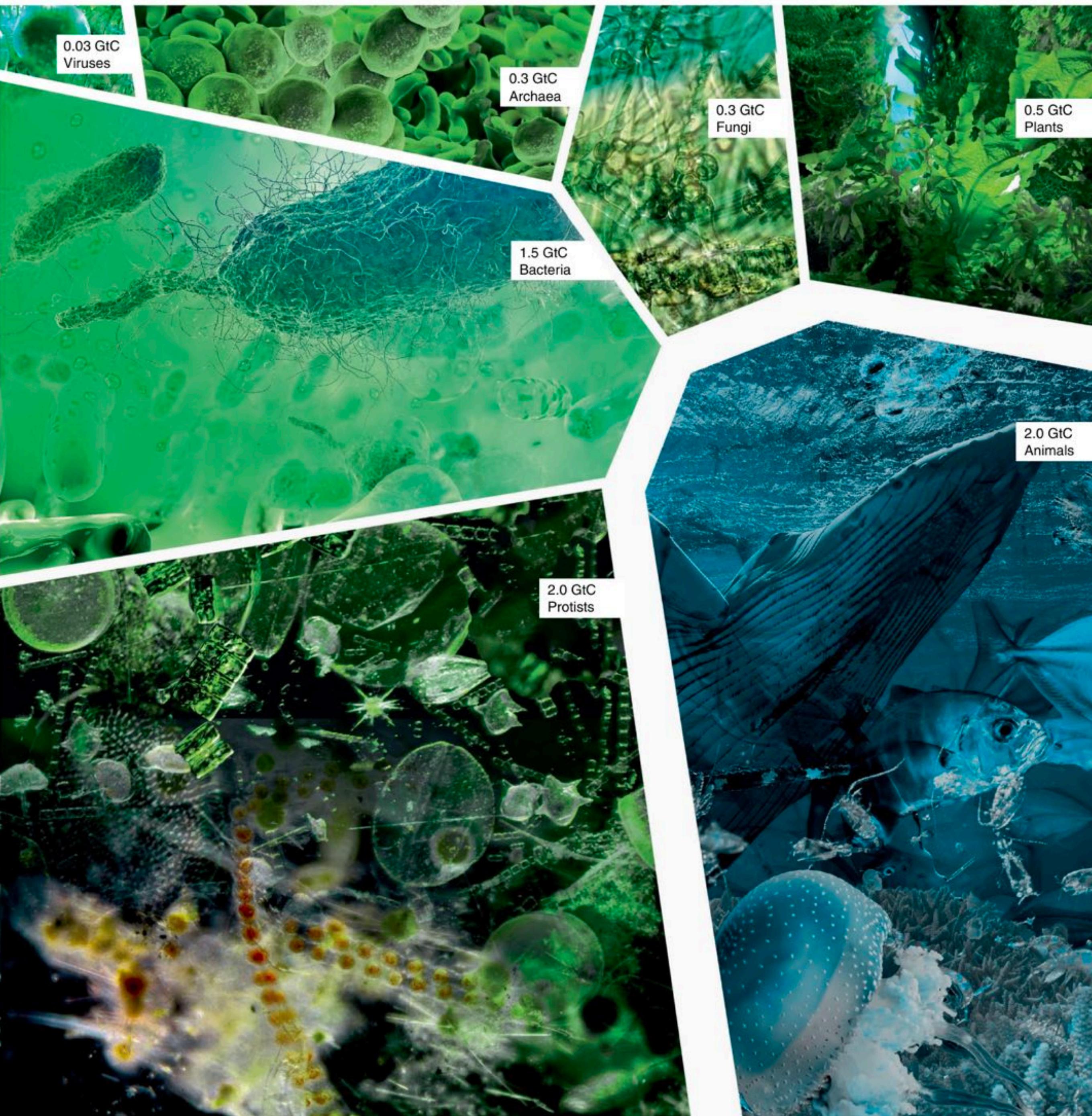
Microbes, functions, metabolisms, ...

Bioinformatics, Biostatistics, ML

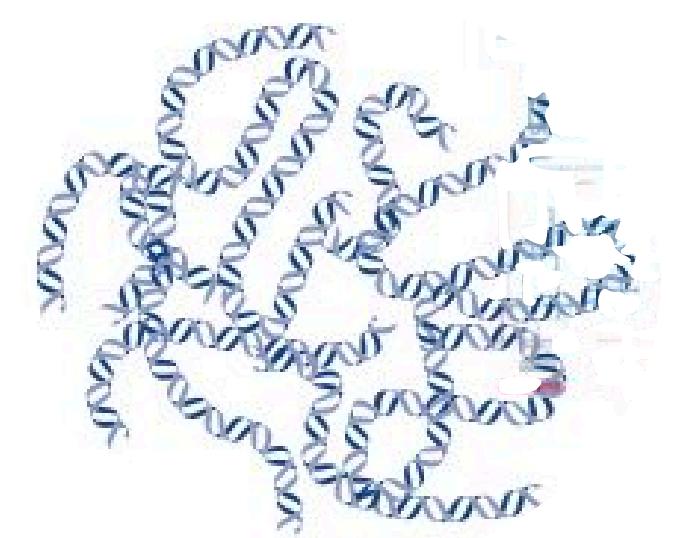
Biodiversity

Ecosystem services

- O₂ release, CO₂ fixation
- Base of food webs



(meta) omics



MOLECULES



ECOSYSTEMS

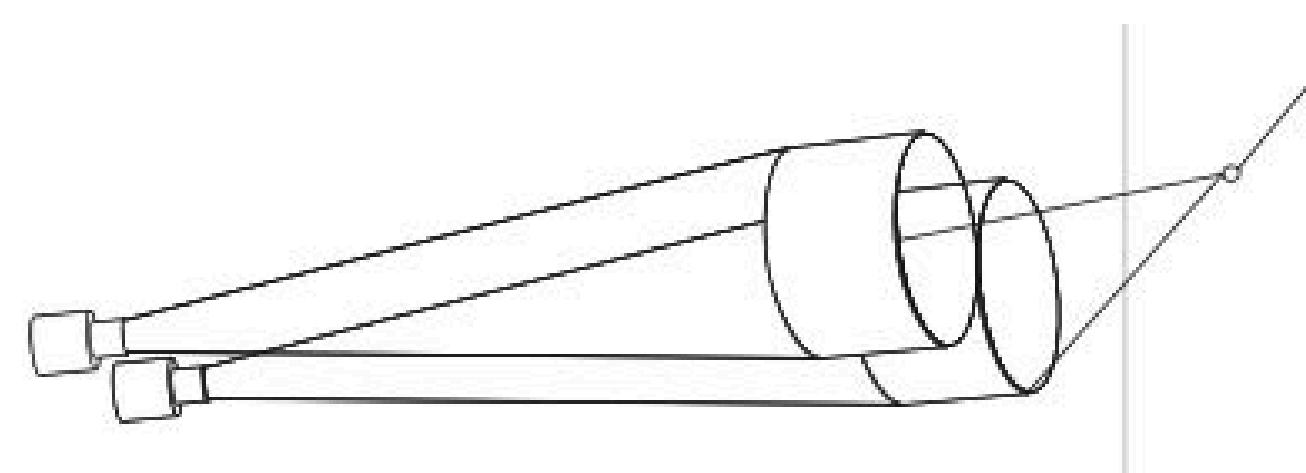


@Lucie_Bittner

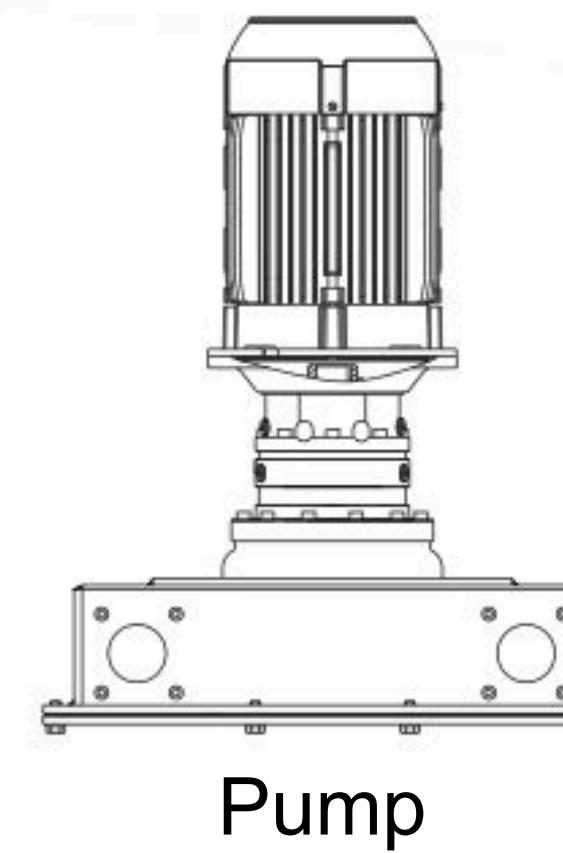
How do we obtain molecular sequences?



