Climbing the omics complexity ladder

-

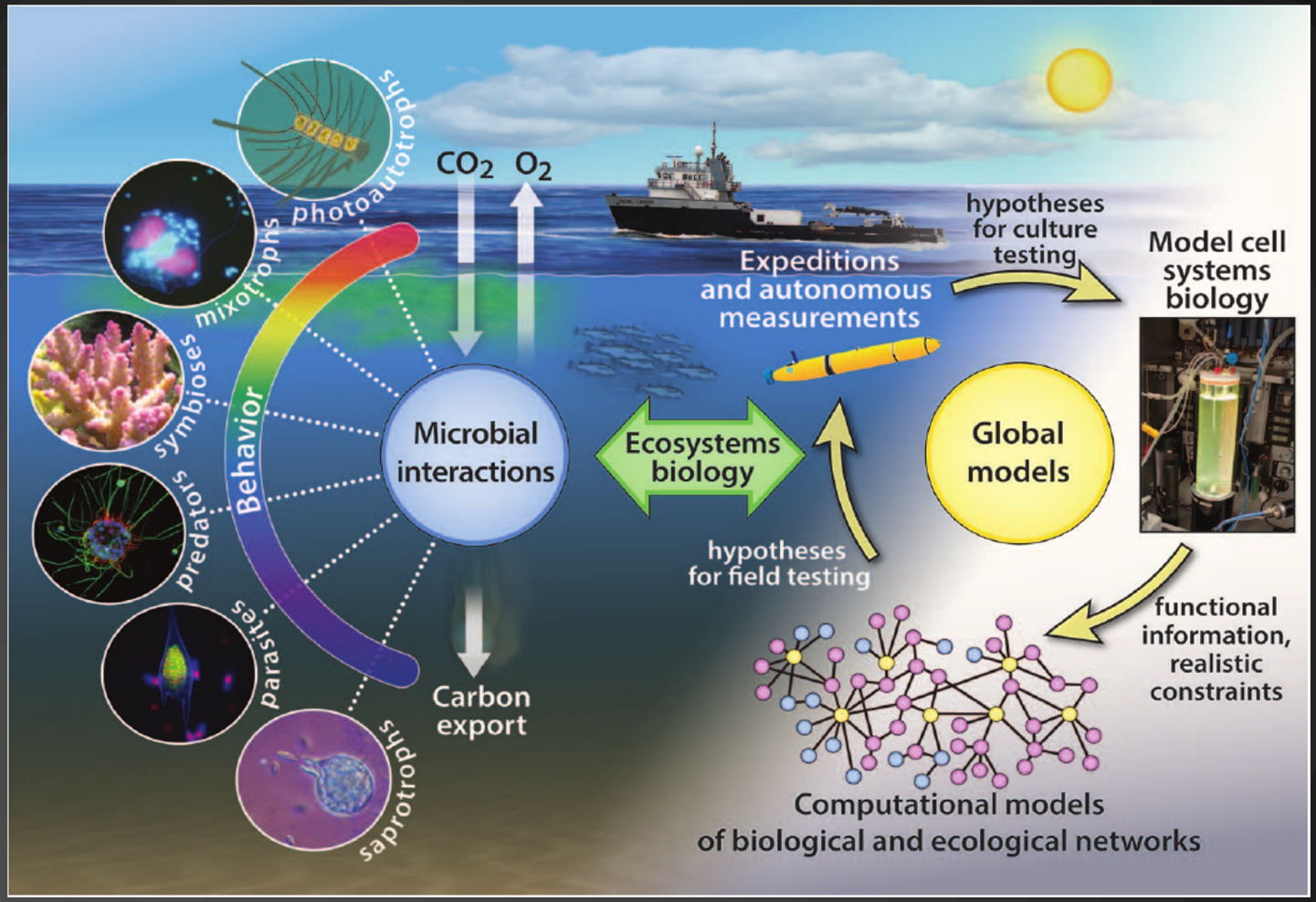
Functional marine protists ecogenomics with *Tara* Ocean

- or -

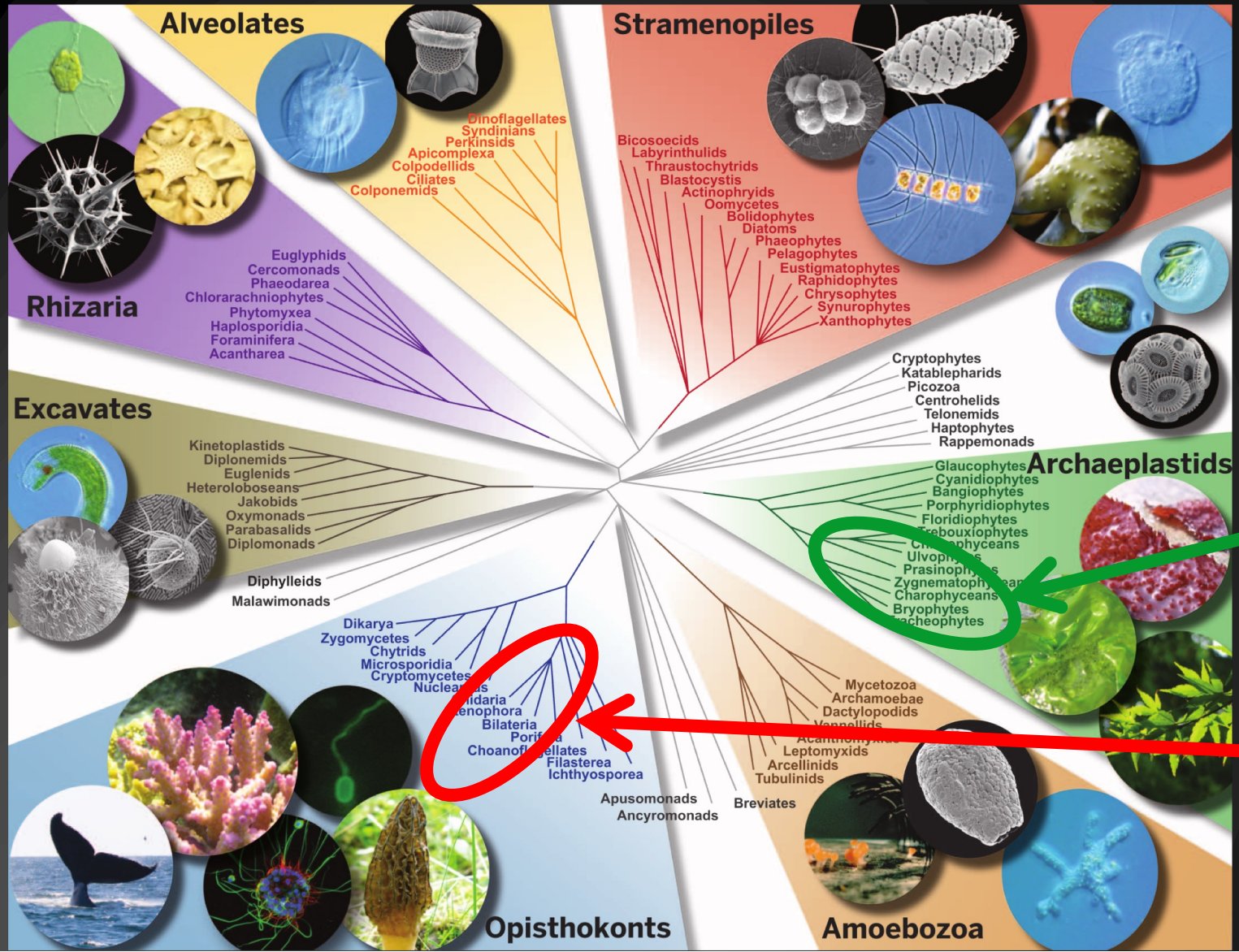
When the sweet turns to sour

Eric Pelletier – CEA / Genoscope – CNRS – Université Paris Saclay

A highly complex ecosystem driving the most



Worden *et al.*, Sci



A huge

Terrestrial plants

Animals

Warden et al. (20

1. From sequence reads

2. To Genes

3. To Organisms

4. To Communities

5. Biogeochemistry

1- From sequence reads

Global omics strategy

MetaG readsets (~30 Gb)

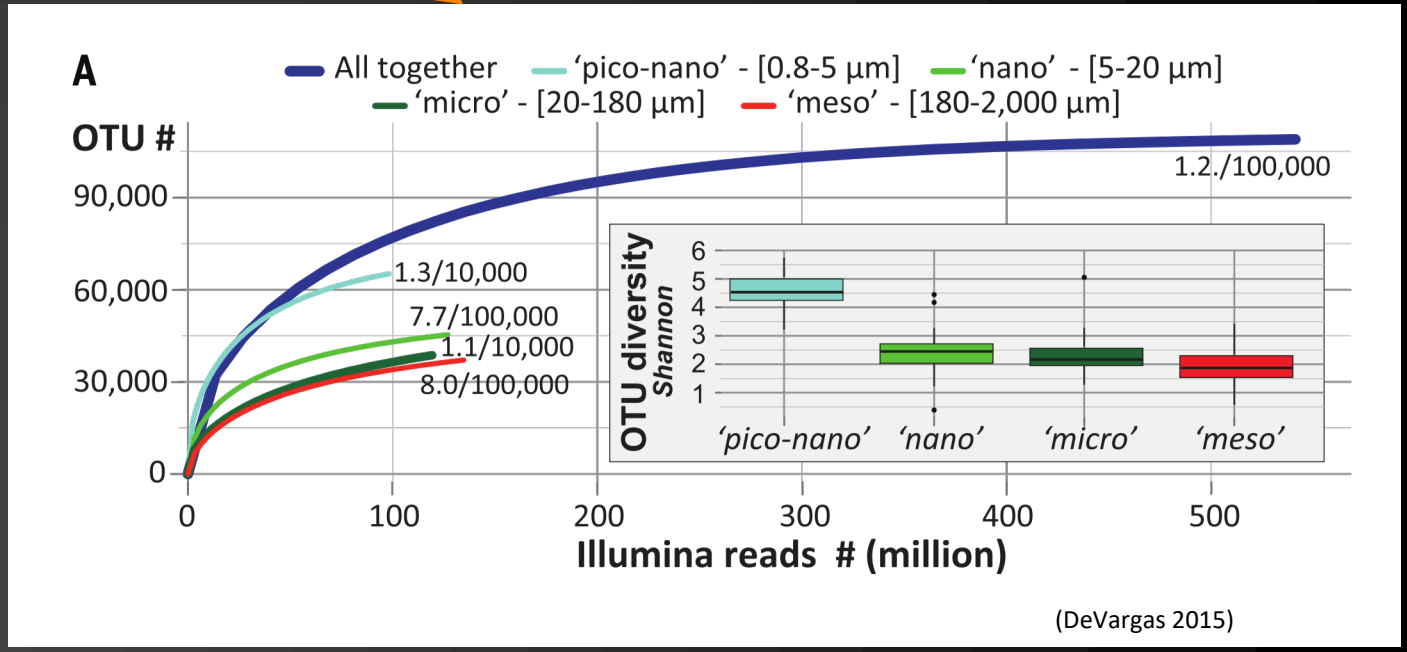
MetaT readsets (~30 Gb)

