

Méthodes d'analyses à grande échelle des métatranscriptomes eucaryotes dans le projet TARA Océans.

Quentin Carradec

30-juin-2015

The logo for TARA OCEANS. The word 'TARA' is in a large, bold, orange sans-serif font, and the word 'OCEANS' is in a large, bold, blue sans-serif font, positioned directly below 'TARA'.

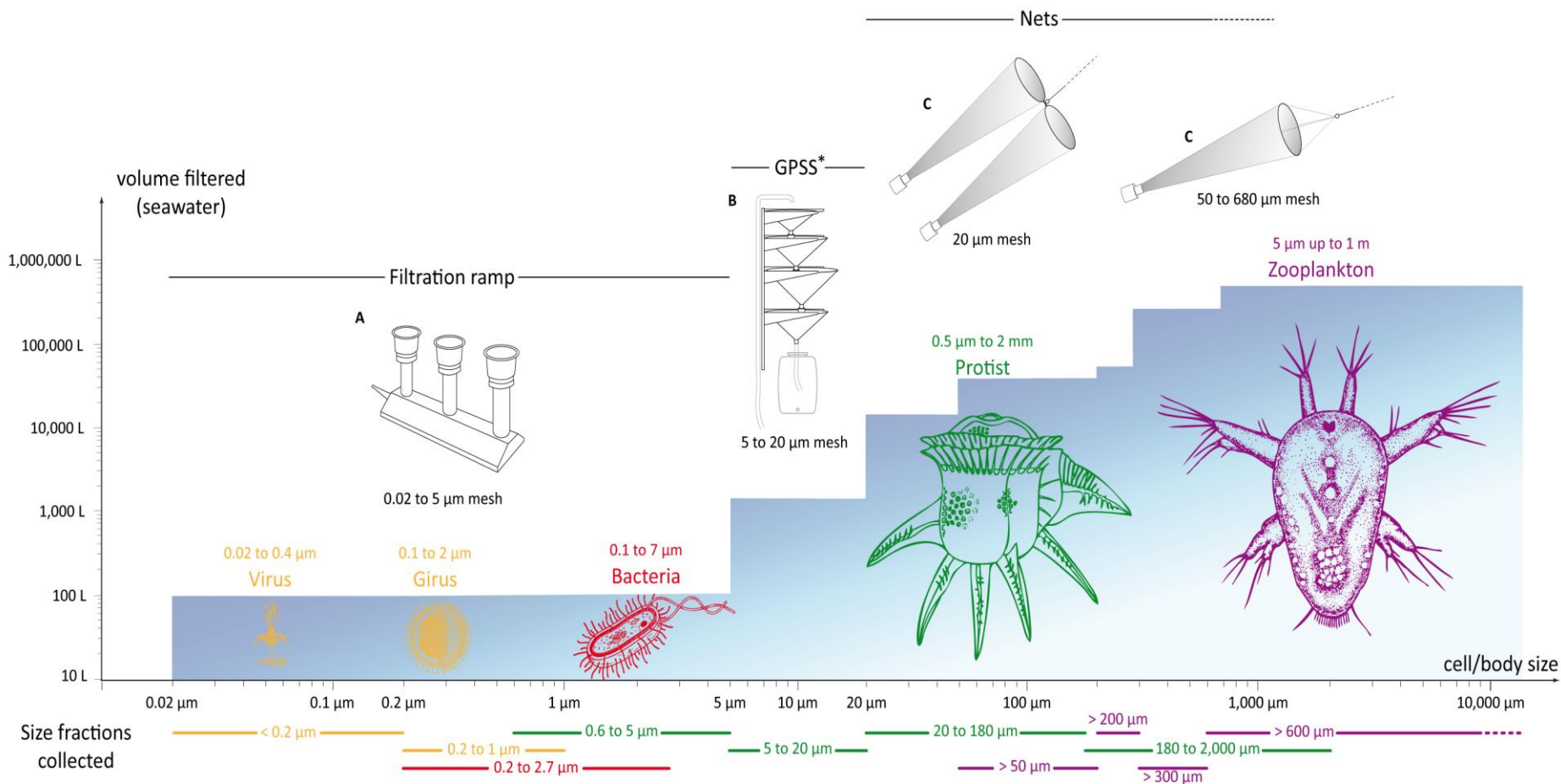
Le projet TARA Océans



Interdisciplinaire :

- séquençage au débit.
- imagerie haute résolution
- données physico-chimiques des océans.

