# Ressources informatiques et bioinformatiques pour le traitement d'un jeu de données de génomique environnementale

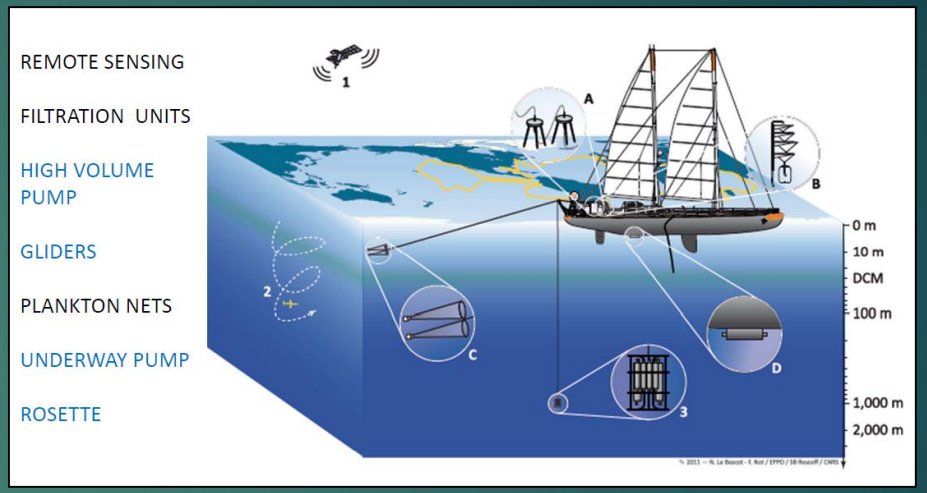

Magali Lescot, Pascal Hingamp

IGS - CNRS Marseille

# Plan

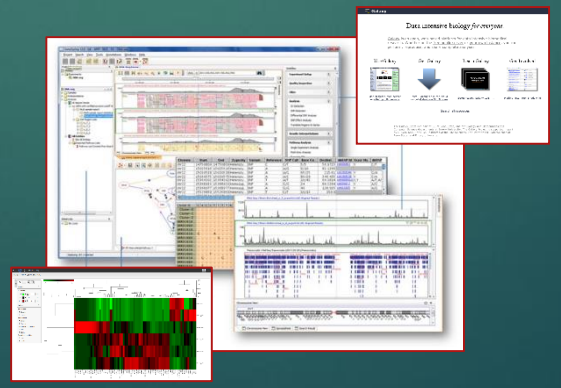
1) L'expédition *Tara* OCEANS

2) La plateforme scientifique *Tara* OCEANS

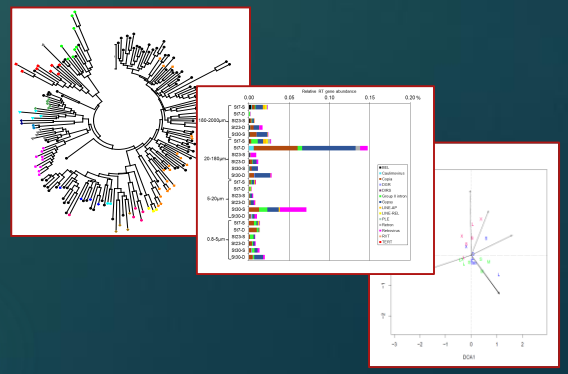



3) Ressources informatiques

5) Ressources bioinformatiques



6) Un cas d'étude



# 1 – L'expédition *Tara* OCEANS

# Cartes de l'expédition

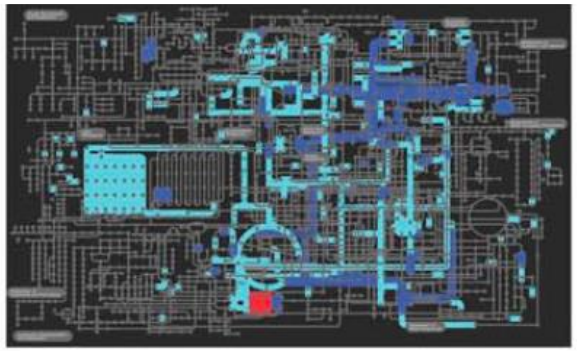

**Tara Oceans / Polar circle 2009-2013**
Effet du réchauffement planétaire sur les systèmes planctoniques et coralliens

- 938 jours en mer
- 250 membres d'équipage
- 140 scientifiques pluridisciplinaires de 40 nationalités
- 210 stations


25 000 km en six mois

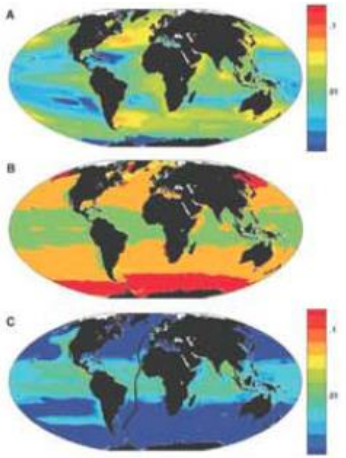
# Holistique pourquoi faire ?

Réseau métabolique du plancton
→ 50% $O_2$, séquestration $CO_2$

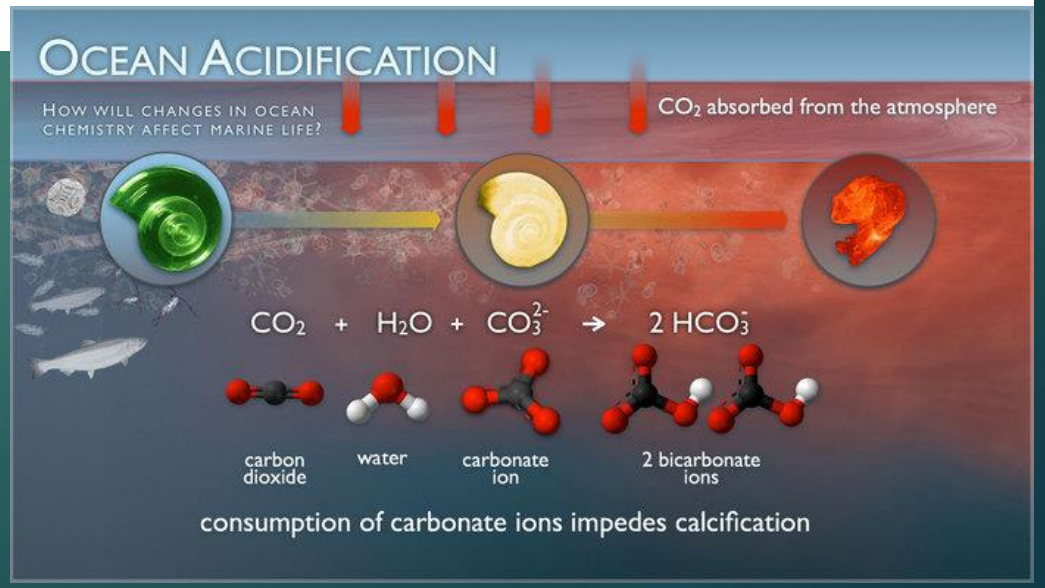


Réseau d'interactions entre espèces de plancton

Modèlisation de l'écosystème



→ Anticiper les changements :
ex. $CO_2$ anthropogenique



# 2 – La plateforme scientifique *Tara* OCEANS

# *Tara* platforme scientifique

**3 – Labo “ sec ”**

**4 – Labos “ terrestres ”**

**2 – Labo “ humide ”**

**1 – Instruments océanographiques**

# A – Instruments océanographiques

REMOTE SENSING

FILTRATION UNITS

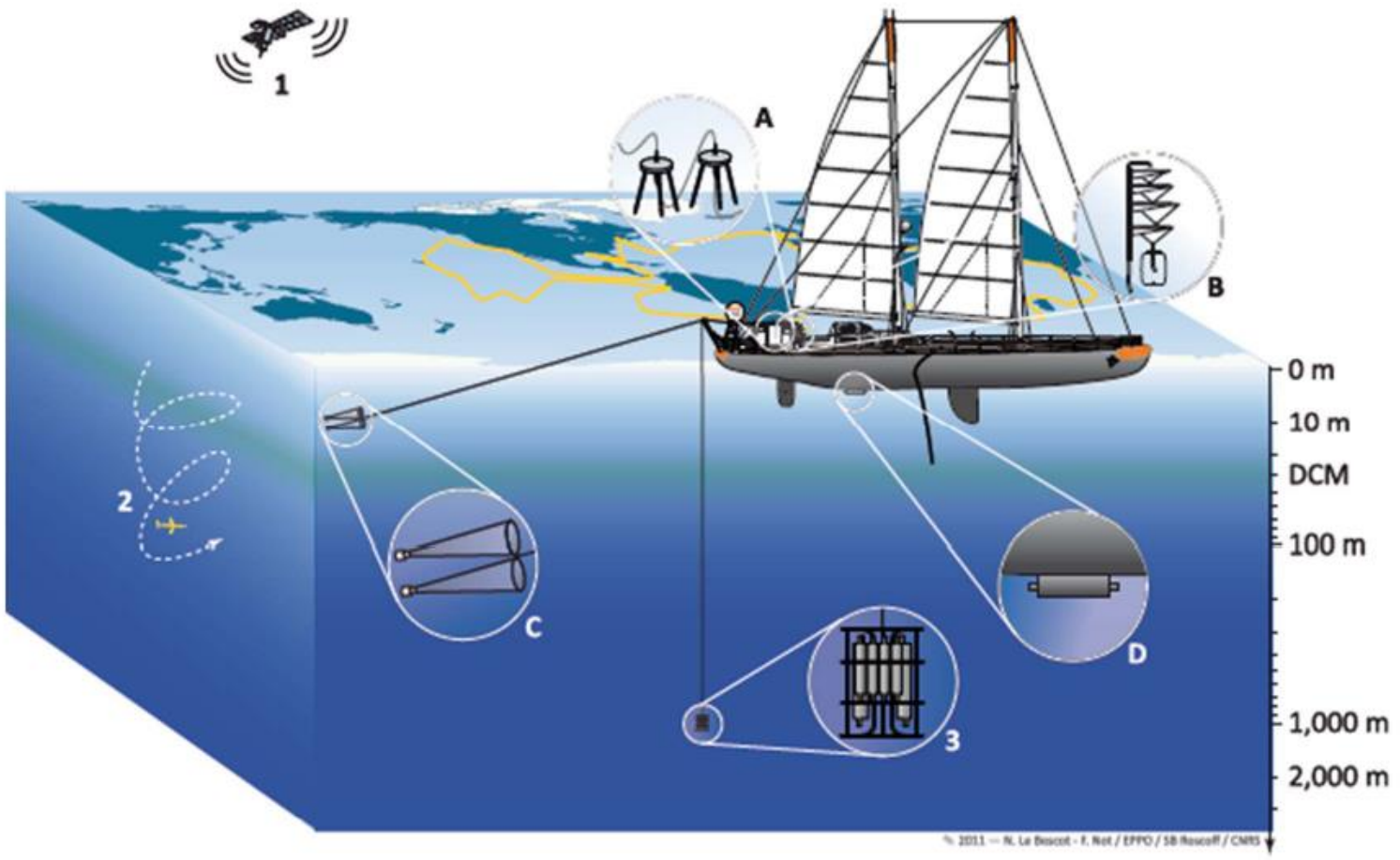
HIGH VOLUME PUMP

GLIDERS

PLANKTON NETS

UNDERWAY PUMP

ROSETTE



**Pompe péristaltique**
(5-120 m profondeur)

Filets plancton (5-600µm)

**Rosette**
(10 bouteilles Niskin
CTD, 0-2000m prof.)

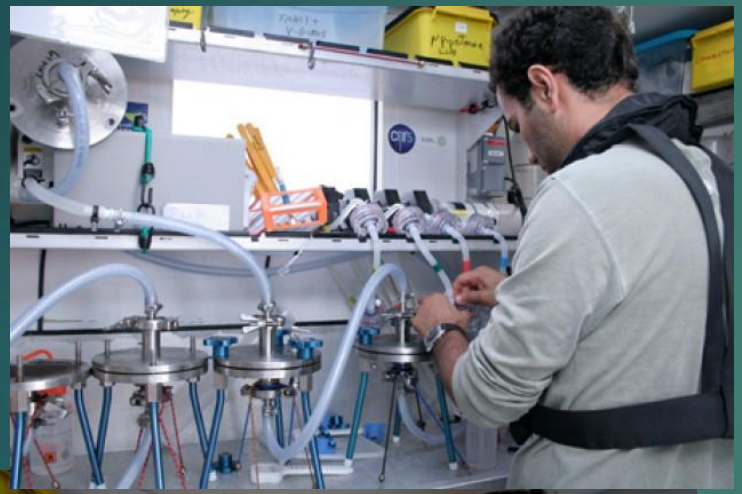
**GPSS**
« Système de filtration
Gravitationnel »