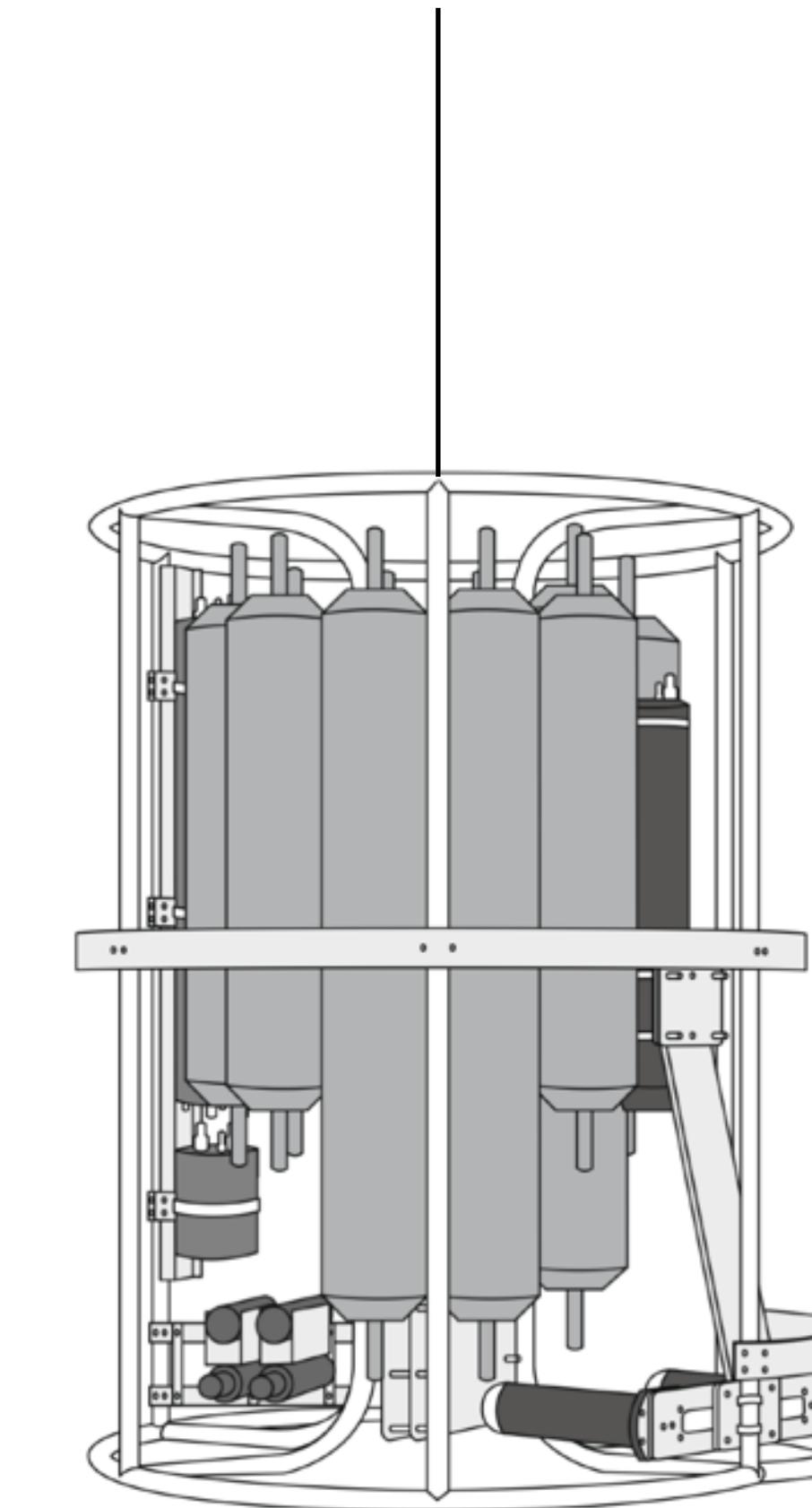
→
SAMPLING



Nets

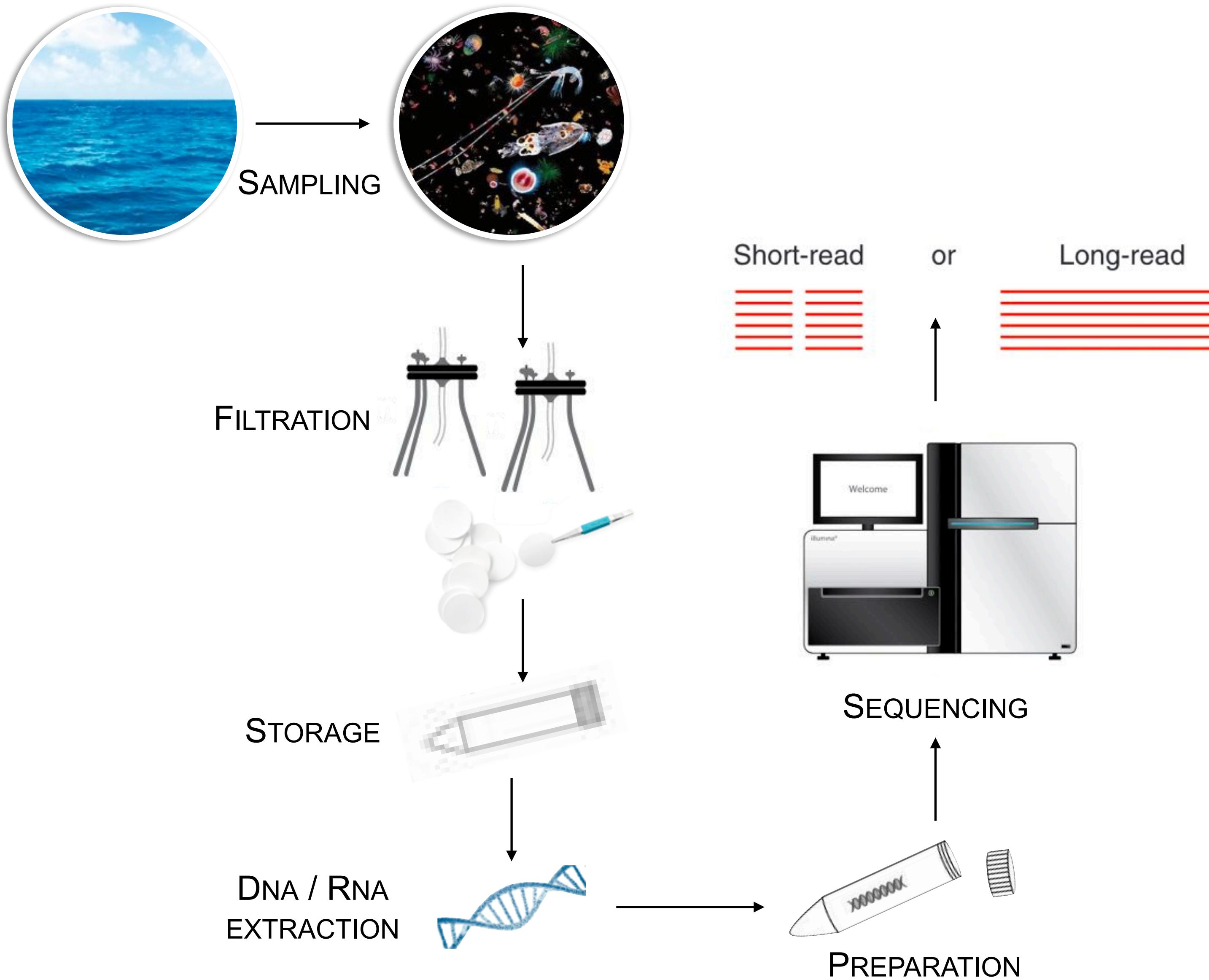


Pump



Niskin bottles on a
'Rosette'

How do we obtain molecular sequences?