L'échantillonnage



2011 N. Le Bescot, F. Not

- En surface
- DCM (Deep Chlorophyll Maximum)
- Mésopelagique (200 m – 1 000 m)

➡ Séquençage au Génoscope

Premières publications

Structure and function of the global ocean microbiome ;
Sunagawa, Coelho, Chaffron, et al. DOI:
10.1126/science.1261359

Global patterns and ecological drivers of ocean viral communities ; Brum, Ignacio-Espinosa, Roux et al. DOI:
10.1126/science.1261498

Determinants of community structure in the global plankton interactome ; Lima-Mendez, Faust, Henry et al.
DOI: 10.1126/science.1262073

Environmental characteristics of Agulhas rings affect inter-ocean plankton transport ; Villar, Farrant, Follows et al. DOI: 10.1126/science.1261447

Eukaryotic plankton diversity in the sunlit global ocean ;
De Vargas, Audic, Henry, et al. DOI:
10.1126/science.1261605

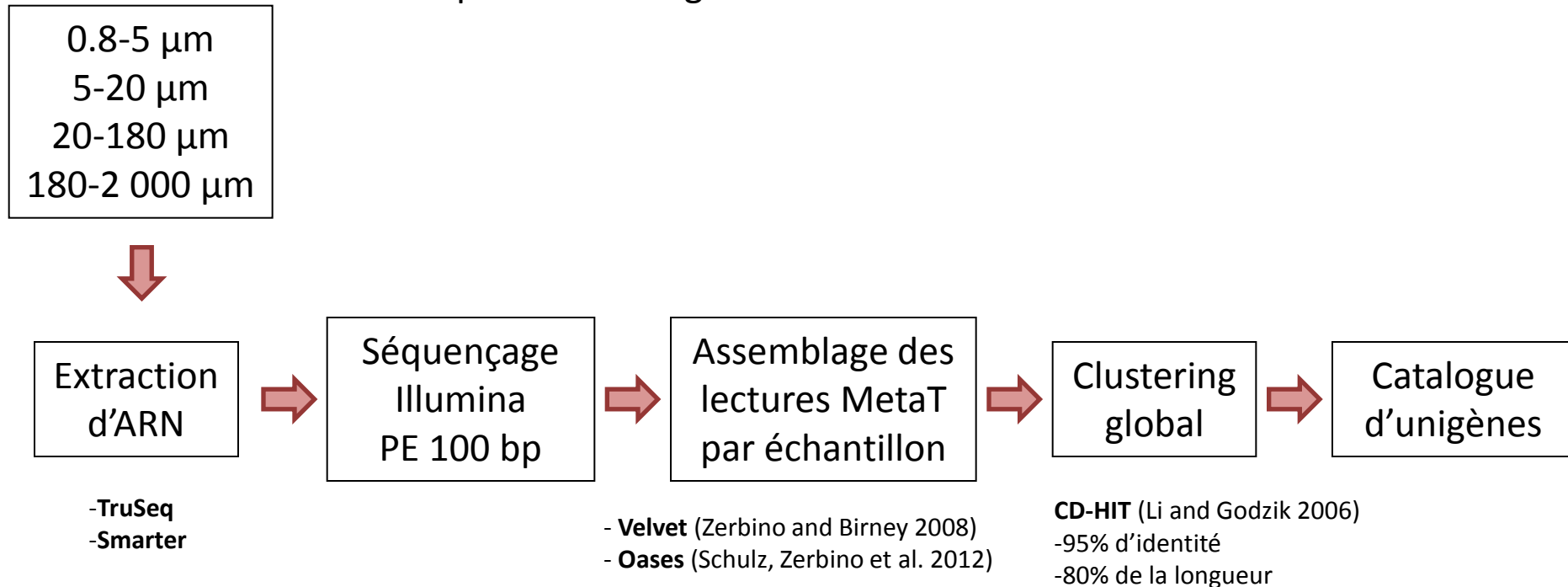


Etude des Métatranscriptomes eucaryotes

- **Etude des Métagénomes eucaryotes compliquée.**
 - assemblage difficile.
 - prédiction de gènes *de novo*.
- **Etude des Métatranscriptomes.**
 - assemblage plus facile.
 - aucune prédiction de gène nécessaire.