Global omics strategy

MetaG readsets (~30 Gb)

MetaT readsets (~30 Gb)

~ 120,000

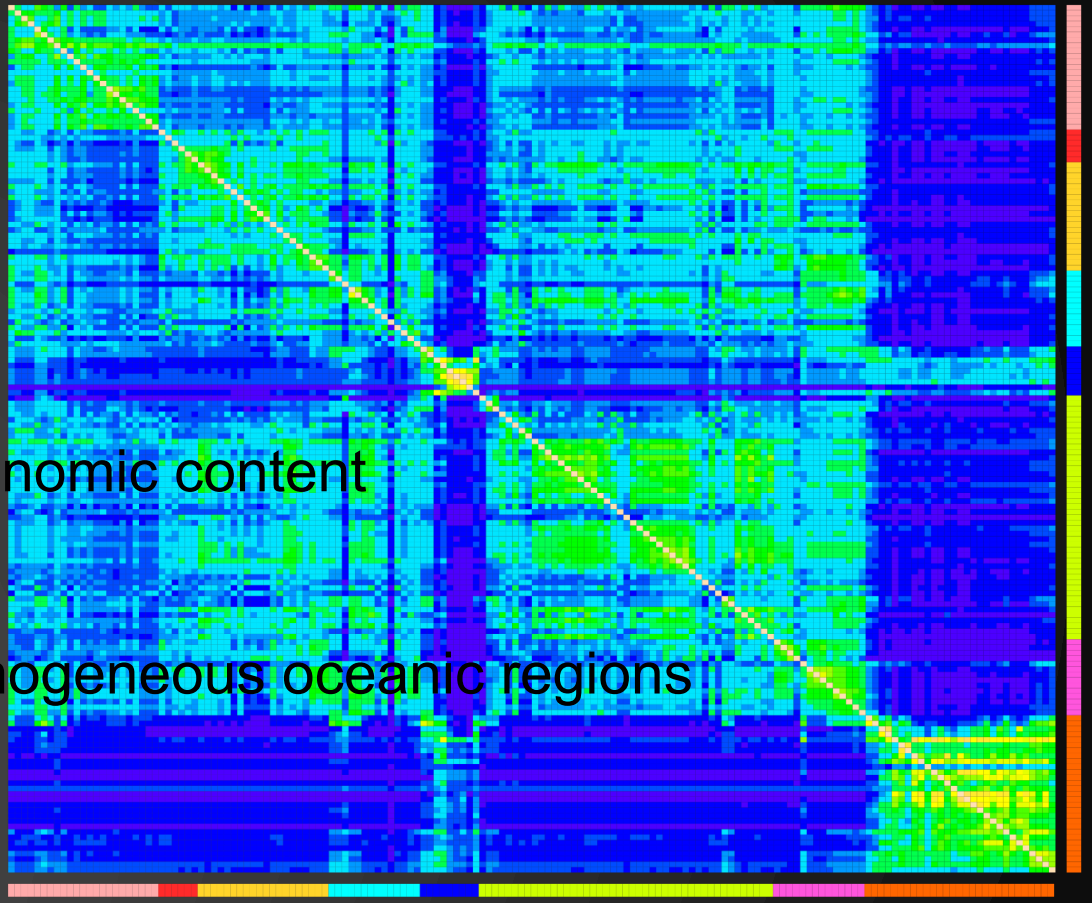


(DeVargas 2015)

Global metagenomic comparisons (ex : 0.8 – 5 µ

Direct comparison of raw genomic content

identification of globally homogeneous oceanic regions



2- To genes

Challenges

Eukaryotes specificities :

- Can have (very) large genomes (up to Gb sizes)
- Potentially lot of non-coding DNA
- Genes are fragmented (introns)
- Genes are (very) difficult to identify de novo

To access the gene content, you may :

- Go through edious analysis
- or
- Use meta-transcriptomics (cDNA focused)

Global omics strategy

MetaG readsets

MetaT readsets

Assembly

Global omics strategy

MetaG readsets

MetaT readsets

Reference
Gene
Catalog

Assembly

Clustering

The Marine Atlas of Tara Oceans Unigenes

	MATOU-v1	MATOU-v2
Stations	**68**	**89**
Filters	**441**	**569**
Unigenes	**116 M**	**154 M**

Corinne da Silva
Quentin Carradec

Global omics strategy

MetaG readsets

MetaT readsets

Reference
Gene
Catalog

Functio
&
Taxonom

Reference proteins

Protein domains

Assembly

Clustering

Global omics strategy

MetaG readsets

MetaT readsets

Reference
Gene
Catalog

Functio
&
Taxonom

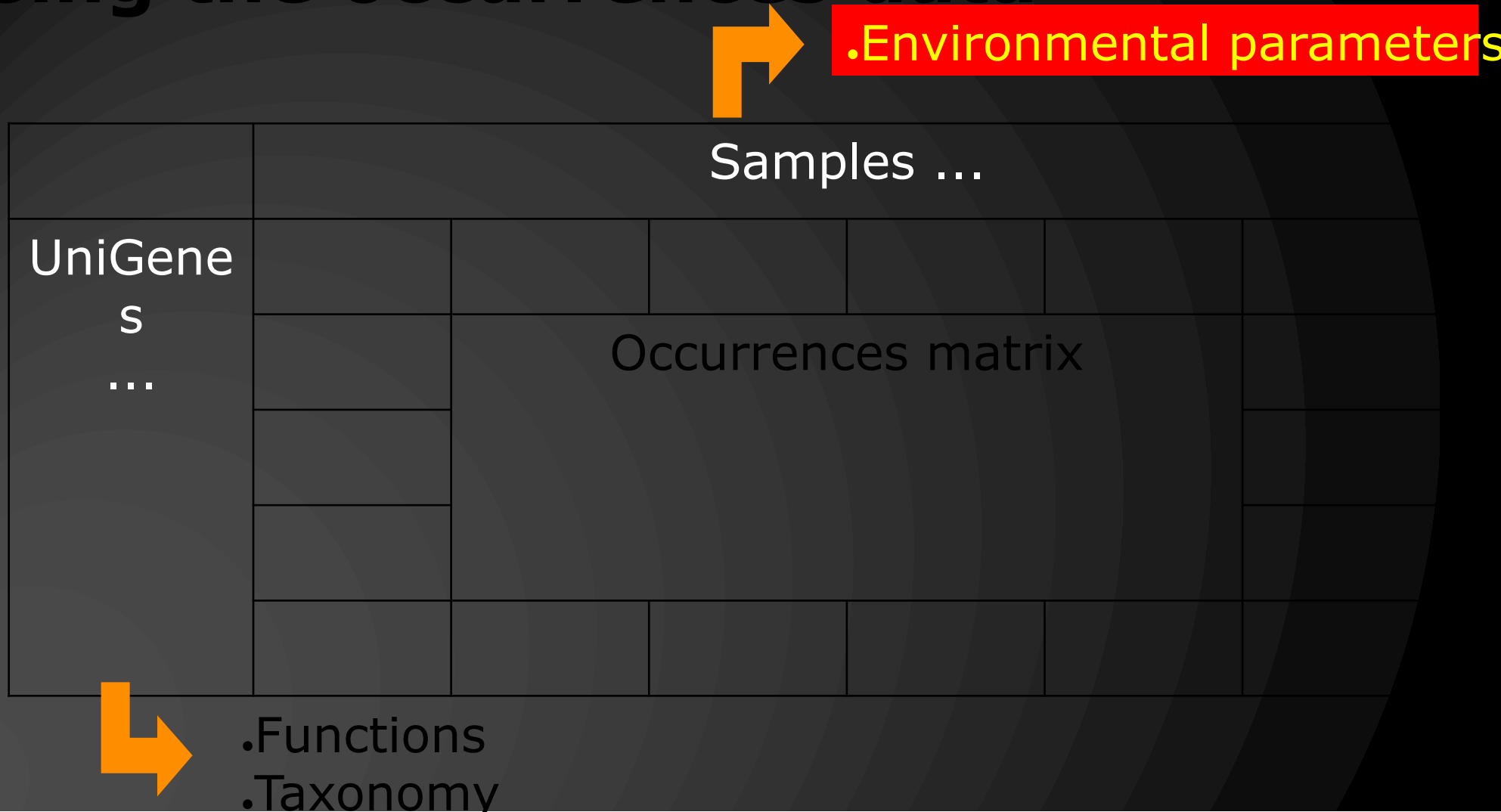
Reference proteins

Protein domains

Assembly

Clustering

Using the occurrences data



Using the occurrences data

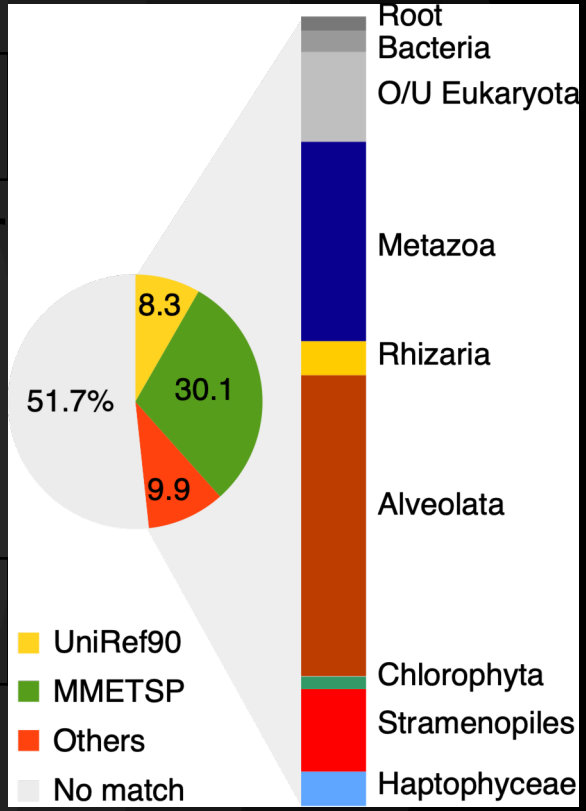
.Environmental parameters

Samples ...

UniGenes ...