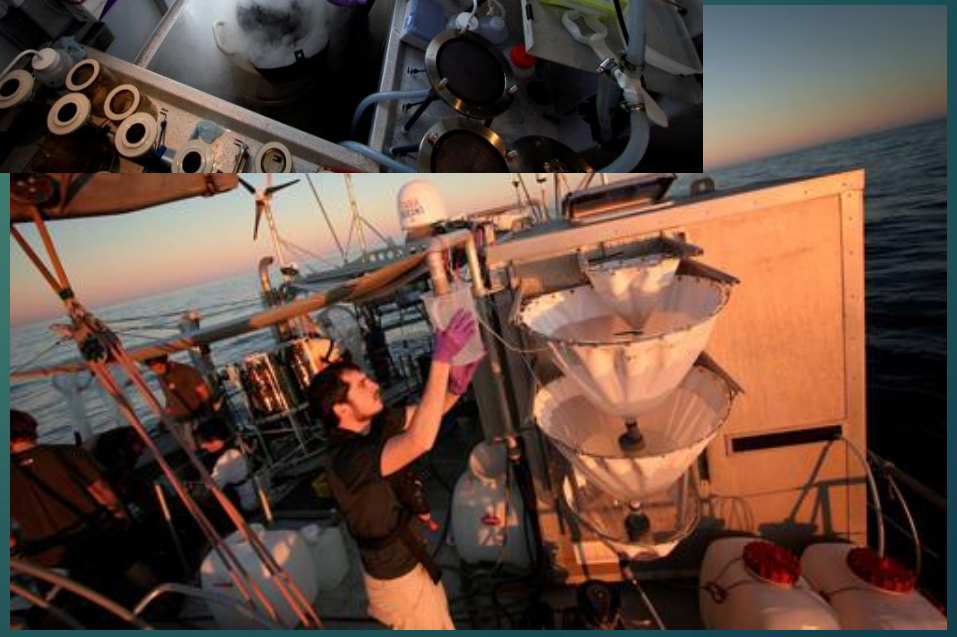
# Paramètres environnementaux

| Environmental Parameters | Rosette | High Volume Pump | Underway Pump | Stand-alone Instruments |
|---|---|---|---|---|
| Lat., Long. position and time | ✓(GPS) | ✓(GPS) | ✓(GPS) | ✓(GPS) |
| Pressure (depth) | ✓(CTD) | ✓(Ecotriplet) | ✓(fixed @ 5m) | ✓(Glider) |
| Conductivity (salinity) | ✓(CTD) | ✓(Ecotriplet) | ✓(TSG) | ✓(Glider) |
| Temperature (water) | ✓(CTD) | ✓(Ecotriplet) | ✓(TSG) | ✓(Glider) |
| Fluorescence (Chl *a* ) | ✓( WET Labs) | ✓( WET Labs) | ✓( WET Labs) | ✓(Glider) |
| Oxygen | ✓(DO) | | | |
| Nitrate | ✓(ISUS) | | | |
| CDOM | ✓(CSTAR) | | | |
| Absorption | ✓(CSTAR) | | ✓(ACs) | |
| Attenuation | ✓(CSTAR) | | ✓(ACs) | |
| Particle size spectrum | ✓(UVP5) | | ✓(ACs) | |
| Photosynthetic activity | | | ✓(FRRF) | |
| Upwelling Radiance | | | | ✓(TSRB) |
| Surface Irradiance | | | | ✓(TSRB) |

# B – Labo « humide »

Rampes de filtration (0.1µm – 5µm)
Fixation des échantillons (formol, éthanol, Azote liquide, -20°C ...)





# C – Labo « sec »

Im Labor unter Deck, mit dem Mikroben-zähler (links) und unterm Mikroskop (rechts), erfassen die Forscher die Zusammen-setzung des Planktons

# D – Labos « terrestres »

En mer

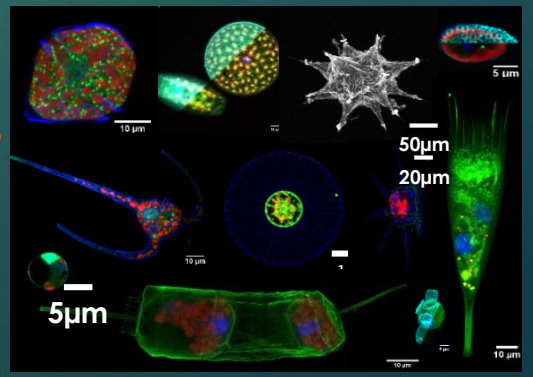
A terre

Echantillonnage

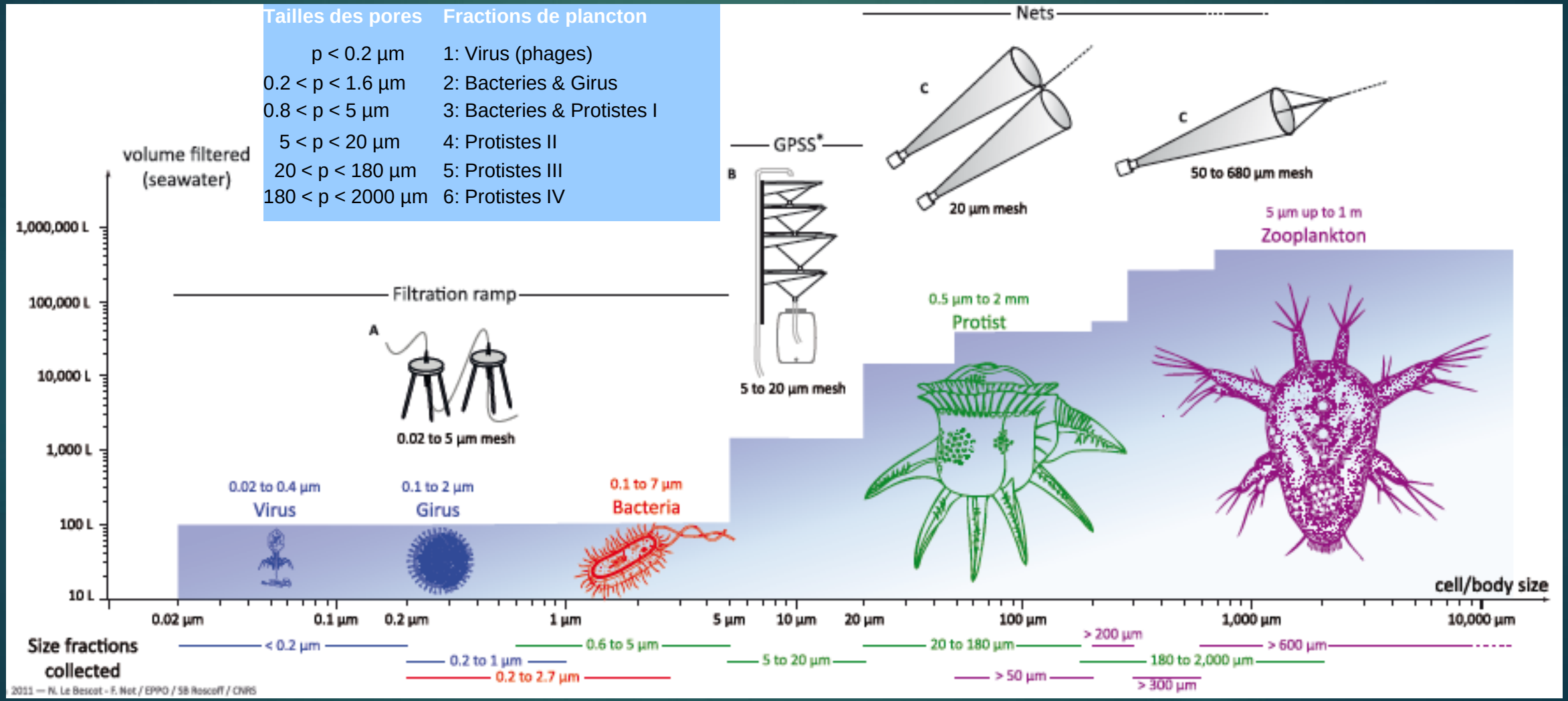
Echantillons fixés

Analyses moléculaires

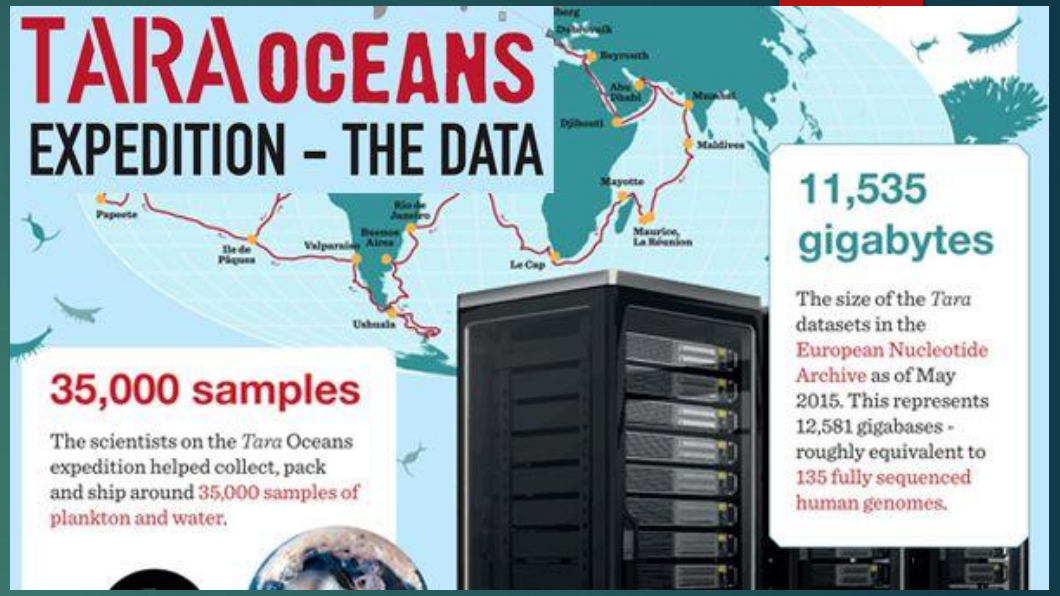

Analyses morphologiques



# Stratégie d'échantillonage



# *Tara* OCEANS (2009-2013)




- 210 stations à travers tous les océans
- 35 000 échantillons collectés
- Un milliard de codes-barres génétiques eucaryotes
- 40 millions de gènes microbiens
- 35 000 espèces de bactéries planctoniques différentes (EMBL)
- ~13,000 mesures contextuelles à 3 profondeurs
  - Métagenomes/ Métabarcodes
  - Bases de données d'images quantitatives et haute résolution
- Données génomiques publiées et validées à l'EMBL-EBI, corrélées avec les données environnementales stockées Pangaea.

# Integration for data sets (L. Stemmann)

**Cell/organism imaging**

Classification of vignettes (taxonomy)

Calculations of biovolume for autotrophs and heterotrophs

**Metagenomics**

Extraction of metabarcodes and OTUs

Taxonomic annotation

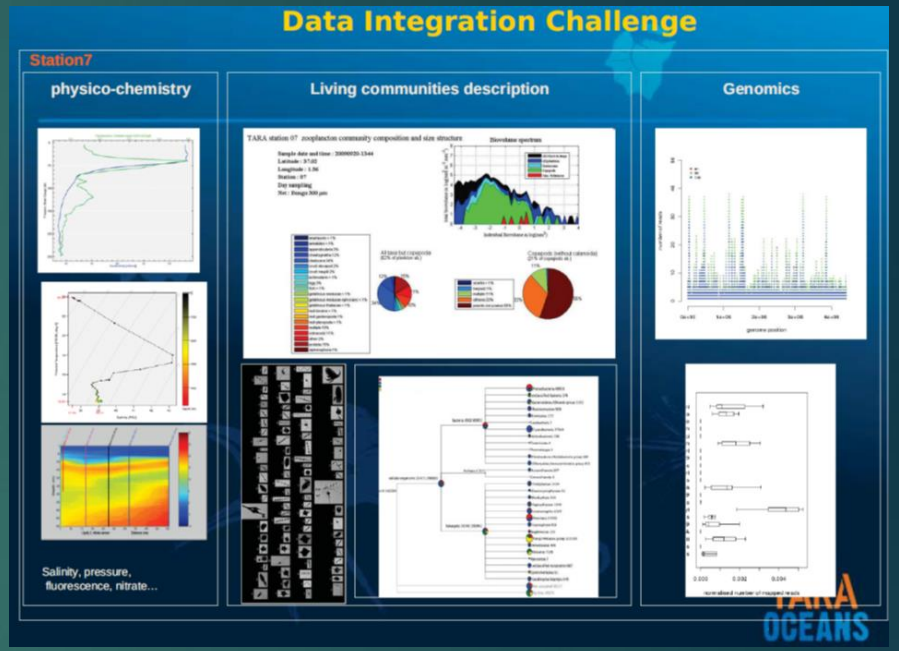
**Environmental variables**

Calculate descriptors

ECOLOGY

- Diversity
- Distribution patterns
- Community structure
  -Co-occurrence

Statistical analyses



# *Tara* OCEANS metagenomic approaches

Less than1% of the marine microorganisms are cultivable

→ which species are present ?