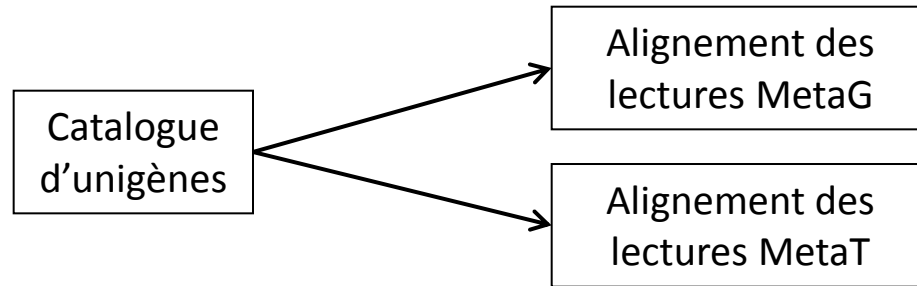
Etude des Métatranscriptomes eucaryotes

- **Etude des Métagénomes eucaryotes compliquée.**
 - assemblage difficile.
 - prédiction de gènes *de novo*.
- **Etude des Métatranscriptomes.**
 - assemblage plus facile.
 - aucune prédiction de gène nécessaire.



- **Annotation fonctionnelle (CDD search)**
- **Affiliation taxonomique**
- **Occurrences (niveaux de couverture)**

Occurrences des unigènes

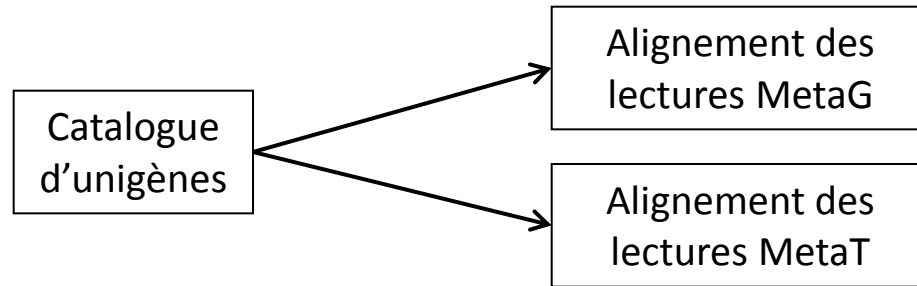


BWA (Li and Durbin 2009)
95% d'identité 80% de la longueur

Pour un unigène :

- Occurrence MetaG ➡ Abondance (nombre de copies du gène)
- Occurrence MetaT / Occurrence MetaG ➡ Expression (par copie)

Occurrences des unigènes



BWA (Li and Durbin 2009)
95% d'identité 80% de la longueur

Pour un unigène :

- Occurrence MetaG ➡ Abondance (nombre de copies du gène)
- Occurrence MetaT / Occurrence MetaG ➡ Expression (par copie)

Pour un sous-set d'unigènes spécifiques d'une espèce :

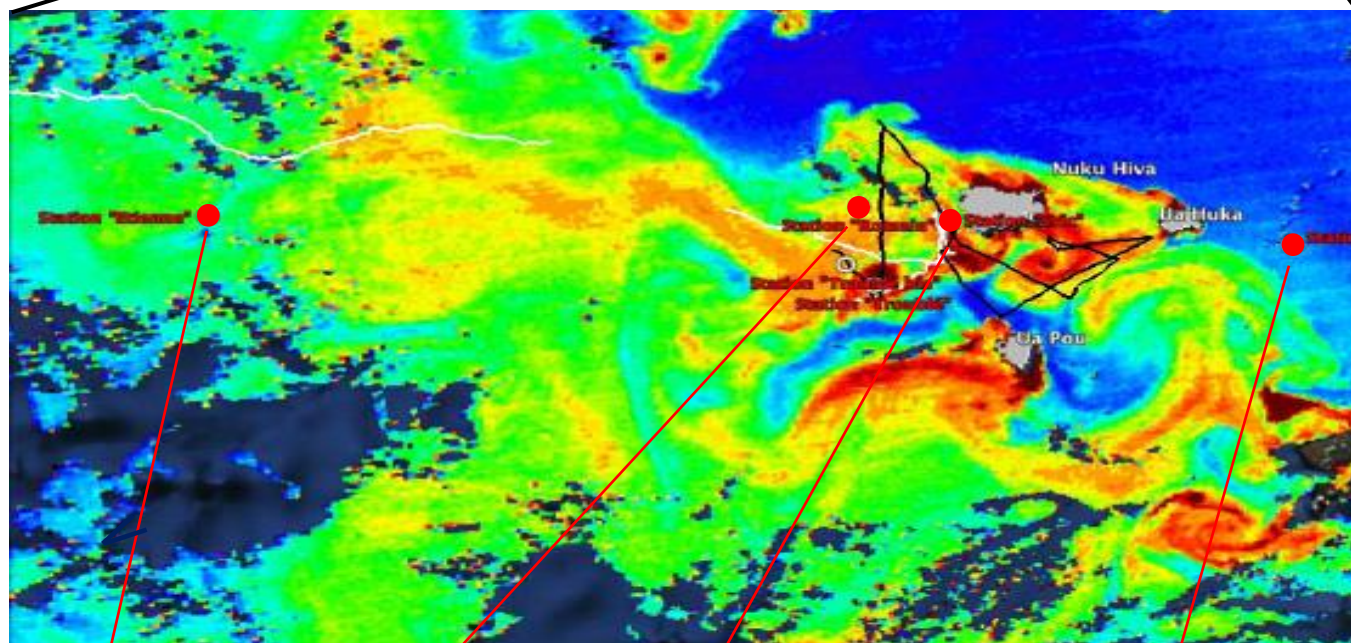
- Obtention des unigènes:
 - Homologie avec un transcriptome connu
 - Affiliation taxonomique
- Spécificité des unigènes
 - Cross-mapping pour les gènes très conservés ➡ Etude de la distribution des occurrences MetaG
- $\text{RPKM metaT} / \text{somme (RPKM metaT)}$ ➡ expression relative de chaque unigène (%)