Occurrences matr

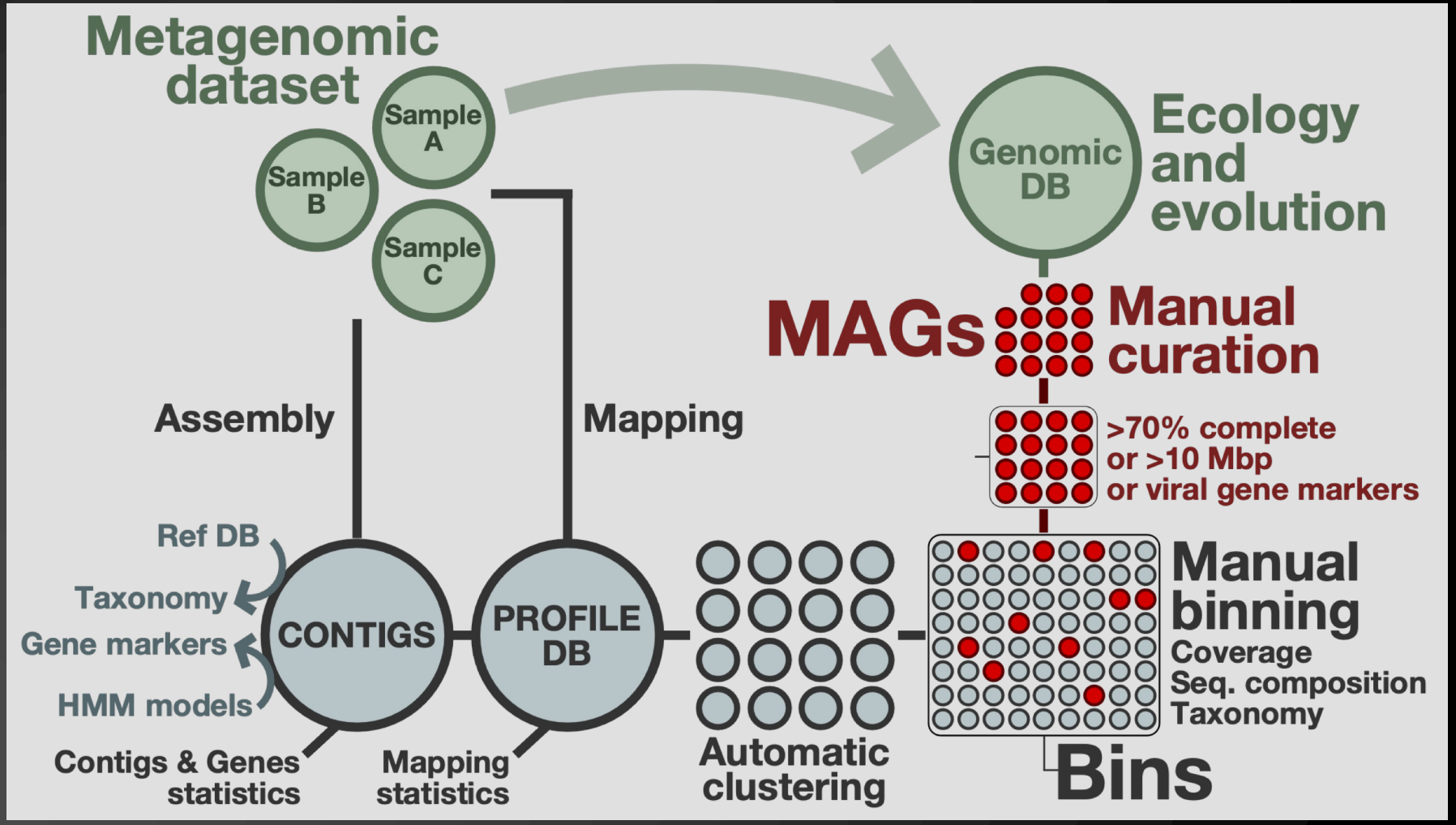
.Functions
.Taxonomy



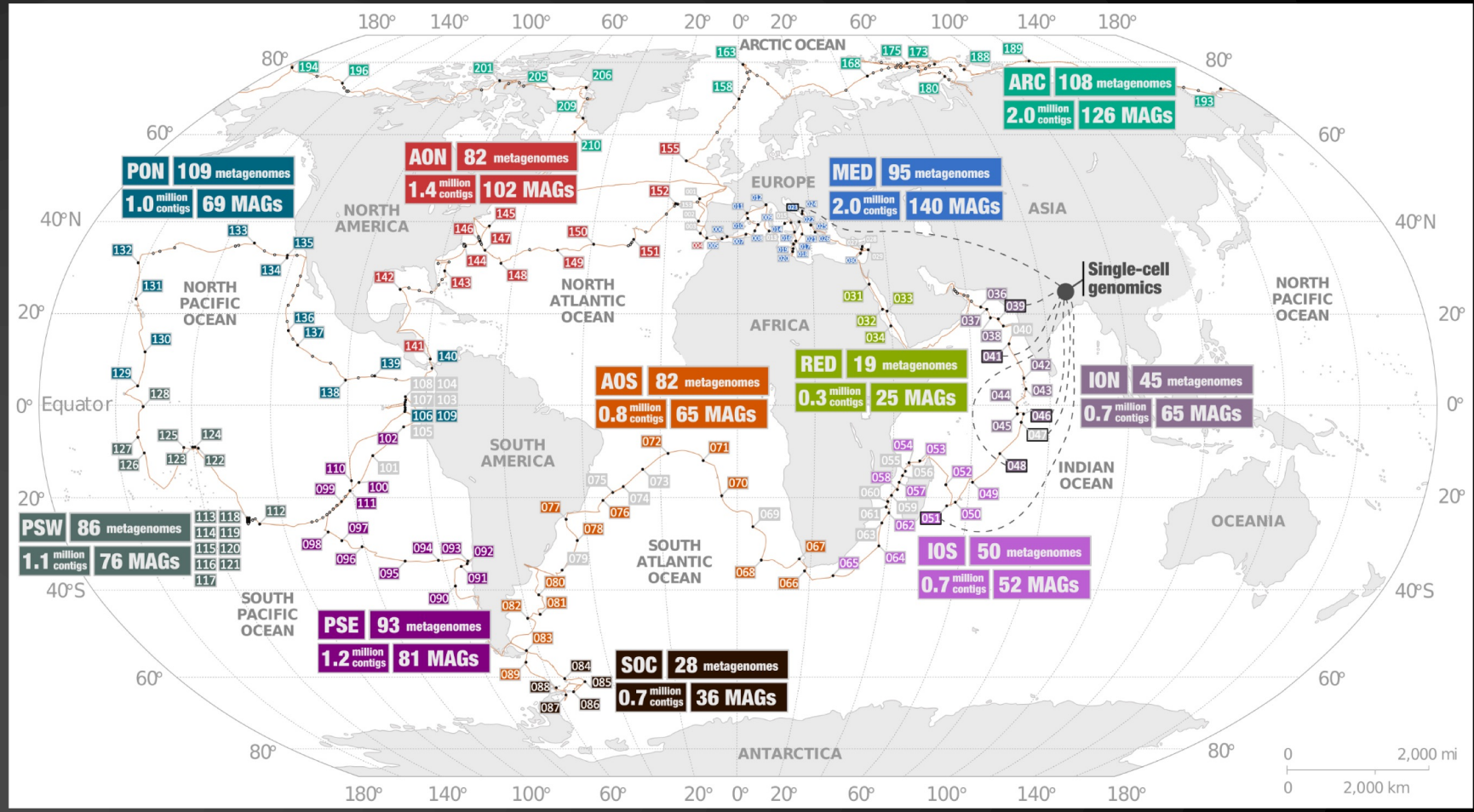
3- To organisms

MAGs

Metagenome-based Assembled Genomes (MAGs)



Metagenome Assembled Genomes (MAGs)



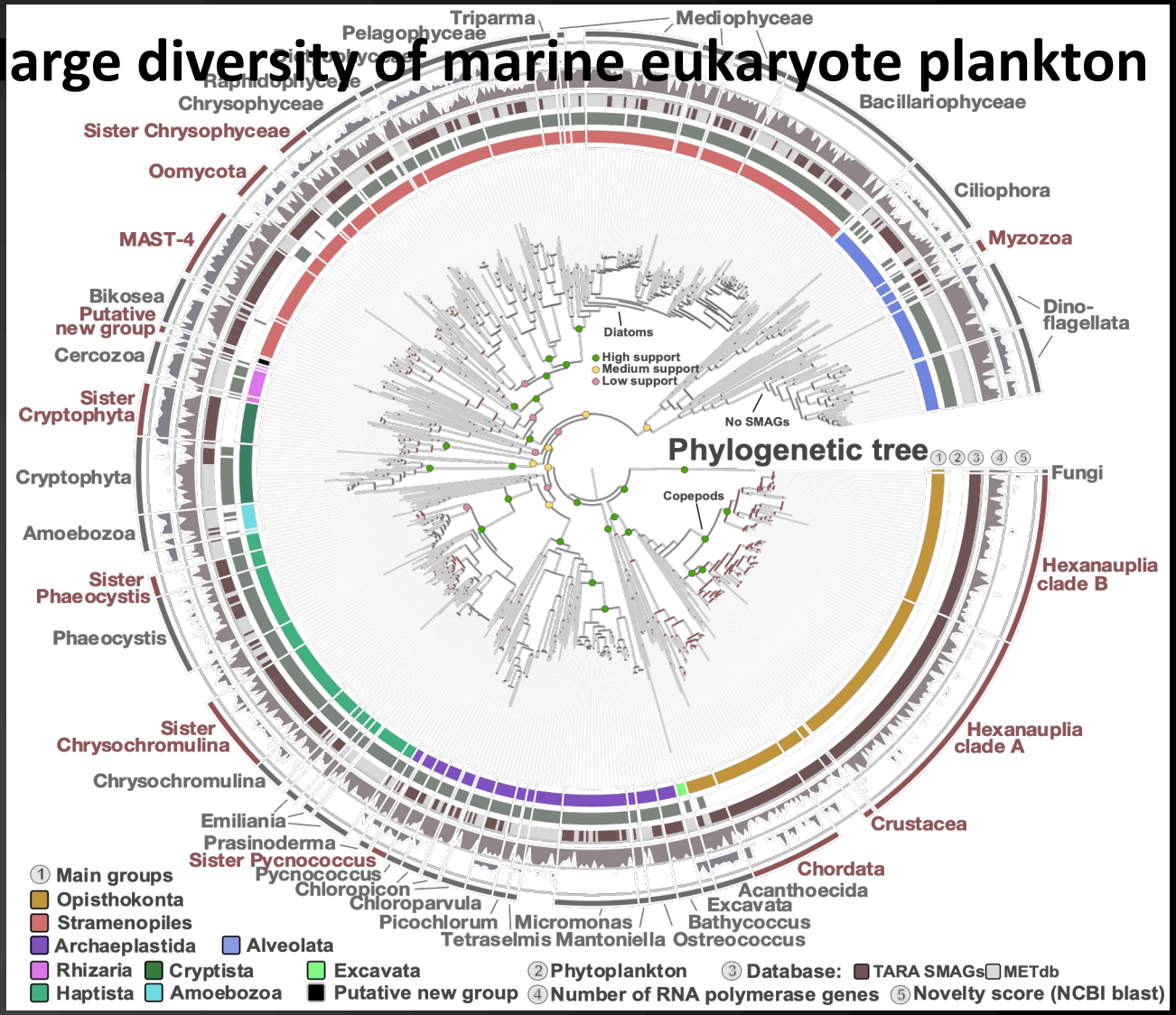
713 eukaryote SAGs and MAGs from *Tara* Oceans samples