Metabarcodes

Amplify marker genes

Extract ecosystem RNA

→ which functions are active now ?

Metatranscriptome

Cell sort plankton

Single Amplified Genomes (SAGs)

→ full genome for representative uncultured species

Extract ecosystem DNA

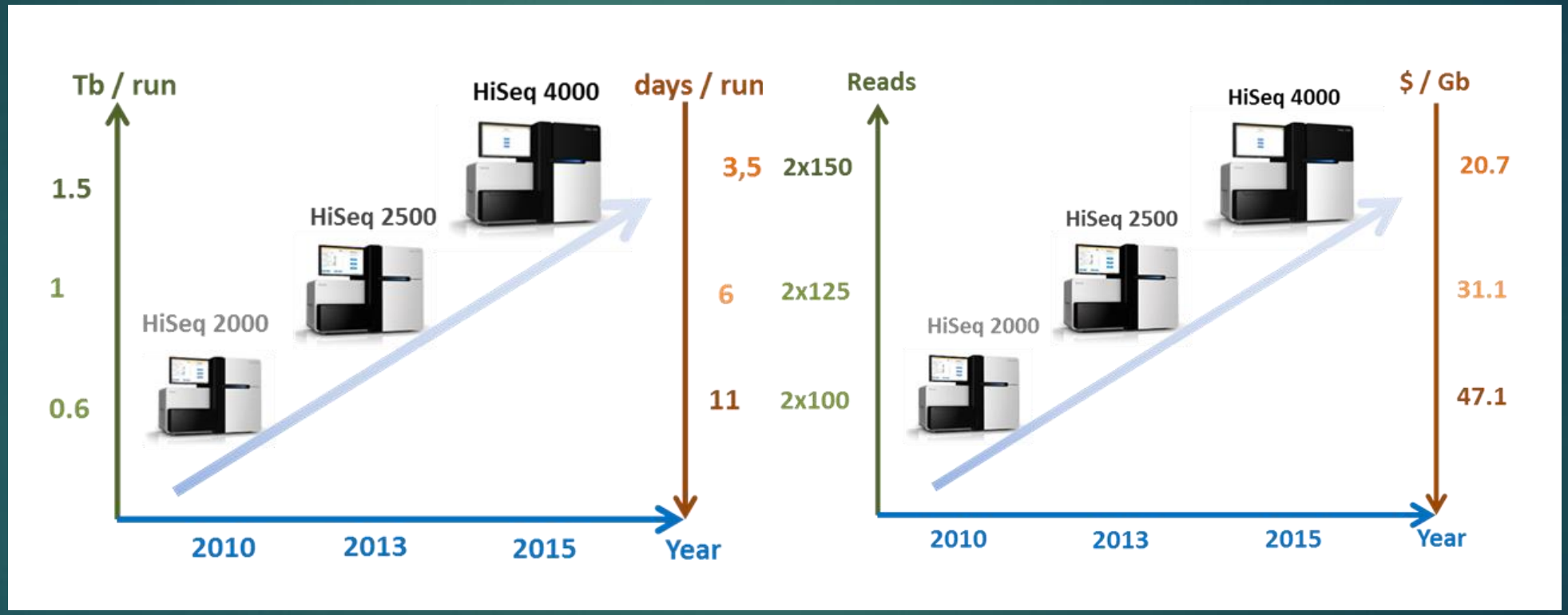
→ what biological functions are present ?

Metagenome

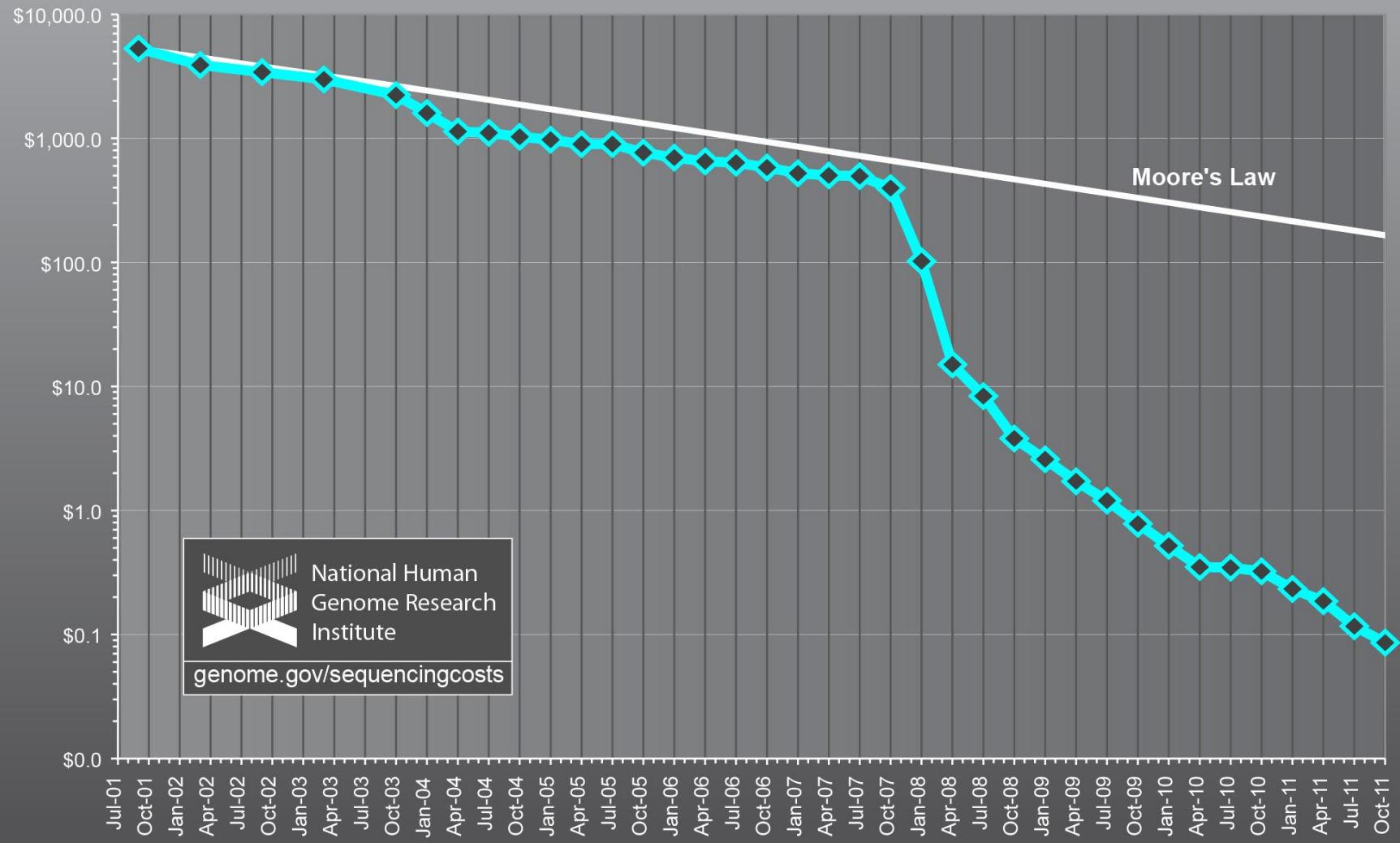
DNA fragments from several genomes in the ecosystem

# 3 – Les ressources informatiques

# Evolution des NGS depuis le début de *Tara* OCEANS



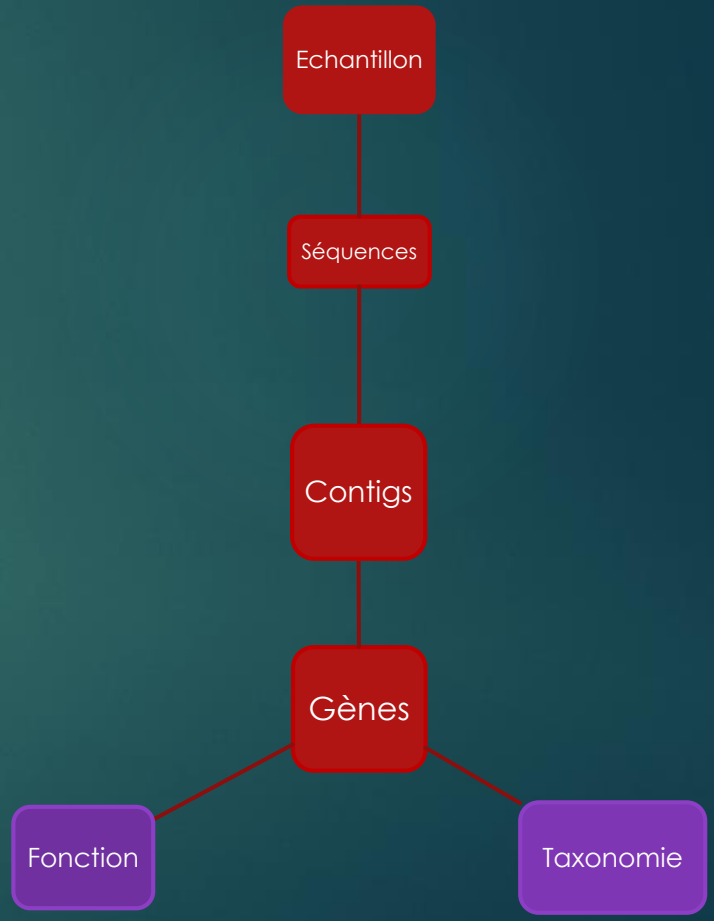
# Cost per Raw Megabase of DNA Sequence



# Principaux points de l'analyse de métadonnées

Quantité considérable de données de séquences !!

- Regrouper les lectures similaires
  - Séquences très fragmentées (efficacité d'assemblage)
- Construire des fragments génomiques
  - Pas de lien avec la cellule d'origine
- Prédire des gènes
- Identifier les espèces
  - Organismes inconnus => pas de génome de référence
- Prédire les fonctions
- Comprendre les interactions



# Global Ocean Microbiome
## (Sunagawa et al, 2015)


243 samples and 68 stations (epipelagic and mesopelagic waters)
7.2 terabases of metagenomic data
72 492 220 288 reads (mean length: 90 nucl.)
> 40 million reference gene catalog (virus, prokaryotes and picoeukaryotes)

Read Trim Filter of raw data (solexaqa/fastx): 20 cpu h / sample => 120 000 cpu.hours
Adapter Screen (USEARCH): 5 cpu h / sample => 30 000 cpu.hours
Assembly (SOAPdenovo)*: 350 cpu h / sample => 2.1M cpu.hours
Map vs Reference DB (SOAPalign)**: 150 cpu h / sample = 900 000 cpu.hours
**Total:** 4.95M cpu.hours

