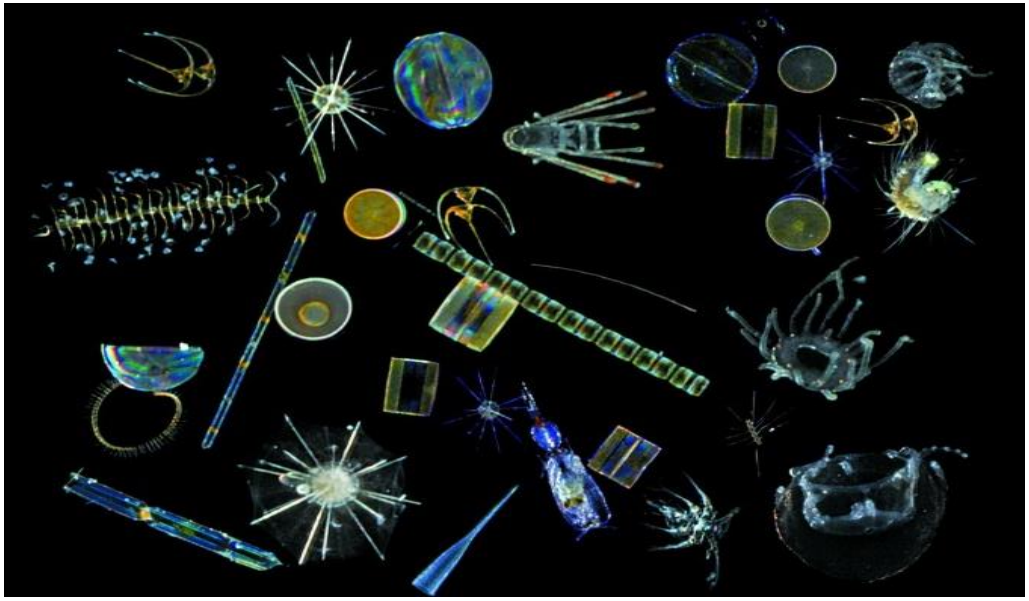
Meta-omics to explore and monitor ocean ecosystems biology in a changing environment

Quentin Carradec

Laboratoire d'analyses génomique des eucaryotes, Genoscope, Evry



Outline

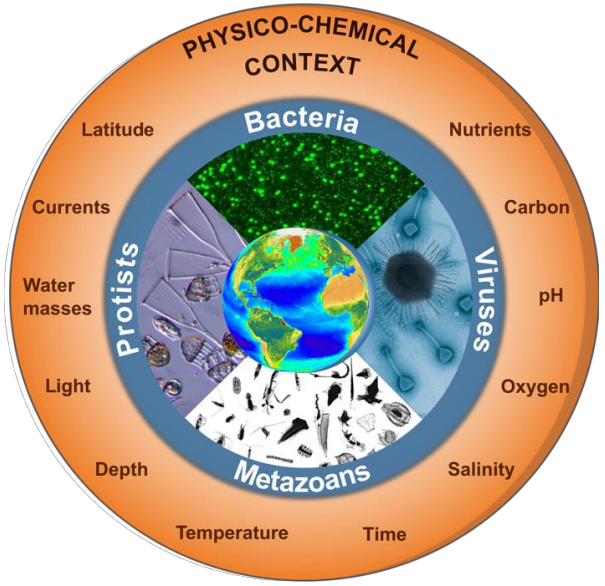
- **Metatranscriptomic approaches to study gene diversity and expression variations of complex planktonic communities**

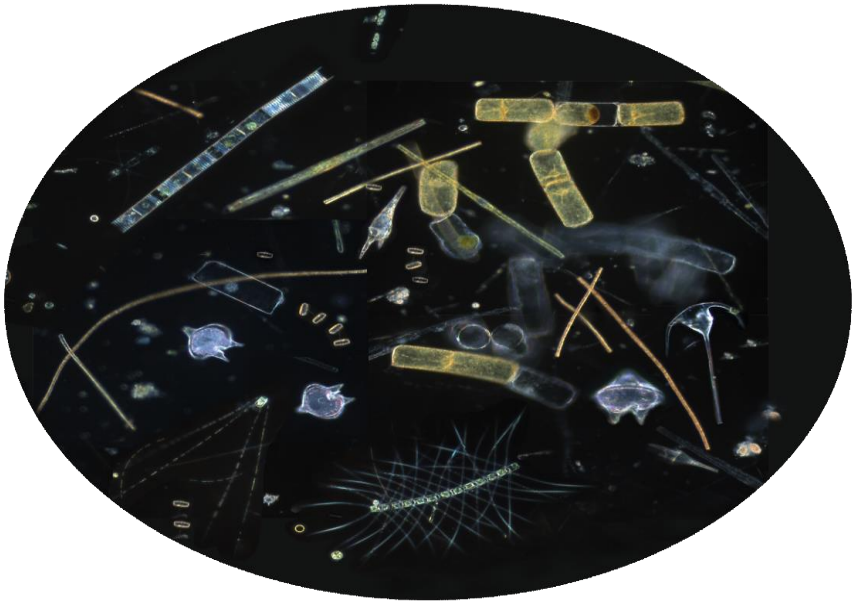
- **Species diversity monitoring with the Nanopore technology**

Marine plankton communities: an essential and complex ecosystem

- 100x to 1000x more complex than a human microbiota.
- Very large 3D ecosystem
- Time variability (daily seasonally)
- Impacted by physico-chemical context

- Basis of oceanic trophic webs
- Production of 50% of the oxygen we breath
- Sequester atmospheric CO_2 in the ocean : carbon pump
- Largely understudied
  - → 98% of biomass is unicellular.
  - → most of species are not cultivable in laborator

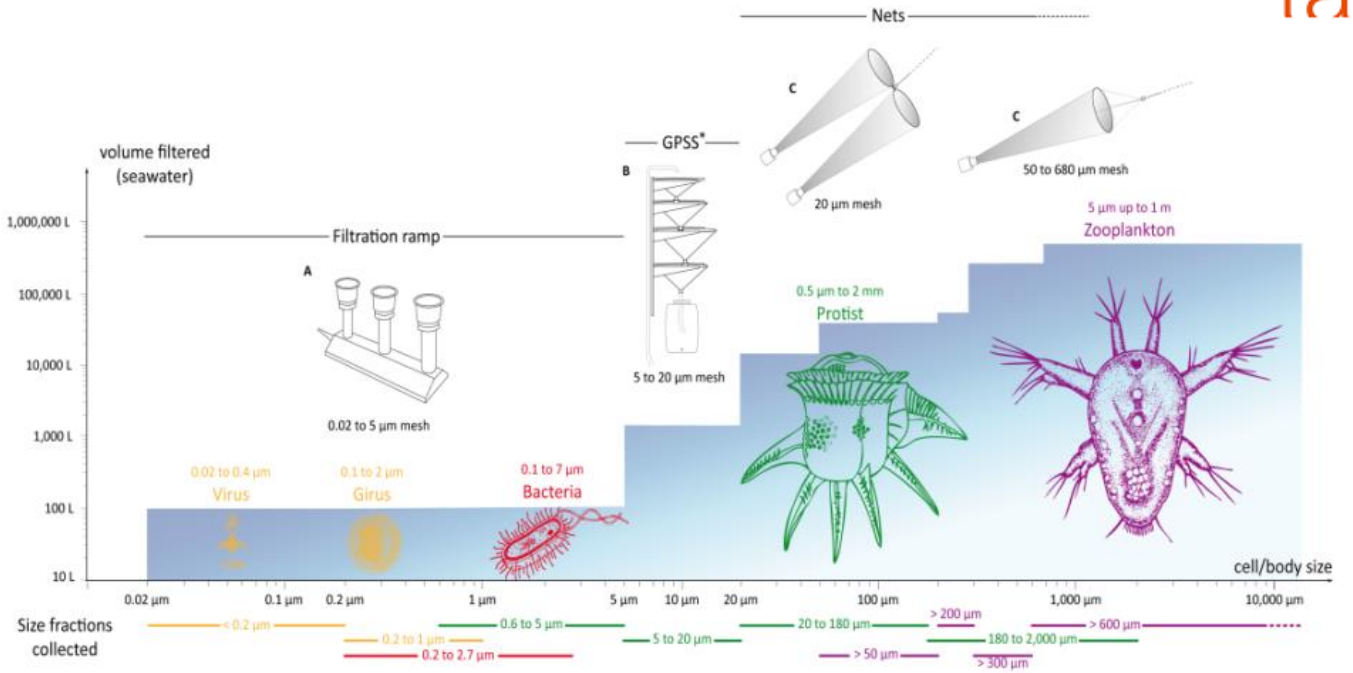




Tara expeditions to explore plankton communities (2009-2012)

Fondation tara océan
explorer et partag

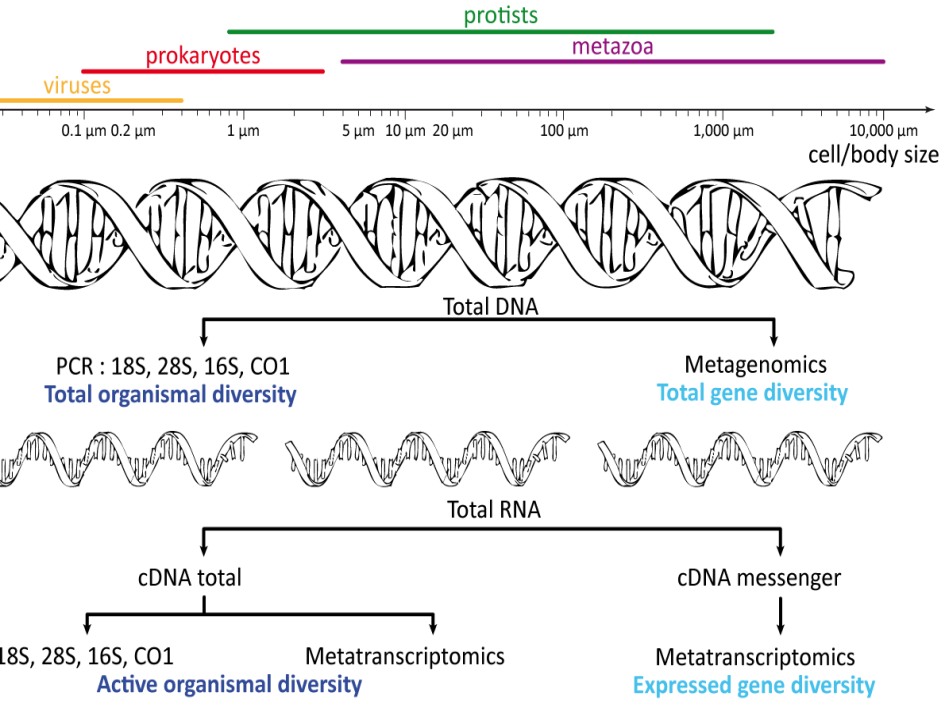
Same protocol of plankton sampling and filtering



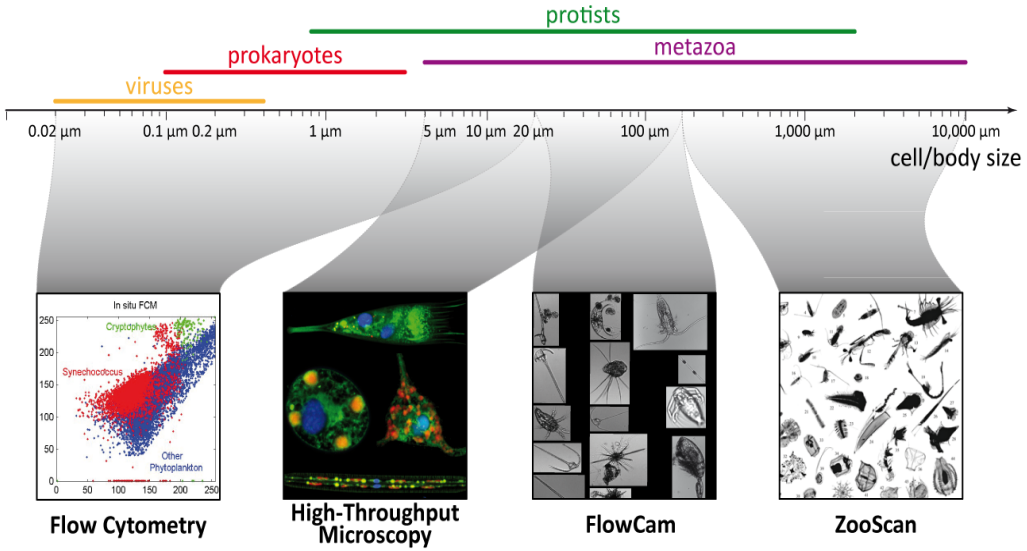
quentin.carradec@genoscope.cns.fr

Holistic approach

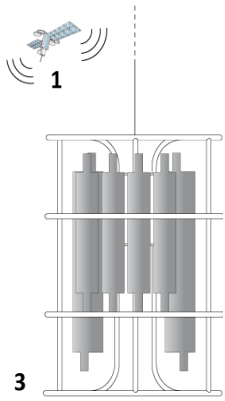
High Throughput Sequencing



High Throughput Imaging



Physico-chem



From seawater sample to genomic catalogs

eawater sampling

68 marine stations

Filtration

5 size-fractions (between 0.8 µm and 2000 µm)
3 depths (Surface, deep chlorophyll maximum, mesopelagic)
441 samples

Cryogenic grinding

DNA extraction

RNA extraction

polyA+ selection for eukaryote size-fractions

Sequencing : Illumina and Nanopore technologies

40 Gbases per sample

Metagenomic reads

Metatranscriptomic reads

40 Gbases per sample

co-assembly per oceanic basin (Anvio)

Assembly sample per sample (Velvet/Oasis)
Clustering of all samples at 95% identity (CD-HIT)