MAGs genomes capture a large diversity of marine eukaryote plankton

~ 700 MAGs

In brown :

→ new taxonomic groups



MGTs

Canopy clustering

Alexey Vorobev

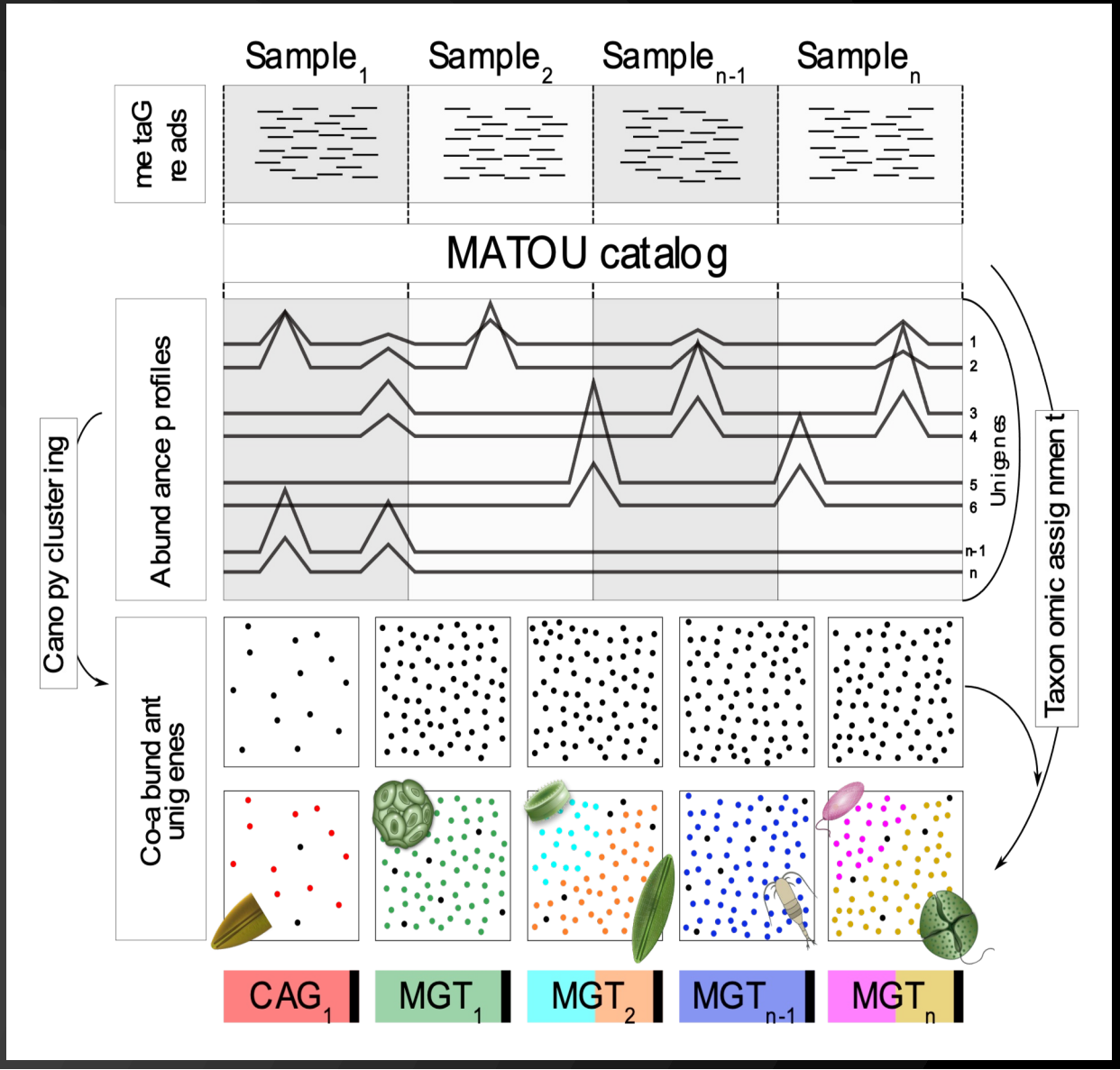
Sequencing &
Genes assembly

Non redundant genes catalog

Abundance profiles

Canopy clustering
by co-abundances

+ Marion Dupouy
and Anita Annamalé



Global taxonomy diversity of MGT

~ 900 MGTs

924 MGTs generated from the MATOU-v1 gene catalogue

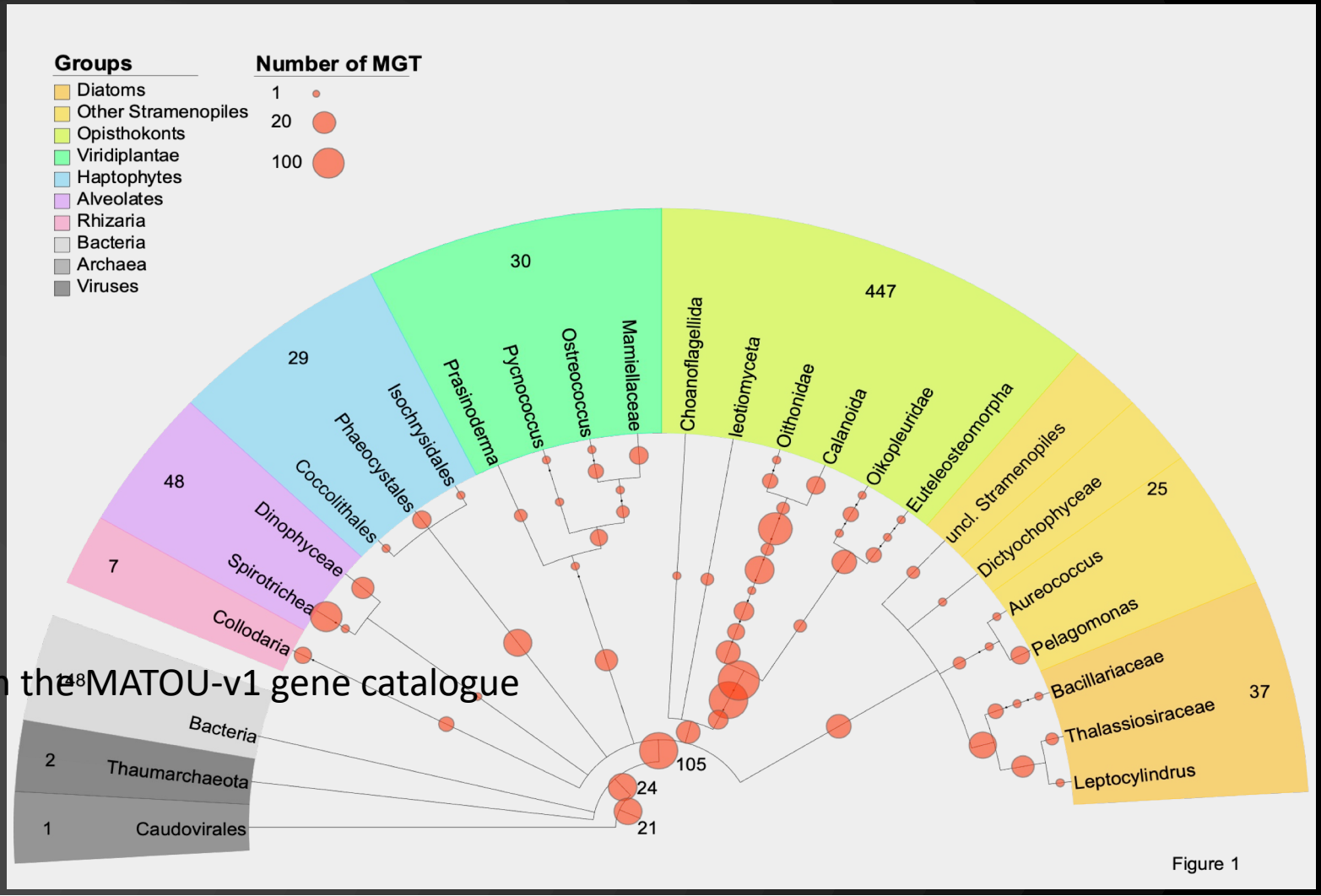
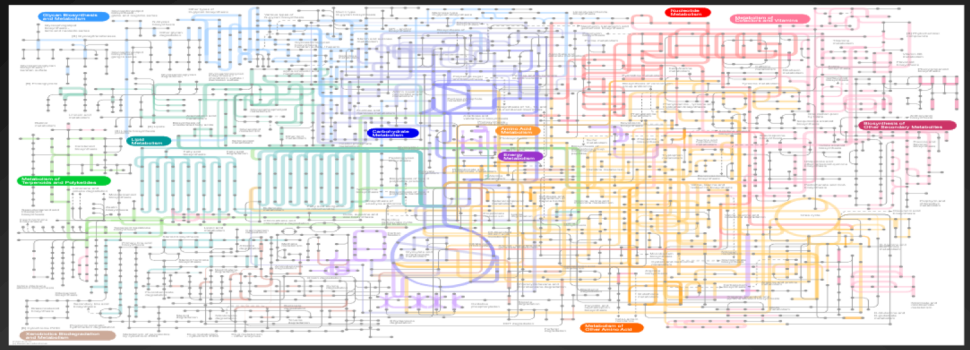
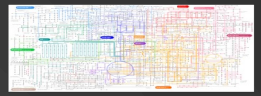