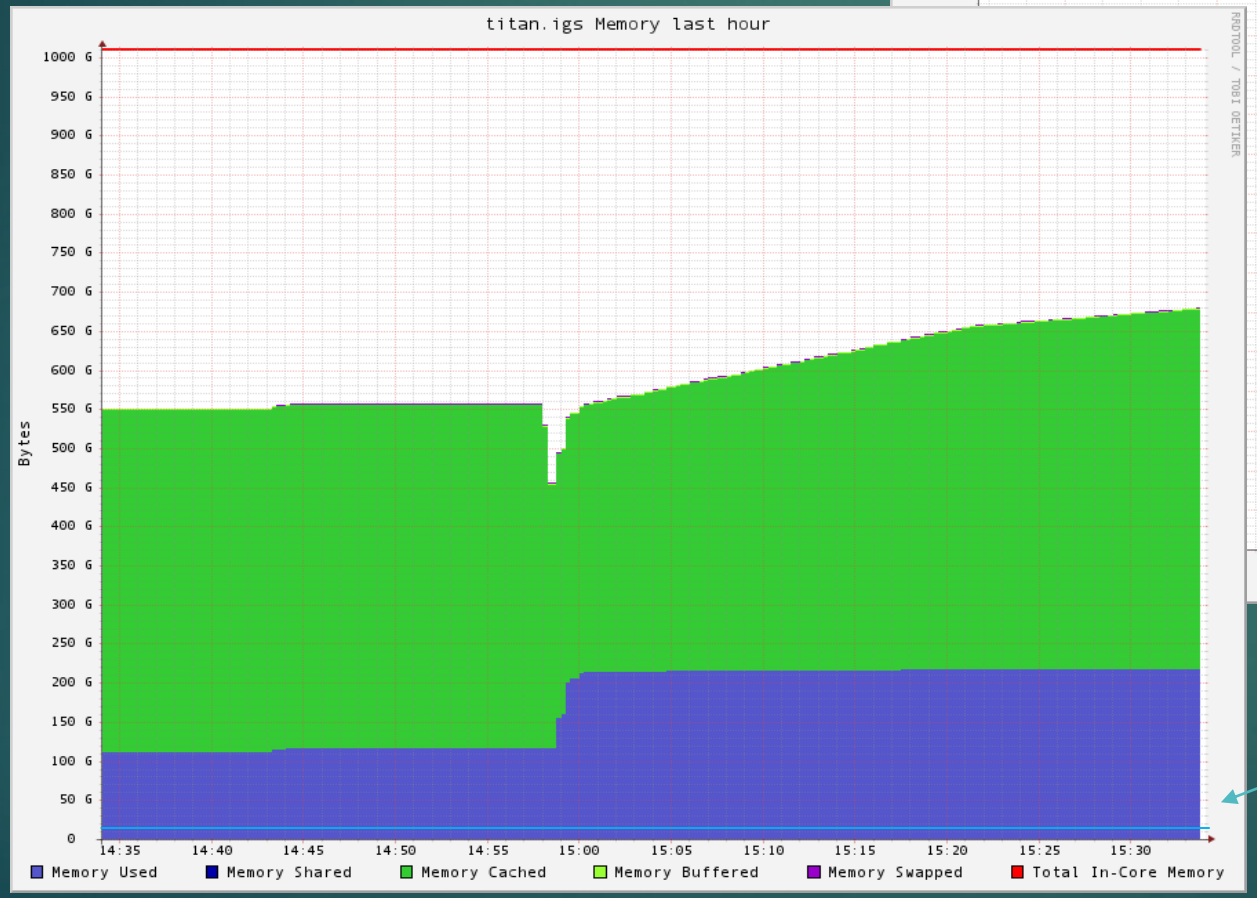
* some samples require hundreds (up to 800) of GB RAM ( >1TB RAM may be required)
** for each DB used for mapping (3 databases used)
cpu h=hours of computation by a single cpu

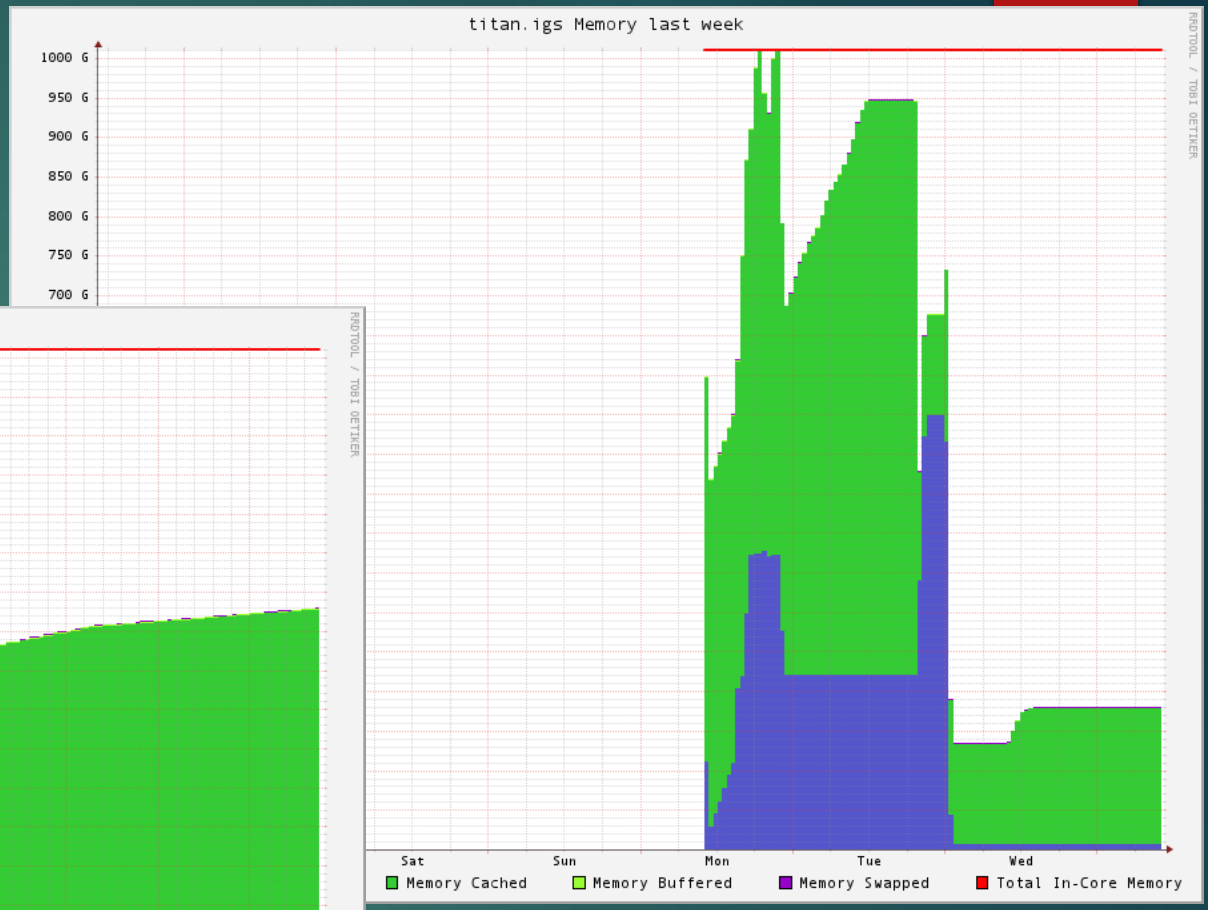
# Ressources en calcul et stockage @IGS

- 624 Cores
- 3348 Go RAM
- 98 To de stockage de données

# Projet permafrost

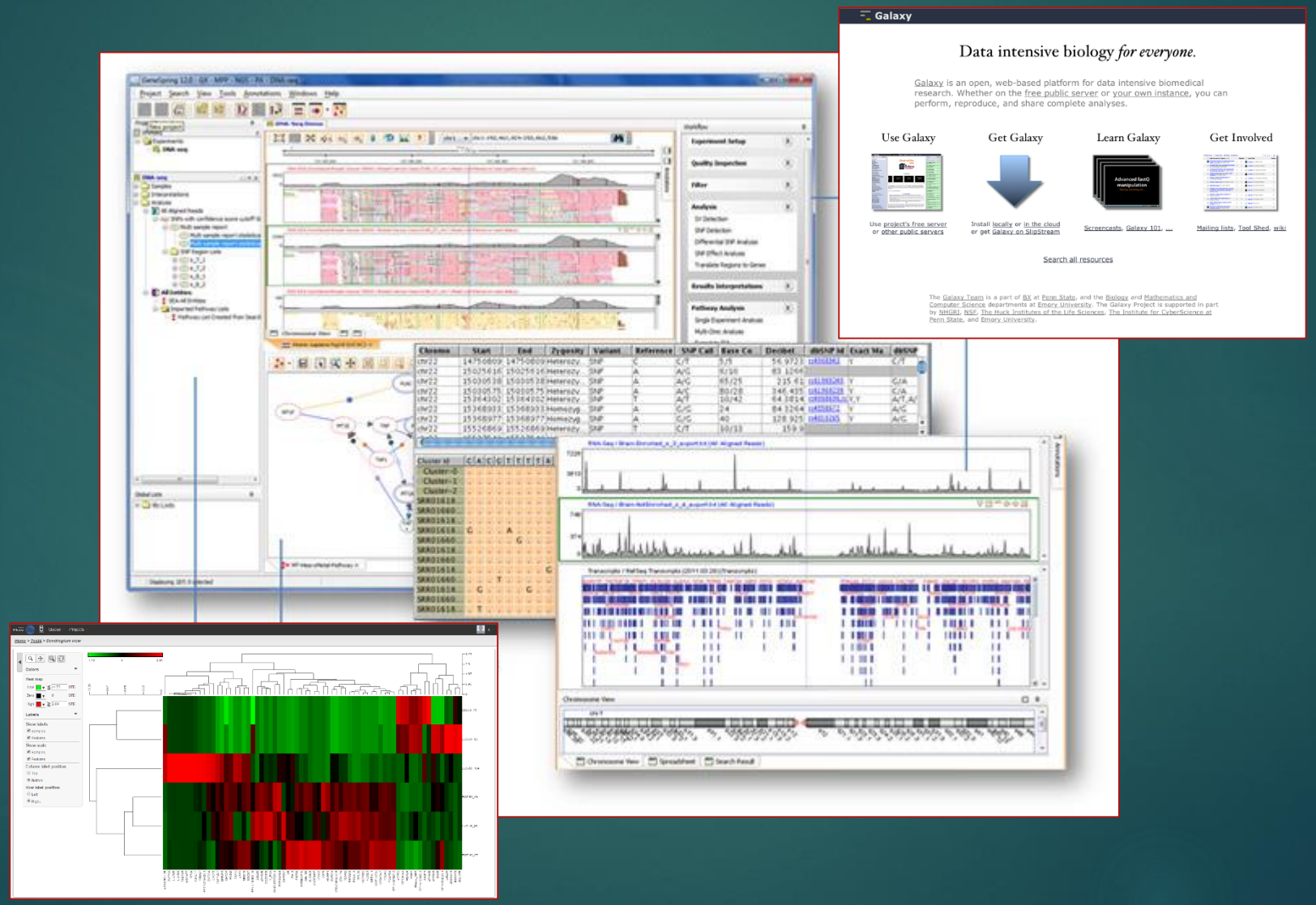
2 fichiers fastq: 17Go
2 fichiers fastq: 24Go




Station d'accueil

# 4 – Les ressources bioinformatiques



The alignment, assembly and utility bioinformatic tools for NGS.