Impact du fer sur la population planctonique

Le projet STEFI

4 stations autour des îles marquises

Concentration en
chlorophylle



Stations :

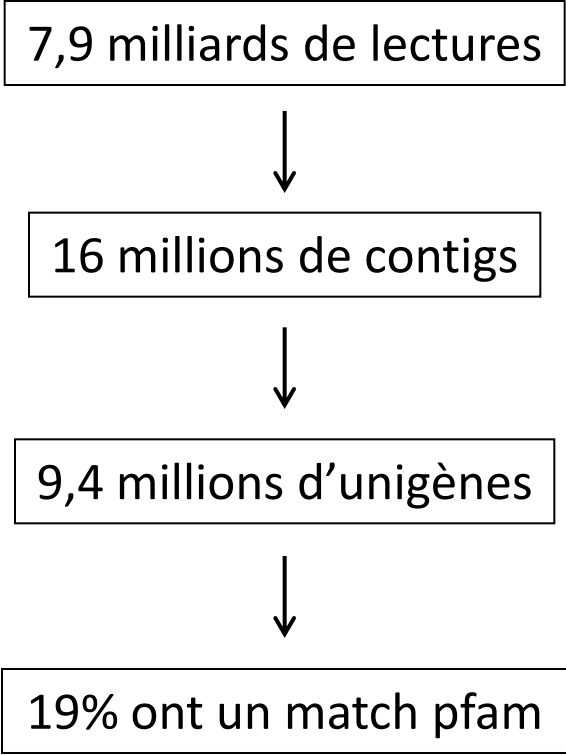
125

124

123

122 (HNLC)