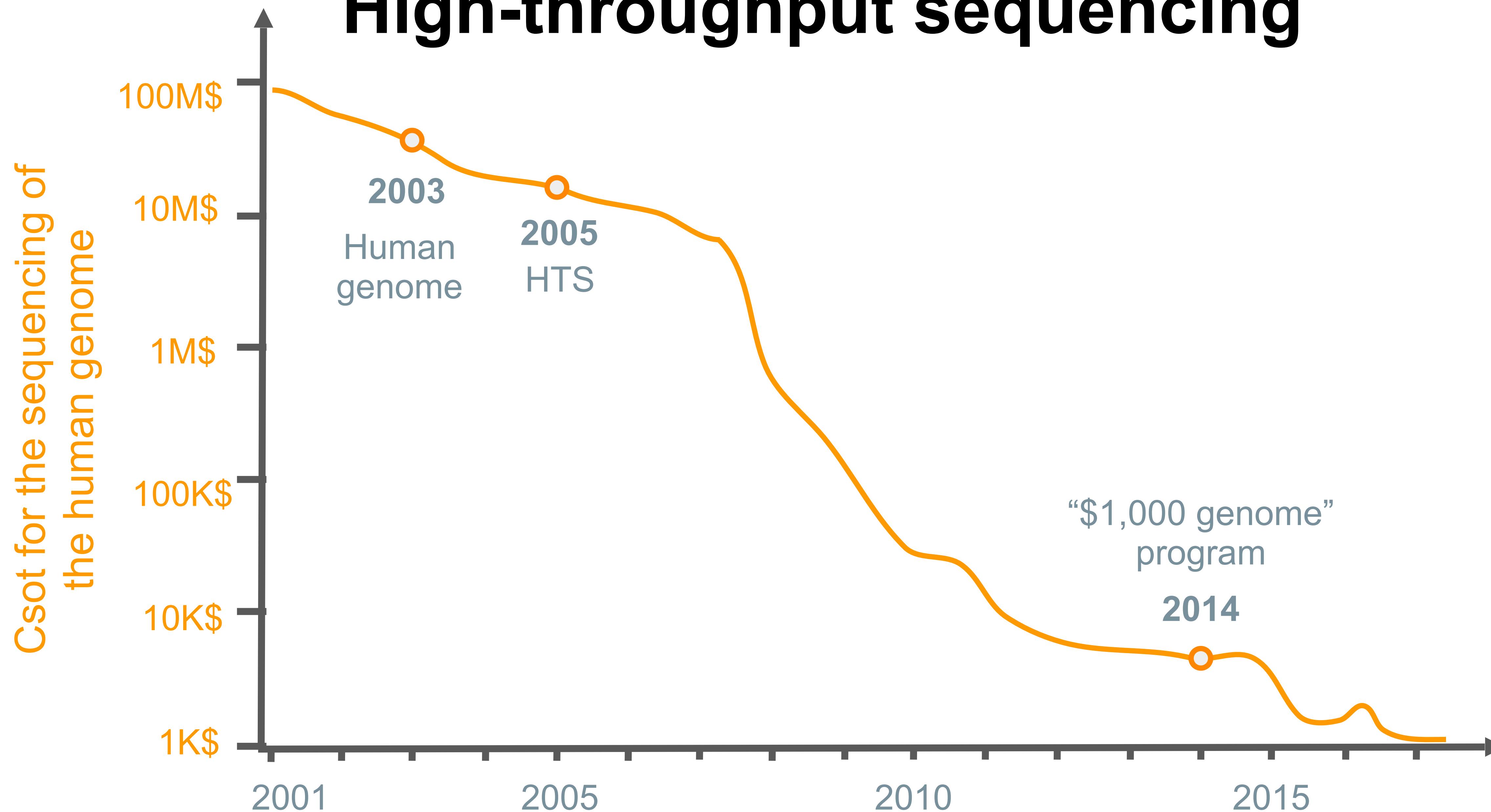
The omics deluge

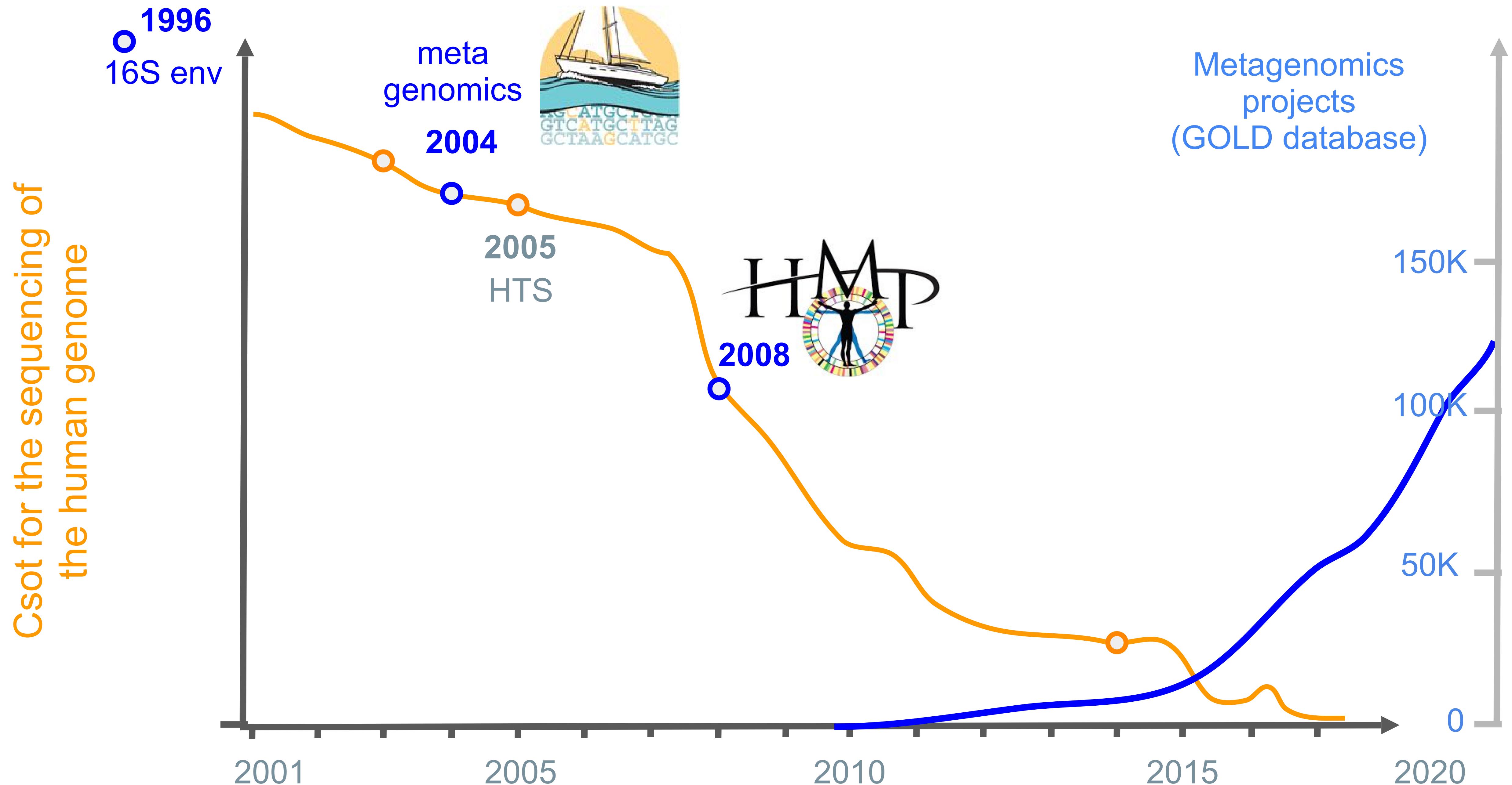


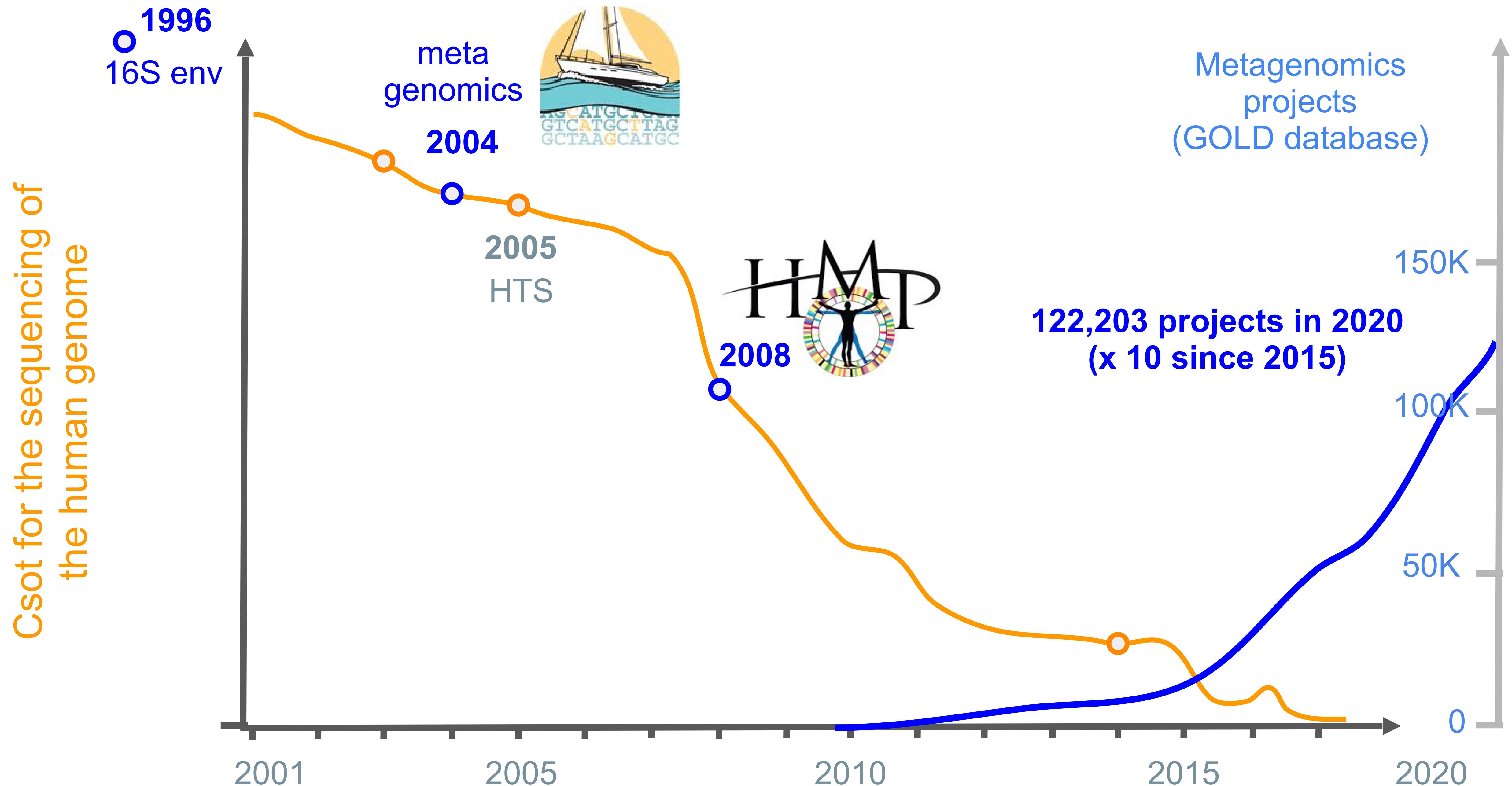
The omics deluge

 @Lucie_Bittner

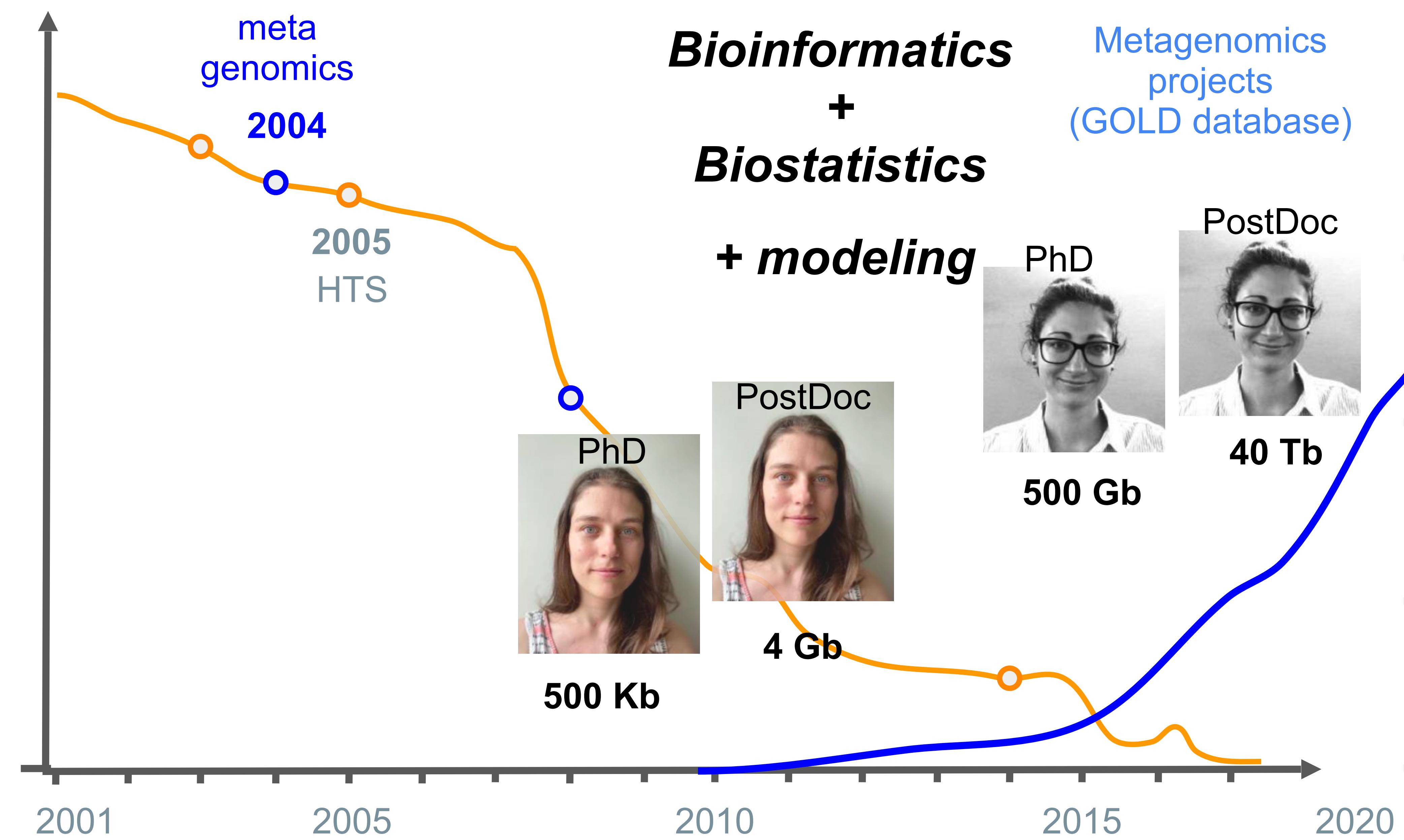
High-throughput sequencing



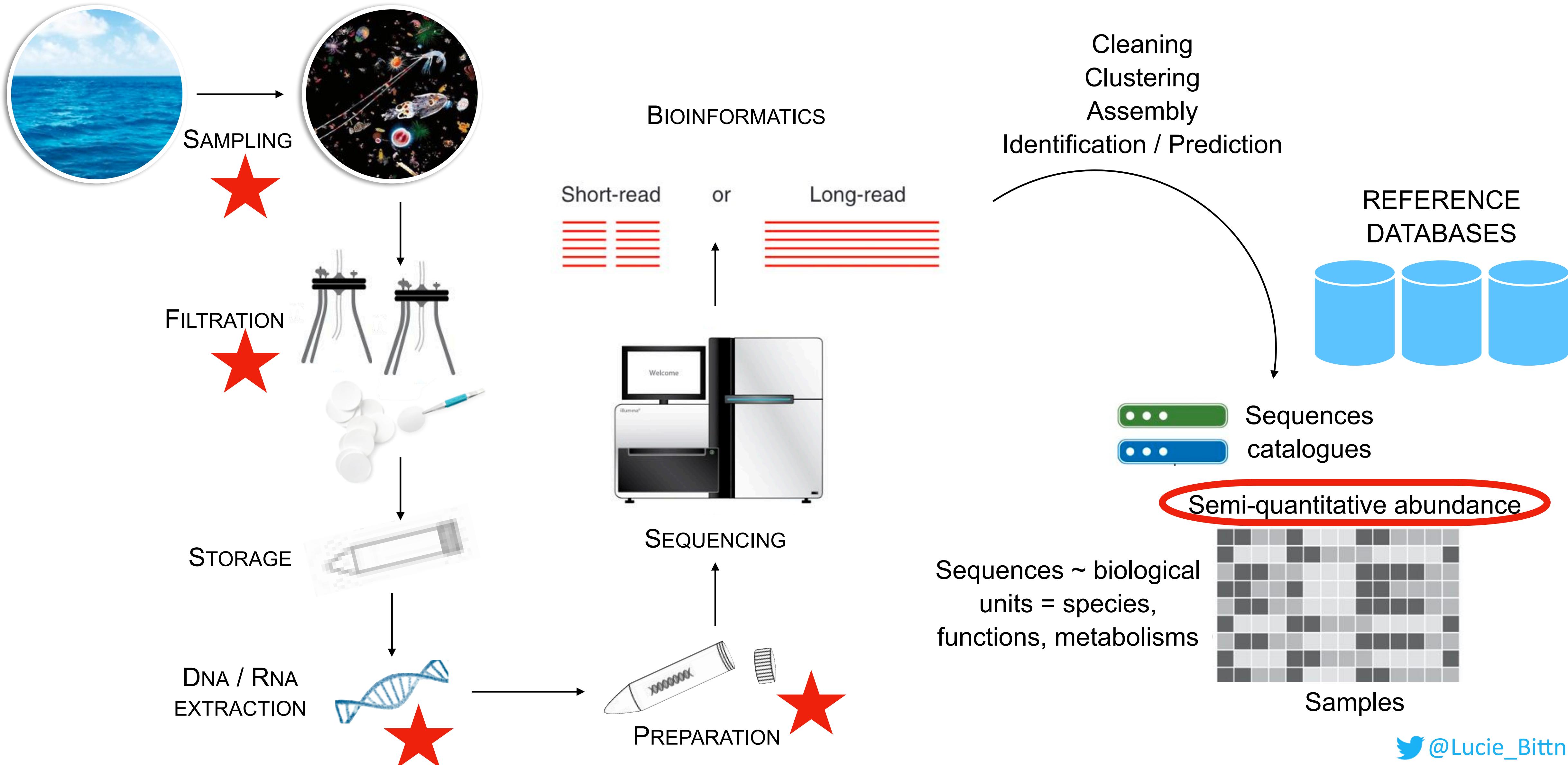




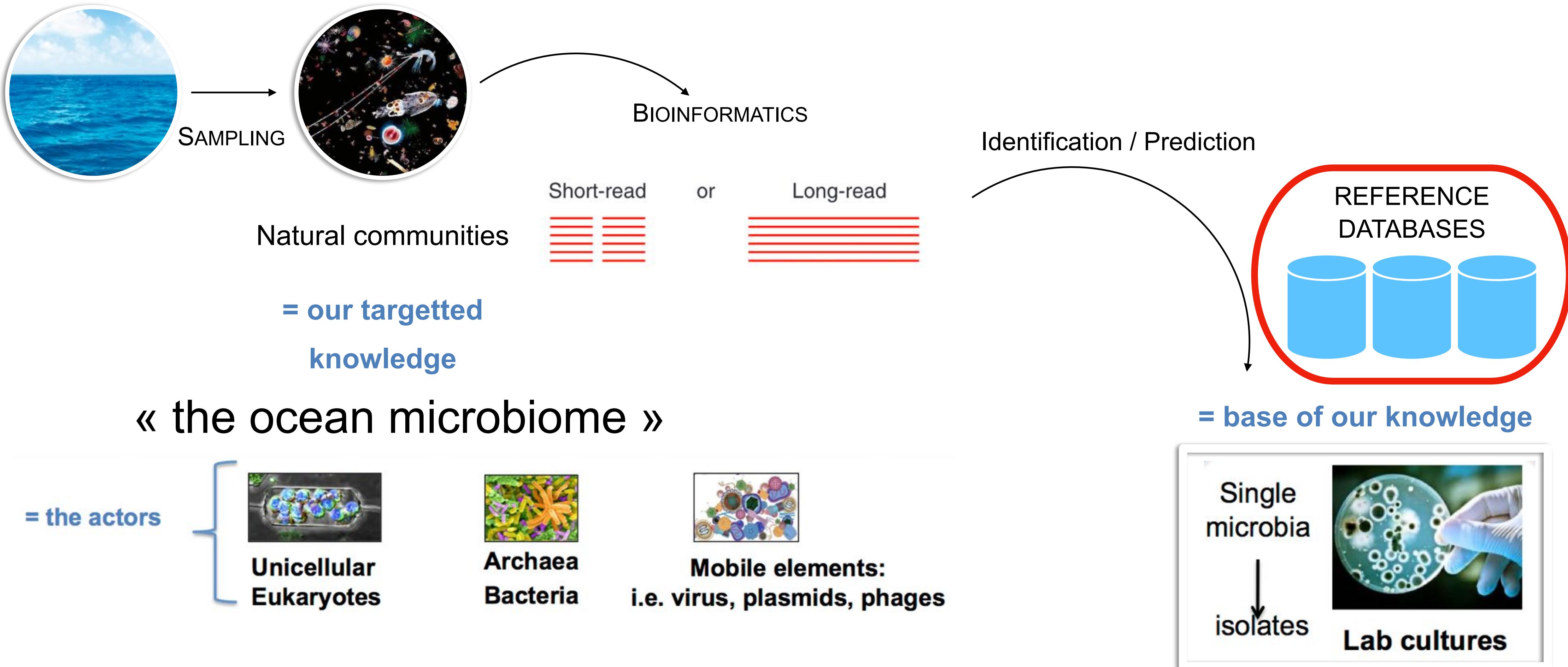
Cost for the sequencing of
the human genome