| Program | Function | Platform | Website |
|---|---|---|---|
| *De novo assembly* | | | |
| Abyss | Alignment/assembly | Illumina | http://www.bcgsc.ca/platform/bioinfo/software/abyss |
| ALLPATHS | Alignment/assembly | Illumina | http://www.broadinstitute.org/science/programs/genome-biology |
| AMOScmp | Alignment/assembly | Roche | http://sourceforge.net/projects/amos/files/ |
| ARACHNE | Alignment/assembly | Roche | http://www.broadinstitute.org/science/programs/genome-biology |
| CAP3 | Alignment/assembly | Roche | http://pbil.univ-lyon1.fr/cap3.php |
| consensus/Seq-Cons | Alignment/assembly | Roche | http://www.seqan.de/downloads/projects.html |
| Curtain | Alignment/assembly | Illumina/Roche/ABI | http://code.google.com/p/curtain/ |
| Edena | Alignment/assembly | Illumina | http://www.genomic.ch/edena |
| Euler-SR | Alignment/assembly | Illumina/Roche | http://euler-assembler.ucsd.edu/portal/?q=team |
| FuzzyPath | Alignment/assembly | Illumina/Roche | ftp://ftp.sanger.ac.uk/pub/zn1/fuzzypath/fuzzypath_v3.0.tgz |
| IDBA | Alignment/assembly | Illumina | http://www.cs.hku.hk/~alse/idba/ |
| MIRA/MIRA3 | Alignment/assembly | Illumina/Roche | http://chevreux.org/projects_mira.html |
| Newbler | Alignment/assembly | Roche | roche-applied-science.com/ |
| Phrap | Alignment/assembly | Illumina/Roche | http://www.phrap.org/consed/consed.html#howToGet |
| RGA | Alignment/assembly | Illumina | http://rga.cgrb.oregonstate.edu/ |
| QSRA | Alignment/assembly | Illumina | http://qsra.cgrb.oregonstate.edu/ |
| SHARCGS | Alignment/assembly | Illumina | http://sharcgs.molgen.mpg.de/ |
| SHORTY | Alignment/assembly | ABI | http://www.cs.sunysb.edu/~skiena/shorty/ |
| SHRAP | Alignment/assembly | Roche | By request |
| SOAPdenovo | Alignment/assembly | Illumina | http://soap.genomics.org.cn |
| SOPRA | Alignment/assembly | Illumina/ABI | http://www.physics.rutgers.edu/%7Eanirvans/SOPRA/ |
| SR-ASM | Alignment/assembly | Roche | http://bioserver.cs.put.poznan.pl/sr-asm-short-reads-assembly-alg |
| SSAKE | Alignment/assembly | Illumina/Roche | http://www.bcgsc.ca/platform/bioinfo/software/ssake |
| Taipan | Alignment/assembly | Illumina | http://sourceforge.net/projects/taipan/files/ |
| VCAKE | Alignment/assembly | Illumina/Roche | http://sourceforge.net/projects/vcake |
| Velvet | Alignment/assembly | Illumina/Roche/ABI | http://www.ebi.ac.uk/%7Ezerbino/velvet |
| *Reference-based assembly* | | | |
| BFAST | Alignment/assembly | Illumina/ABI | http://sourceforge.net/apps/mediawiki/bfast/index.php?title=Mai |
| Bowtie | Alignment/assembly | Illumina/Roche/ABI | http://bowtie-bio.sourceforge.net |
| BWA | Alignment/assembly | Illumina/ABI | http://bio-bwa.sourceforge.net/bwa.shtml |
| CoronaLite | Alignment/assembly | ABI | http://solidsoftwaretools.com/gf/project/corona/ |
| CABOG | Alignment/assembly | Roche/ABI | http://wgs-assembler.sf.net |
| ELAND/ELAND2 | Alignment/assembly | Illumina/ABI | http://www.illumina.com/ |
| EULER | Alignment/assembly | Illumina | http://euler-assembler.ucsd.edu/portal/ |
| Exonerate | Alignment/assembly | Roche | http://www.ebi.ac.uk/~guy/exonerate |
| EMBF | Alignment/assembly | Illumina | http://www.biomedcentral.com/1471-2105/10?issue=S1 |
| GenomeMapper | Alignment/assembly | Illumina | http://1001genomes.org/downloads/genomemapper.html |
| GMAP | Alignment/assembly | Illumina | http://www.gene.com/share/gmap |





| Program | Function | Platform | Website |
|---|---|---|---|
| gnumap | Alignment/asse | | |
| ICON | Alignment/asse | | |
| Karma | Alignment/asse | | |
| LAST | Alignment/asse | | |
| LOCAS | Alignment/asse | | |
| Mapreads | Alignment/asse | | |
| MAQ | Alignment/asse | | |
| MOM | Alignment/asse | | |
| Mosaik | Alignment/asse | | |
| mrFAST/mrsFAST | Alignment/assembly | Illumina | http://mrfast.sourceforge.net/ |
| MUMer | Alignment/assembly | ABI | http://mummer.sourceforge.net/ |
| nexalign | Alignment/assembly | Illumina | http://genome.gsc.riken.jp/osc/english/dataresource/ |
| Novocraft | Alignment/assembly | Illumina | http://www.novocraft.com/ |
| PerM | Alignment/assembly | Illumina/ABI | http://code.google.com/p/perm/ |
| RazerS | Alignment/assembly | Illumina/ABI | http://www.seqan.de/projects/razers.html |
| RMAP | Alignment/assembly | Illumina | http://rulai.cshl.edu/rmap |
| segemehl | Alignment/assembly | Illumina/Roche | http://www.bioinf.uni-leipzig.de/Software/segemehl/ |
| SeqCons | Alignment/assembly | Roche | http://www.seqan.de/projects/seqcons.html |
| SeqMap | Alignment/assembly | Illumina | http://biogibbs.stanford.edu/*jiangh/SeqMap/ |
| SHRiMP | Alignment/assembly | Illumina/Roche/ABI | http://compbio.cs.toronto.edu/shrimp |
| Slider/SliderII | Alignment/assembly | Illumina | http://www.bcgsc.ca/platform/bioinfo/software/slider |
| SOCS | Alignment/assembly | ABI | http://solidsoftwaretools.com/gf/project/socs/ |
| SOAP/SOAP2 | Alignment/assembly | Illumina/ABI | http://soap.genomics.org.cn |
| SSAHA/SSAHA2 | Alignment/assembly | Illumina/Roche | http://www.sanger.ac.uk/Software/analysis/SSAHA2 |
| Stampy | Alignment/assembly | Illumina | http://www.well.ox.ac.uk/~marting/ |
| SXOligoSearch | Alignment/assembly | Illumina | http://synasite.mgrc.com.my:8080/sxog/NewSXOligoSearch.php |
| SHORE | Alignment/assembly | Illumina | http://1001genomes.org/downloads/shore.html |
| Vmatch | Alignment/assembly | Illumina | http://www.vmatch.de/ |
| *Diagnostics/utilities* | | | |
| Artemis/ACT | Visualization tool | Illumina/Roche | http://www.sanger.ac.uk/resources/software/artemis/ |
| CASHX | Pipeline | Illumina | http://seqanswers.com/wiki/CASHX |
| Consed | Visualization tool | Illumina/Roche | http://www.genome.washington.edu/consed/consed.html |
| EagleView | Visualization tool | Illumina/Roche | http://bioinformatics.bc.edu/marthlab/EagleView |
| FastQC | Quality assessment | Illumina/ABI | http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/ |
| Gambit | Visualization tool | Illumina/Roche | http://bioinformatics.bc.edu/marthlab/Gambit |
| Goby | Data management | Illumina/Roche/ABI | http://campagnelab.org/software/goby/ |
| G-SQZ | Data management | Illumina/ABI | http://public.tgen.org/sqz |
| Hawkeye | Visualization tool | Illumina/Roche | http://amos.sourceforge.net/hawkeye |
| Hybrid-SHREC | Error Correction | Illumina/Roche/ABI | http://www.cs.helsinki.fi/u/lmsalmel/hybrid-shrec/ |
| IGV | Visualization tool | Illumina | http://www.broadinstitute.org/igv/?q=home |
| LookSeq | Visualization tool | Illumina/Roche | http://lookseq.sourceforge.net |
| MagicViewer | Visualization tool | Illumina | http://bioinformatics.zj.cn/magicviewer/ |



Published in final edited form as:
*J Genet Genomics*. 2011 March 20; 38(3): 95–109. doi:10.1016/j.jgg.2011.02.003.

## The impact of next-generation sequencing on genomics

**Jun Zhang**[a,b,*], **Rod Chiodini**[c], **Ahmed Badr**[a], and **Genfa Zhang**[d]

[a] COE for Neurosciences, Department of Anesthesiology, Texas Tech University Health Sciences Center El Paso, TX 79905, USA

[b] Department of Biomedical Sciences, Texas Tech University Health Sciences Center El Paso, TX 79905, USA

[c] Internal Medicine, Texas Tech University Health Sciences Center El Paso, TX 79905, USA

[d] College of Life Sciences, Beijing Normal University, Beijing 100875, China

| Program | Function | Platform | Website |
|---|---|---|---|
| MapView | Visualization tool | Illumina | http://evolution.sysu.edu.cn/mapview/ |
| NGSView | Visualization tool | Illumina/ABI | http://ngsview.sourceforge.net |
| PIQA | Quality assessment | Illumina | http://bioinfo.uh.edu/PIQA |
| Reconciliation | Assembly pipeline | Illumina | http://www.genome.umd.edu/software.htm |
| RefCov | Sequence coverage | Illumina/Roche | http://genome.wustl.edu/tools/cancer-genomics |
| SAM Tools | Utilities | Illumina/Roche | http://sourceforge.net/projects/samtools/files/ |
| Savant | Visualization tool | Illumina/Roche | http://compbio.cs.toronto.edu/savant/ |
| ShortRead | Quality assessment | Illumina/Roche | http://bioconductor.org/packages/2.6/bioc/html/ShortRead.html |
| SHREC | Error Correction | Illumina/Roche | http://www.informatik.uni-kiel.de/jasc/Shrec/ |
| Staden Tools (GAP5) | Pipeline | Illumina/Roche | http://sourceforge.net/projects/staden/files/ |
| Tablet | Visualization tool | Illumina/Roche | http://bioinf.scri.ac.uk/tablet |
| TagDust | Data cleaning | Illumina | http://genome.gsc.riken.jp/osc/english/software/. |
| TileQC | Quality assessment | Illumina | http://www.science.oregonstate.edu/~dolanp/tileqc |
| XMatchView | Visualization tool | Illumina/Roche | http://www.bcgsc.ca/platform/bioinfo/software/xmatchview |
| Yenta | Visualization tool | Illumina | http://genome.wustl.edu/tools/cancer-genomics |
| Geneus | Data management | Illumina/ABI | http://www.genologics.com/solutions/research-informatics/ |

- ✓ Rechercher des programmes adaptés aux jeux de données
- ✓ Et les tester …….

# Galaxy pour l'analyse de NGS

Galaxy est une plateforme web
« open-source » pour l'intégration de divers
outils et bases de données pour la génomique
dans un espace de travail cohérent.


- Manipuler les donnés dans différents formats
- Outils pour divers analyses dont l'analyse statistique

# Galaxy pour…

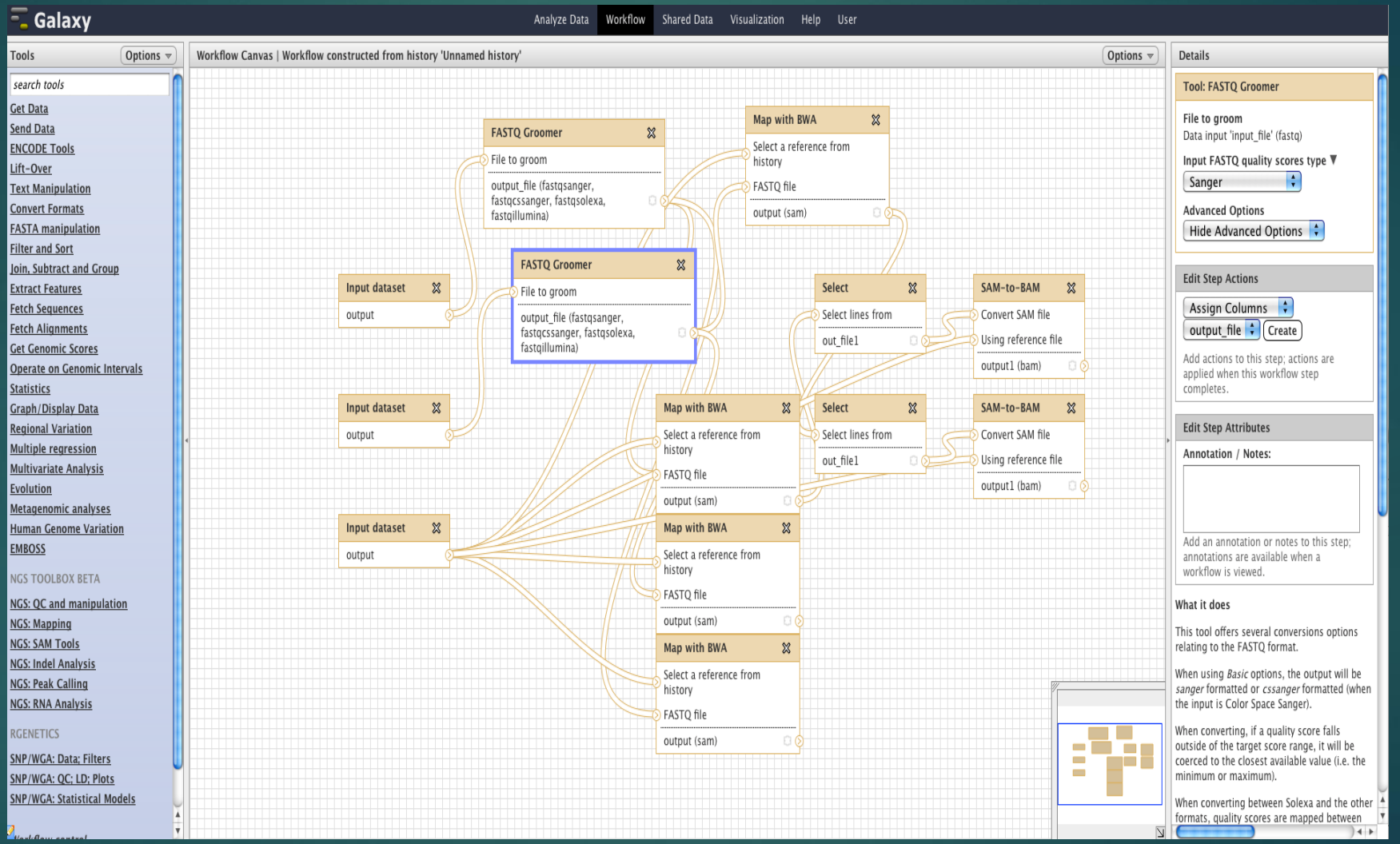
- Bioinformaticiens
  - Programme installable
  - Modulable
    - Ajouter de nouveaux outils
    - Intégrer des nouvelles données
  - Installer votre propre server Galaxy privé