Analyse globale



Analyse globale

7,9 milliards de lectures



16 millions de contigs

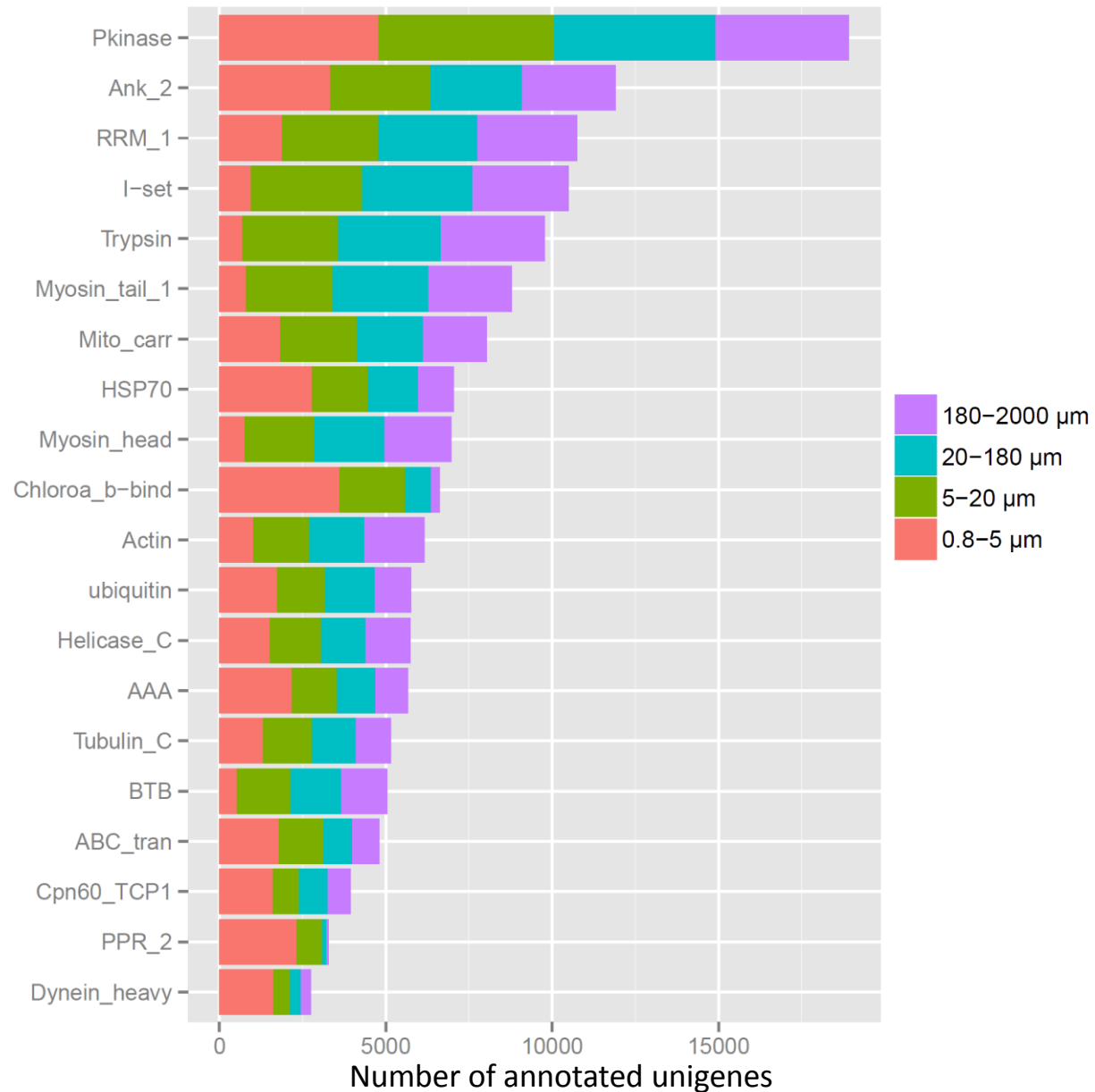


9,4 millions d'unigènes



19% ont un match pfam

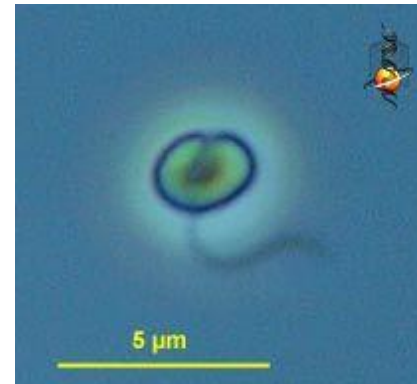
Quels sont les domaines de fonction les plus rependus au sein du plancton ?



Impact du fer sur la prolifération d'un stramenopile

Pelagomonas calceolata

stramenopile autotrophe



Impact du fer sur la prolifération d'un stramenopile

3 transcriptomes de référence dans
le pacifique (21 768 gènes)



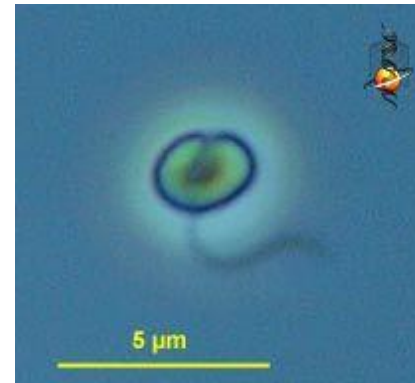
BLAST contre les unigènes
(95% Identity 80% length)



Identification de 14092 unigènes

Pelagomonas calceolata

stramenopile autotrophe



Impact du fer sur la prolifération d'un stramenopile

3 transcriptomes de référence dans
le pacifique (21 768 gènes)



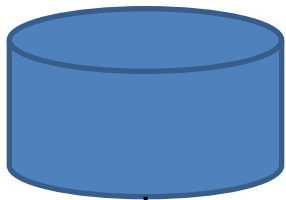
BLAST contre les unigènes
(95% Identity 80% length)



Identification de 14092 unigènes



Base de donnée



RPKM MetaG (0.8-5µm)

→ Elimination de 408 outliers

RPKM MetaT (0.8-5µm)