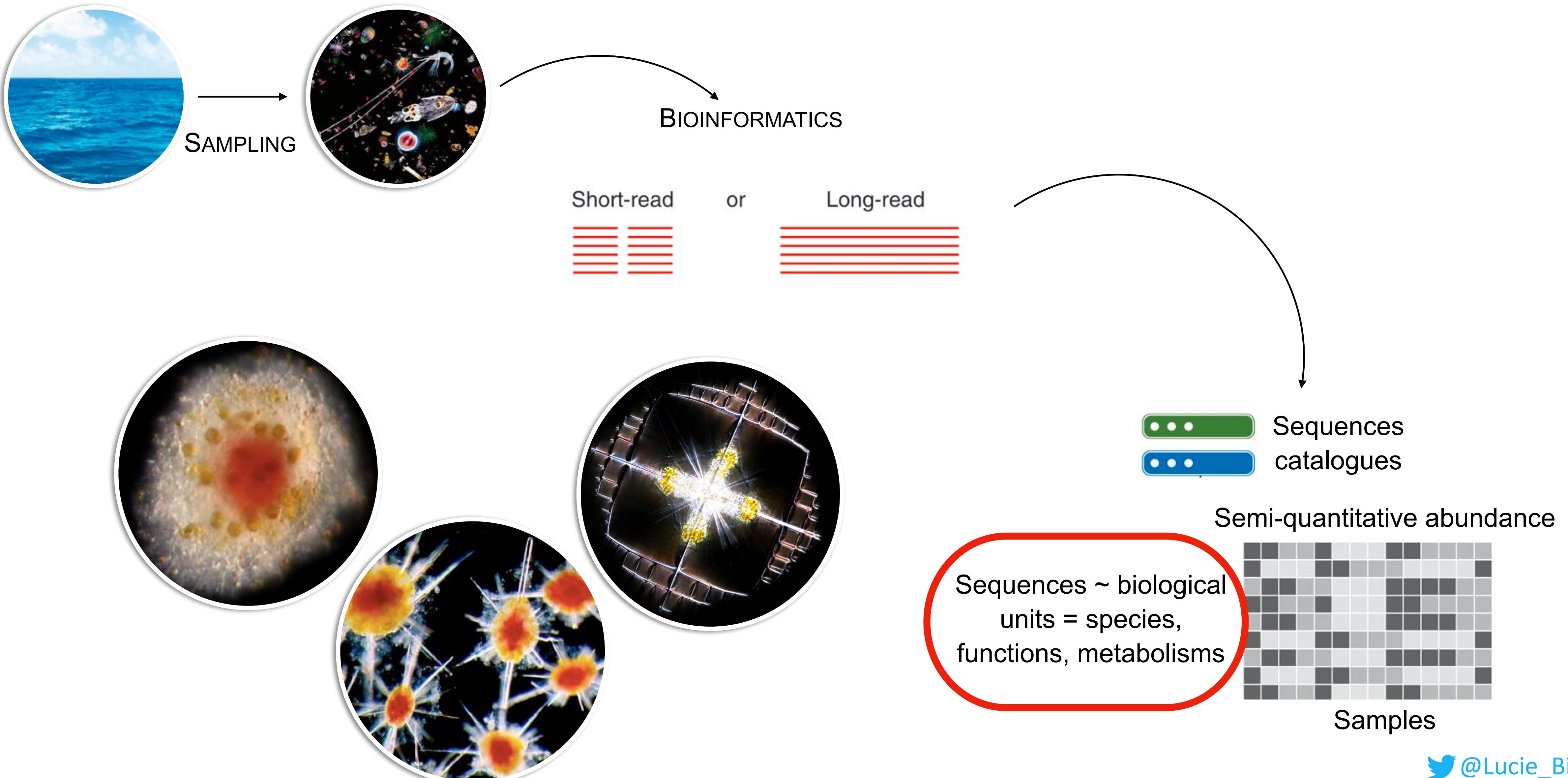
How do we obtain molecular sequences?



How do we obtain molecular sequences?



How do we obtain molecular sequences?



(Meta)omics and ocean ecosystems biology

Who is there?

What is done?

Photosynthesis

Growth

Grazing

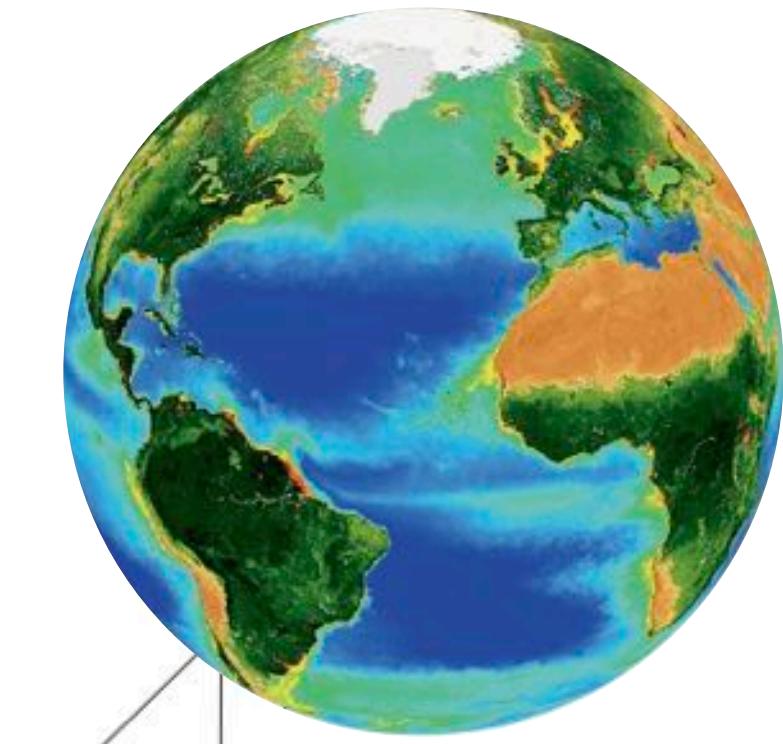
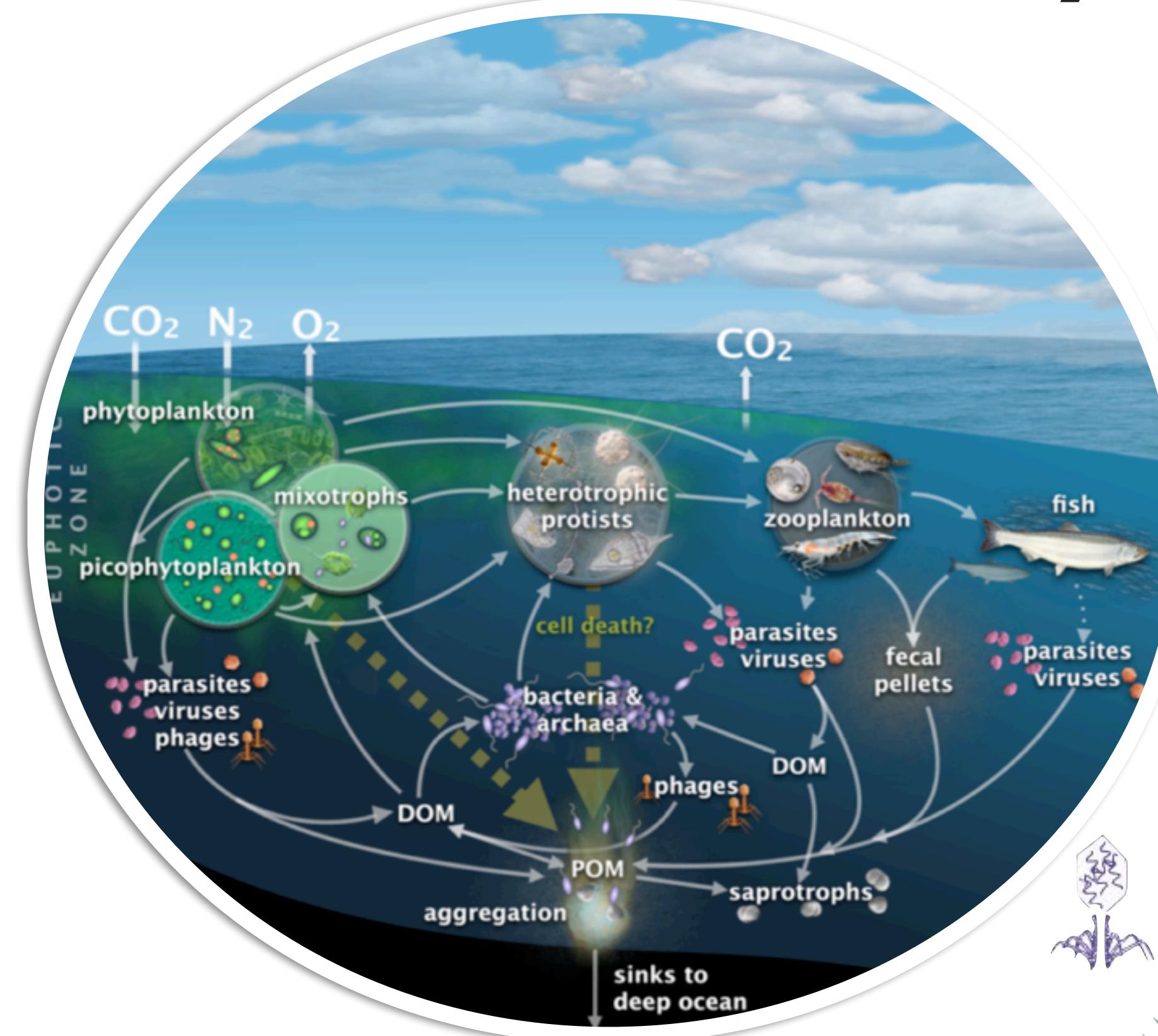
Virulence