- Biologistes
  - Analyser ses propres données
  - Récupérer des bases de données ou sa propre DB
  - Manipuler des données génomiques
  - Visualiser
  - Publier et partager
    - Pipelines personnalisables (Workflows) ou protocoles

# Galaxy pour les  NGS

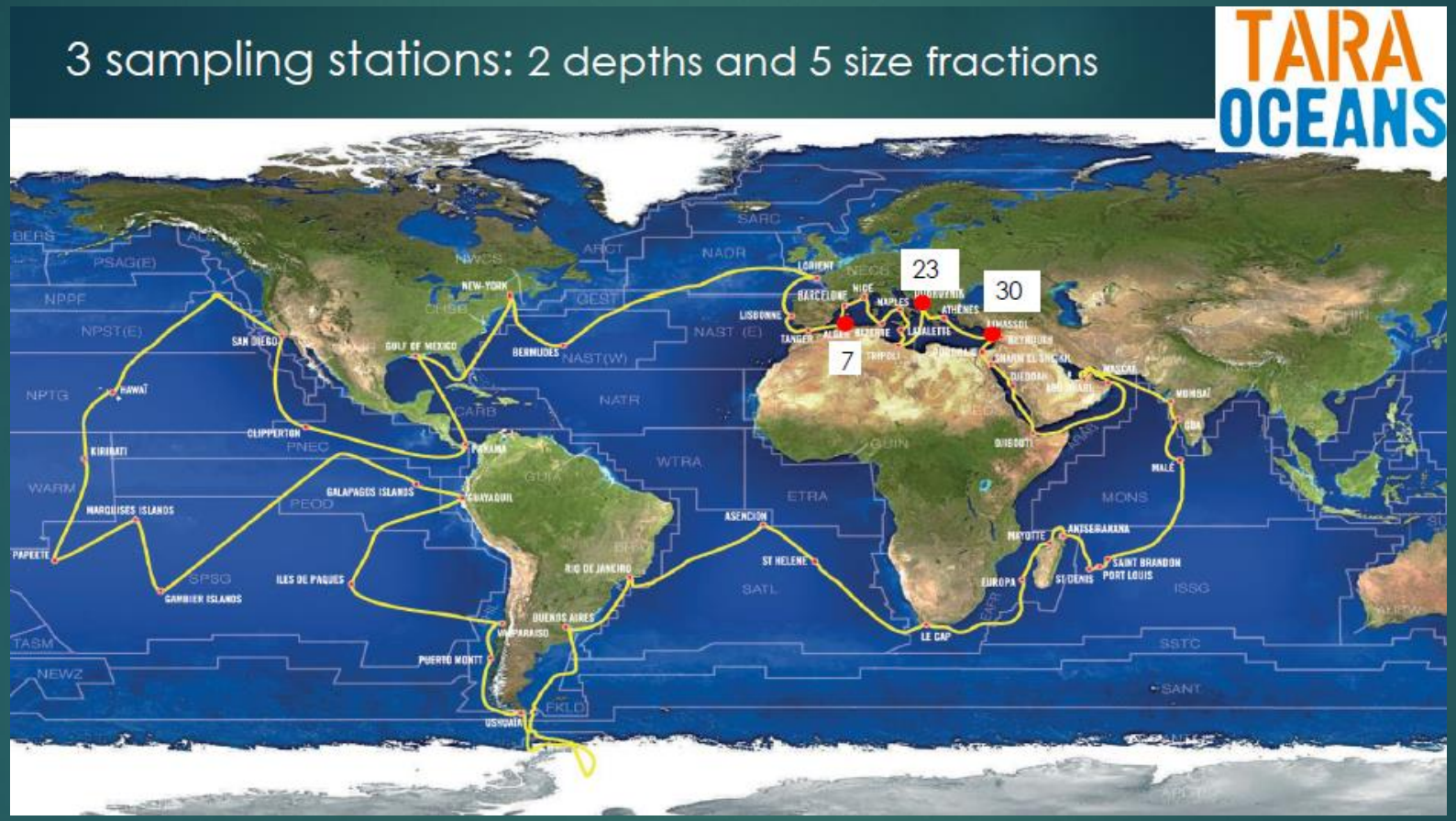
- Données brutes: Sequencing Reads (FASTQ)
- Alignements contre des génomes de reference (SAM/BAM)
- Annotations (GFF/GTF, BED)
- Préparer, regarder la qualité et manipule les lectures FASTQ
- Mapping
- SAMTools
- Analyse de SNP and INDEL
- Analyse de RNAseq

# Workflow




Créer un “workflow”, permet à l’utlisateur de répéter l’analyse avec différents jeux de données

# 5 – Un cas d'étude




What are the most abundant genes in marine plankton?

# Metagenomes and metatranscriptomes

Extract ecosystem DNA/RNA

Raw DNA or RNA sequences HiSEQ & GAIIx

Assemble

Gene prediction

Similarity search

No hits = unknown

Public Database of **known proteins**

Similarity detected = known

Taxonomy

Function

## Metagenomes

- 29 DNA samples
- 4.1 Gb sequences
- 4.9 million contigs 760 bp on average
- 7,355,407 ORFs

## Metatranscriptomes

- 23 RNA samples
- 8 Gb sequences
- 66 million contigs 120 bp on average
- 170,785,174 ORFs

# NCBI/CDD Search in MetaG

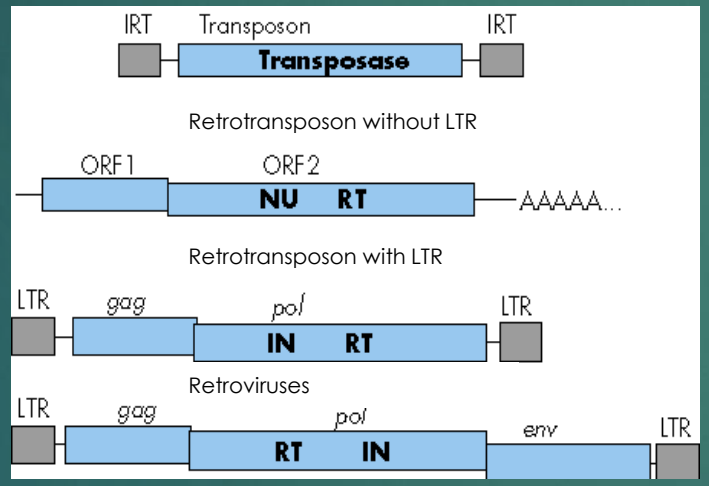
| Accession/Name | Description | Number of assigned ORFs |
|---|---|---|
| cd01650/RT_nLTR_like* | Non-LTR (long terminal repeat) retrotransposon and non-LTR retrovirus reverse transcriptase (RT). This subfamily contains both non-LTR retrotransposons and non-LTR retrovirus RTs. | 2782 |
| cd01647/RT_LTR* | Reverse transcriptases (RTs) from retrotransposons and retroviruses which have long terminal repeats (LTRs) in their DNA copies but not in their RNA template. | 1207 |
| pfam00665/rve* | Integrase core domain. Integrase mediates integration of a DNA copy of the viral genome into the host chromosome. Integrase is composed of three domains. | 988 |
| pfam03372/Exo_endo_phos* | Endonuclease/Exonuclease/phosphatase family. This large family of proteins includes magnesium dependent endonucleases and a large number of phosphatases involved in intracellular signaling. | 556 |
| cd01644/RT_pepA17* | Reverse transcriptase (RTs) in retrotransposons. This subfamily represents the RT domain of a multifunctional enzyme. | 551 |
| cd00204/ANK | Ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins. | 438 |
| pfam05380/Peptidase_A17*† | Pao retrotransposon peptidase. Corresponds to Merops family A17. | 431 |
| KOG2462/KOG2462 | KOG2462, C2H2-type Zn-finger protein [Transcription] | 342 |
| cd00190/Tryp_SPc | Trypsin-like serine protease; Many of these are synthesized as inactive precursor zymogens that are cleaved during limited proteolysis to generate their active forms. | 321 |
| pfam07727/RVT_2* | Reverse transcriptase (RNA-dependent DNA polymerase). A reverse transcriptase gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. | 260 |
| pfam05970/DUF889 | PIF1 helicase. The PIF1 helicase inhibits telomerase activity and is cell cycle regulated. | 249 |
| pfam00078/RVT_1* | Reverse transcriptase (RNA-dependent DNA polymerase). | 222 |
| KOG3623/KOG3623 | KOG3623, Homeobox transcription factor SIP1 [Transcription] | 211 |
| cd06222/RNaseH* | RNase H (RNase HI) is an endonuclease that cleaves the RNA strand of an RNA/DNA hybrid in a not sequence-specific manner. | 186 |
| KOG1214/KOG1214 | KOG1214, Nidogen and related basement membrane protein | 179 |

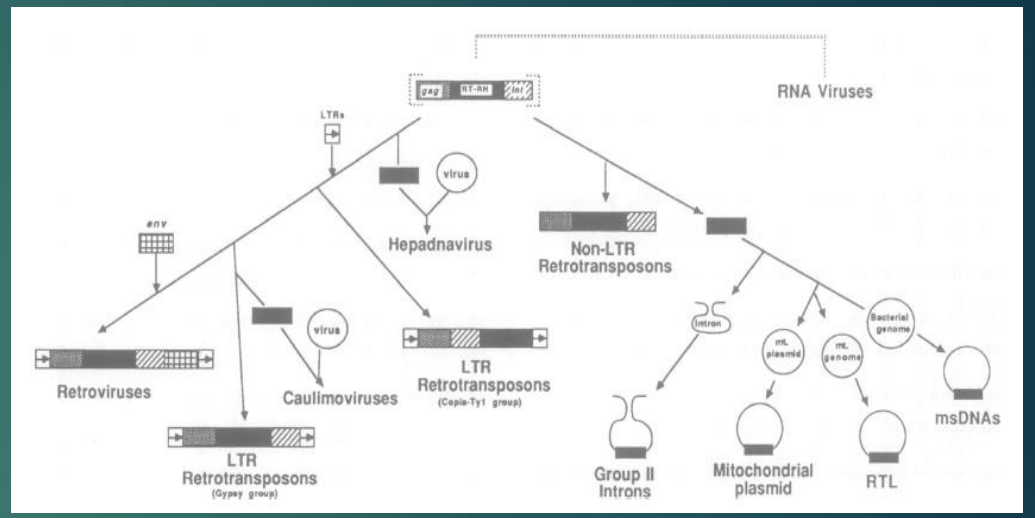
# Transposable elements

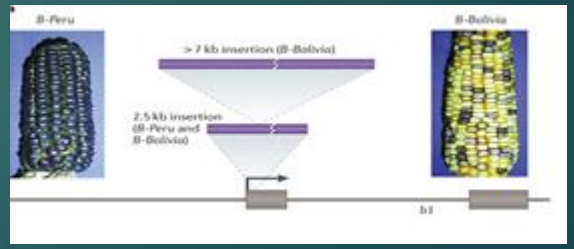
Two types of mobile elements

Class II
DNA transposon

Class I
retrotransposon



- Genomic rearrangement
- Regulation of gene expression
- Retrotransposons responsible for genome expansion (45% of human genome, 80% of maize, 55% of *Chondrus crispus*)

# Pipeline

MetaG

MetaT

ORFs prediction

getorf (EMBOSS)

HMM search
RT HMM profiles

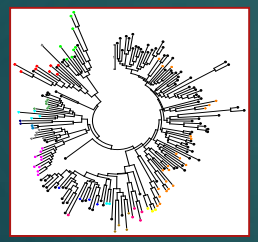
58 HMM profiles
(from GyDB and (Gladyshev and Arkipova, 2011))

BLASTP
against RT classes

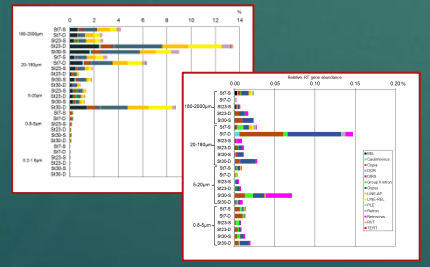
343 RT proteins database

RT-like sequences

Phylogeny analysis (PhyML)