11 millions contigs

Contigs

Transcript catalog

117 millions

Carradec Q , Pelletier E et al 2018 Nat Comm

binning (Anvio)

Binning (canopy clustering)

700 genomes

Delmont T et al 2022 Cell Genomics

MAGs

MGTs

924 transcriptomes

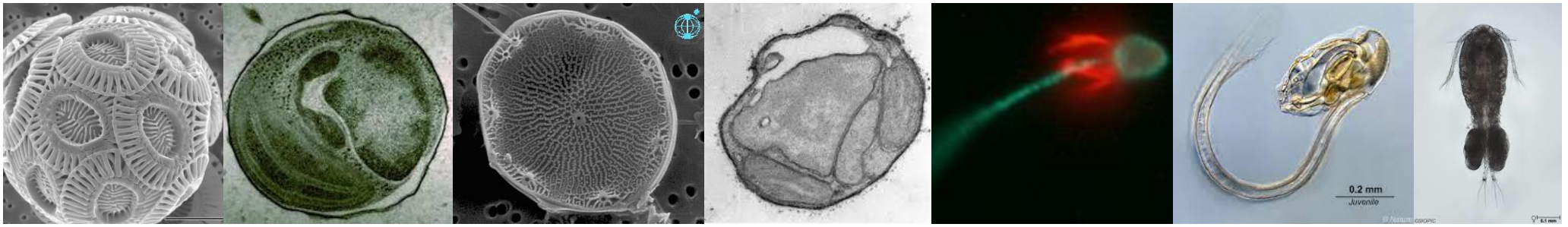
Vorobev A et al 2020 Genome Research

quentin.carradec@genoscope.cns.fr

Transcriptome completion and fragmentation

stimation based on sequenced planktonic genomes

Reference genomes	Phylum	Number of genes	% of identity cutoff	Number of selected genes		Average number of unigenes per gene
Emiliania huxleyi	Haptophyte	38 548	96	17949	(46.6%)	**2.15**
Bathycoccus prasinos	Chlorophyte	7 851	98	6364	(81.1%)	**3.32**
Thalassiosira oceanica	Stramenopiles (Diatom)	34 500	89	9029	(26.2%)	**2.29**
Aureococcus anophagefferens	Stramenopiles (Pelagophyceae)	11 522	87	6765	(58.7%)	**2.01**
Monosiga brevicollis	Metazoa Choanoflagellates	9 175	97	3823	(41.7%)	**1.86**
Oikopleura dioica	Metazoa Tunicata	18 020	86	9670	(53.7%)	**2.40**
Oithona nana	Metazoa Copepoda	17 767	92	10312	(58.0%)	**1.89**



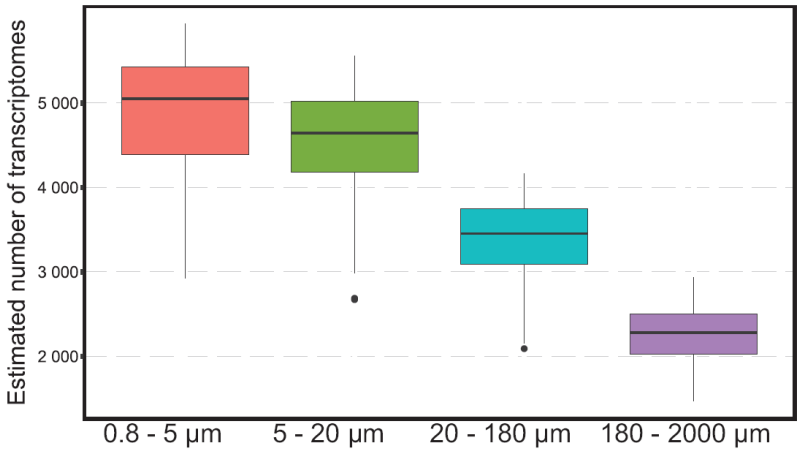
Completion is strongly variable according to the species abundance.

Fragmentation is estimated to 2.2 unigenes per gene

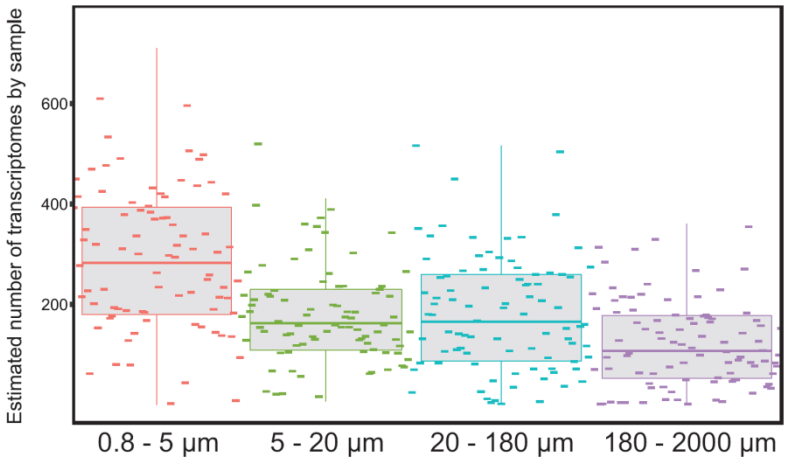
How many transcriptomes are in the gene catalog ?

Number of different ribosomal proteins (single copy) detected by sample or by size fraction (average of 24 proteins)

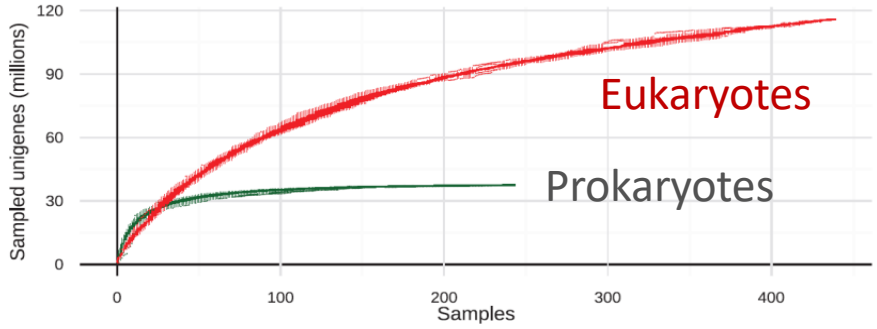
Total diversity per size fraction



Diversity per sample

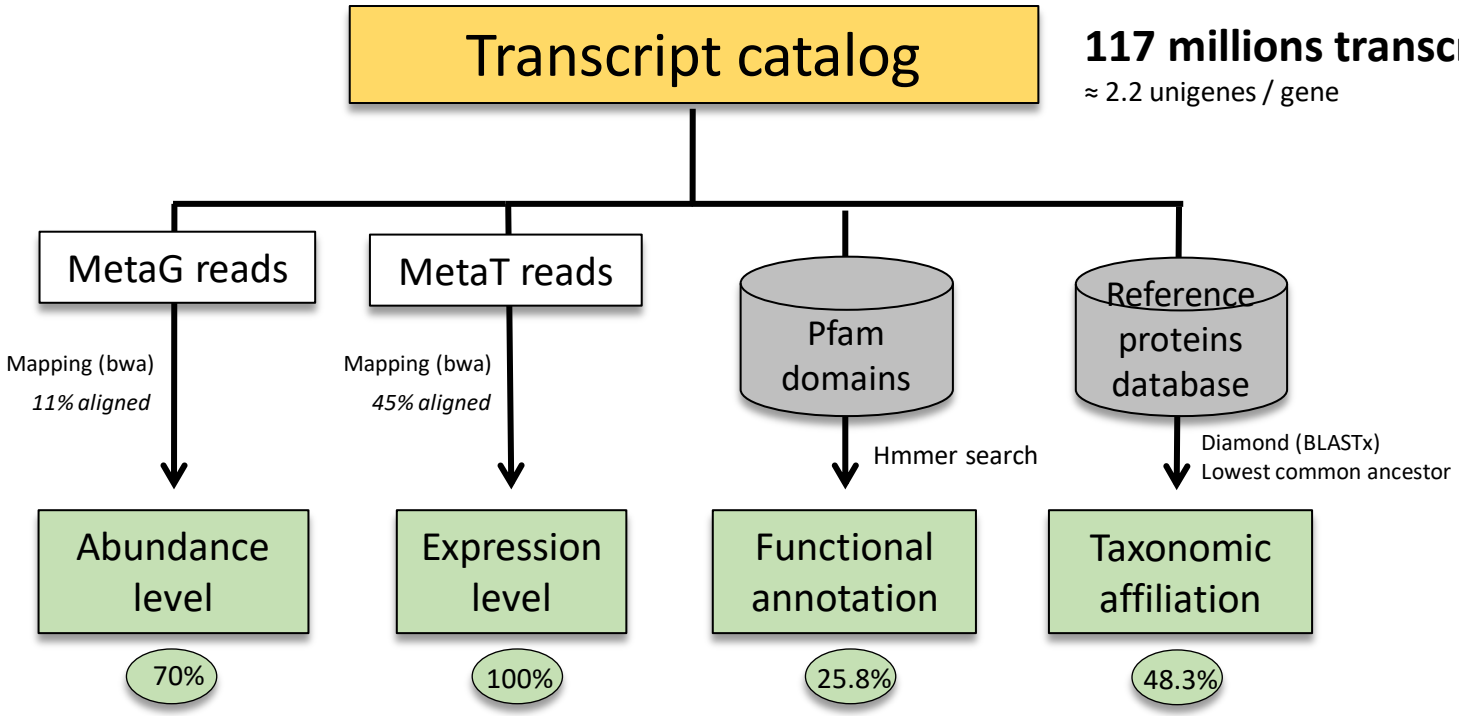


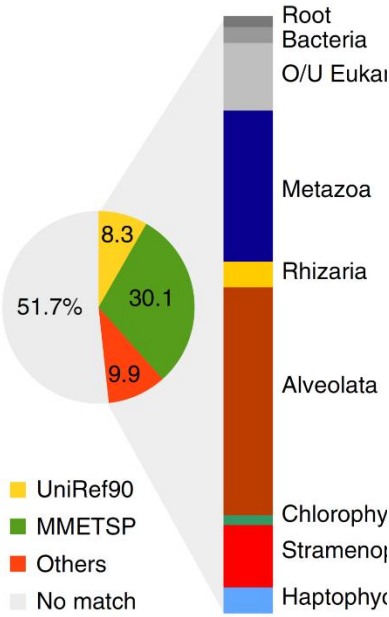
➡ Partial transcriptomes from 9 000 organisms << 150 000 OTUs (18SV9 rRNA)



➡ Between 166M and 190M unigenes in ocean surface

Transcript catalog annotation





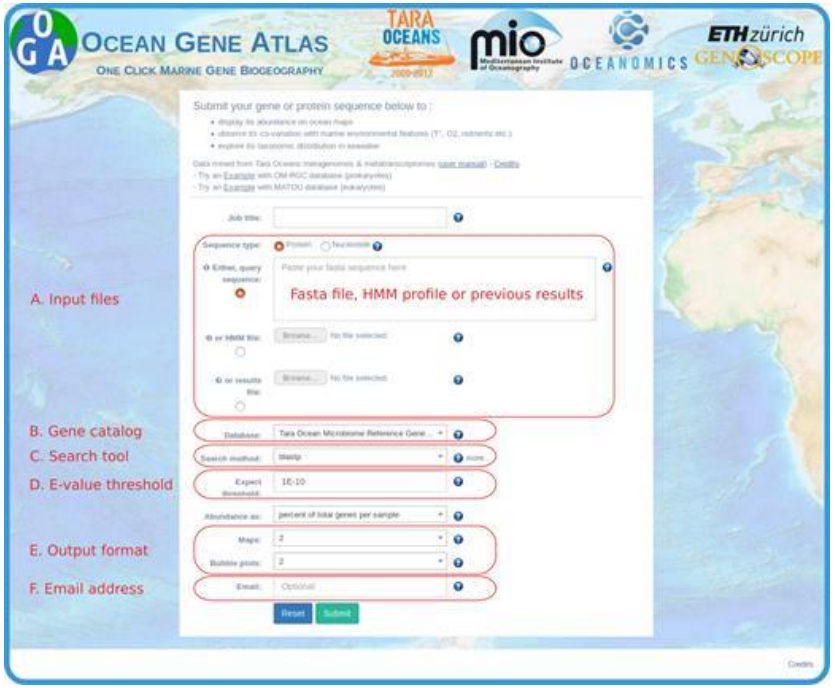
➡ > 50 % of unknown genes

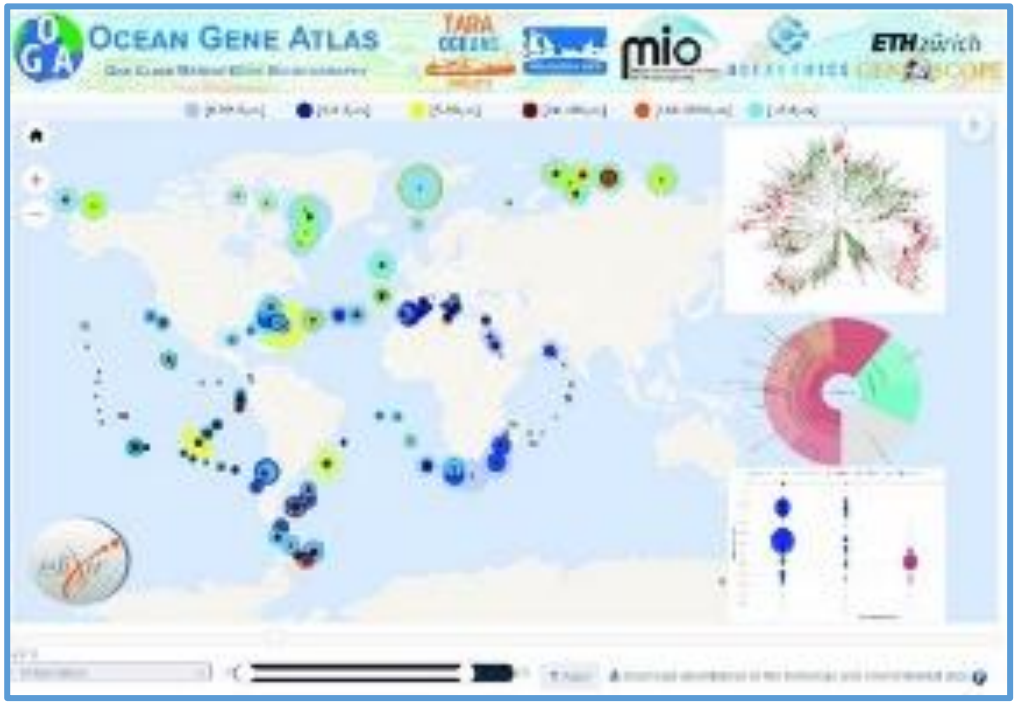
Data sharing

Raw readsets : https://www.ebi.ac.uk/metagenomics/

Gene and genome catalogs : https://www.genoscope.cns.fr/tara/

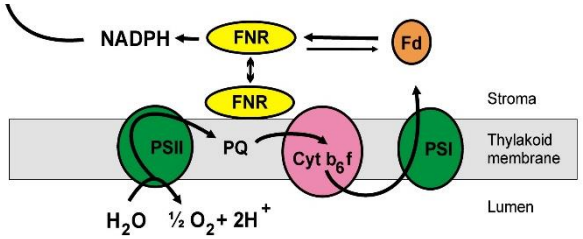
Analysis: http://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas/