Stress

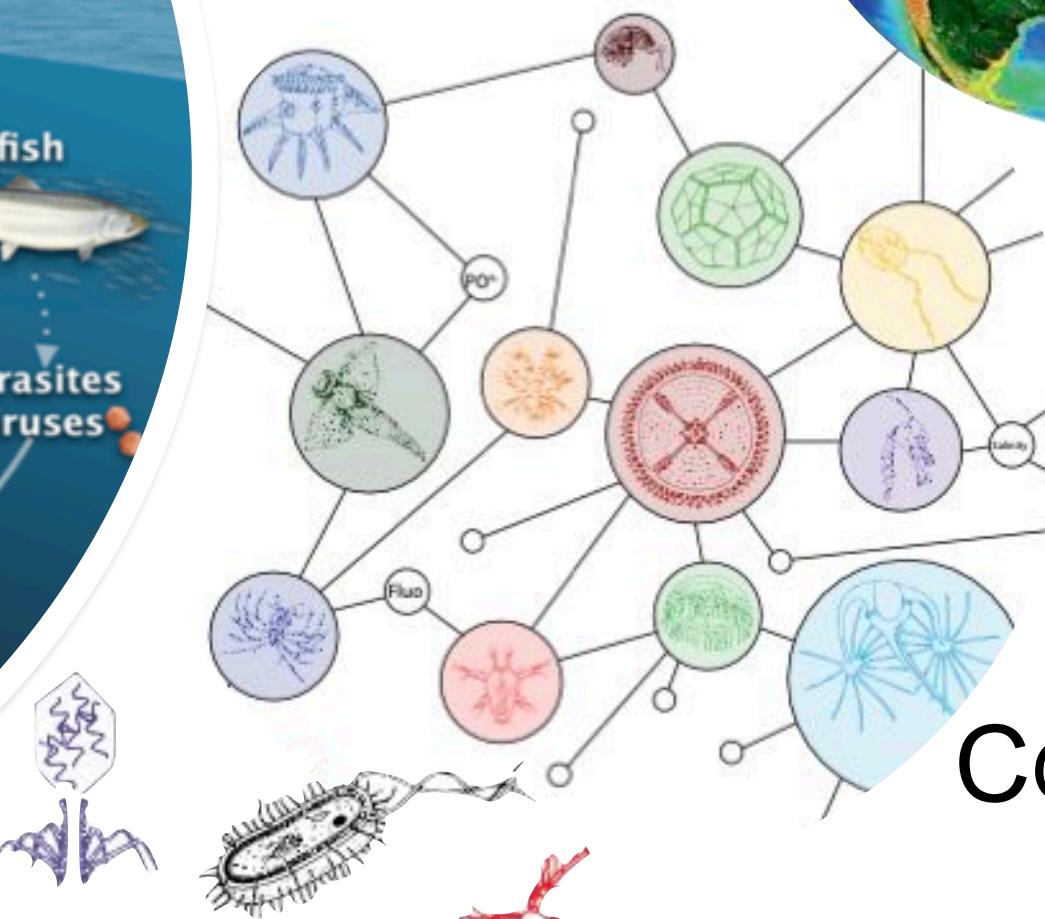
Degradation

C, N, O fixation or release

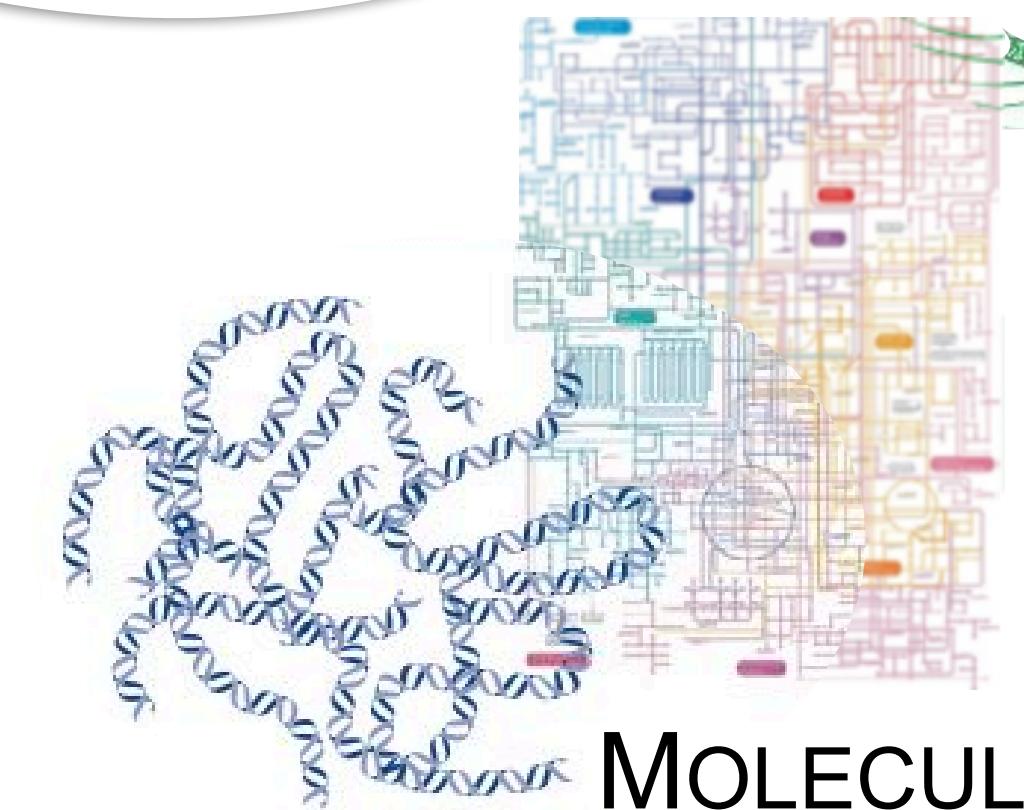
Transporters



ECOSYSTEMS



COMMUNITIES



METABOLISMS

MOLECULES

(Meta)omics and ocean ecosystems biology

Who is there?

What is done?

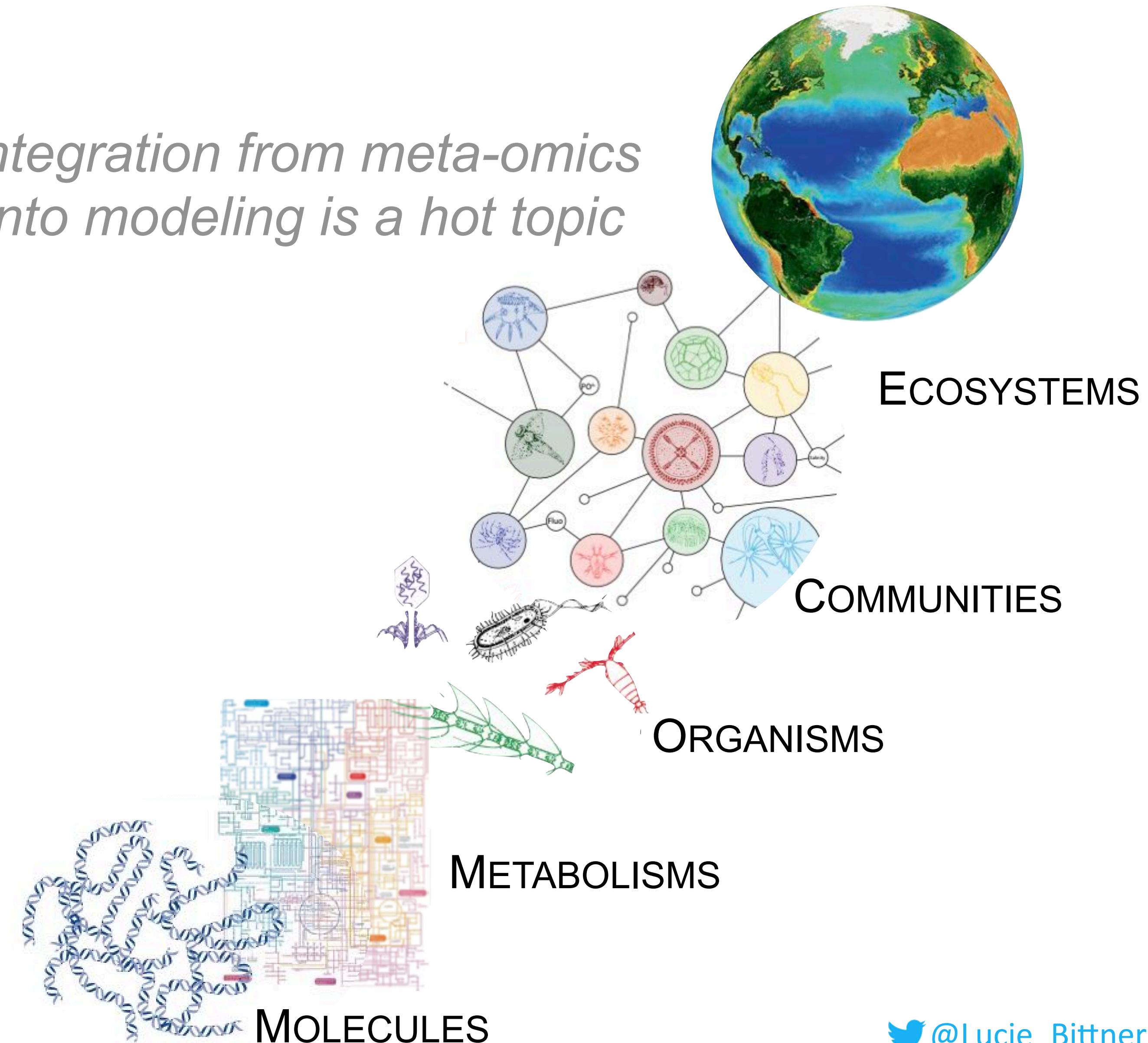
Who is doing what?

In which conditions?

Who is interacting with who?

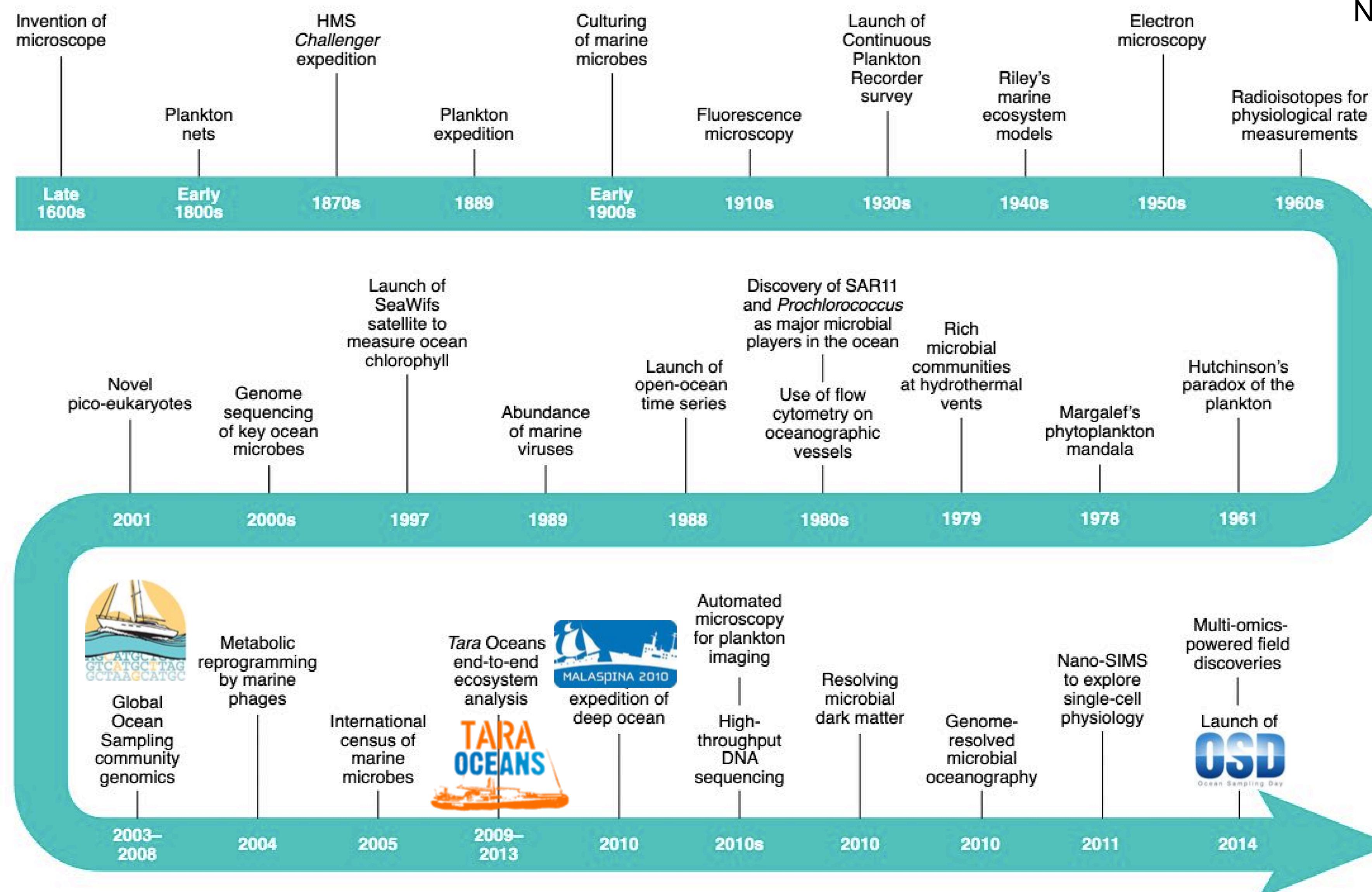
Who is doing what with who?

*Integration from meta-omics
into modeling is a hot topic*



Timeline of ocean microbiome research

Tara Ocean Foundation et al.
Nature Microbiology, 2022



Advantages / drawbacks of (meta)omics

Not yet fully automated

Getting data is relatively :

- fast
- cheap (sampling instrument, sequencing cost)
- many biological scale can be investigated

Which future for meta-omics?

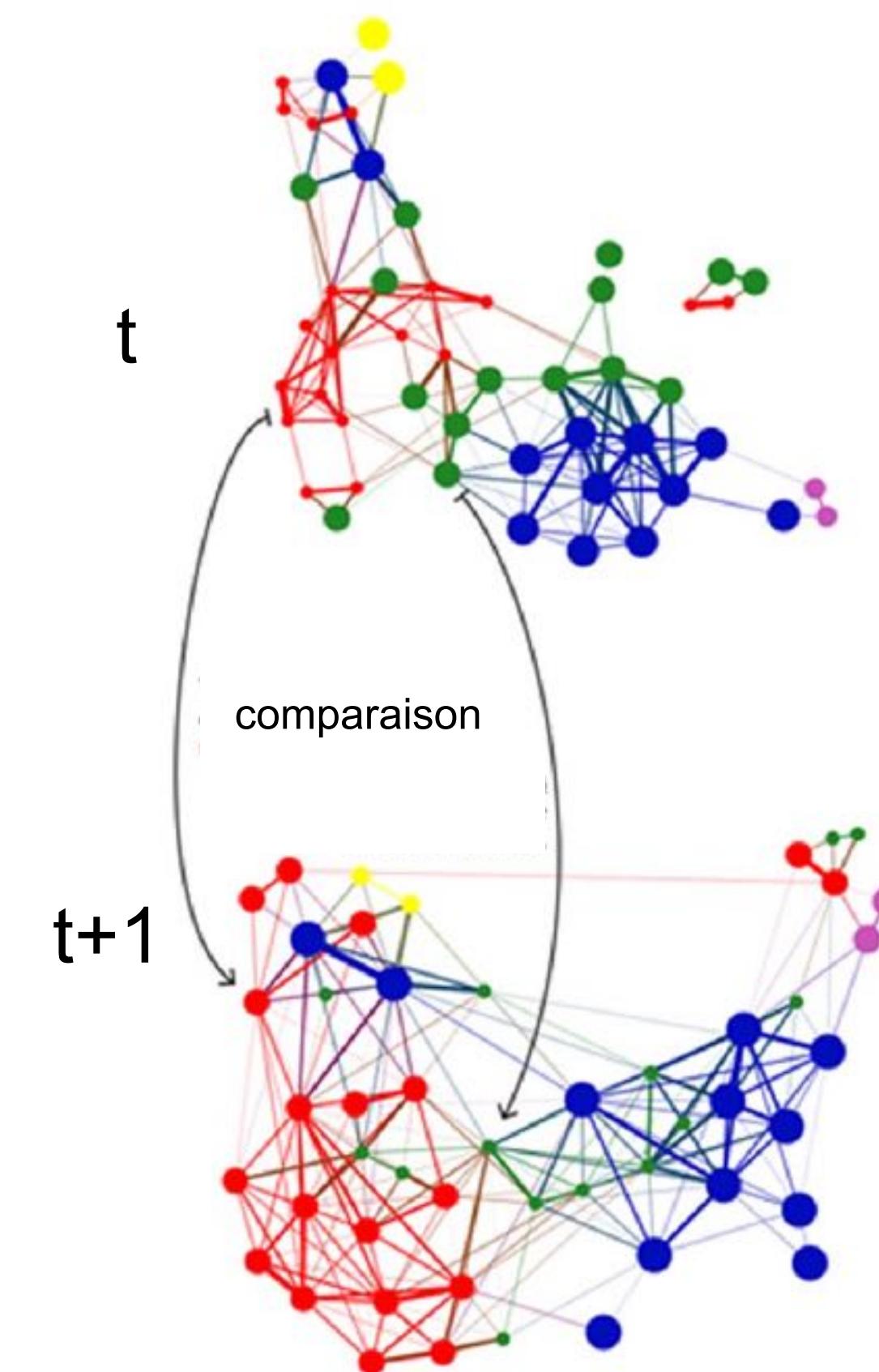
Automated samplers



Global array of samplers and in-cloud network reconstruction



Real-time monitoring of community structure



Next generation biomonitoring
Bohan et al. 2017

Advantages / drawbacks of (meta)omics

Getting data is relatively :

- fast
- cheap (sampling instrument, sequencing cost)
- many biological scale can be investigated

Not yet fully automated

Semi-quantitative data, no link with biomass or rates

Pragmatical units obtained might be far
from the biological concepts

No consensus on sampling protocols or data treatment
(Contrary to consortium like Geoscapes)

Assignment/interpretation is dependent of reference data

Finding statistical signal can be tough !