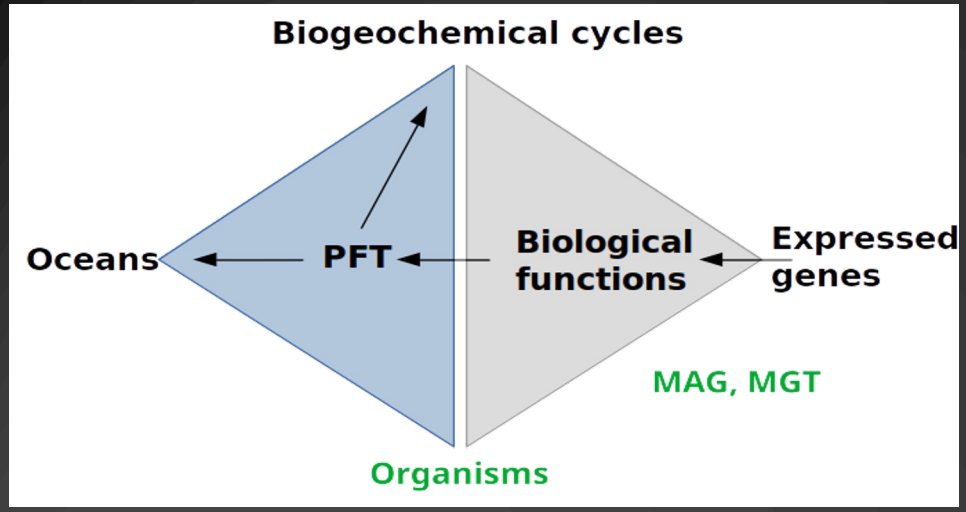
Figure 1

Metabolism

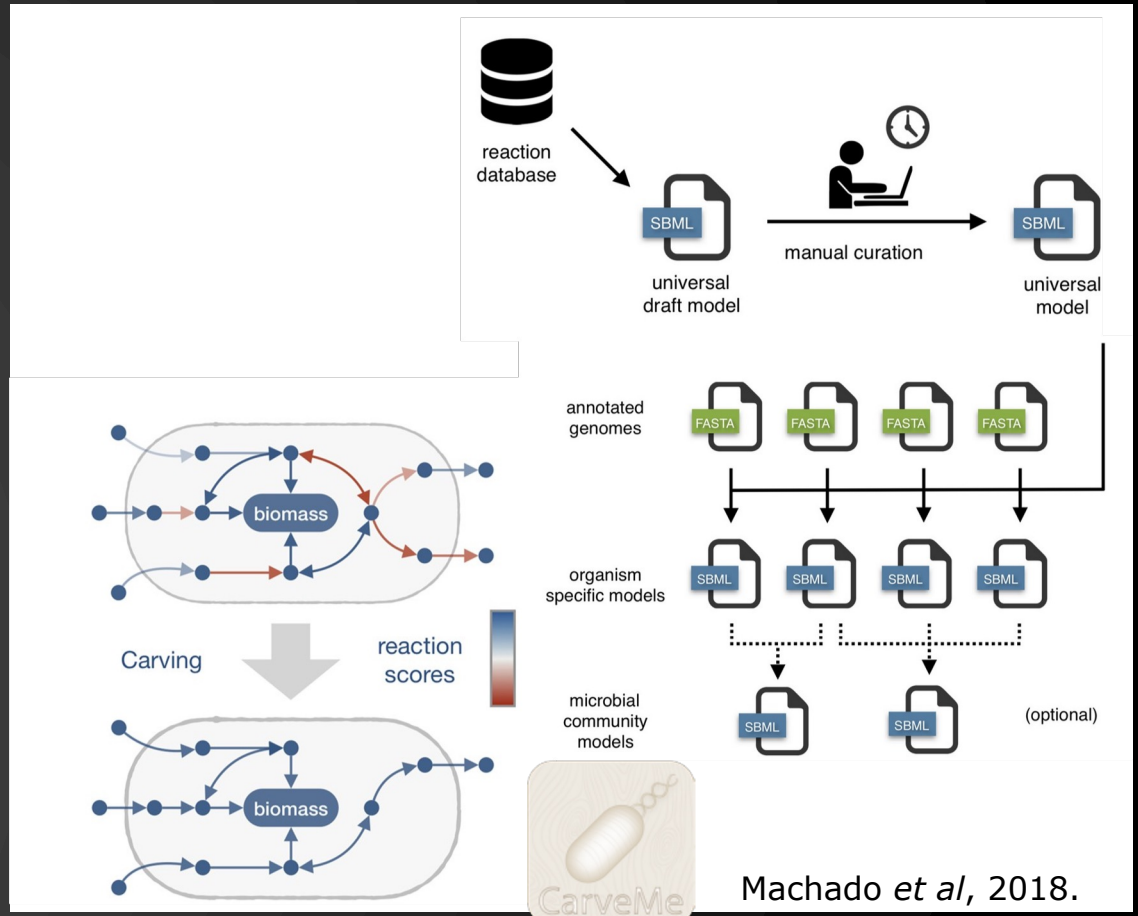




Metabolism modelling, linking omics with environment

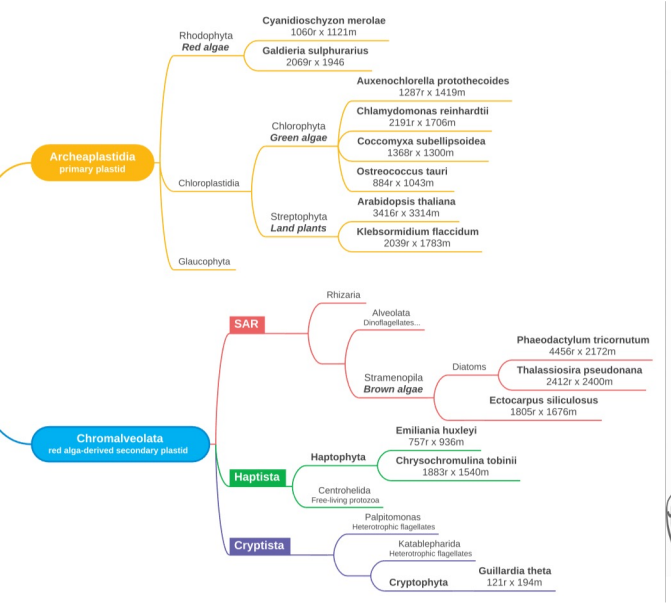


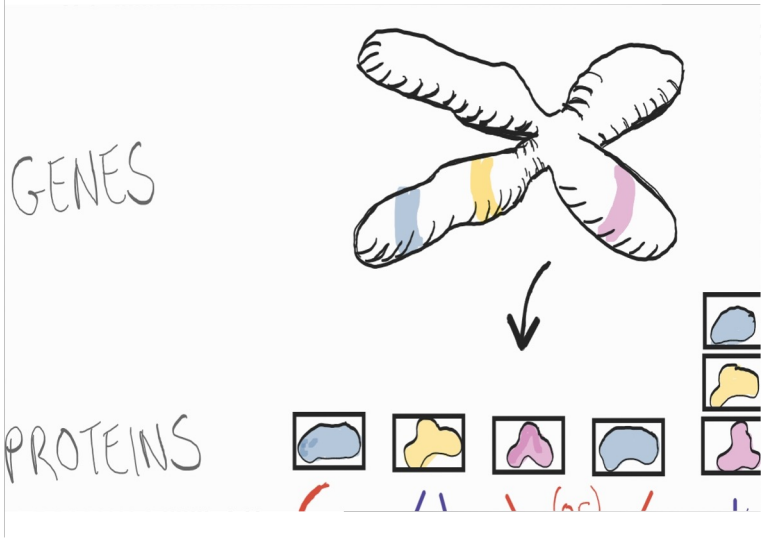
Marie Burel

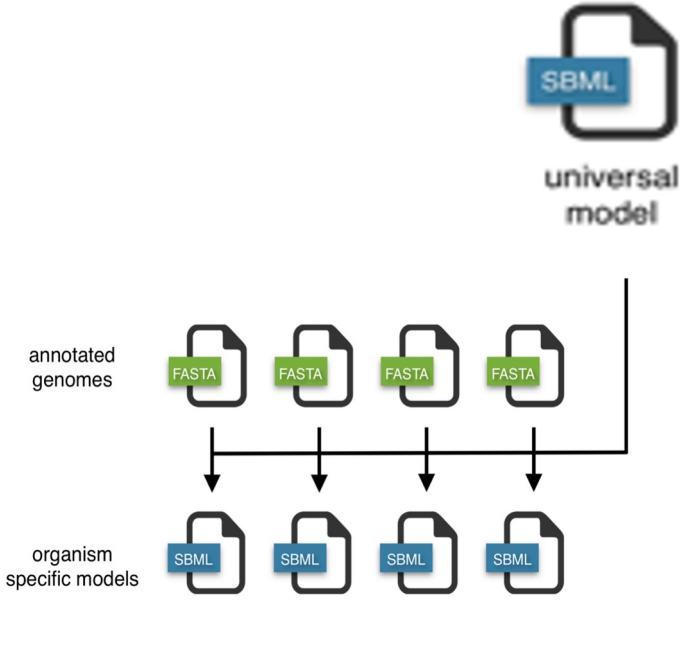


Machado *et al*, 2018.