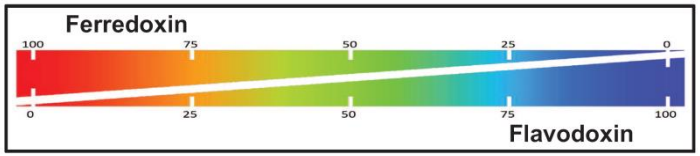




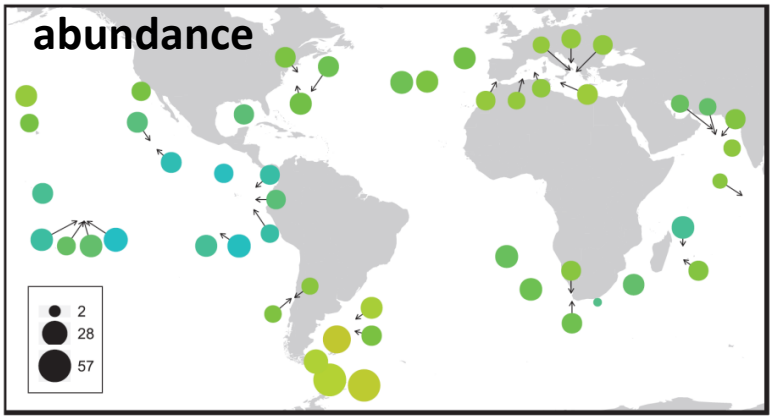
Ocean Gene Atlas: v2.0: online exploration of the biogeography and geny of plankton genes Caroline Vernette *et al 2022*

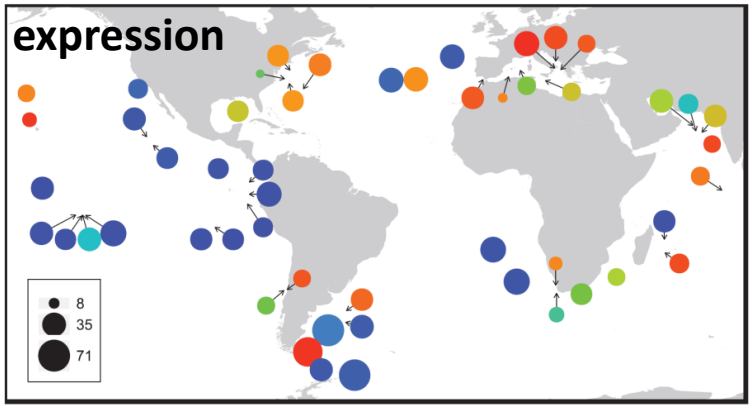
Ferredoxin vs Flavodoxin expression

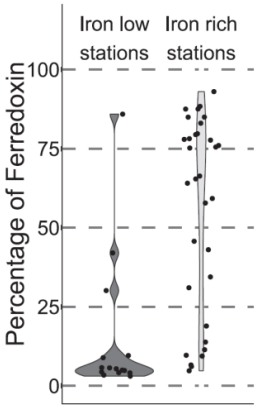




Haptophytes



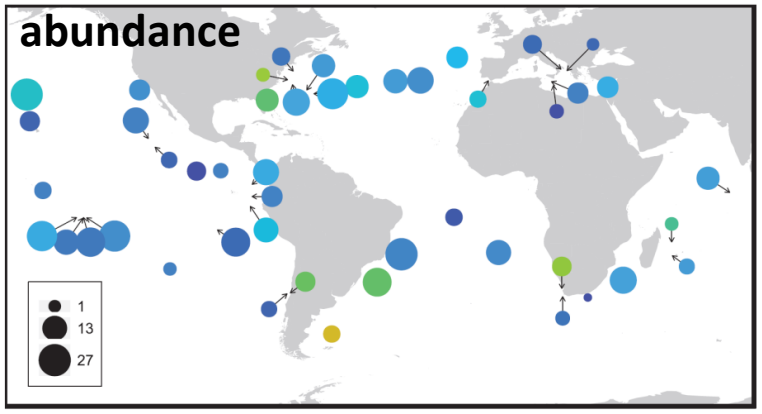


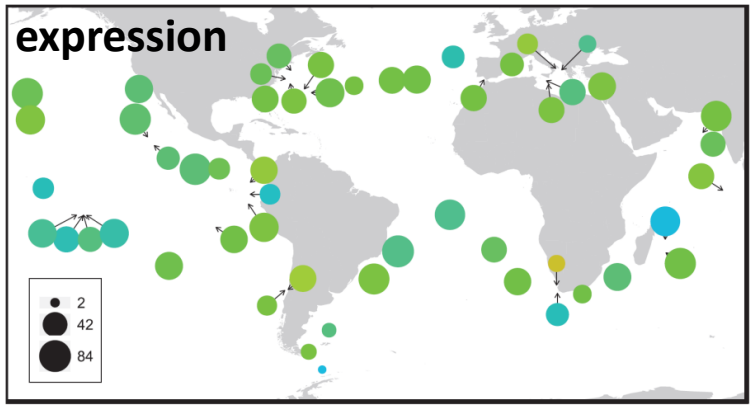


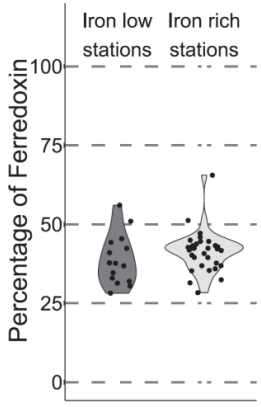
p = 8.37e-06

Transcriptional acclimation

Dinoflagellates







p = 1.49e-01

No transcriptional adaptation

Decelle

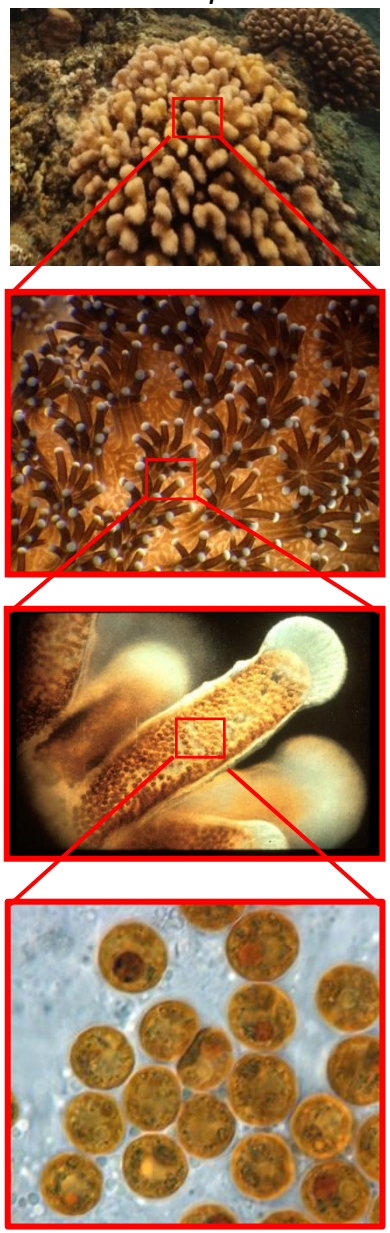
Gene catalogs to study specific organisms

Worldwide occurrence and activity of the reef-building coral symbiont Symbiodinium in the open ocean