Why marine (meta)omics + ML?

Increasing number of environmental samples on various scales (space, depth, time)

Find structure in the data
Highlight markers of traits
Make the dark matter talk ...

> opportunity for data-driven analyses:
to cross and link the infos and scales

*Practical lab this afternoon =
habitat modeling for protein / metabolisms*

Datasets from this afternoon

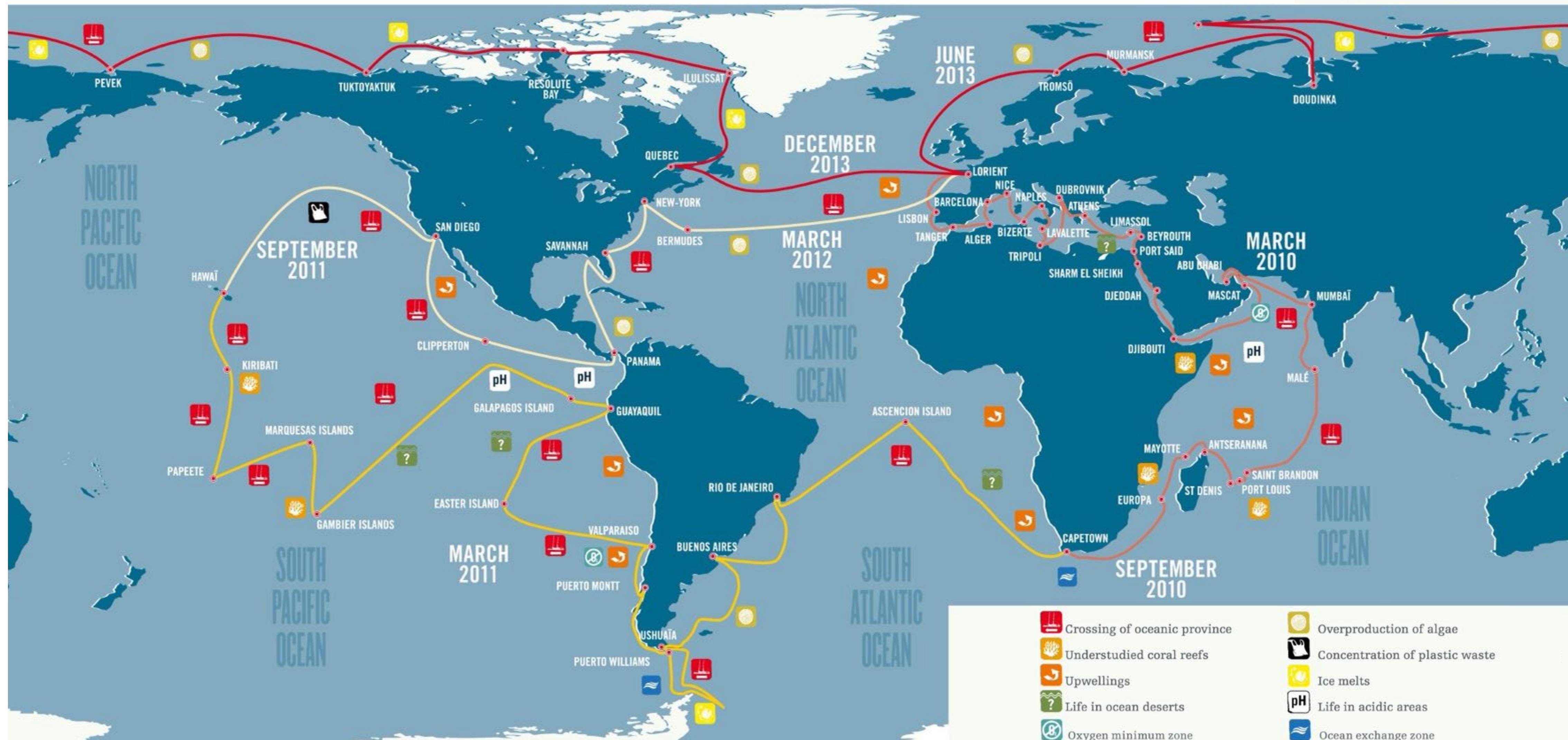


TARA schooner



Expedition 2009-2013: Sailing and Sampling around the world covering 140 000 Km

- **FIRST YEAR:** Lorient - Cape Town
 - **SECOND YEAR:** Cape Town - Honolulu
 - **THIRD YEAR:** Honolulu - Lorient
 - **FOURTH YEAR :** Lorient - Lorient
 - **PORT OF CALL**





Expedition 2009-2013:
Sailing and Sampling around
the world covering 140 000 Km





2009-2013

Sampling strategy (210 stations)



Sampling strategy (210 stations)

OPEN  ACCESS Freely available online

Community Page

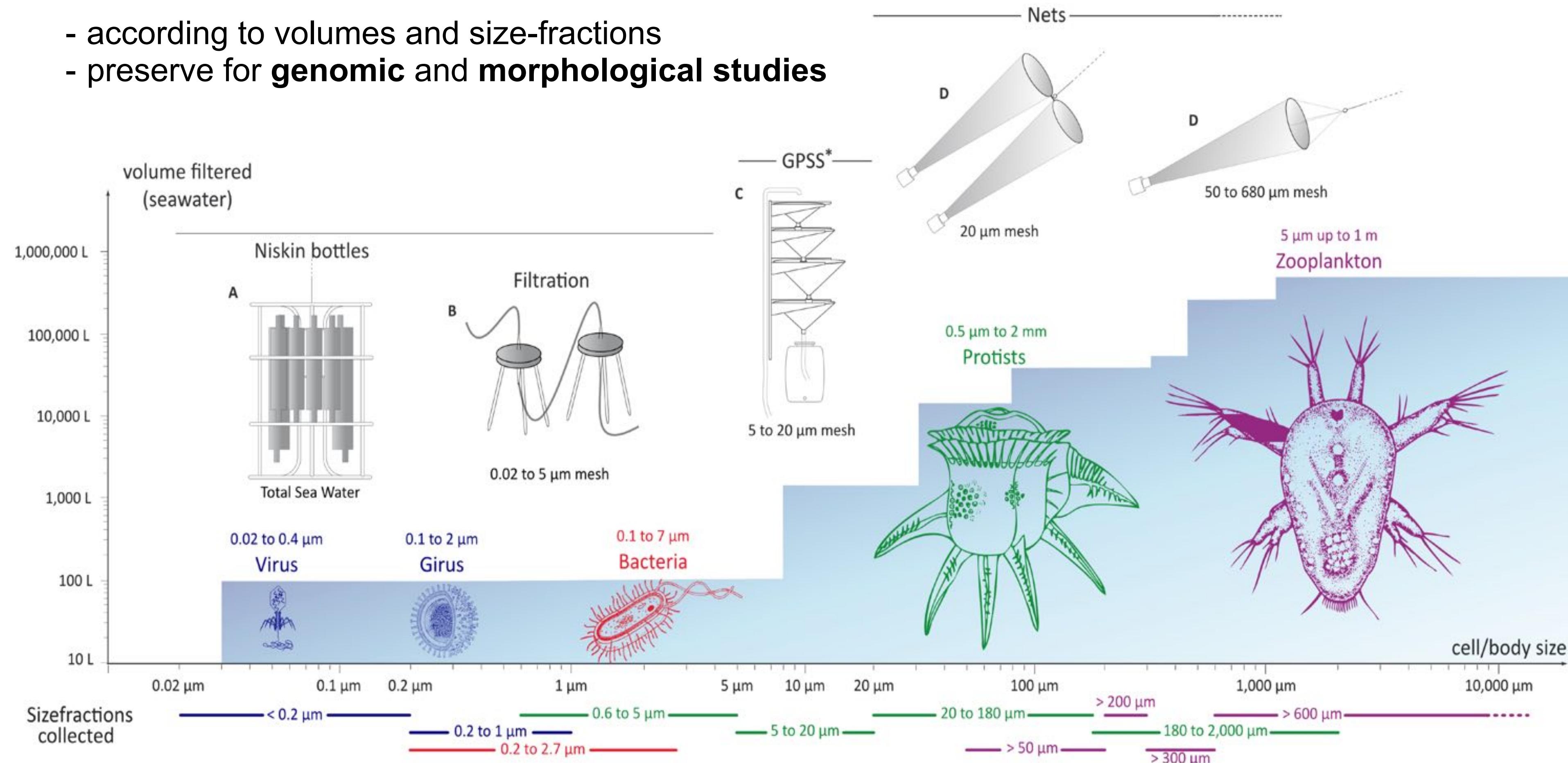


Karsenti et al. (2011)

A Holistic Approach to Marine Eco-Systems Biology

Sampling the organisms and register *in situ* environmental parameters (T°C, salinity, flux, ...)

- according to volumes and size-fractions
- preserve for **genomic and morphological studies**