PhotoEukStein : a generic model for phototrophic microeukaryote plankton



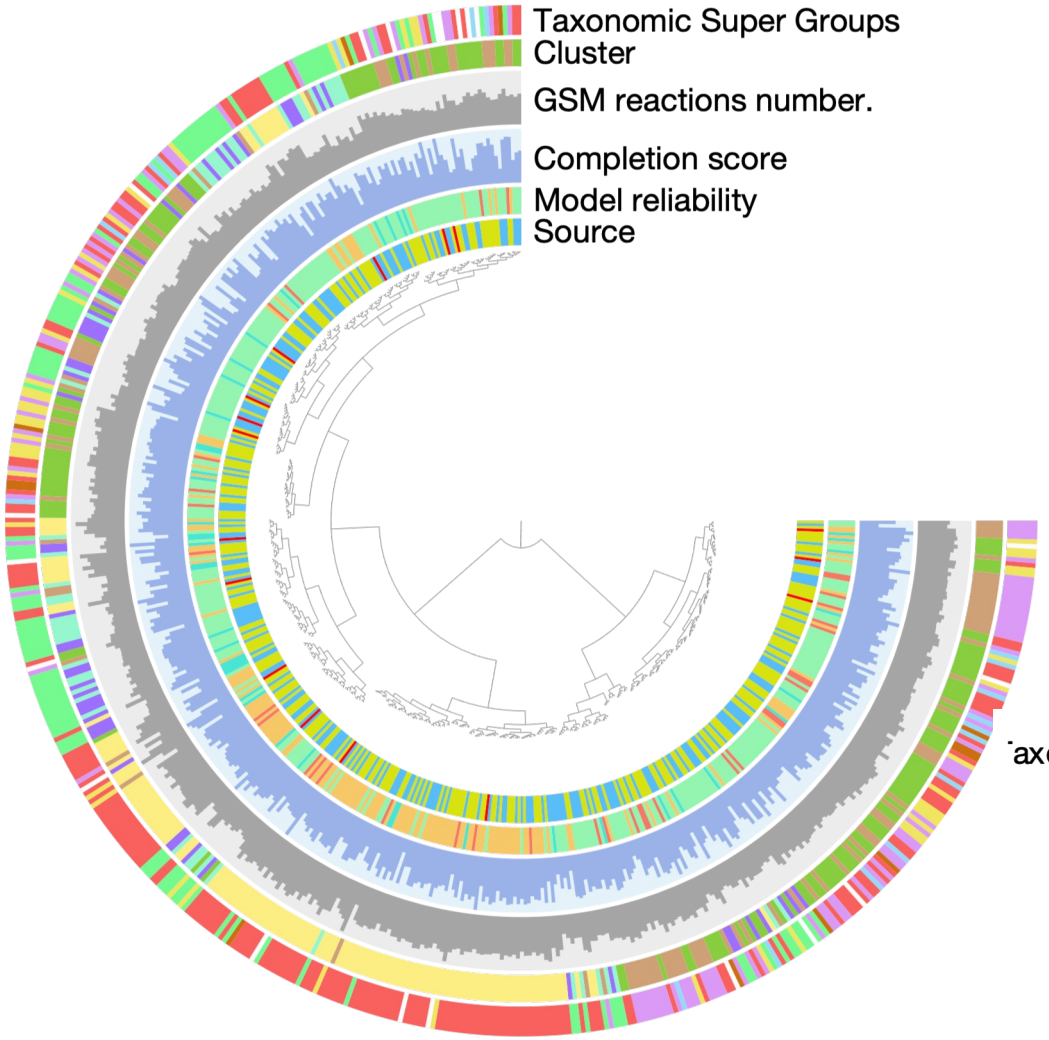




20468 sequences

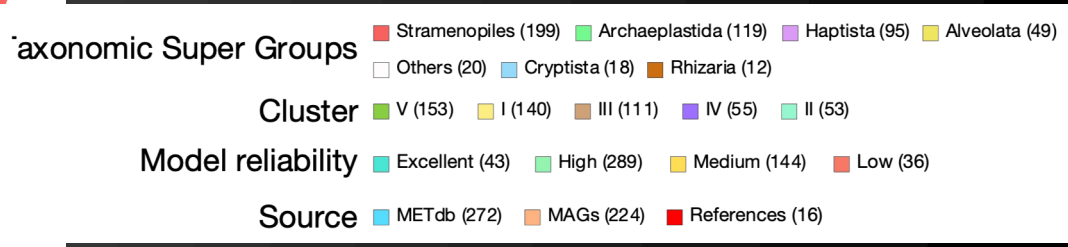
Gene-Protein-Reaction associations

PhotoEukStein
11229 reactions (9162 « internal »)
5831 metabolites



PhotoEukStein : 549 Genome-Scale Metabolic models for microeukaryote phototrophs

Marie Burel, Antoine Régimbeau

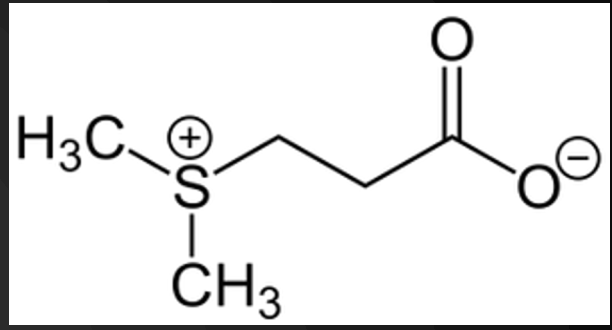


MAGs from Delmont et al., 2022
METdb from Niang, Corre and Pelletier, 2

Biogeochemistry

Sulfur cycle & eukaryote plankton

Dimethylsulfopropionate :



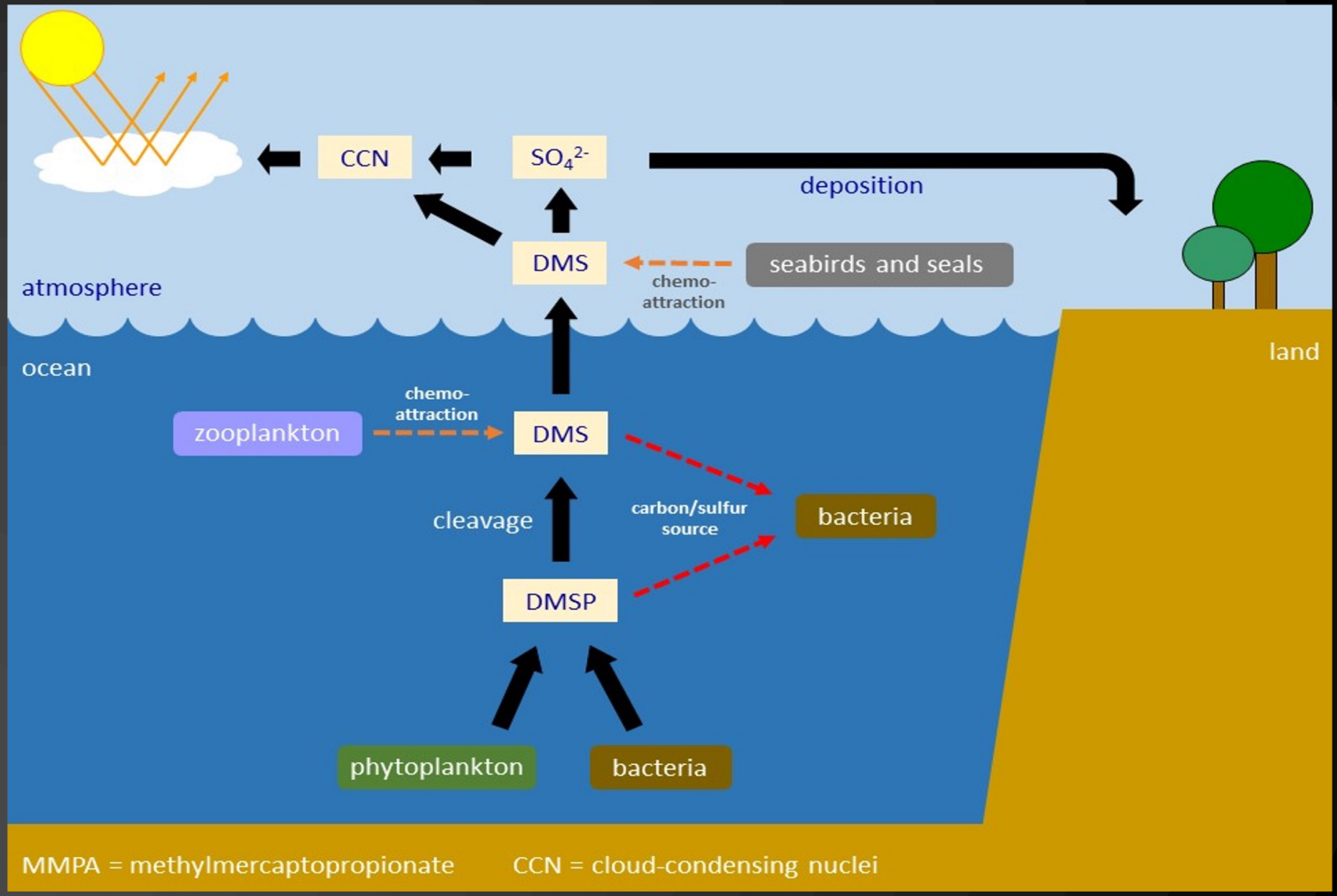
Produced by :

- Plants
- Green algaea
- Bacteria

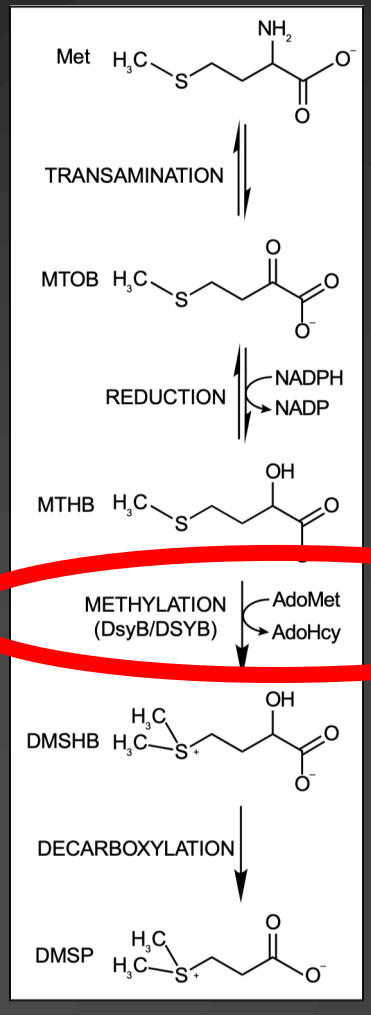
Functions :

- Intracellular osmolyte in bacteria
- Carbon source (1 to 13 % of bacterial carbon demand in ocean)
- Source of sulphur for bacterioplankton proteins
- ...

DMSP / DMS biogechemical cycle



DMSP biosynthesis pathway (bacteria, macroalgeaes, diatoms, haptophyes)



Methionine

4-methylthio-2-oxobutyrate

4-methylthio-2-hydroxybutyrate

4-dimethylsulfonio-2-hydroxybutyrate

Curson *et al.*, 2018

DSYB sequences (Curson *et al.*)

Clustering (MMseq2)