quentin.carradec@genoscope.cns.fr

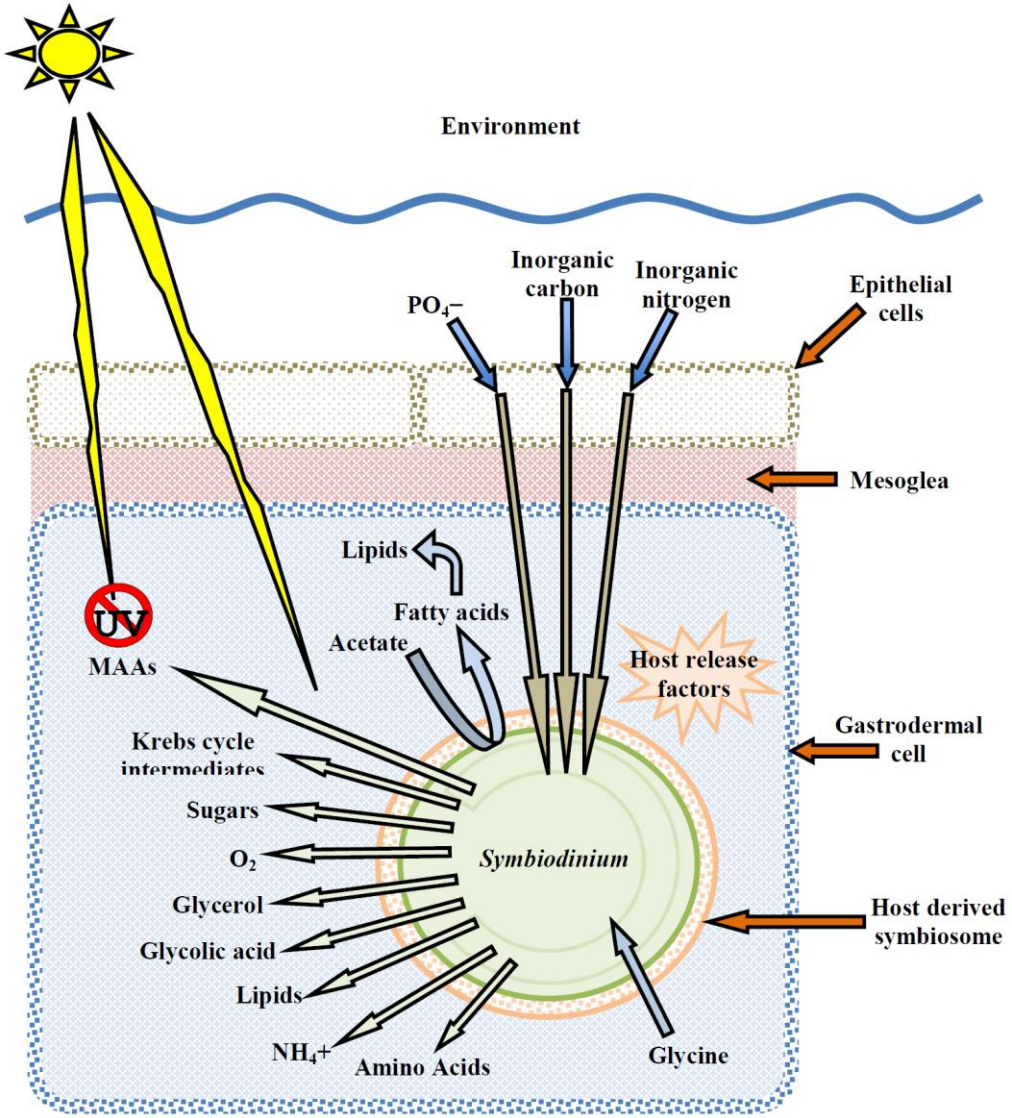
Coral – microalgae symbiosis

Pocillopora



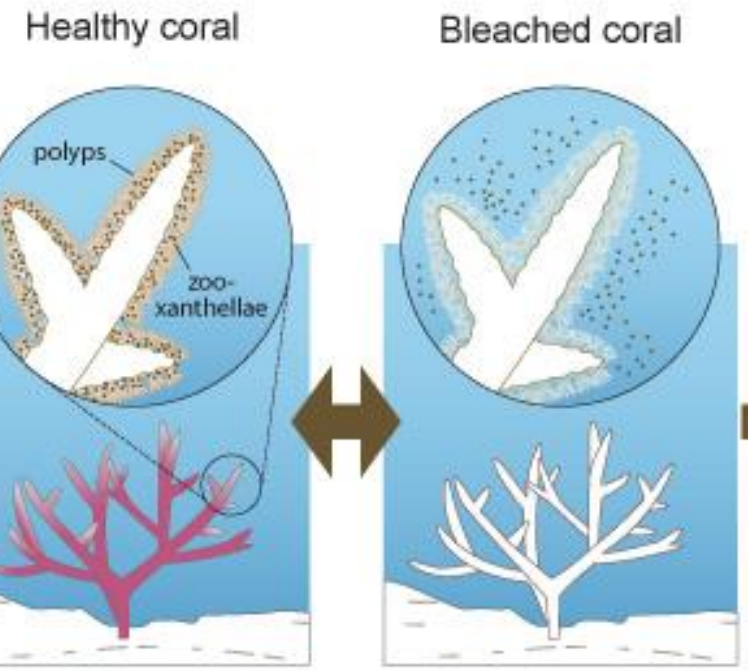
biodiniaceae

~5 µm

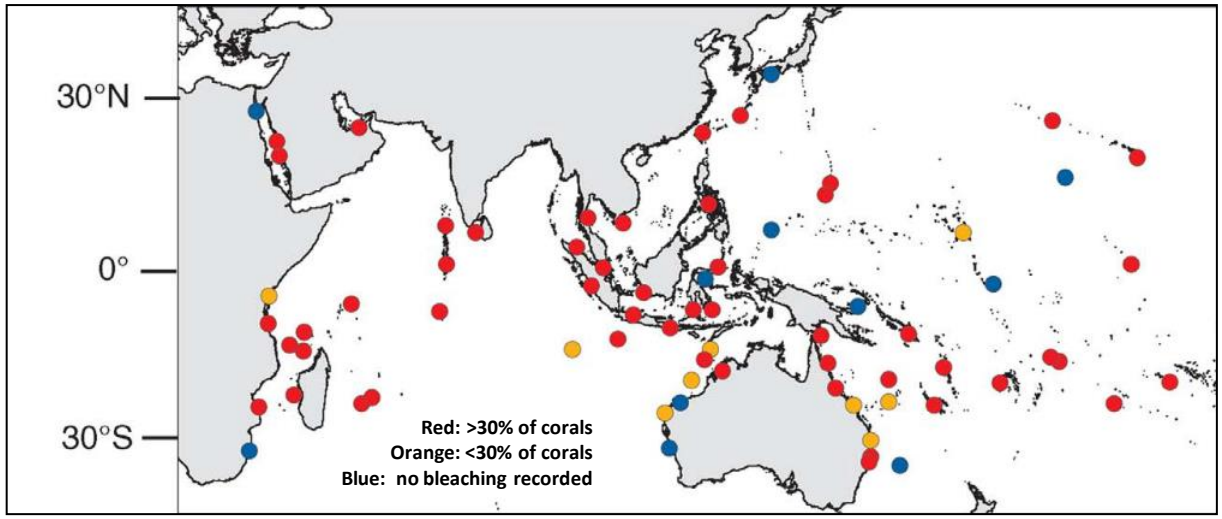


Obligatory endosymbiosis with corals

Symbiosis temperature dependant



2016: mass bleaching in the oceans



Bleaching events are mainly due to global ocean warming

Bleached corals

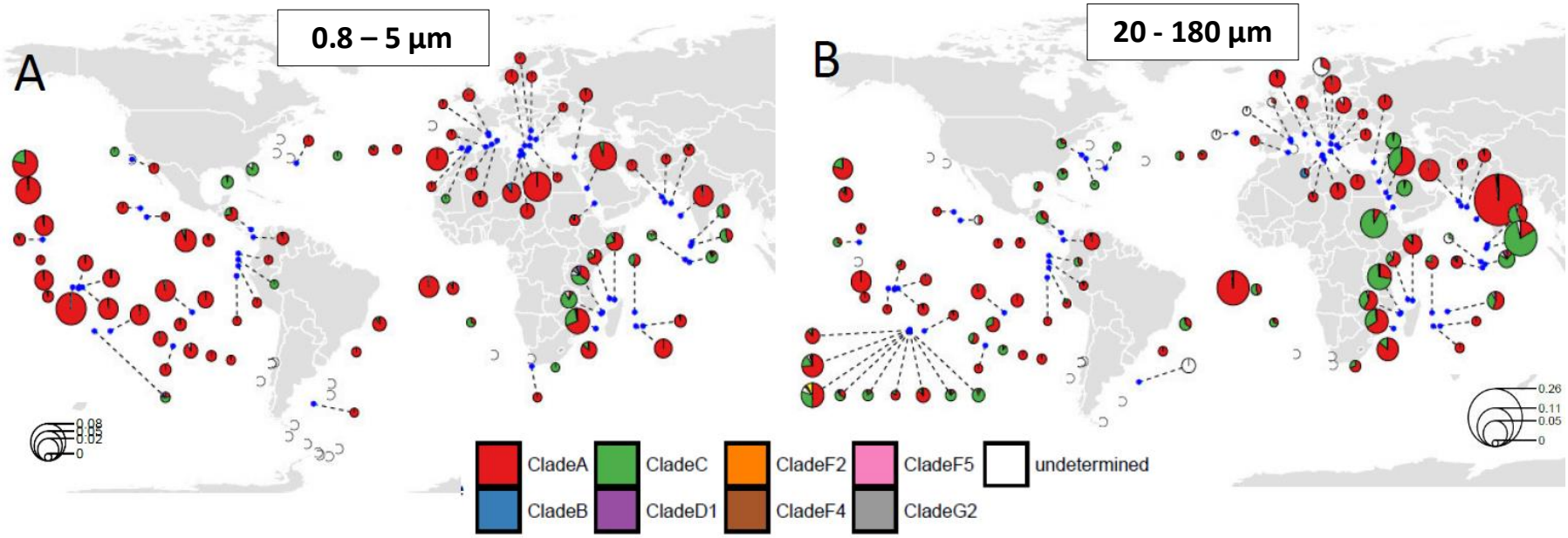
Coral resilience is dependant of *Symbiodinium* presence in the environment.

Is *Symbiodinium* able to travel long distances across coral reefs

*Hughes, T. P. et al. Spatial and temporal patterns of mass bleaching of corals in the Anthropocene. Science **359**, 80-83, doi:10.1126/science.aan8048 (2018).*

Symbiodinium abundance in open oceans

- Metabarcoding V9 (18S)
- 19 different sequences (A, B, C, D1, F2, F3, F4, F5, G1, G2)



➡ ***Symbiodinium*** **is present in all oceans, in small as well as in large size-fractions**