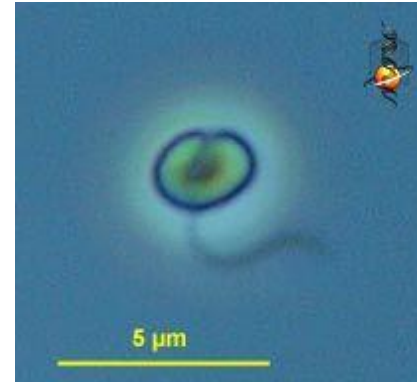
→ Expression des 13684 unigènes inliers

Pfam

→ Etude fonctionnelle

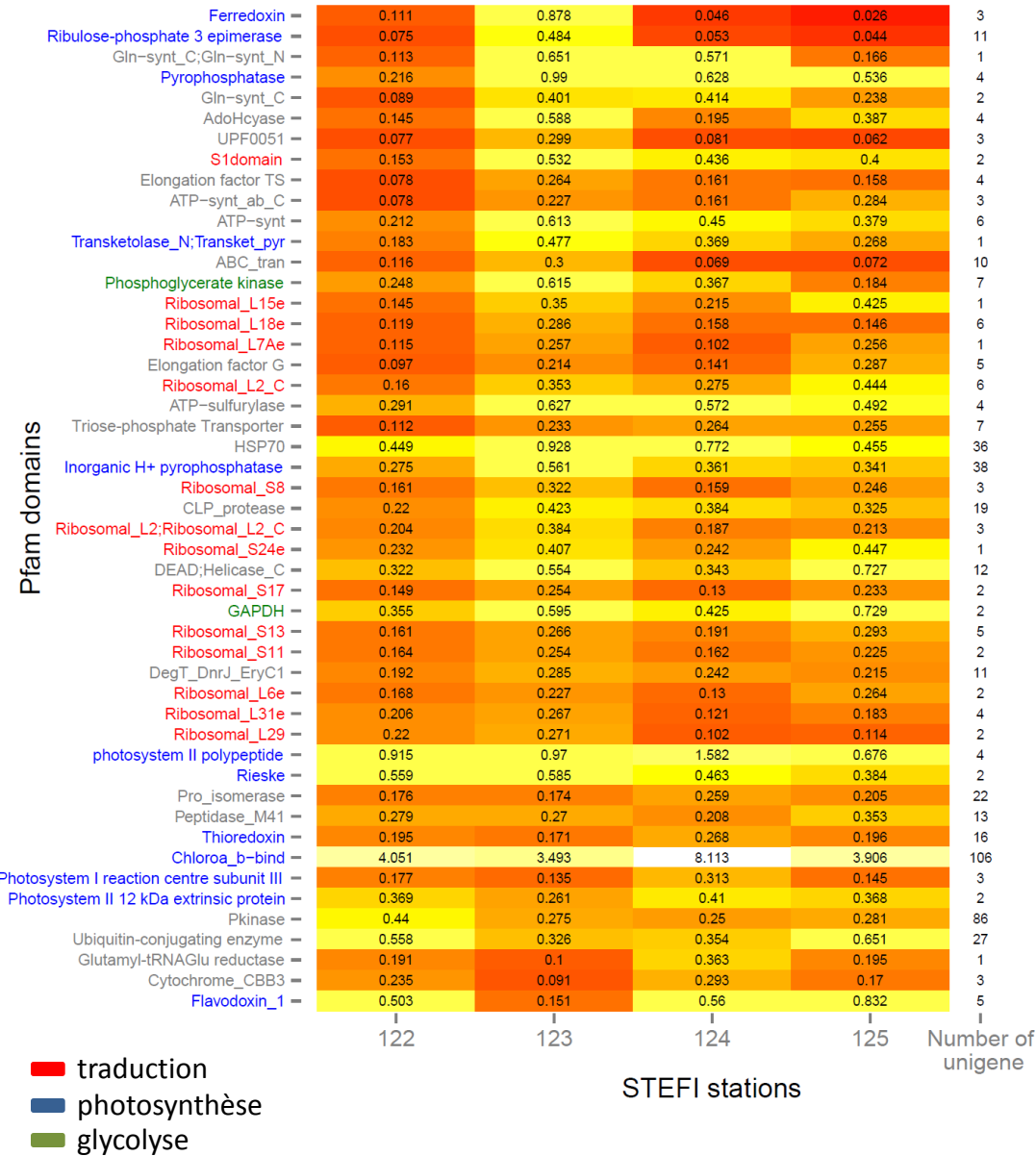