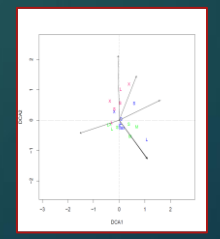


Relative gene abundance



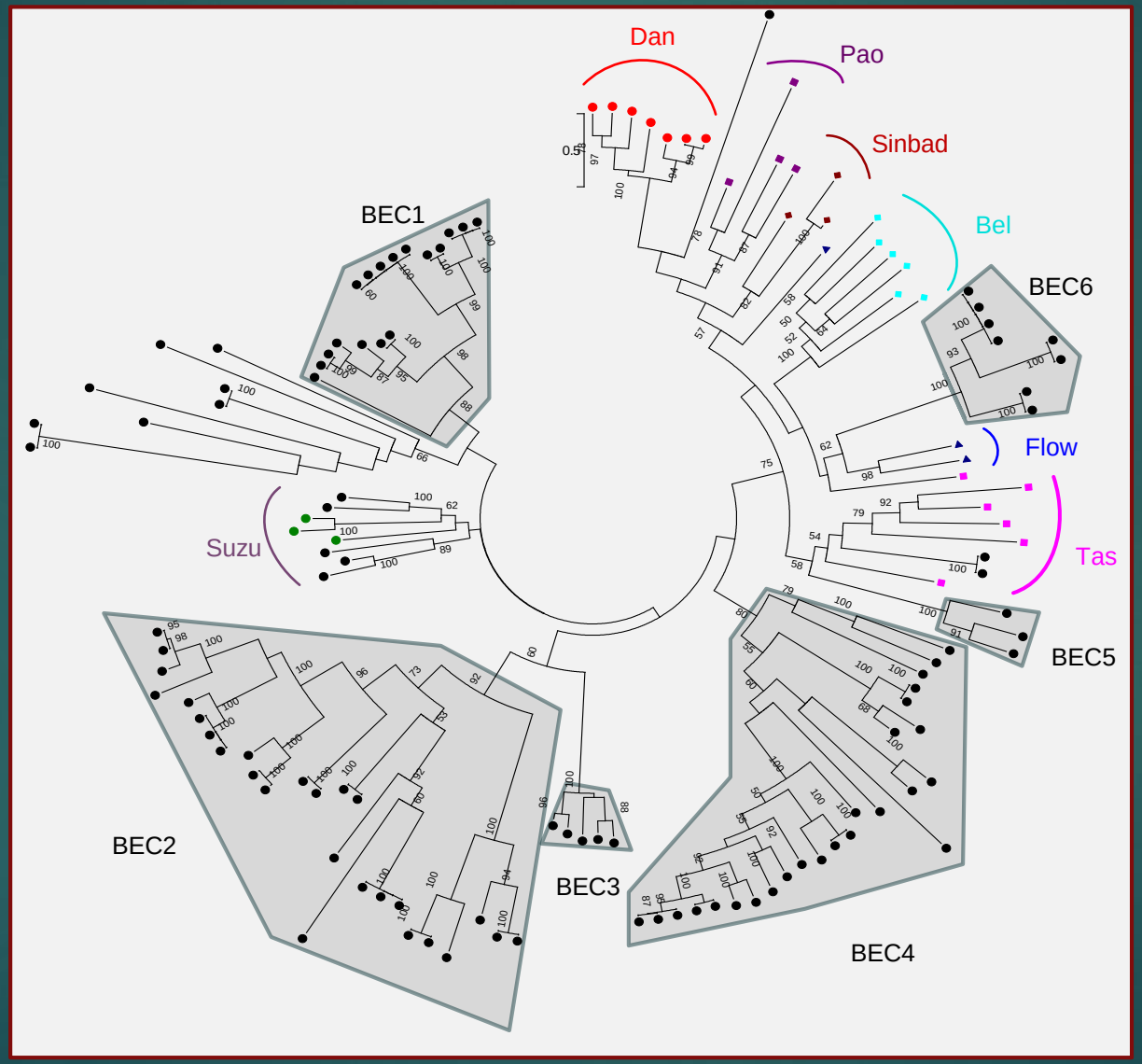
Taxonomy analysis

Correspondence analysis



- Characterize and classify RT genes (retrotransposons) having important roles within plankton populations
- Make a link with environmental conditions

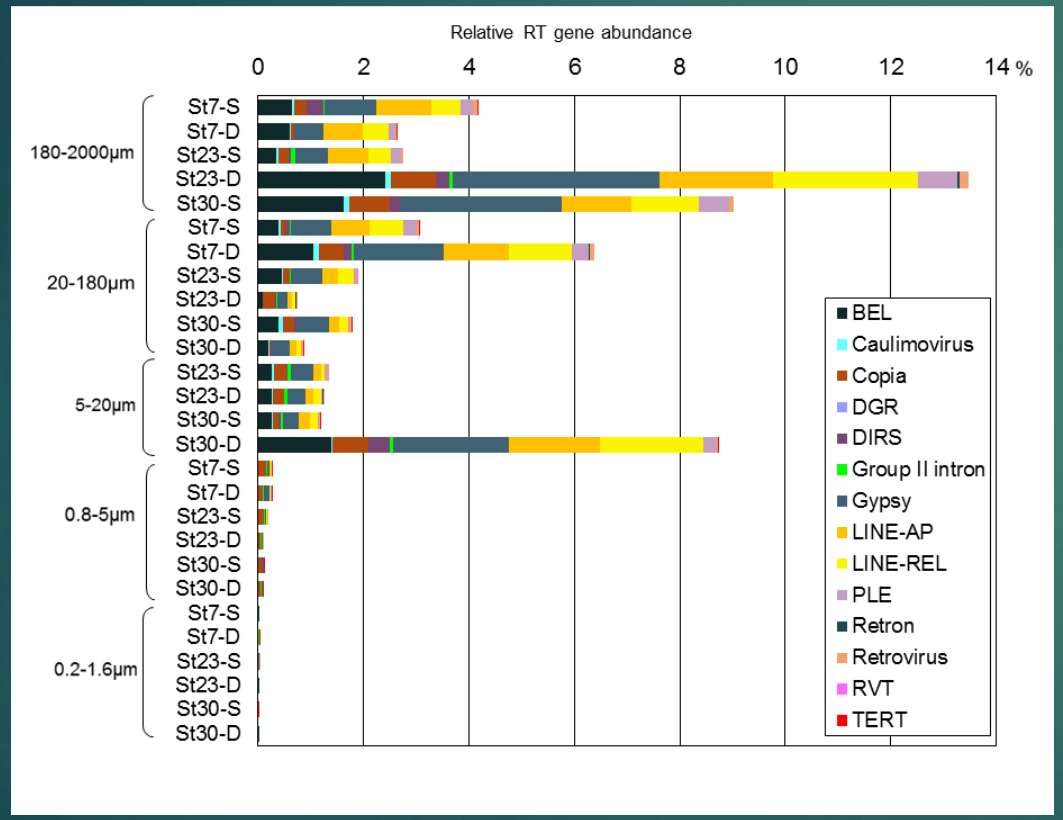
# Identification of novel BEL reverse transcriptase domains