➡ **Clades A and C are the most abundant**

Decelle et al Current biolo

Transcriptomic activity of *Symbiodinium*

Catalog of 117 millions transcripts

27 transcriptomes

blastn

Symbiodinium unigenes

5 119 unigenes

Functional annotation

Expression and abundance level In each station

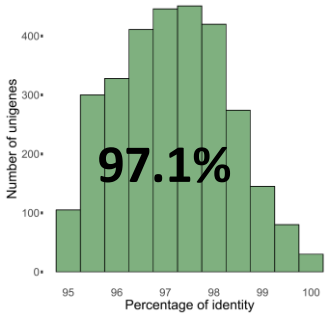
2 136 unigenes

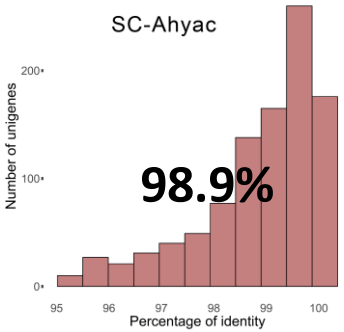
Clade A

2 983 unigenes

Clade C

2 transcriptomes

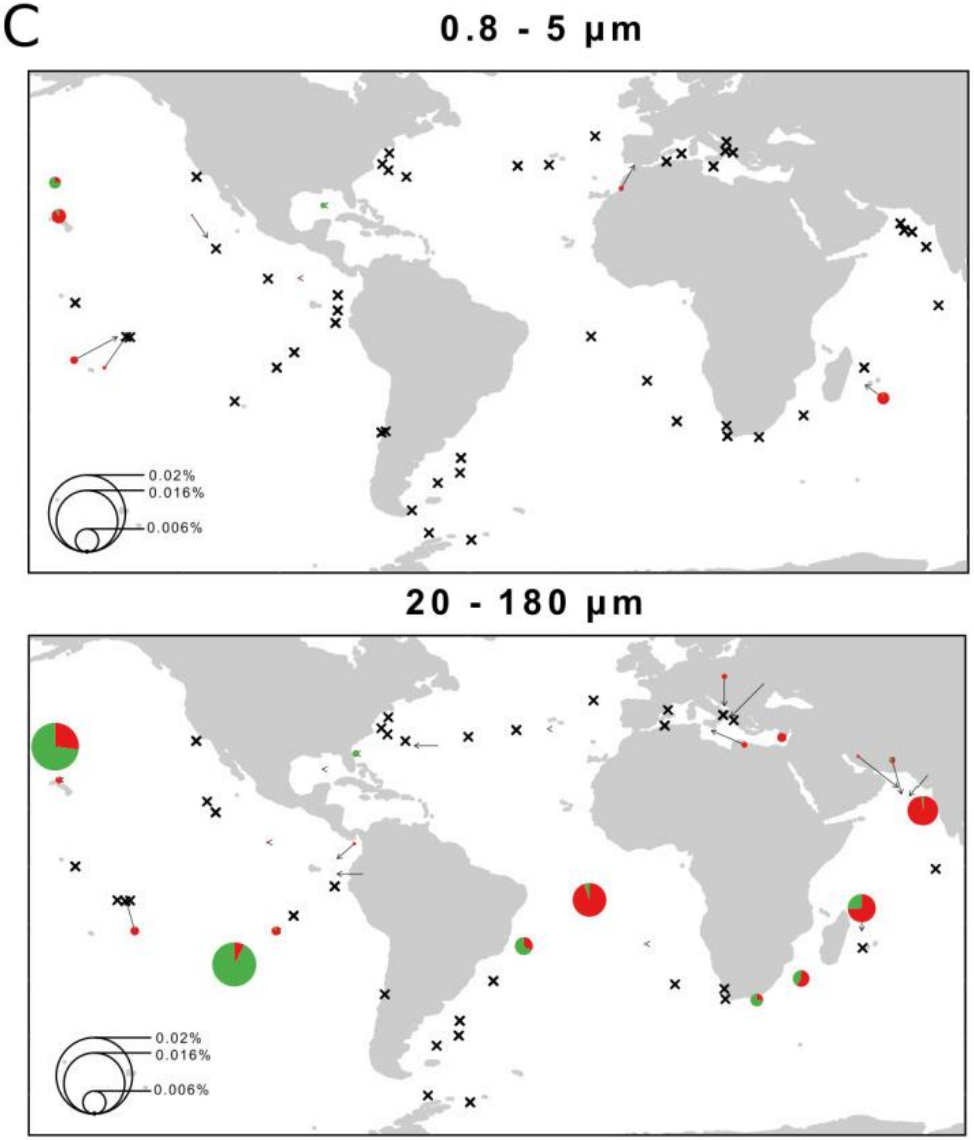




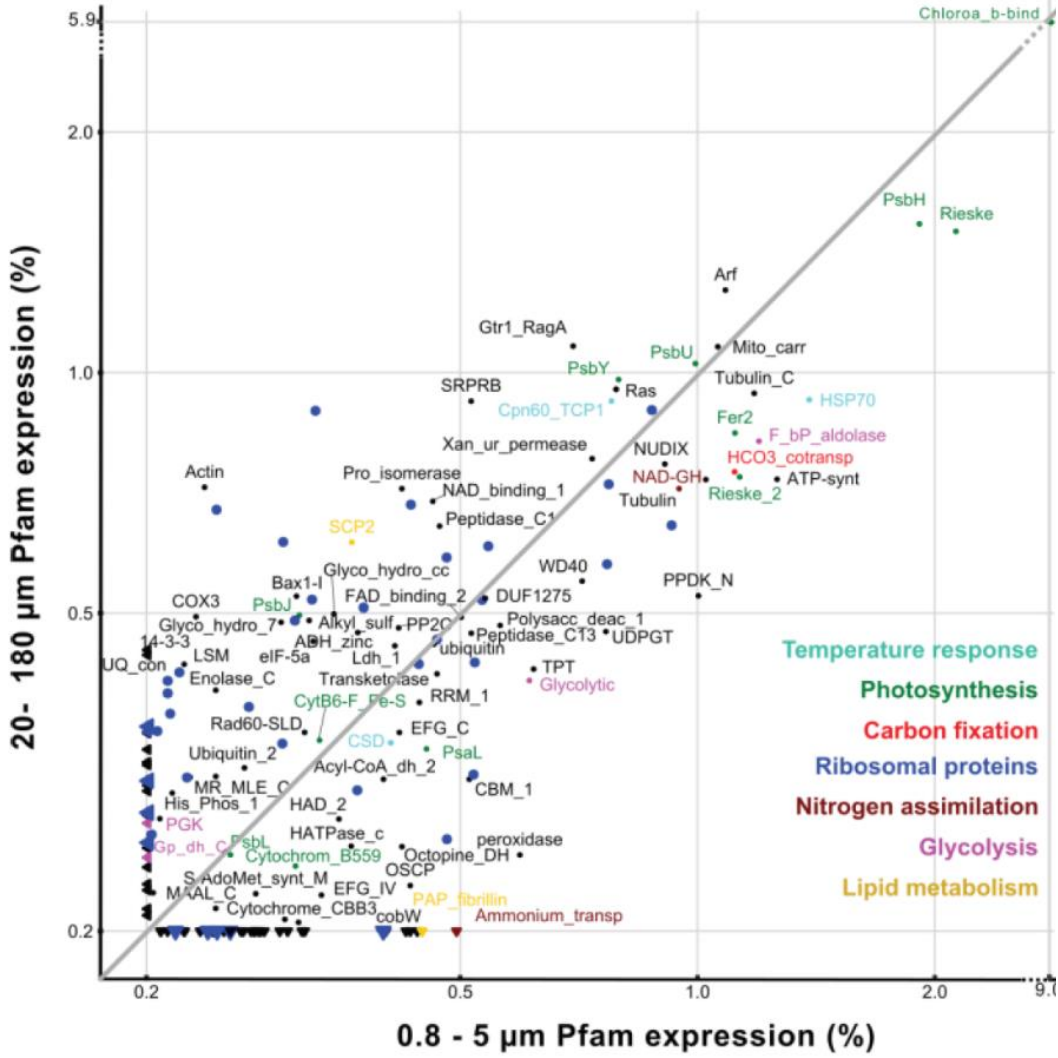
quentin.carradec@genoscope.cns.fr

Most expressed functions in large vs small size-fraction

C



Symbiodinium clade C



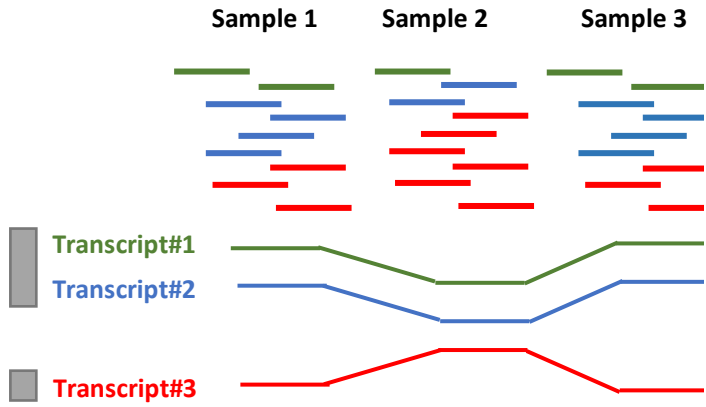
bev Eric Pelletier

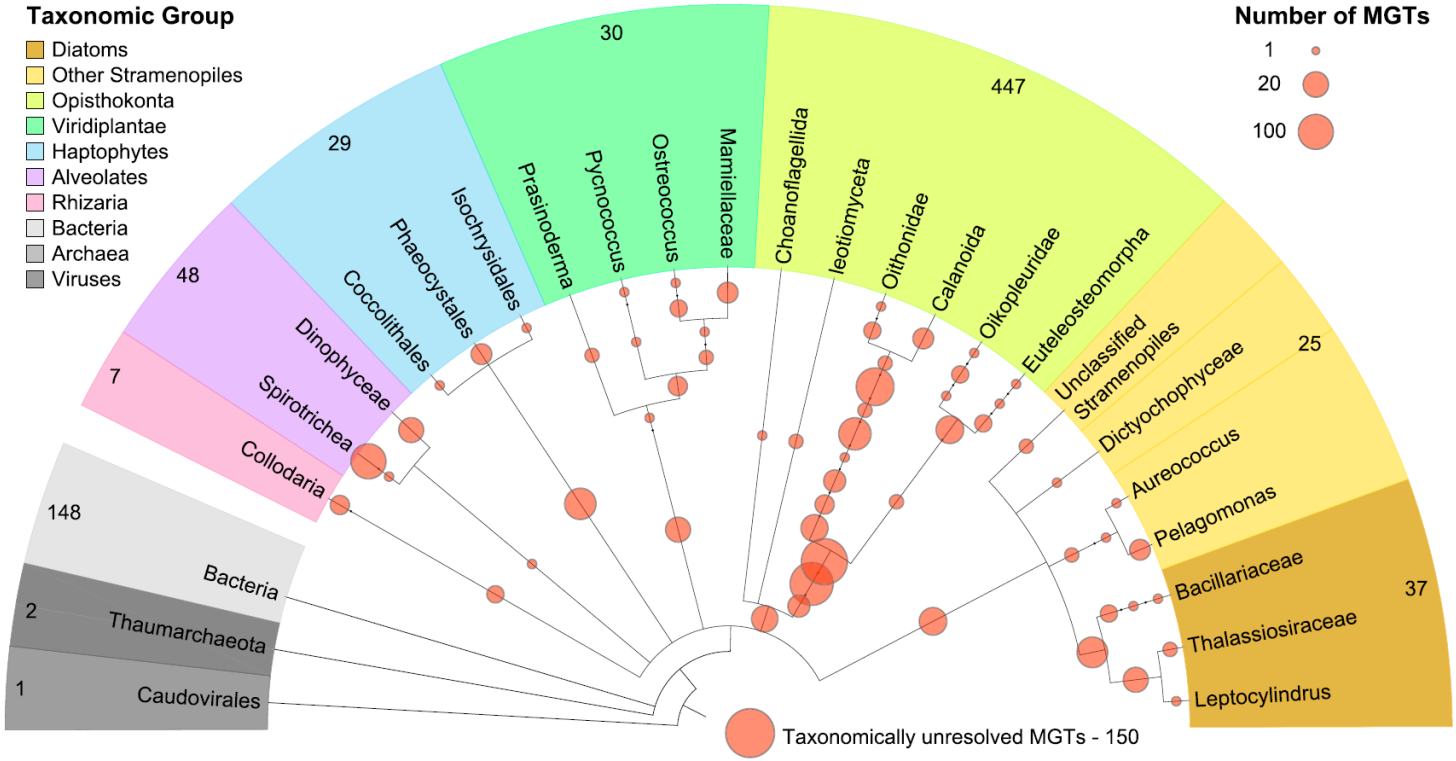
From transcripts to transcriptomes (MGTs)

117 M transcripts

37 M transcripts detected in at least 3 marine stations

7 M transcripts clustered in 924 MGTs (canopy clustering)





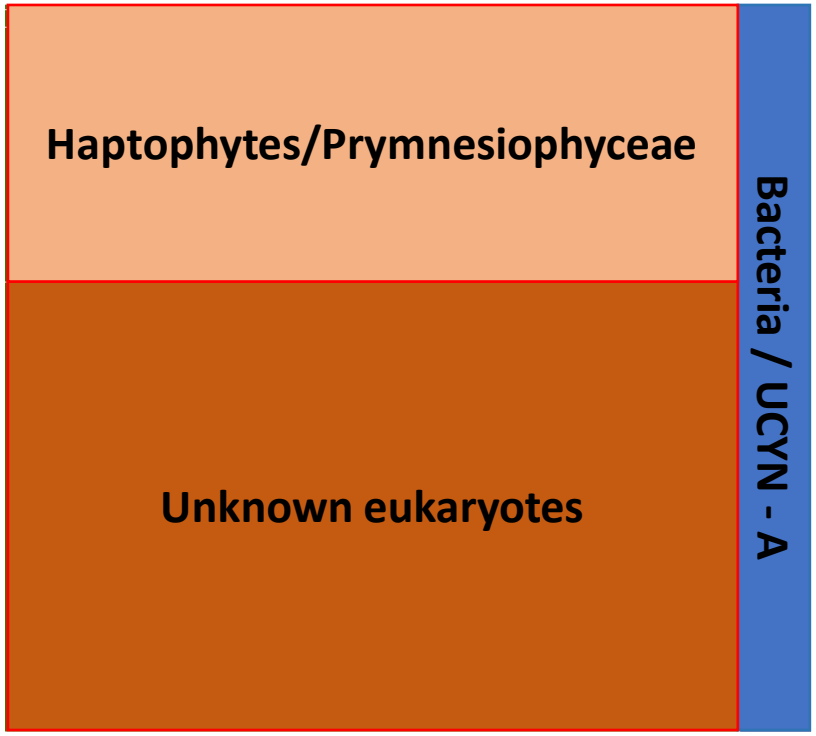
et al 2020 Genome Research

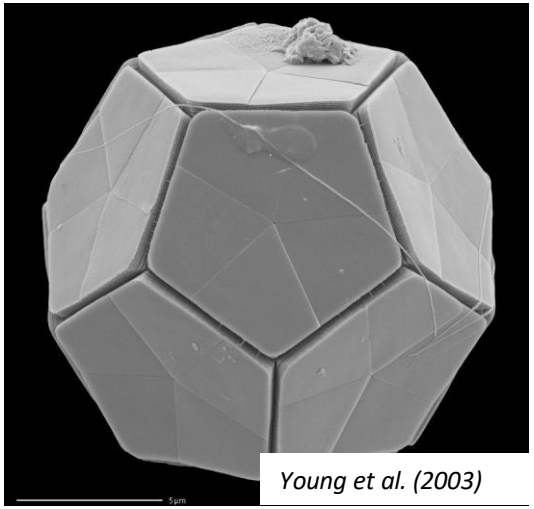
bev Eric Pelletier

From transcripts to transcriptomes

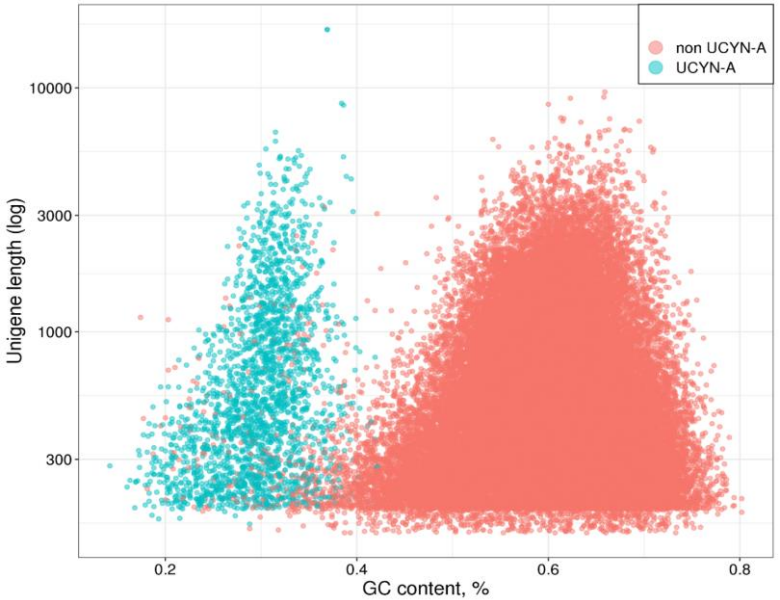
MGT#29 : 48 292 transcripts

24 942 transcripts with a taxonomic affiliations





Braarudosphaera bigelowii (Haptophyte)

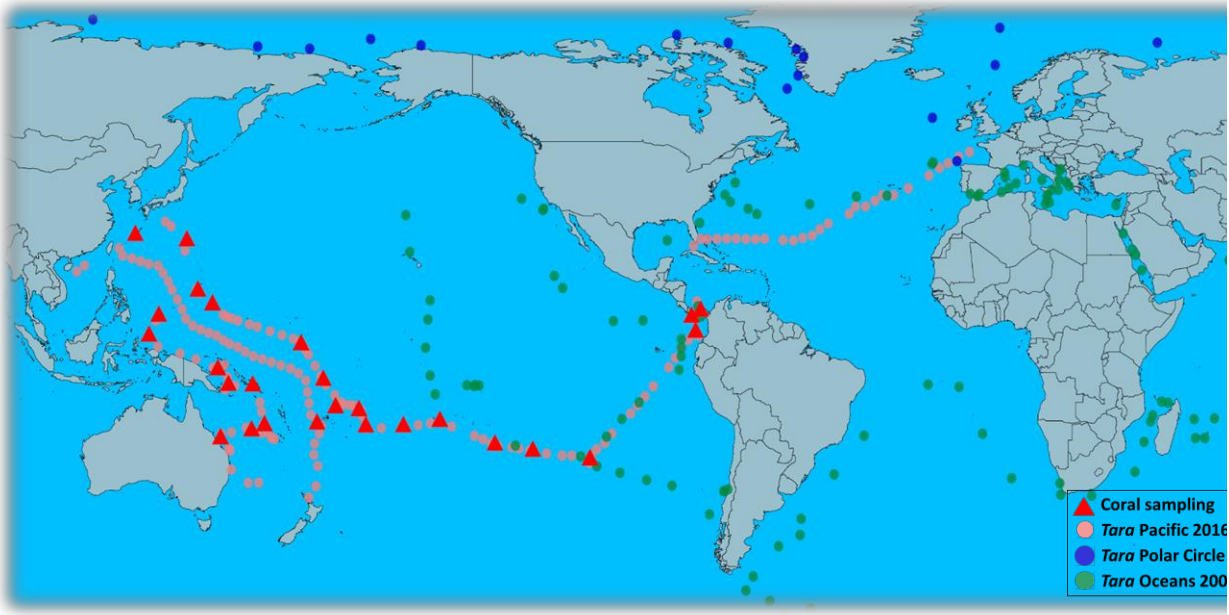


Near complete transcriptome of UCYN-A host

et al 2020 Genome Research

Catalog of transcripts version 2

	V1	V2
Marine stations	68	~200
Samples	441	1 645
Contigs	300 M	1 140 M
Nanopore RNAseq	0	147 M reads