Pelagomonas calceolata

stramenopile autotrophe



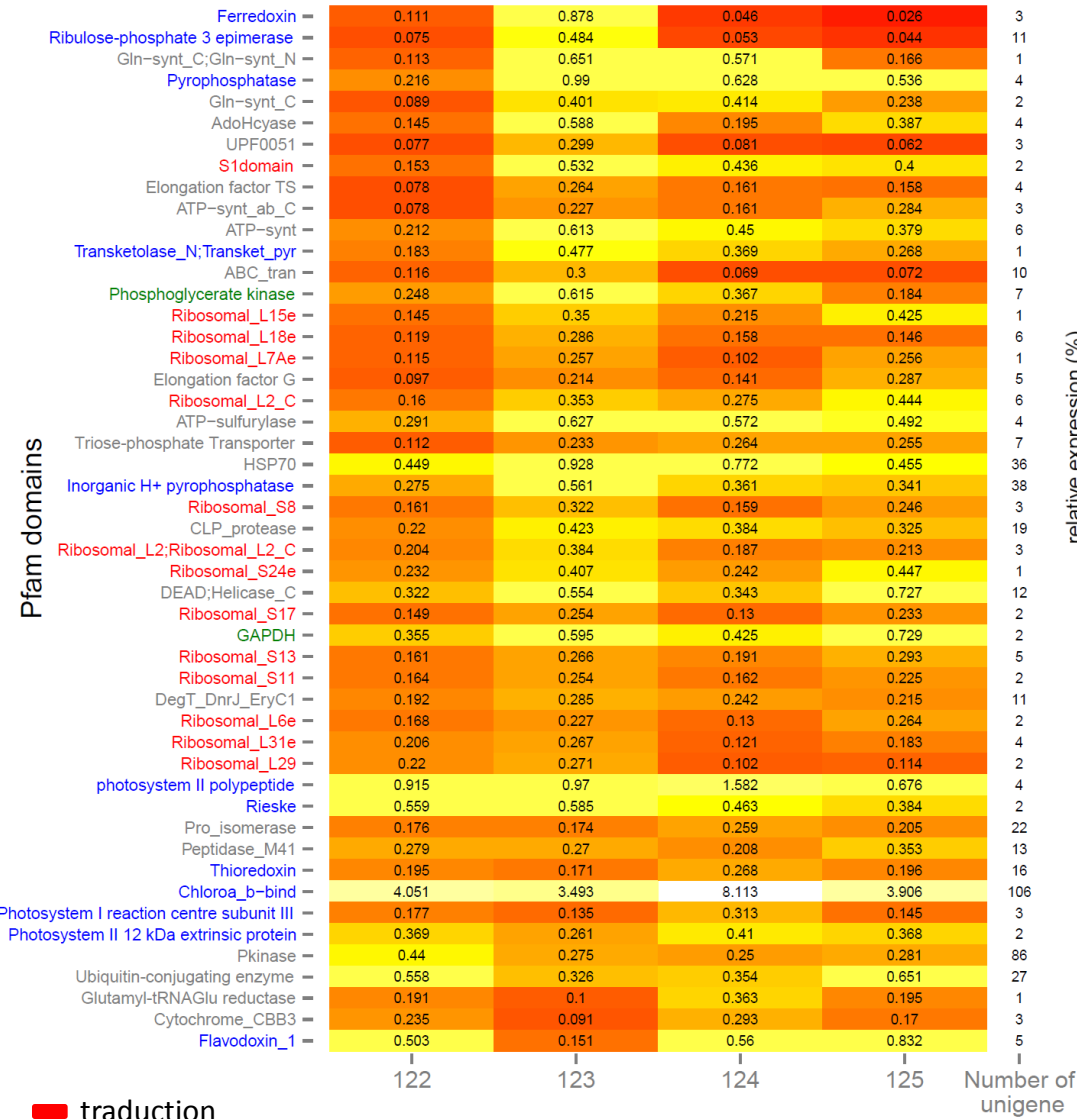
Impact du fer sur la prolifération d'un stramenopile