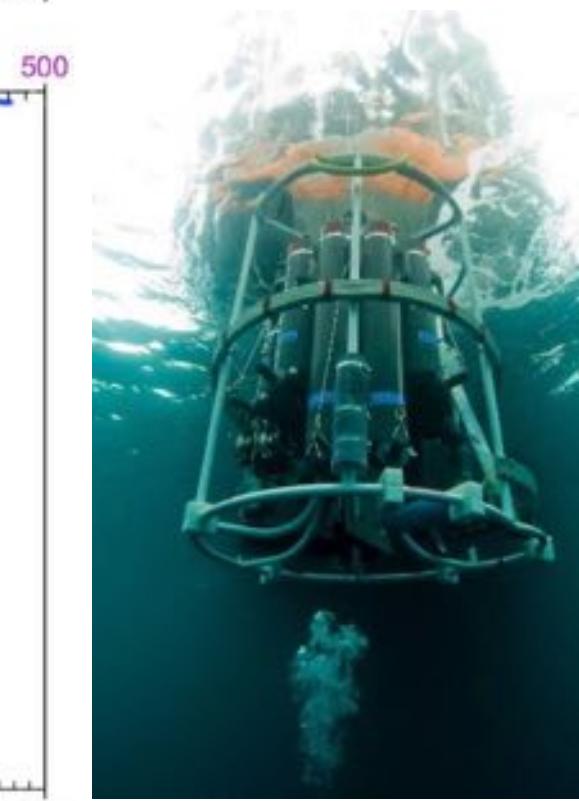
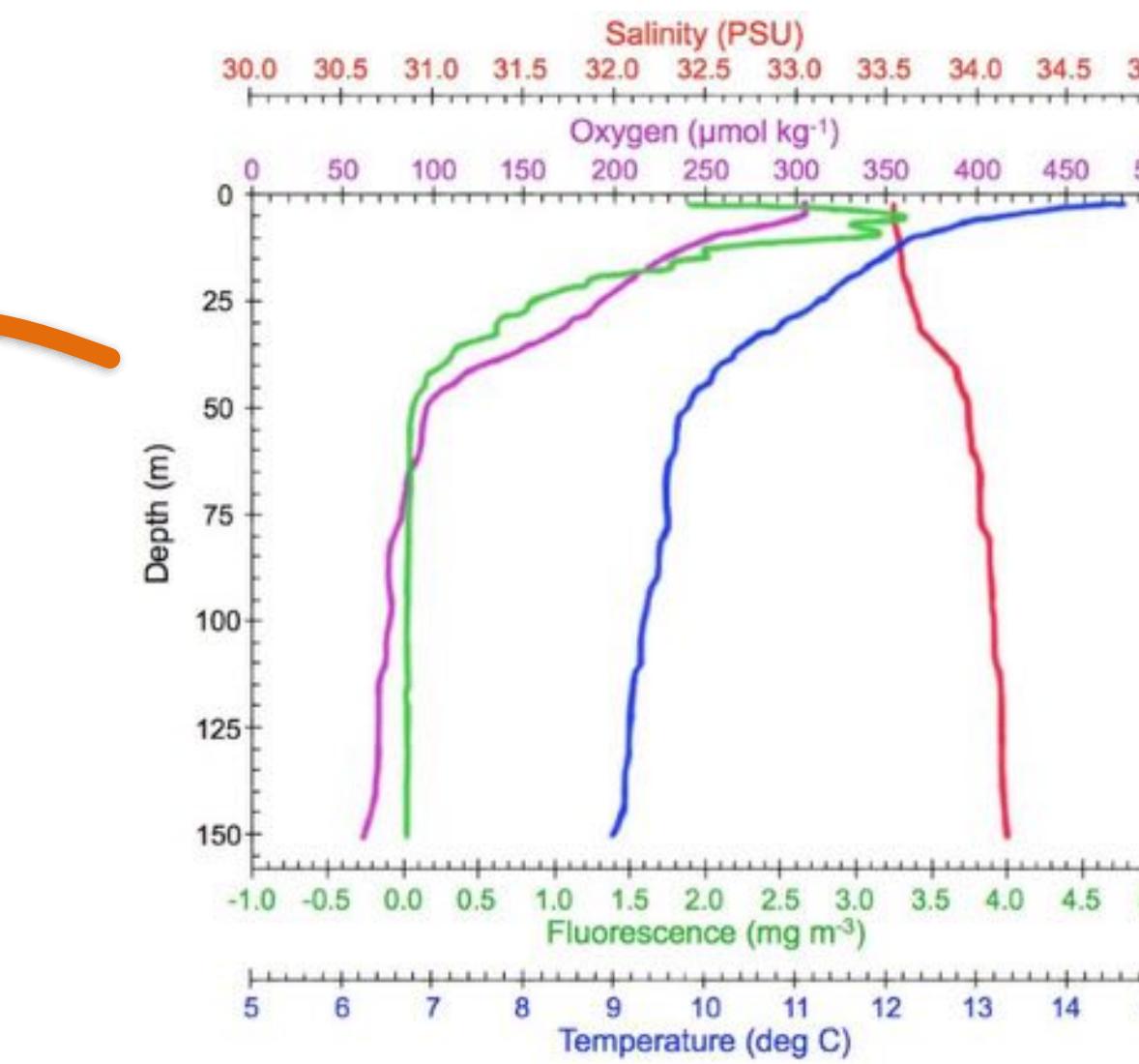
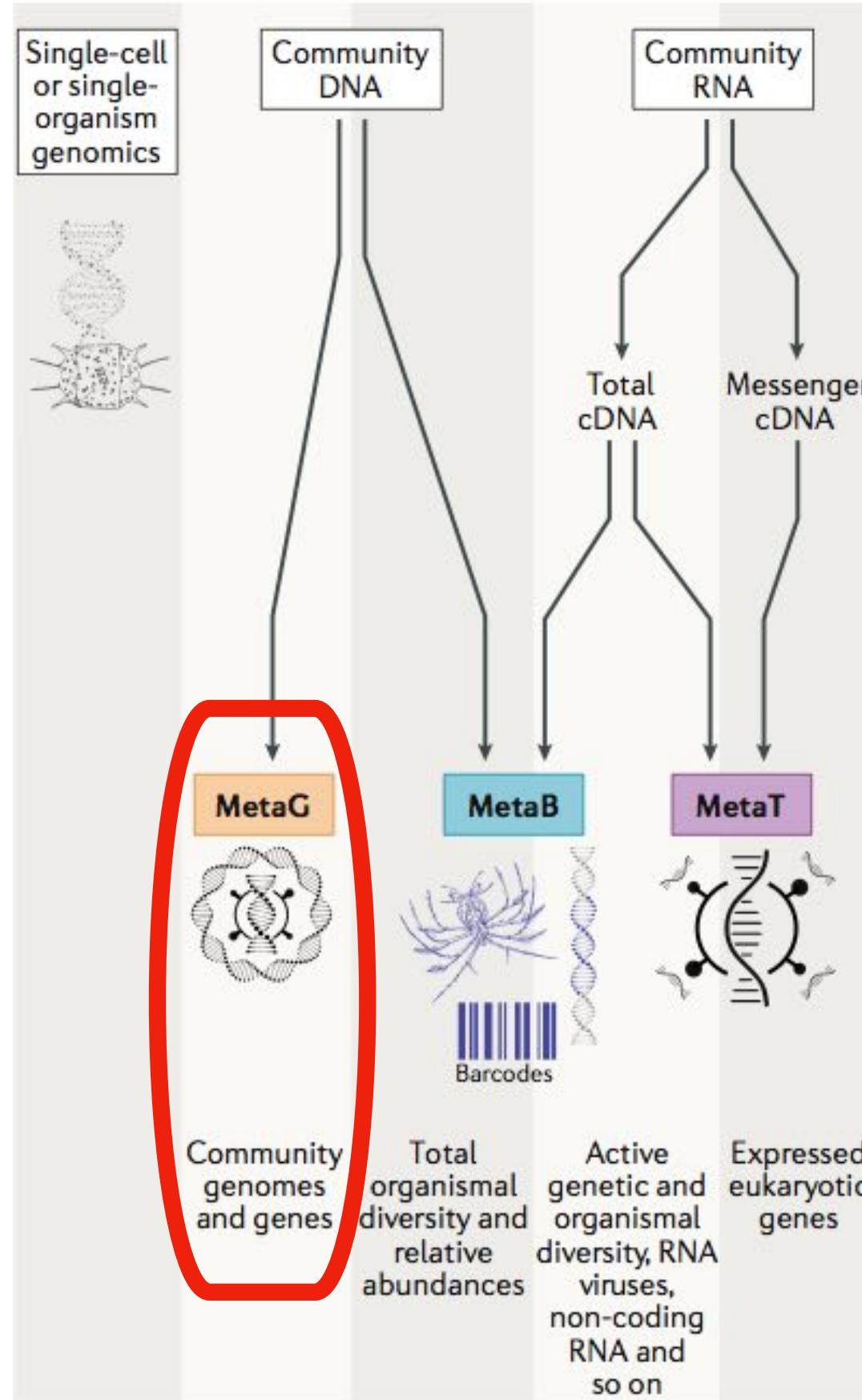


A Holistic Approach to Marine Eco-Systems Biology

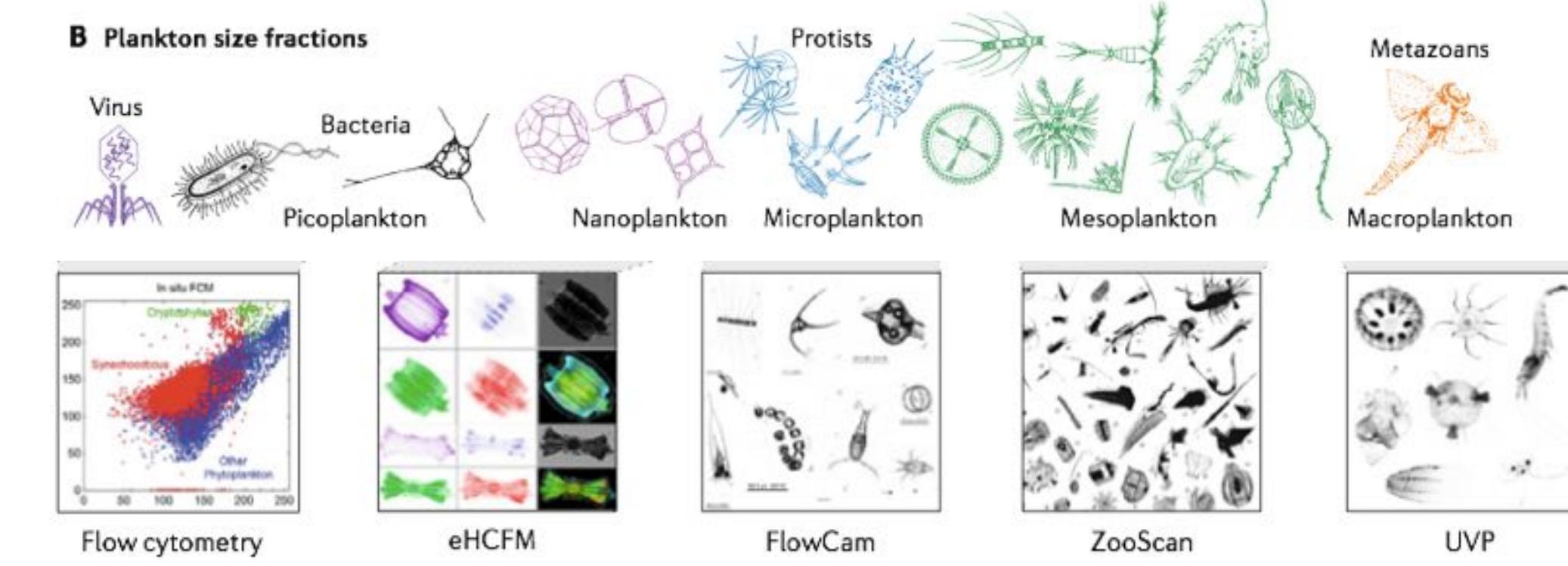
Karsenti et al. (2011)