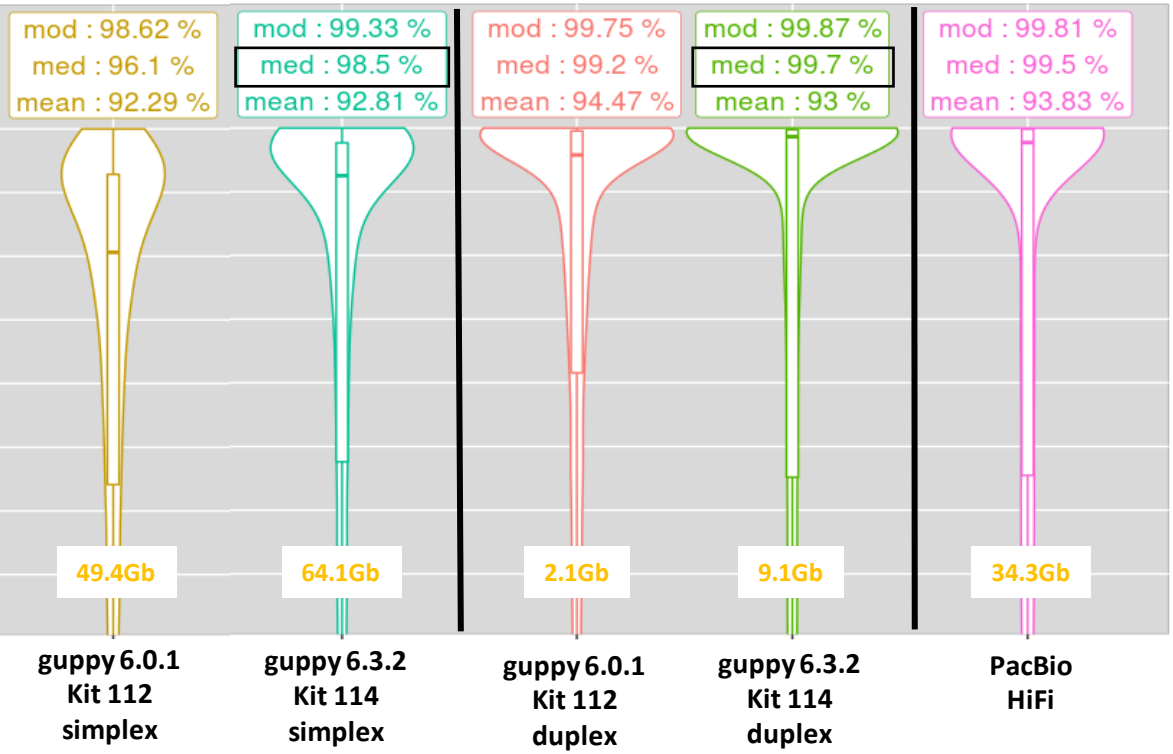
➡ Adaptation of analysis pipelines to

- the increasing number of samples
- the sequencing technology

Da Silva *Jean-Marc Aury*

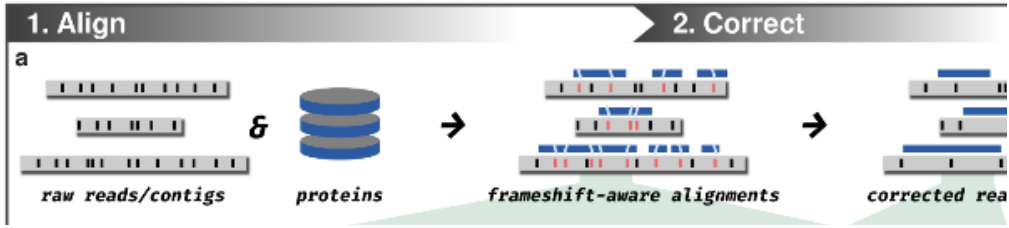
Catalog of transcripts version 2

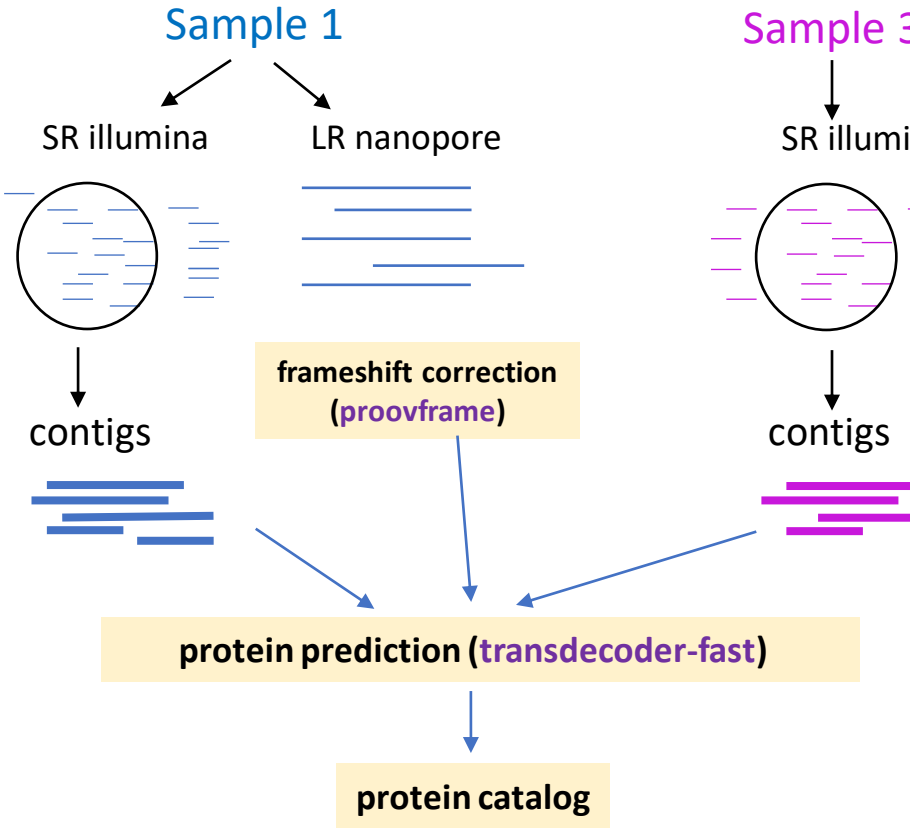
Nanopore sequencing errors



< 1% error

- **Frame-shift correction with proovframe** (*Hackl T et al 2021*)





Take home messages

- Very large diversity of eukaryotic genes in the oceans and more than 50% of them are unknown.
- Potential new biological functions to be explored in plankton communities.
- Various strategies adaptation and acclimation strategies across planktonic lineages.
- Efficient tool to study specific organisms or specific functions
- New version of gene catalogs will complete existing datasets and improve sequence quality.

Acknowledgments

Genoscope teams :

Laboratoire d'analyse génomiques des eucaryotes Patrick Wincker

Eric Pelletier
Betina Porcel
Olivier Jaillon
Tom Delmont
Morgan Gaia
Ardien Thurotte
Janaina Rigonato
Julie Poulain
Paul Frémont
Marie Burel
Sophie Mangenot
Nina Guérin
Lucas Pavlovic
Margot Credeville
Clément Leboine

Laboratoire de Bioinformatique pour la Génomique et la Biodiversité Jean-Marc Aury

Laboratoire de séquençage Pedro Oliveira

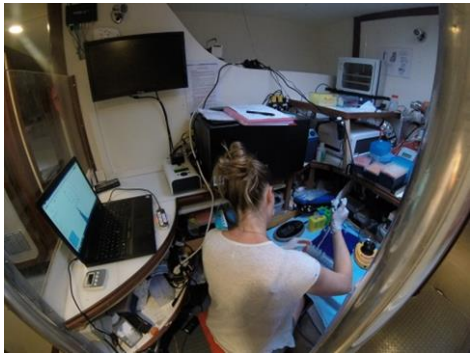
Thank you for your attention

***Tara* Oceans consortium**

Species diversity monitoring with the Nanopore technology

quentin.carradec@genoscope.cns.fr

Nanopore team in the Genoscope



Julie Poulain
Corinne Cruaud
Thomas Guerin
Emilie Payen
Karine Labbadie

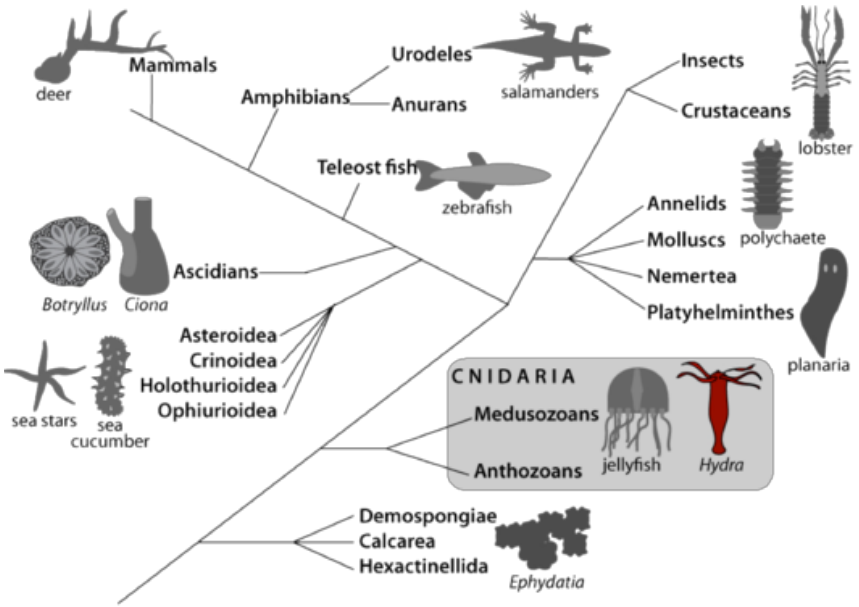
Sophie Mangenot
Lucie Cartairade
Jean – Marc Aury
Stefan Engelen
Benjamin Istace

Sequencing and analysis on the field

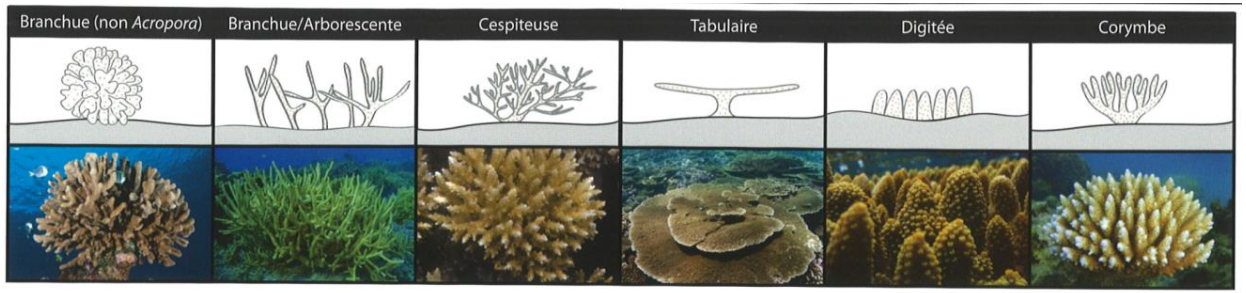
Nanopore sequencing of coral holobionts
Carradec Q, Poulain J et al, Scientific Reports 2020

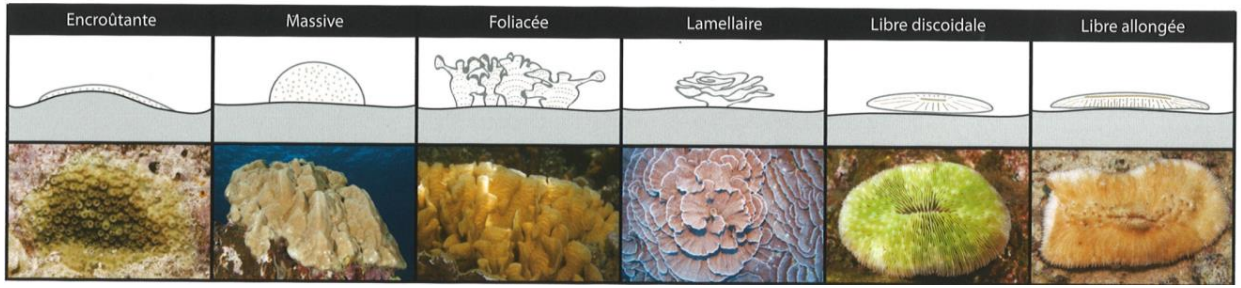
Fish diversity in French Polynesia (eDNA)
Poulain J, Cartaride L, Mangenot S Ongoing work

Coral diversity



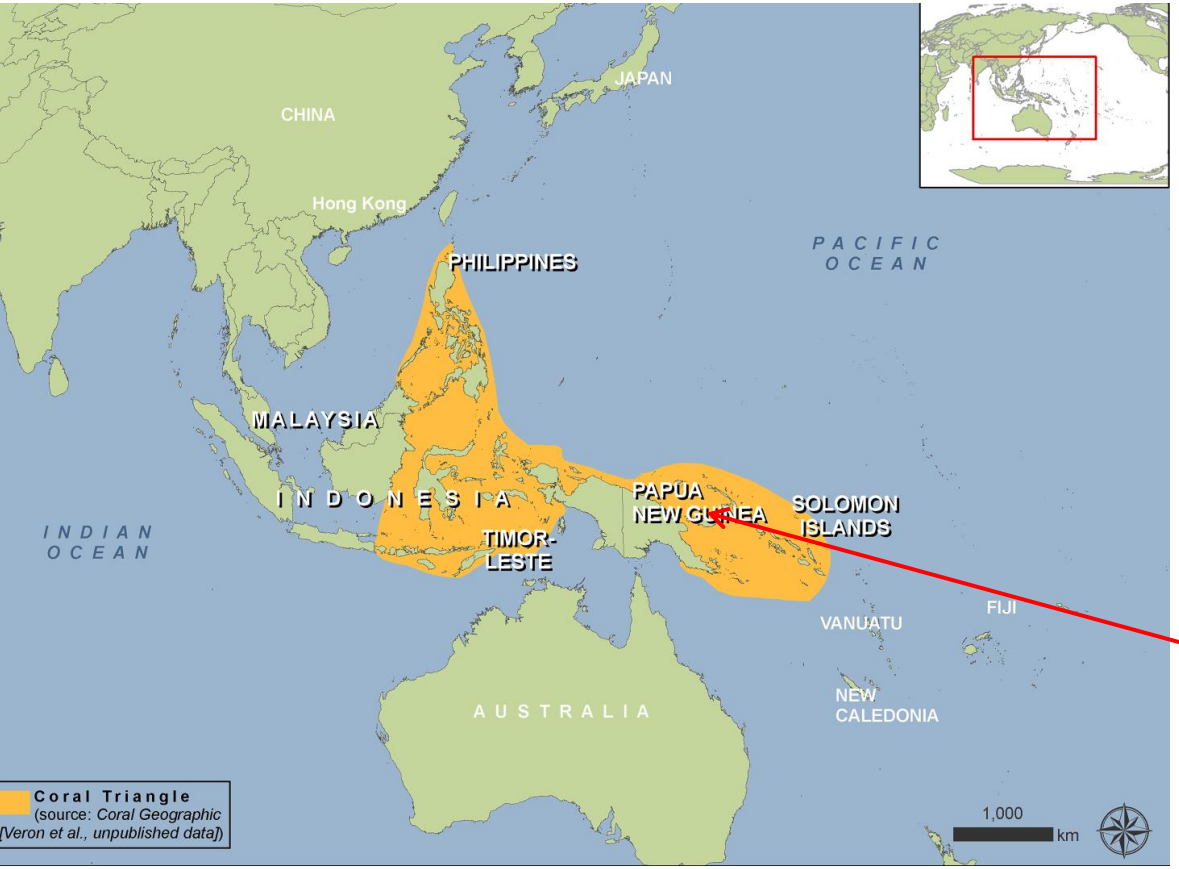
Phylum	Class	Subclass	Order
Cnidaria	Hydrozoa		• Hydroidea (hydroids) • Milleporina (fire corals) • Stylasterina
	Anthozoa	Octocorallia	• Helioporacea (blue corals) • Alcyonacea (soft corals) • Pennatulacea (sea pens)
		Hexacorallia	• Actiniaria (simple sea anemones) • Zoanthidia (colonial anemones) • Corallimorpharia (corallimorpharian • Scleratinia (true stony corals) • Rugosa (Paleozoic coral – now extinc • Tabulata (Paleozoic corals – now exti
		Ceriantipatharia	• Antipatharia (black corals) • Ceriantharia (tube anemones)