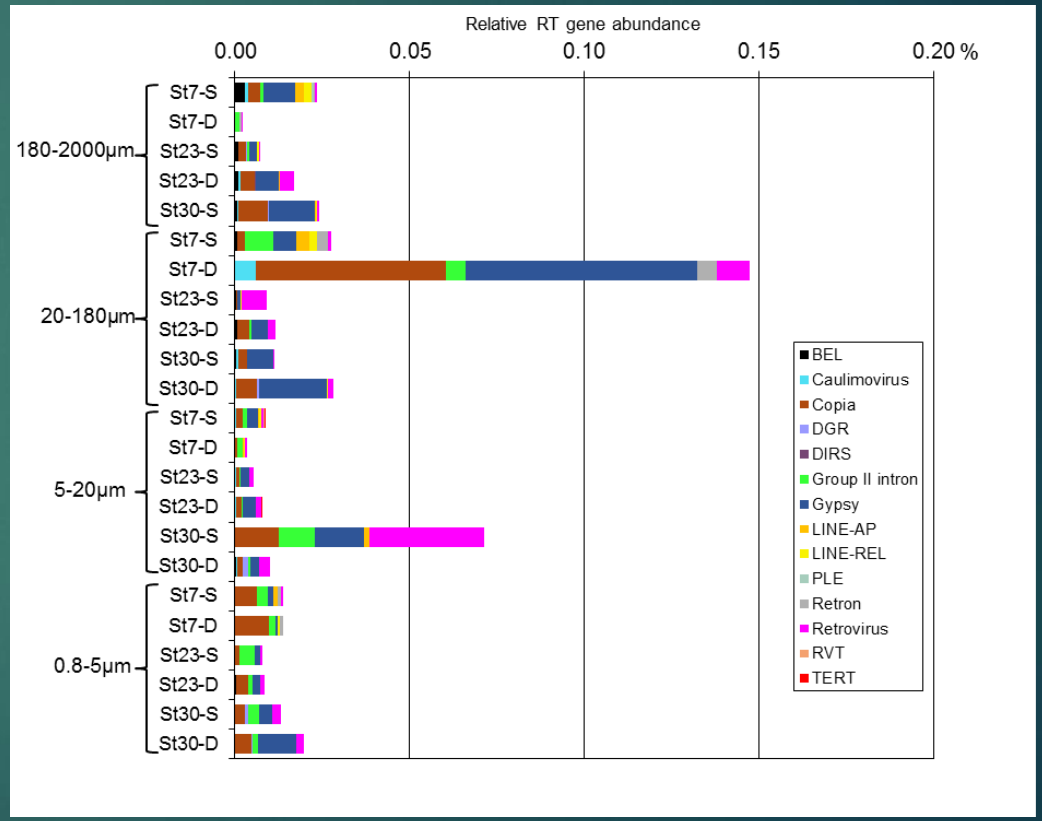

- 6 new clades:
94 environmental RT-like sequences

# Relative abundance of RT from prokaryotes to zooplankton

- MetaG

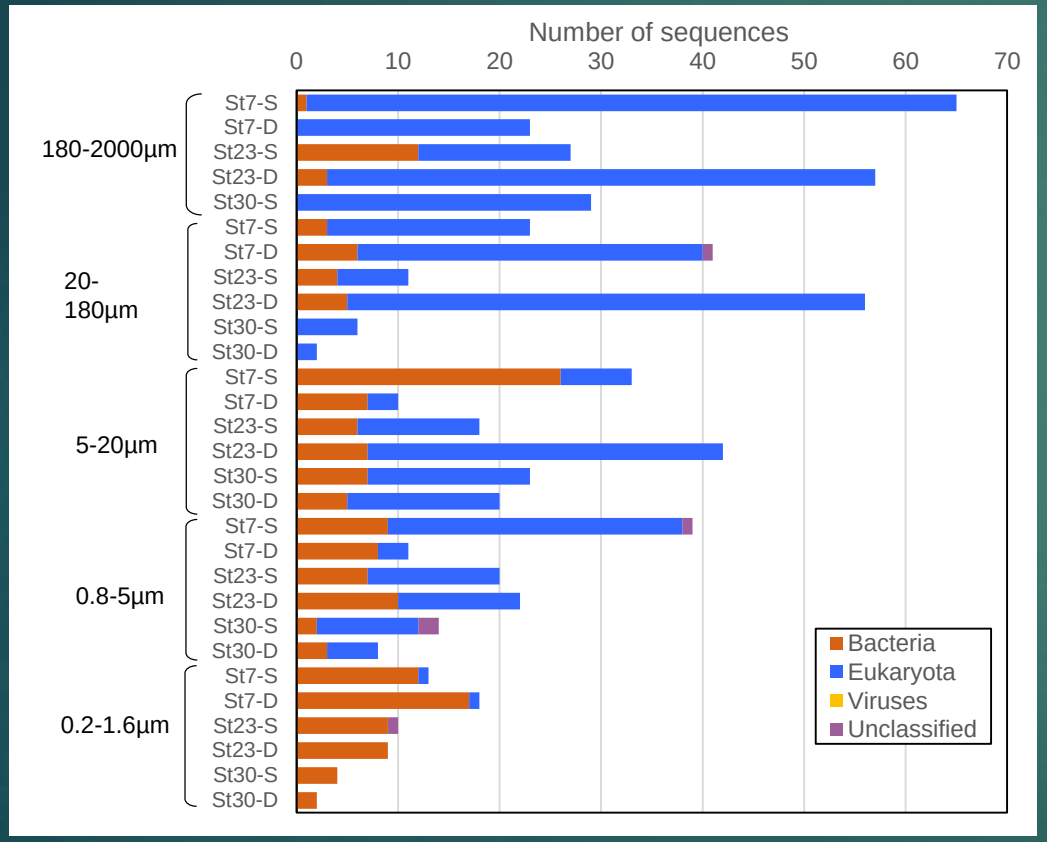



- MetaT

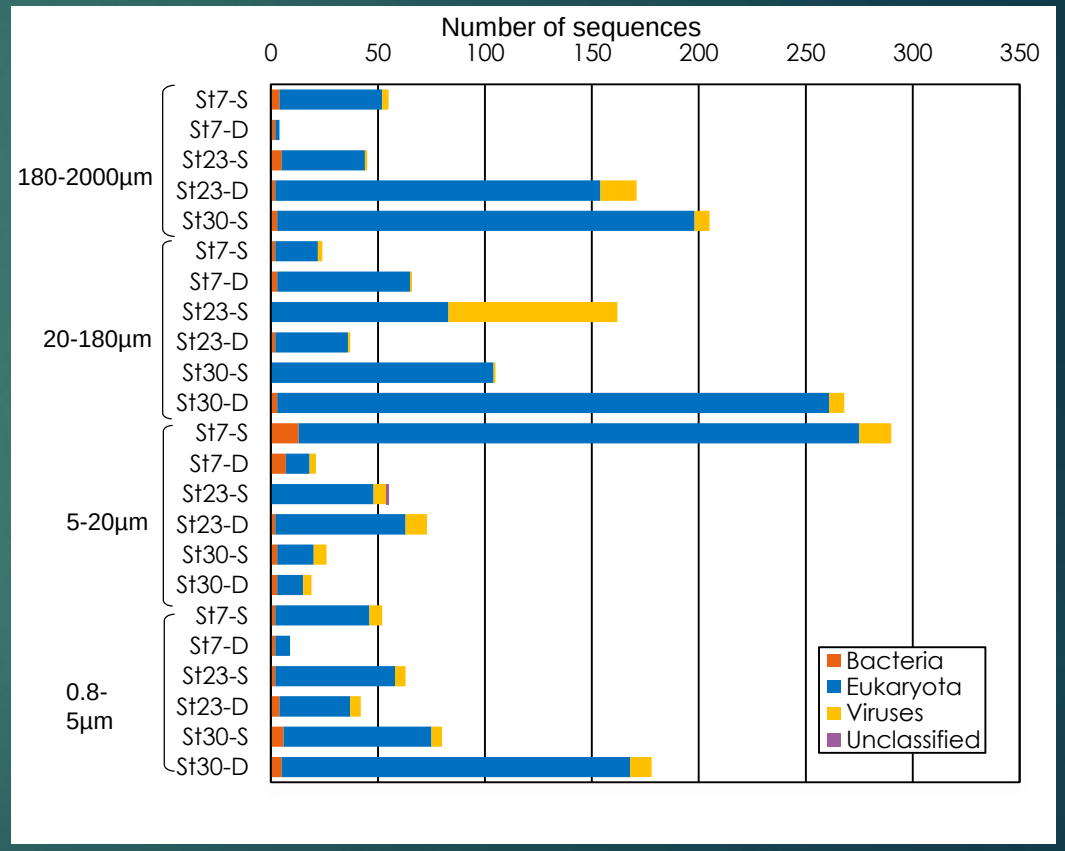


# MetaG and MetaT Taxonomy

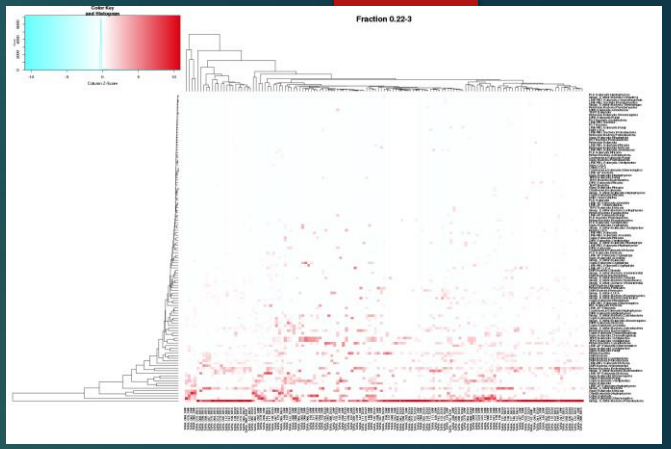
- MetaG

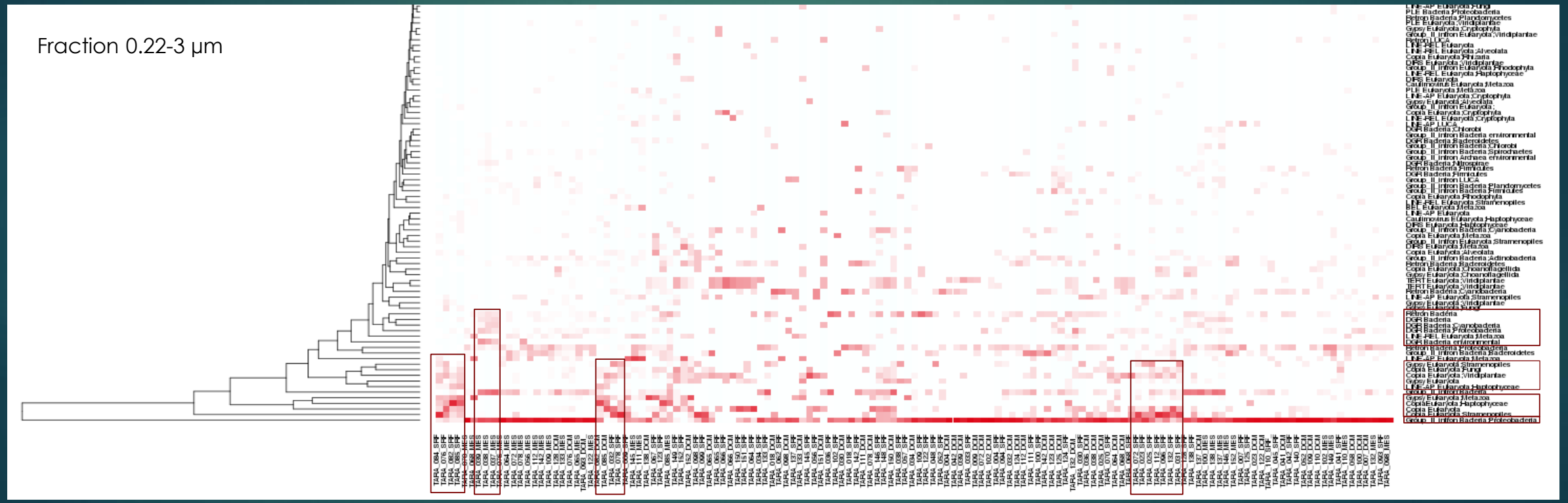



- MetaT



# Clustering of RT abundance pattern and their taxonomy



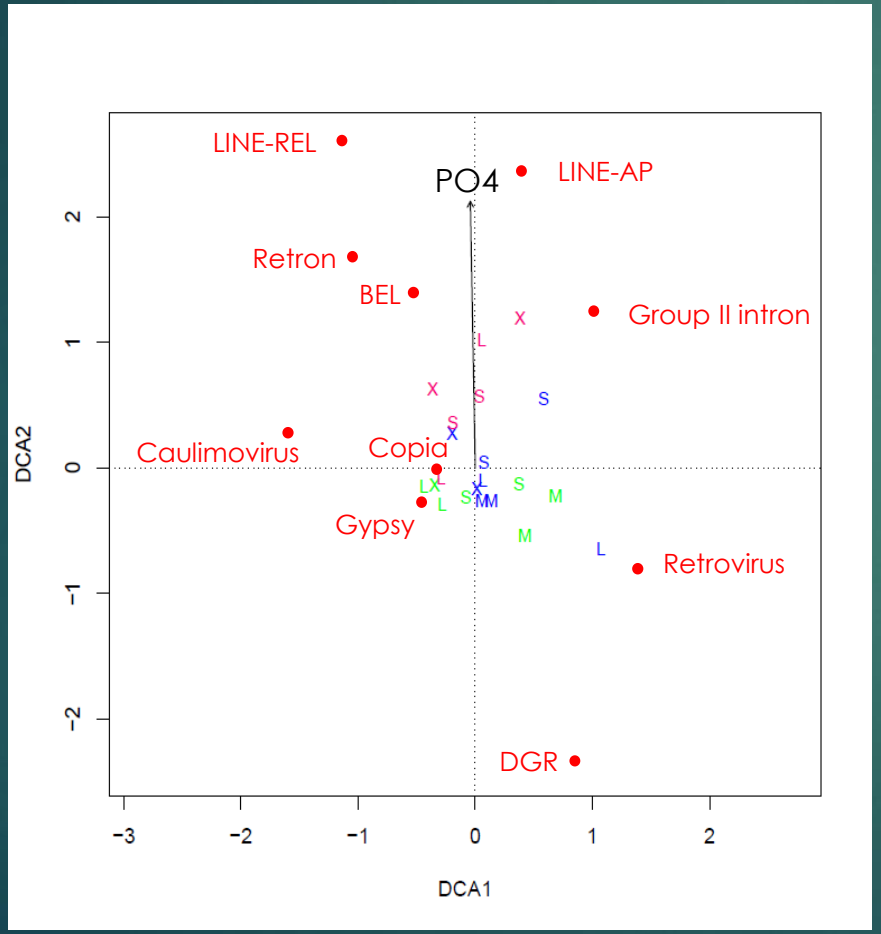

TARA_076_SRF
TARA_082_SRF
TARA_084_SRF
TARA_085_SRF

TARA_037_MES
TARA_038_MES
TARA_039_MES

TARA_032_SRF
TARA_078_SRF
TARA_082_DCM
TARA_085_DCM

TARA_023_SRF
TARA_025_SRF
TARA_031_SRF

# Correlation between the biotic and abiotic variables

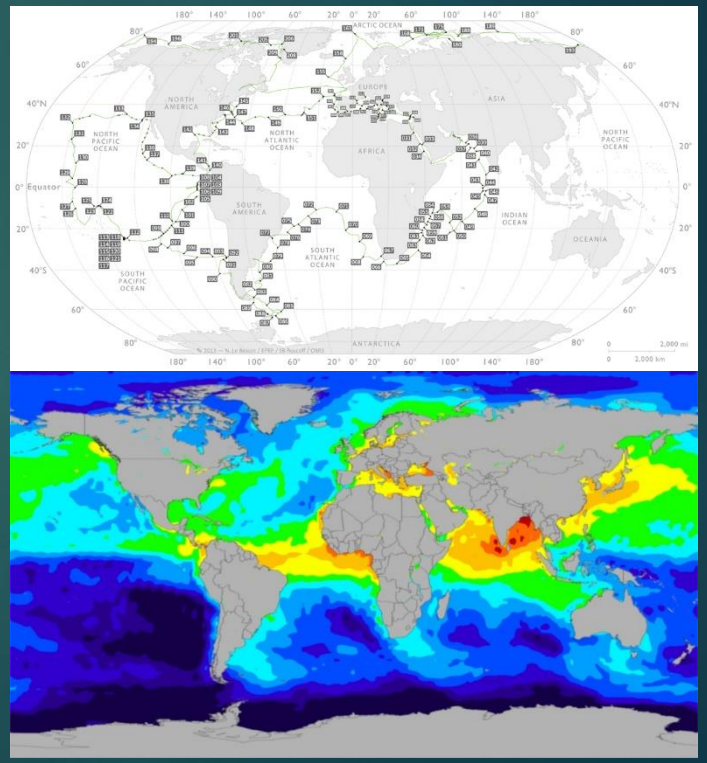



- Detrended correspondence analysis (DCA)
- Association of RT transcriptional pattern with environmental variables
- Test the significance of correlations
  - PO4 most correlated with RT

# Take home message
# Answer to a biological question!

Reverse transcriptases are the most abundant genes in marine eukaryotic plankton

- Expansion of the class I TE as observed in terrestrial animal and plant genomes
  - Plankton adaptation
    - Ecological variables (Phosphates)
  - Genomic adaptation
    - Intense competition between organisms
- Correlation between the biotic and abiotic variables





micro B3
Biodiversity. Bioinformatics. Biotechnology.
TARA OCEANS
GENOSCOPE
Centre National de Séquençage