HIGH THROUGHPUT SEQUENCING



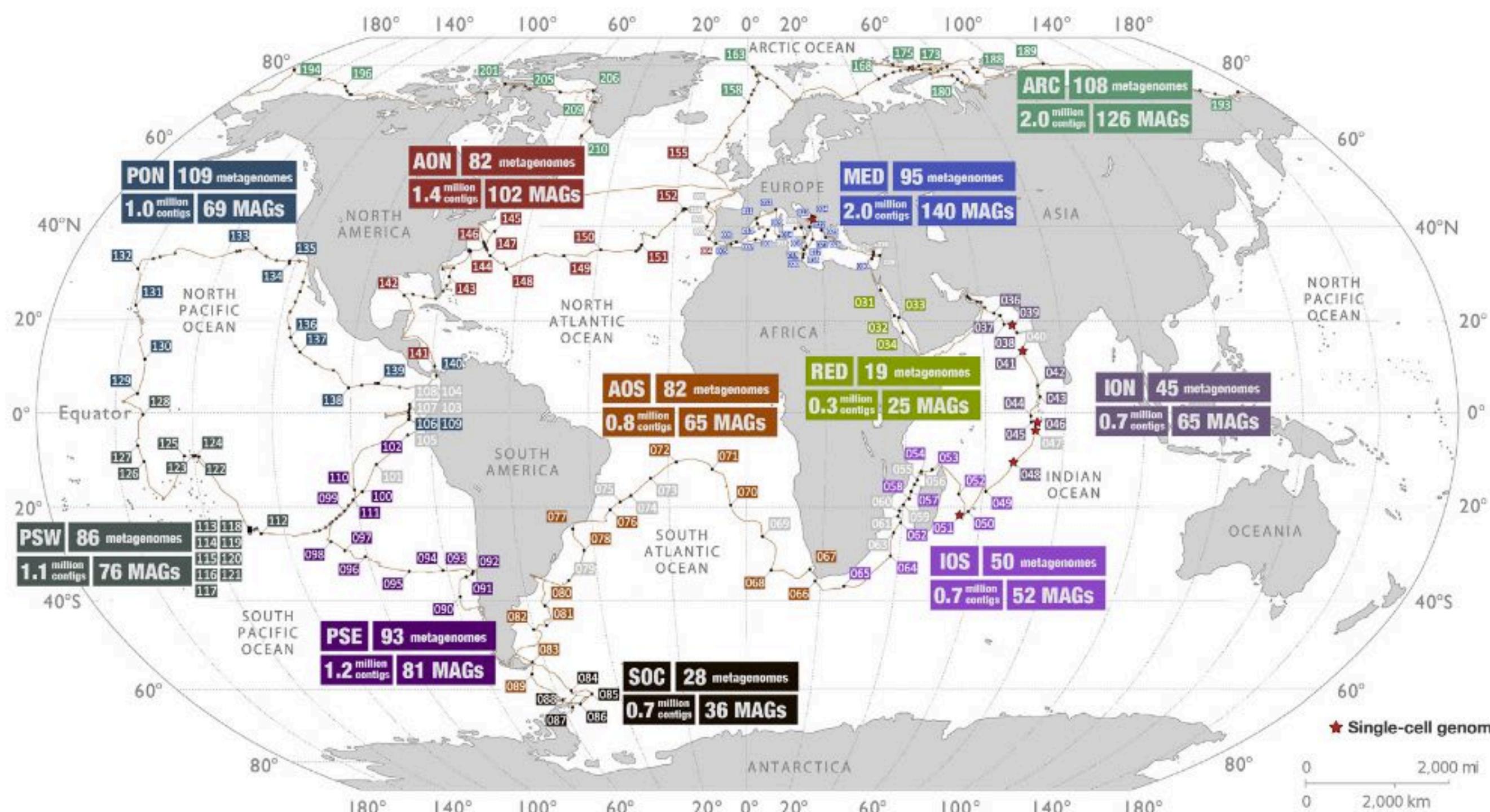
ENVIRONMENTAL PARAMETERS



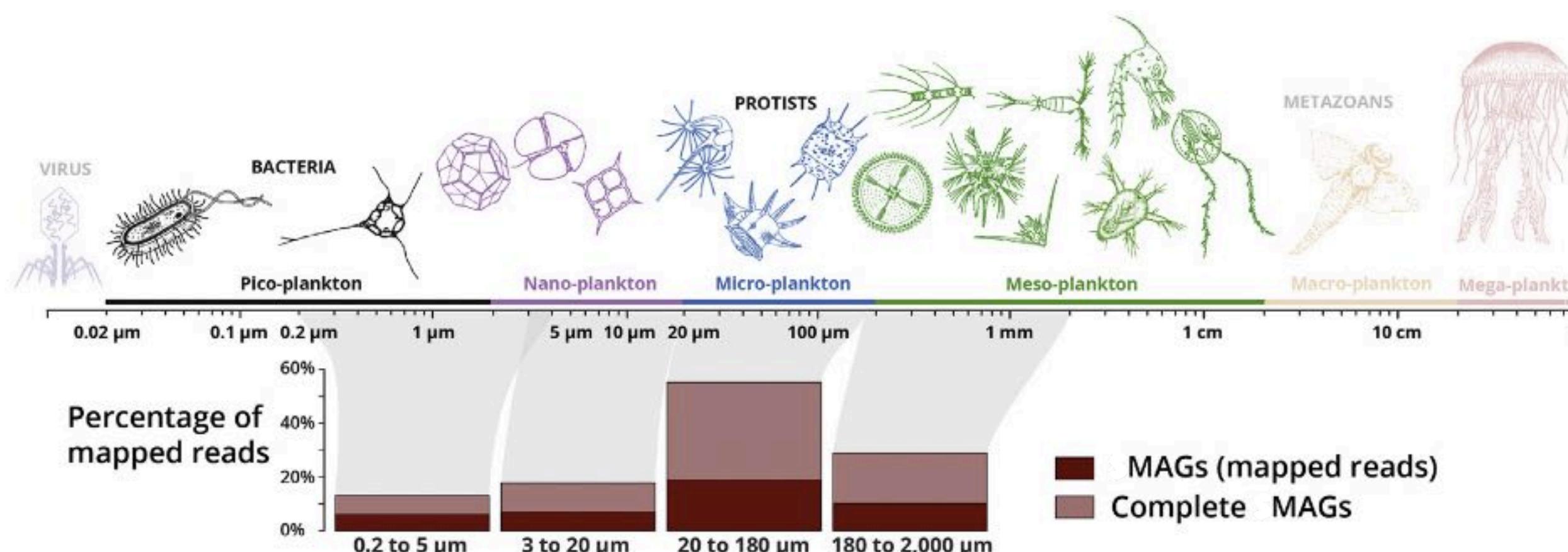
HIGH THROUGHPUT IMAGING

Meta-omics dataset from this afternoon

Delmont et al.
Cell, 2022

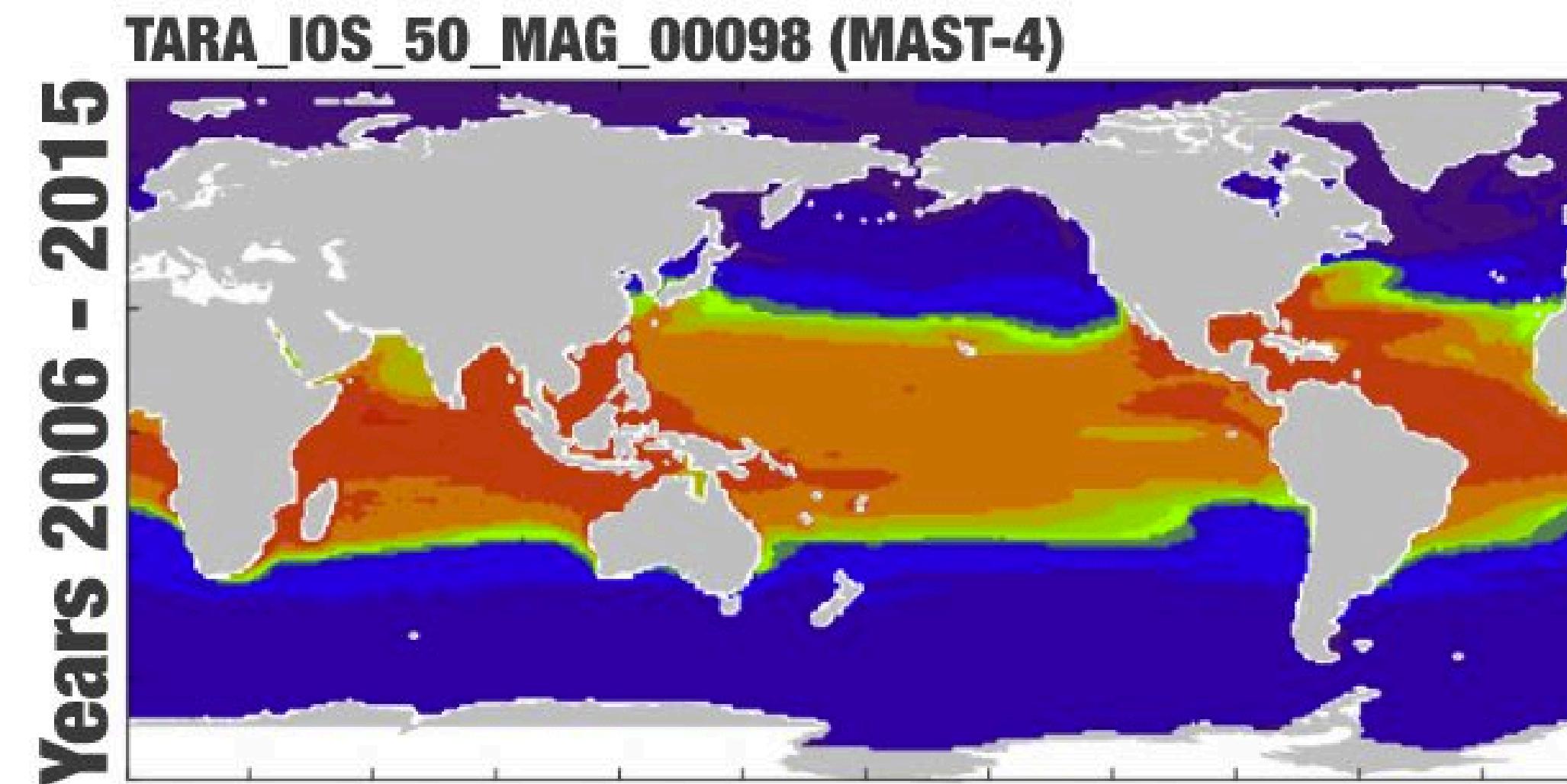


surface and deep chlorophyll maximum layer of 143 stations from the Pacific, Atlantic, Indian, Arctic, Southern Oceans Mediterranean and Red Seas



Meta-omics dataset from this afternoon

Delmont et al.
Cell, 2022

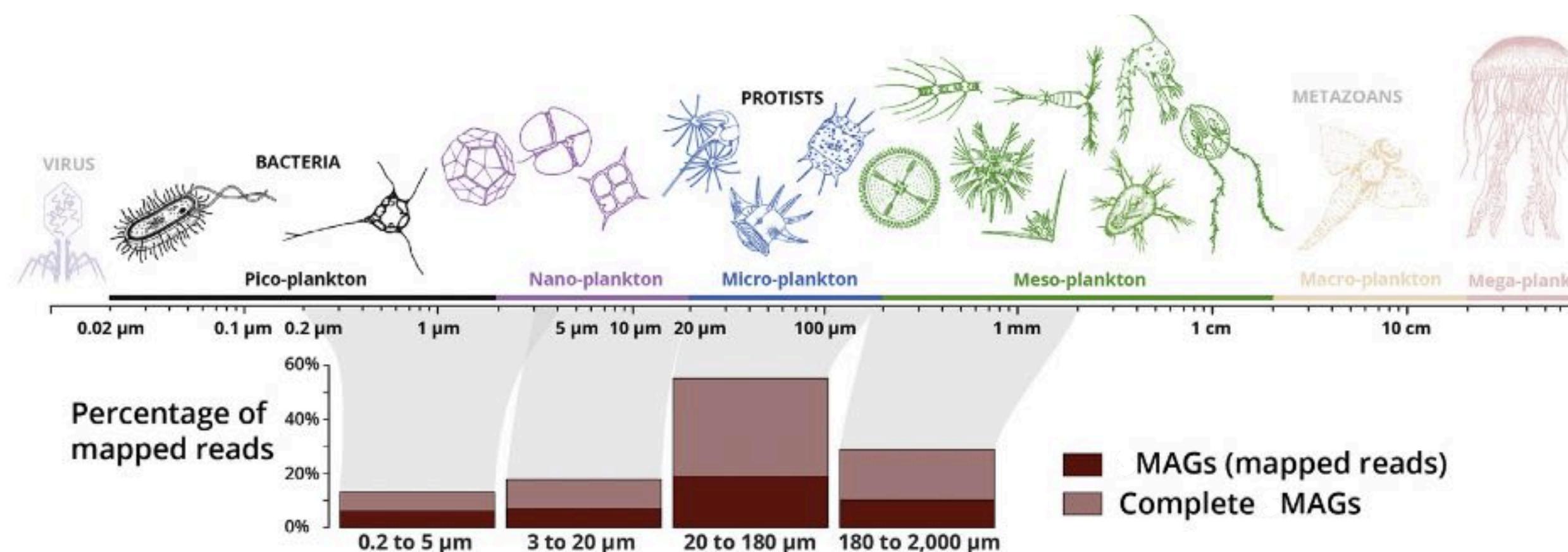


Habitat projection for MAGs

probability of presence
ranges from 0 (purple)
to 1 (red), with green
corresponding to 0.5

280 billion reads from 798 samples
were co-assembled

surface and deep chlorophyll maximum
layer of 143 stations from the Pacific,
Atlantic, Indian, Arctic, Southern Oceans
Mediterranean and Red Seas



10,207,435 eukaryotic proteins
involved in 700 « environmental genomes » (MAGs)

Meta-omics dataset from this afternoon

Delmont et al.
Cell, 2022

HABITAT PROJECTION FOR PROTEINS AND METABOLISMS

Datasets prepared by



Pavla
Debeljak

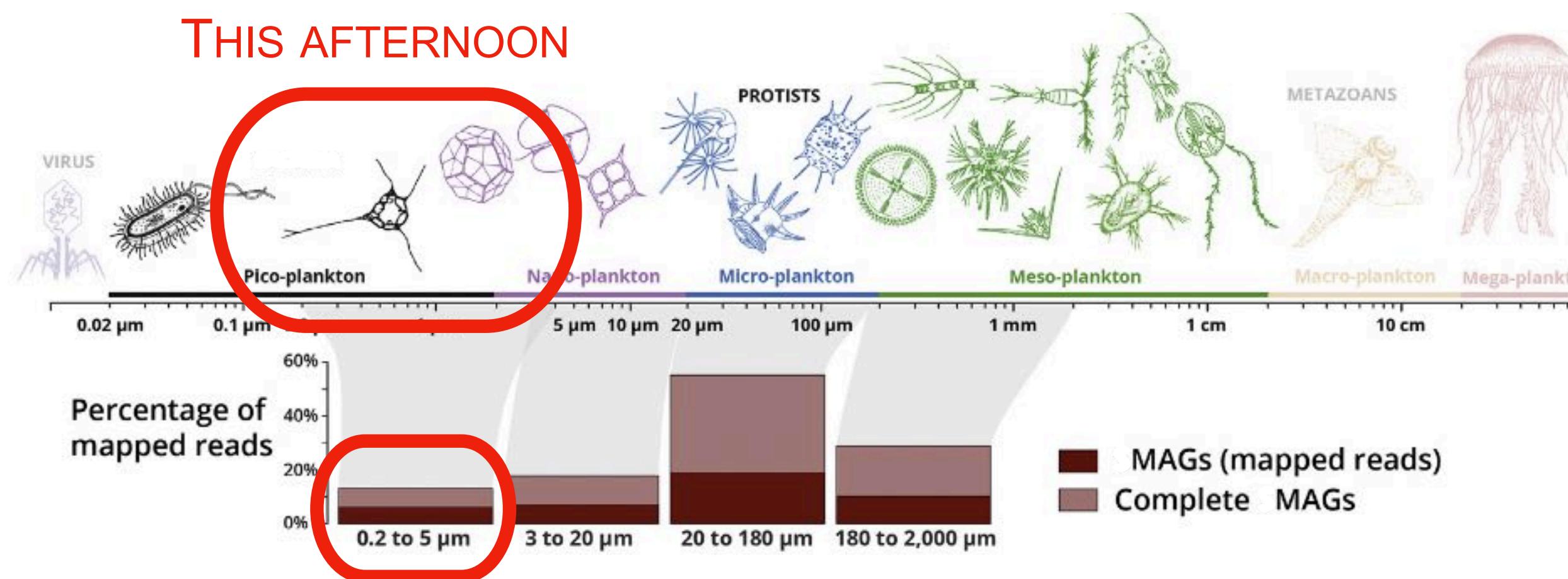


Alexandre
Schickele



280 billion reads from 798 samples
were co-assembled

surface and deep chlorophyll maximum
layer of 143 stations from the Pacific,
Atlantic, Indian, Arctic, Southern Oceans
Mediterranean and Red Seas



10,207,435 eukaryotic proteins
involved in 700 « environmental genomes » (MAGs)