About 1500 species

Triangle of coral diversity

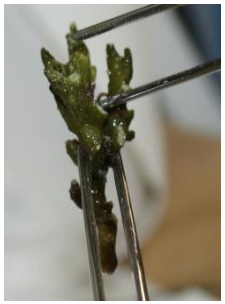


50% of total coral diversity

Sampling, DNA extraction and PCR

coral colonies

Sampling



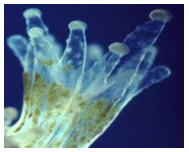
Grinding

Teralyser instrument

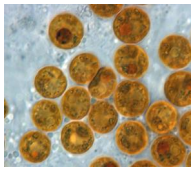
DNA extractio

PCR on three marker genes

Coral



Symbio-dinium



Bacteria

18S rRNA (1.8Kb)
250 ng

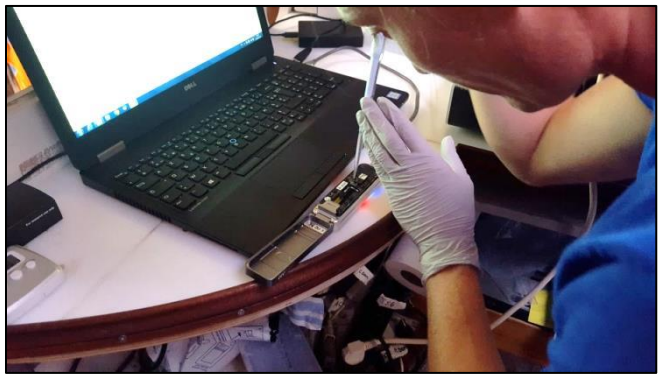
ITS2 sequence (0,3Kb)
2.5 ng

16S rRNA (1.4kb)
250 ng

2nd PCR
(addition of WGP)

Pool of the 3 amplicons

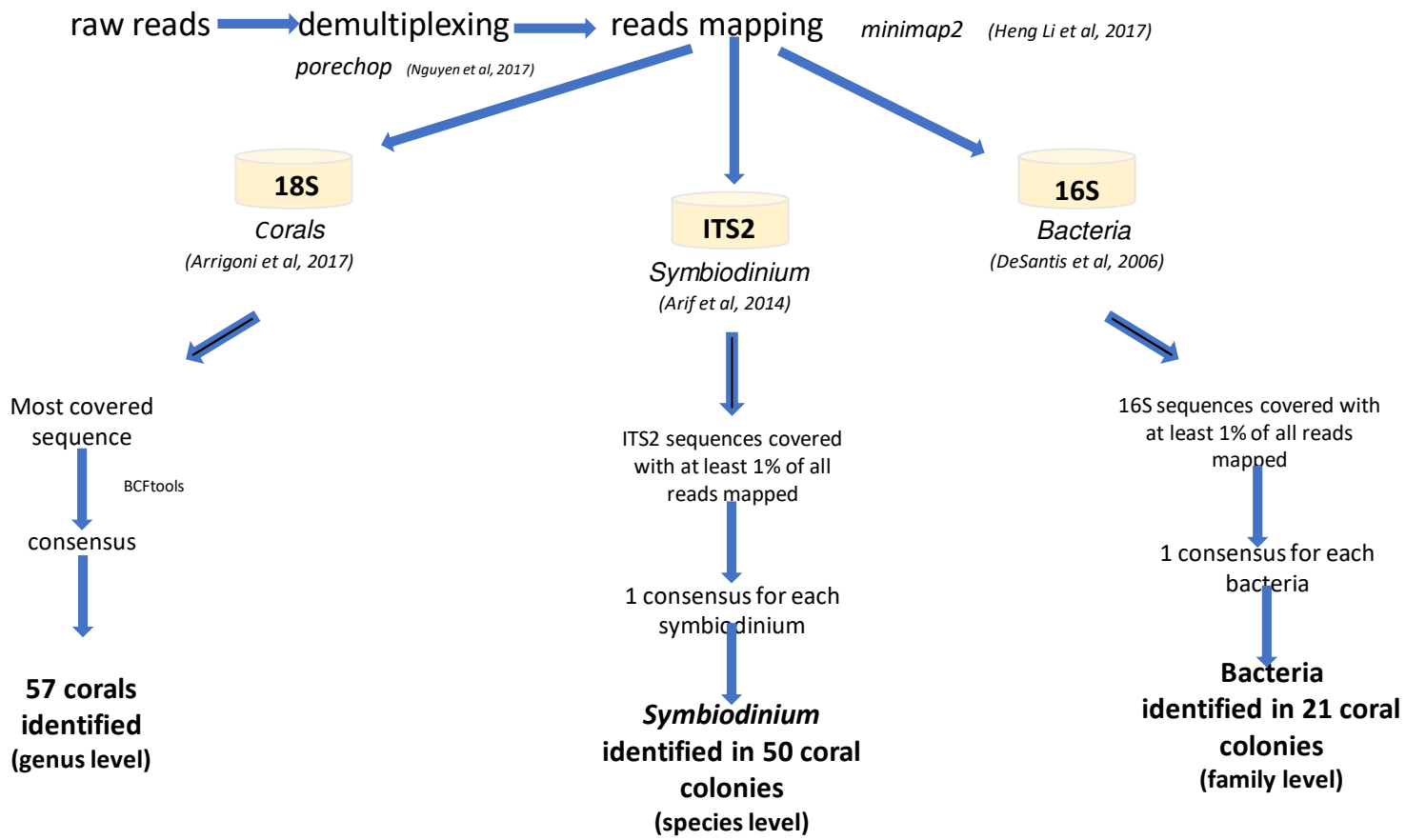
ONT library construction and sequencing



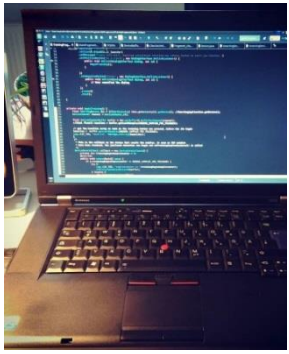
→ About 5h for 12 coral colonies

mean of 400 000 reads per run

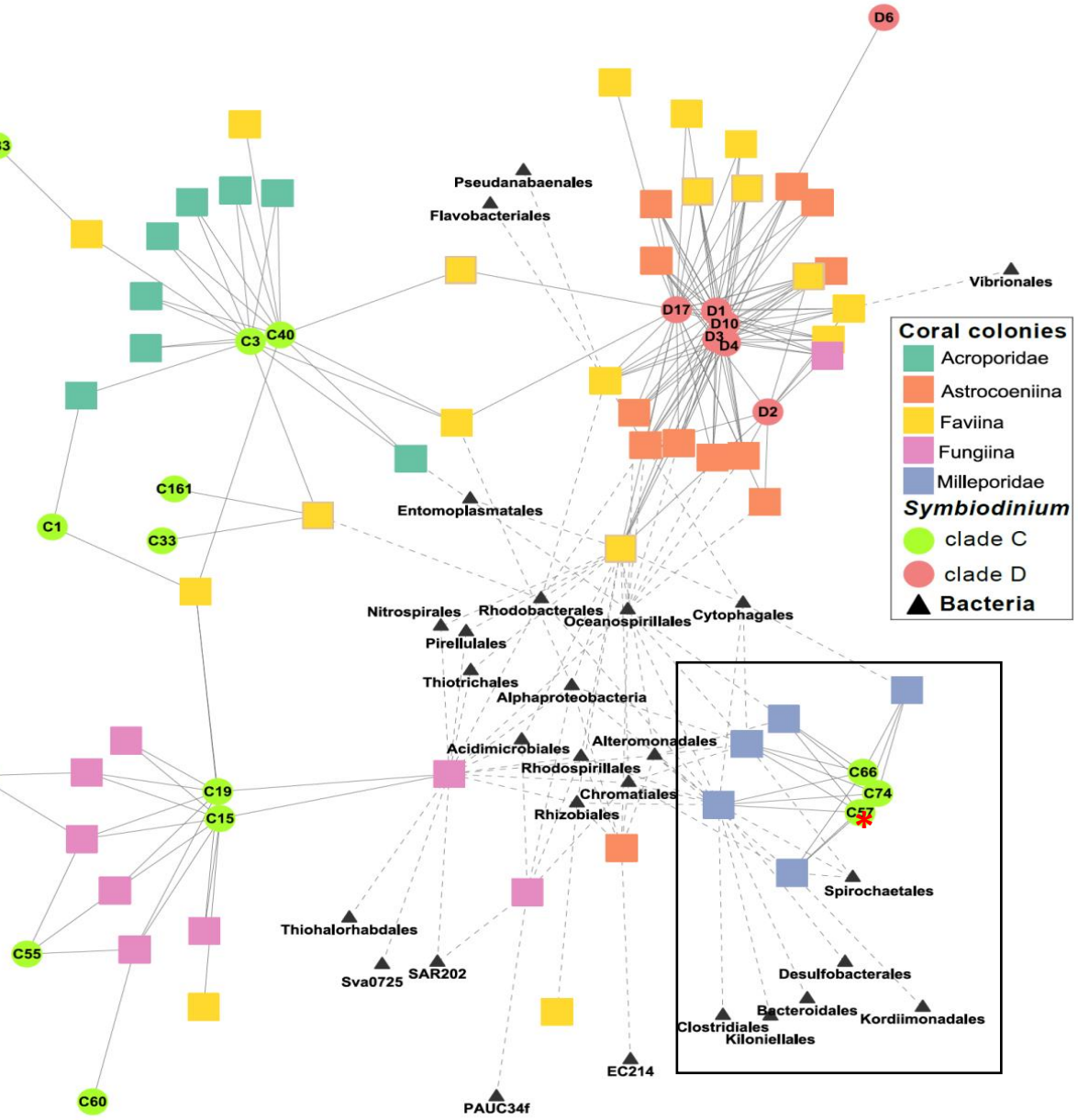
Bioinformatic pipeline adapted for a laptop



2h for one sequencing run on a laptop



Coral holobiont network



- **Most of *Symbiodinium* species are specific to one coral family.**
- **Most of bacteria are shared between several coral families.**

Fish diversity survey in French Polynesia (eDNA)

Moorea island

Serge Planes

Lucie Cartaraide

Julie Poulain

CRIOBE

École Pratique des Hautes Études

quentin.carradec@genoscope.cns.fr

In situ rapid evaluation of fish abundance and diversity using eDNA approaches

> 300 reef fish species

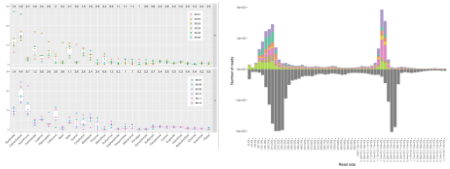
>0.2 µm
or
>3 µm

DNA extraction

mitochondrial 12S

PCR

MinIon
sequencing

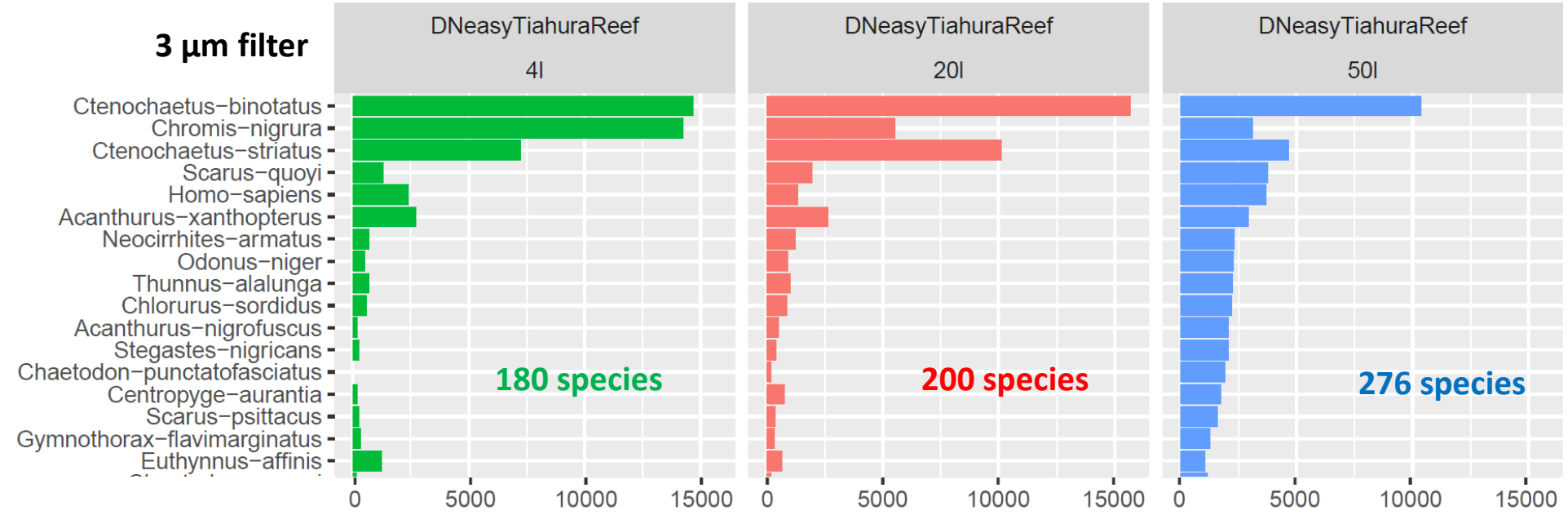


Estimation of fish diversity and their relative abundance

quentin.carradec@genoscope.cns.fr

Fish species detected with 12S primers

-**Database:** NCBI 12S sequences of vertebrates = 79 065 sequences
-**minimap2** mapping of nanopore reads on the database
-**filtering:** minimum coverage of 10 reads; at least 0.1% of all reads mapped