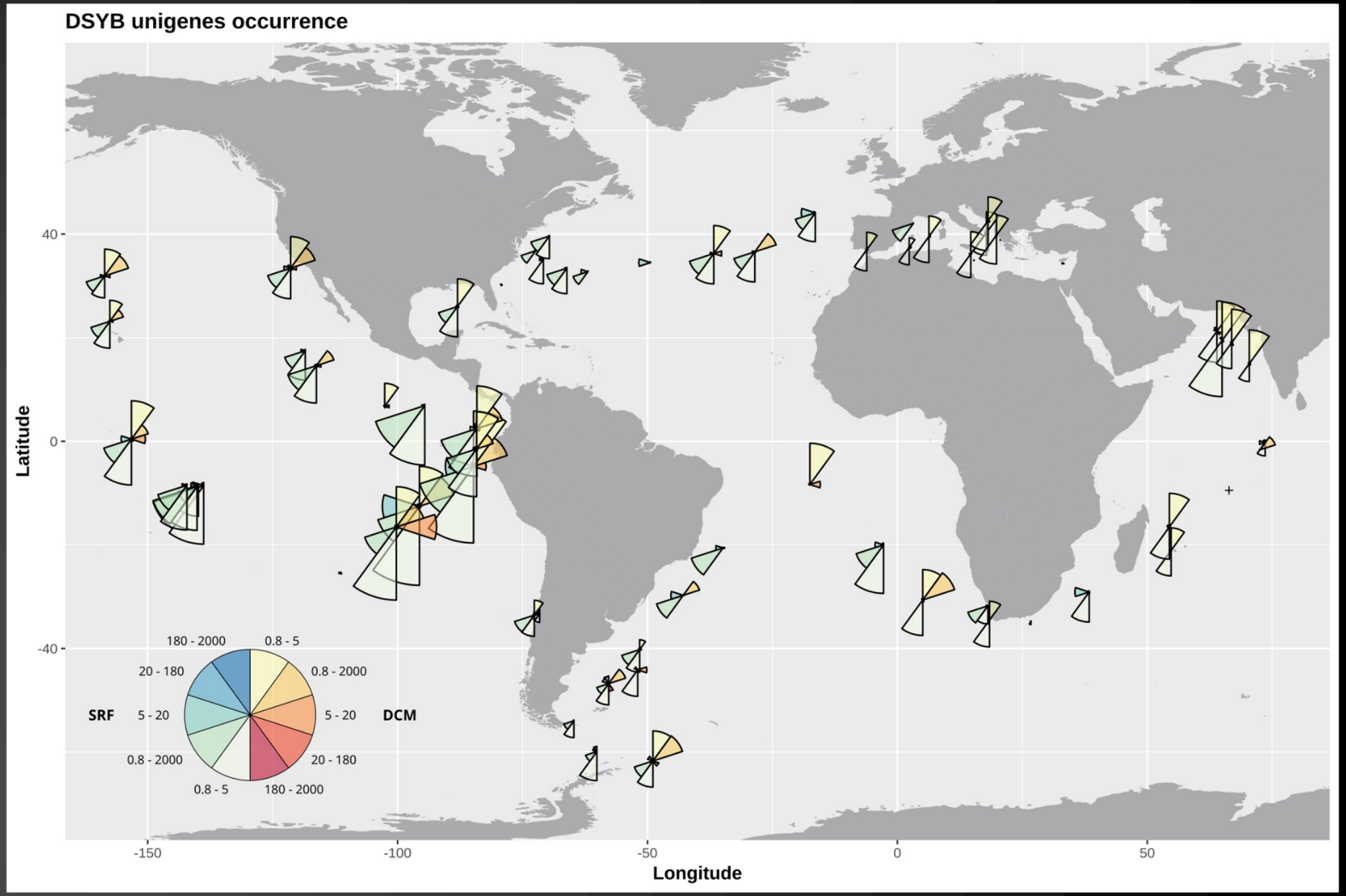
Alignment (MAAFT)

HMM build (HMMer)

MATOU-v1 scan

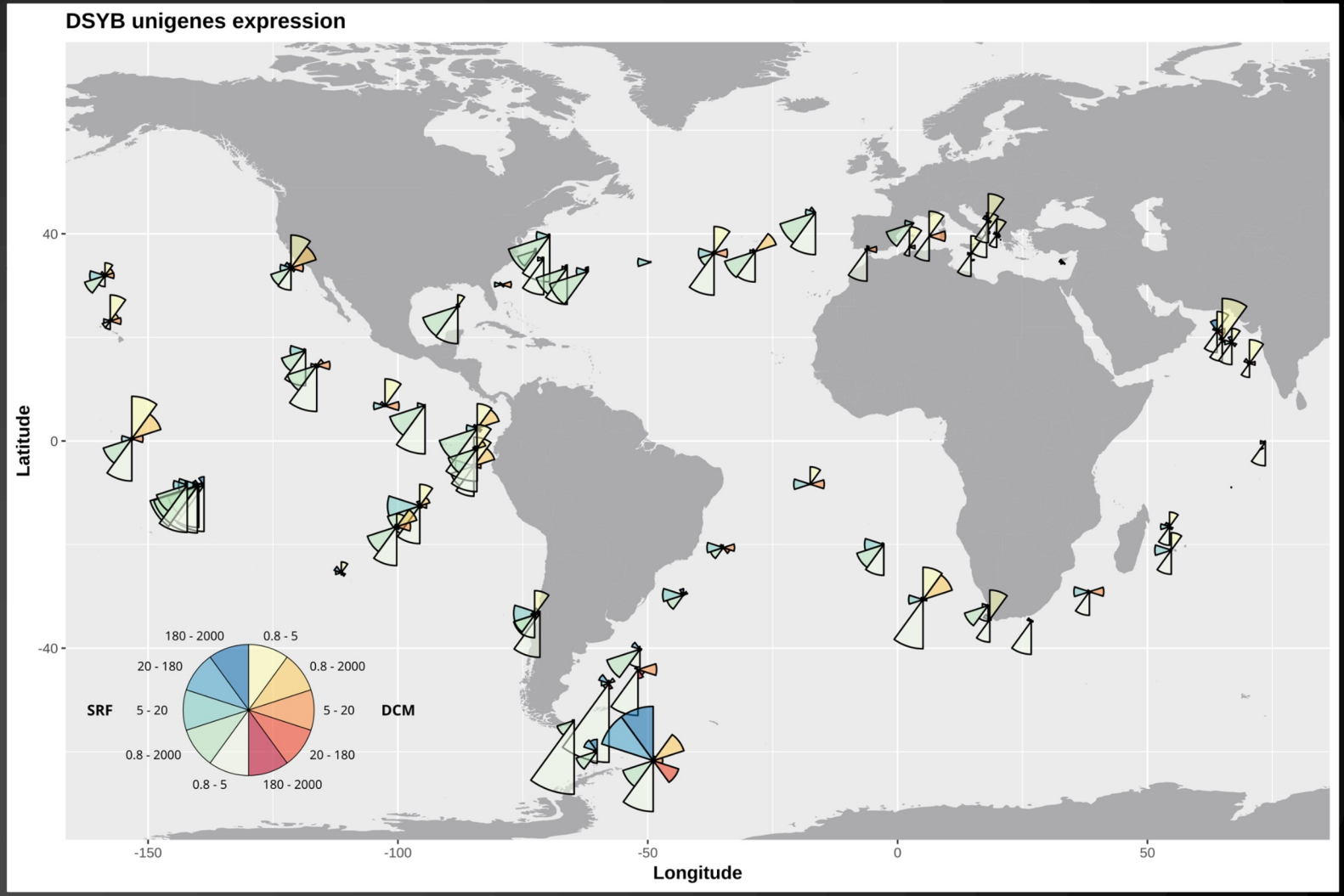
DSYB unigenes occurrence in Tara Ocean samples

1214 eukaryotes DSYB genes



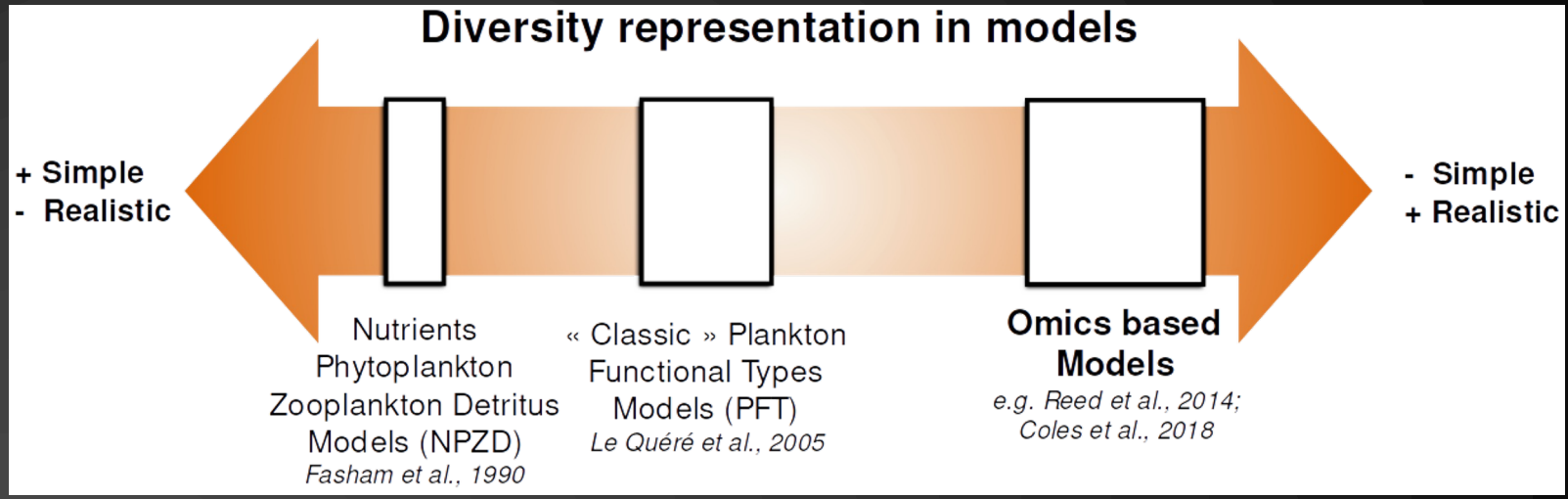
DSYB unigenes expression in Tara Ocean samples

1214 eukaryotes DSYB



Plankton Functional Traits

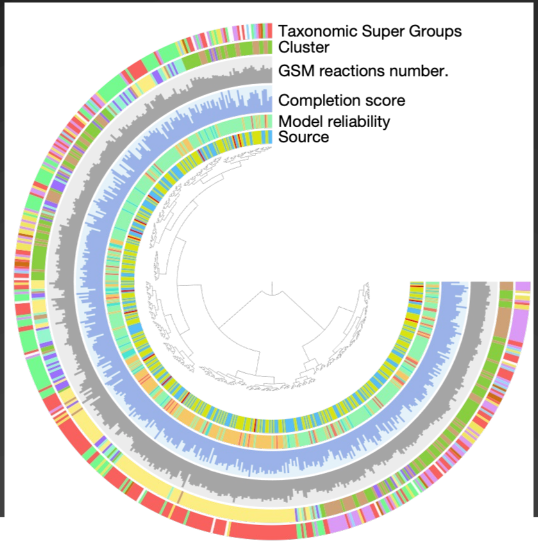
BioGeOmics

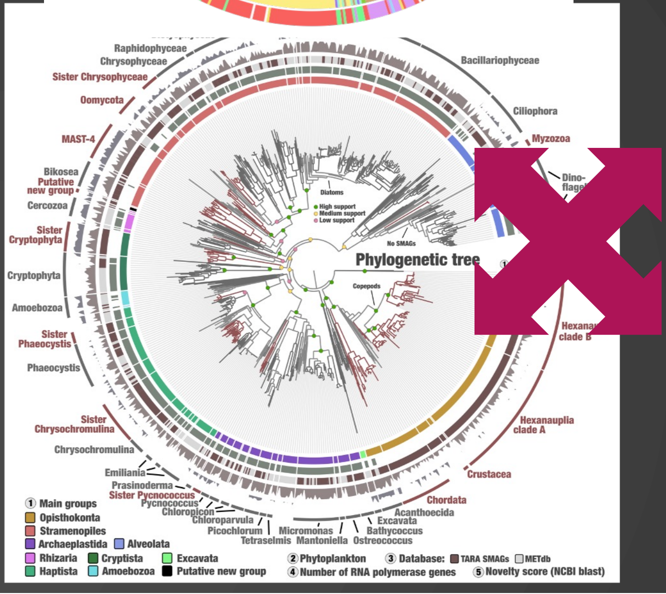


priori select trait of biogeochemical interest

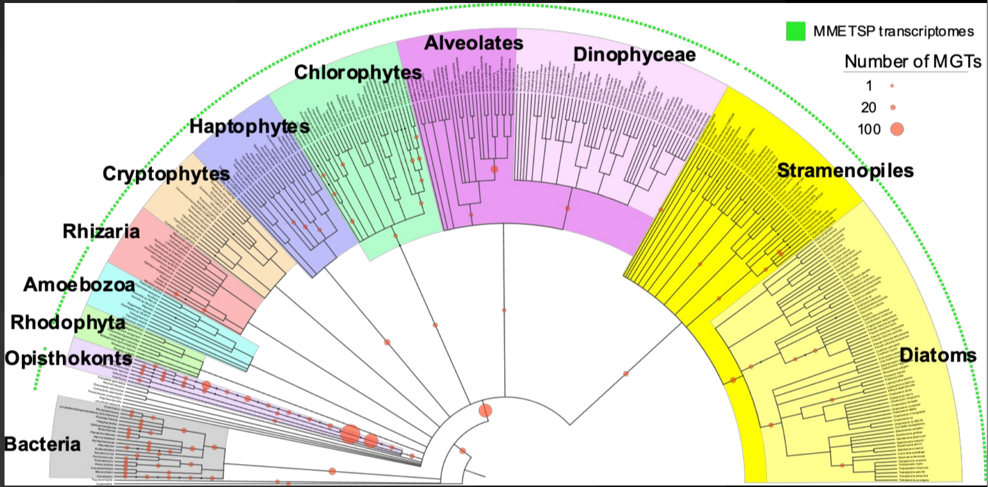
Data driven approaches to identify an

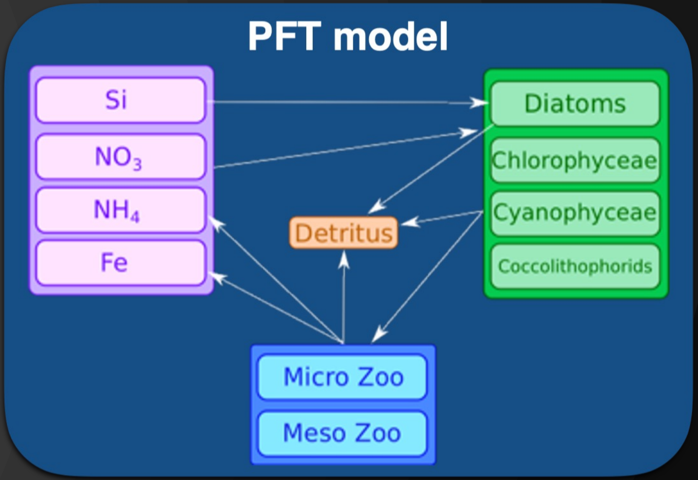
Connecting omics with biogeochemistry





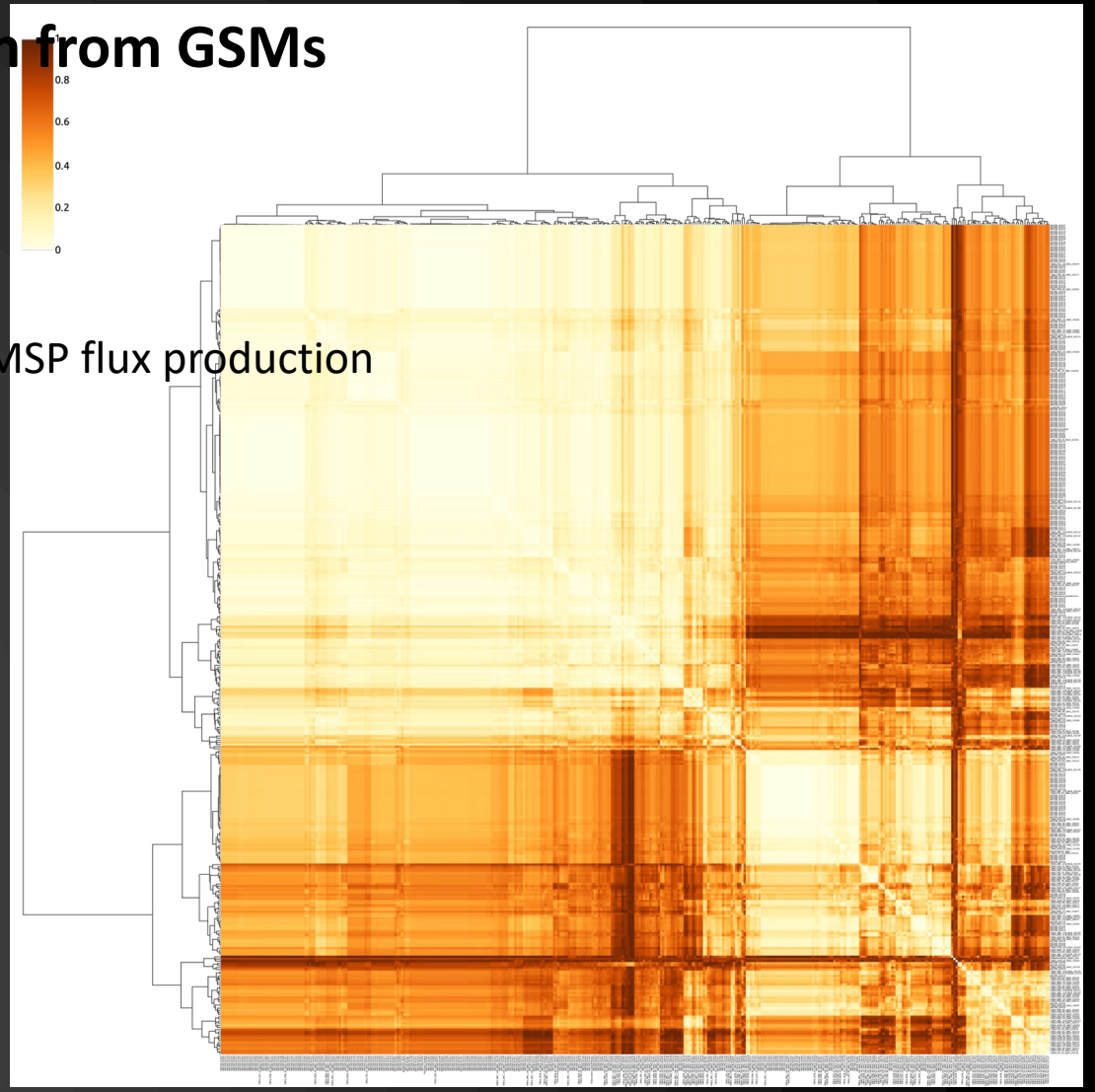
?





Towards functional traits definition from GSMs

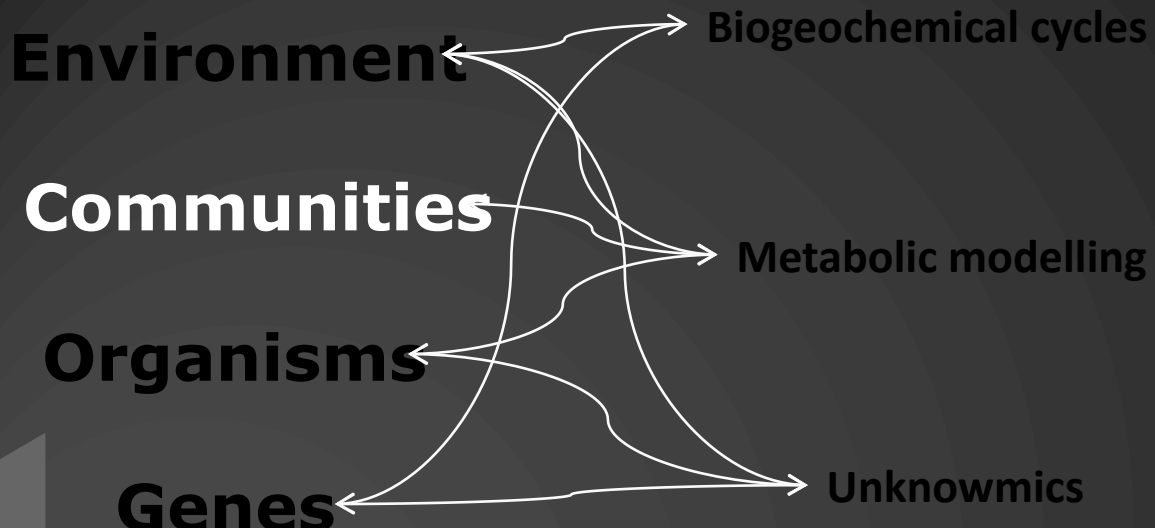
Responses to metabolic niches variations in DMSP flux production

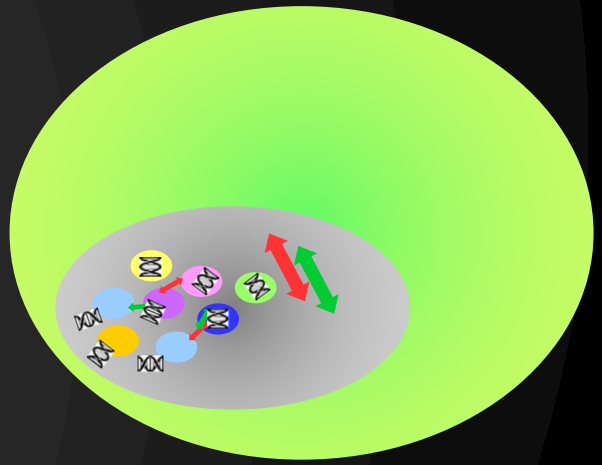


- 387 GSMs producing DMSP
- 2 main « functionally responsive » clusters
- Not linked to taxonomy

Summing things up

A Holistic approach, multi-scale and multi-d





Temporal evolu

(some) challenges :

- Integrating heterogeneous data types
- Low in-depth resolution of communities
- (almost) still missing the organismal integration level
- Functional interpretation at the various levels

Adrianna Alberti
Alizée Amsler
Jean-Marc Aury
Marie Burel
Quentin Carradec
Corinne DaSilva
Tom Delmont
Marion Dupouy
Morgan Gaïa
Damien Hinsinger
Olivier Jaillon
Amos Kirilovsky
Jade Leconte
Amine Madoui
Marie Burel
Samuel Mondy
Mario Néou
Eric Pelletier
Julie Poulain
Gaelle Samson
Yoann Seeleuthner
Thomas Vannier
Alexey Vorobev
Marc Wessner
Patrick Wincker

& Tara Oceans Consortium