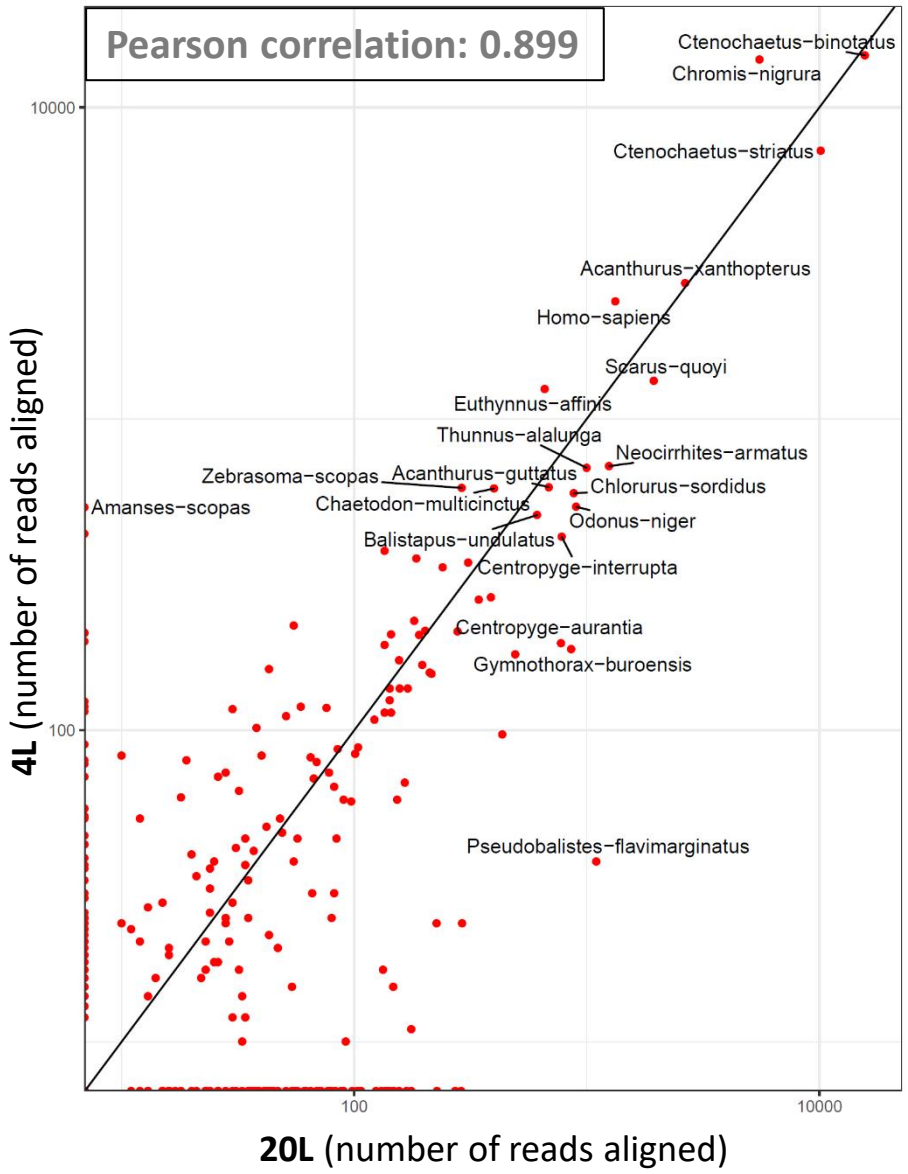


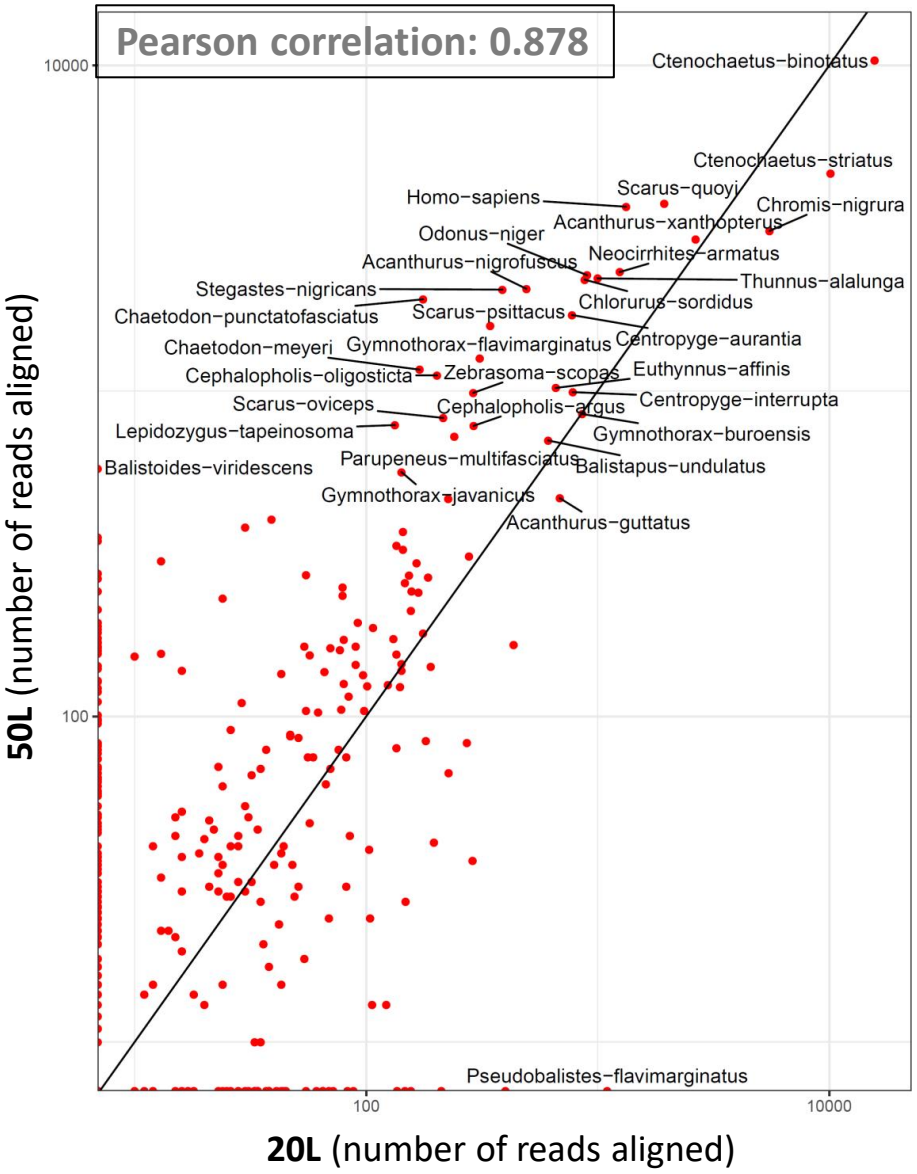
Ctenochaetus binotatus *Chromis nigrua* *Ctenochaetus striatus* *Scarus quoyi* *Acanthurus xanthopterus* *Neocirrhites armatus*

➡ Most abundant fish are detected whatever the volume of water collected.

➡ Relative proportions are variable between different experiments.

Correlation of fish abundances between variable volume of water





Strong correlation between 4L, 20L and 50L of water

Correlation between the abundance estimated with eDNA and fish observation.

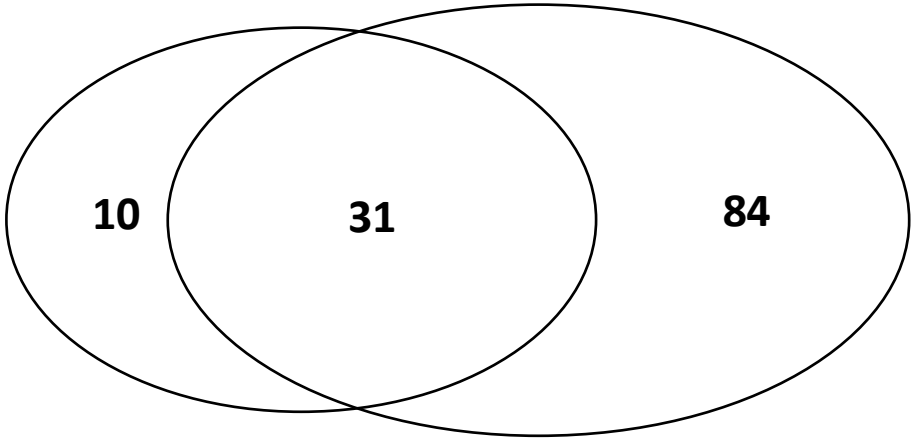
Gilles Siu

1 dive of about 1h.

-68 fish species
-Abundance (889 individuals)
-Approximative size

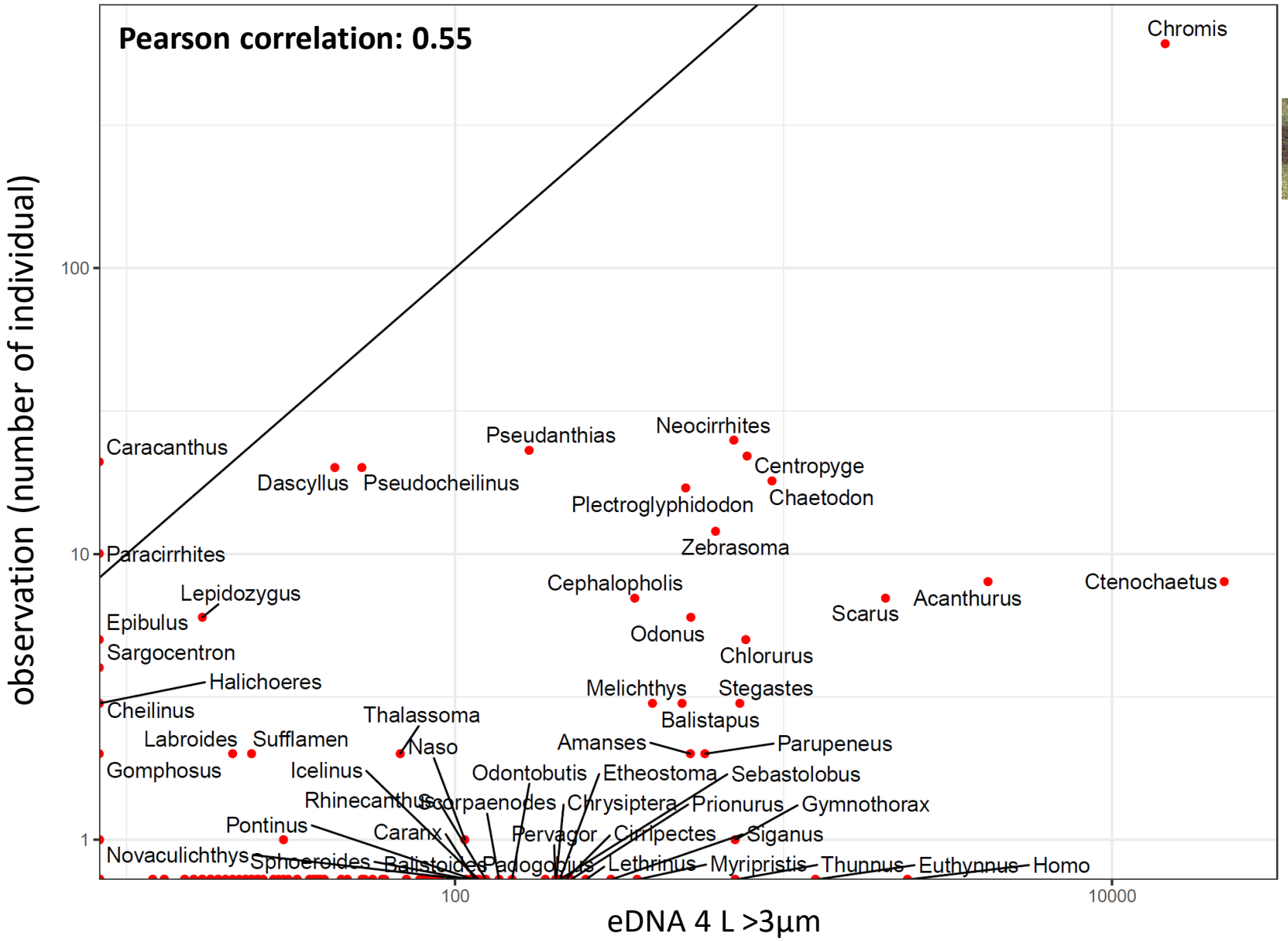
Gilles observation
41 genus

eDNA
115 genus



Number of genus detected

Correlation between eDNA and fish observation



Chromis nigrua

Take home messages

MinIon technology can be used to explore the diversity *in situ* in different environments.

New bio-informatic tools are needed to take into account Nanopore specificities.

This technology progress very quickly and could be used more widely in the future for long read sequencing capacities.