Heatmap des 50 Pfam les plus exprimés

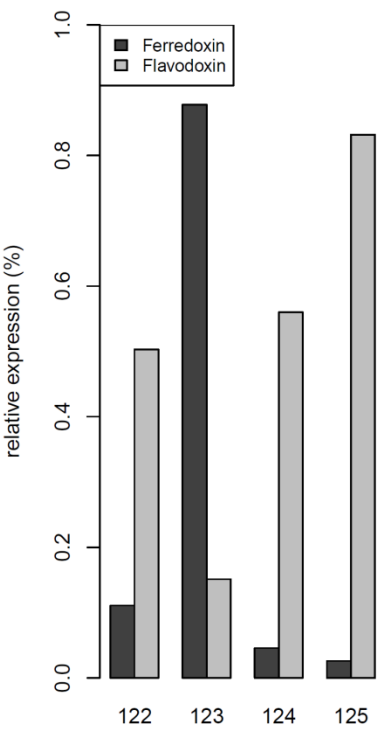


Impact du fer sur la prolifération d'un stramenopile

Heatmap des 50 Pfam les plus exprimés



Photosynthèse

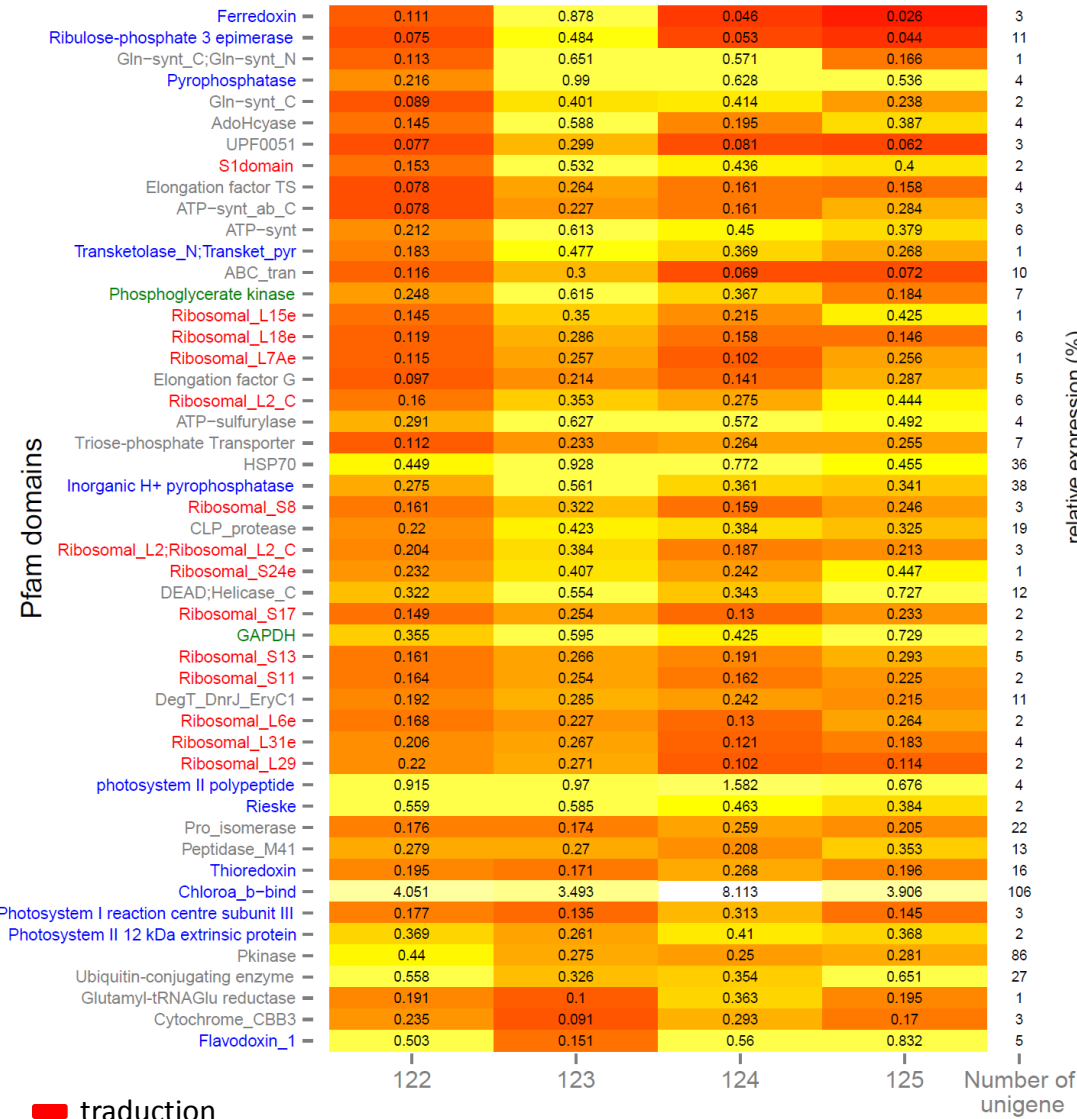


- traduction
- photosynthèse
- glycolyse

STEFI stations

Impact du fer sur la prolifération d'un stramenopile

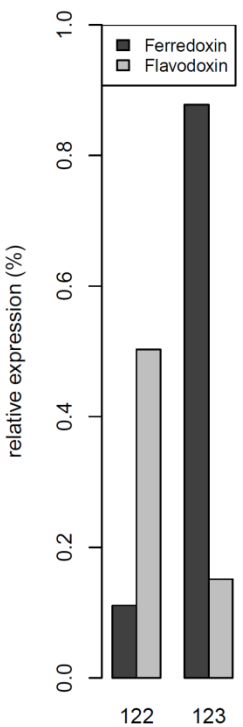
Heatmap des 50 Pfam les plus exprimés



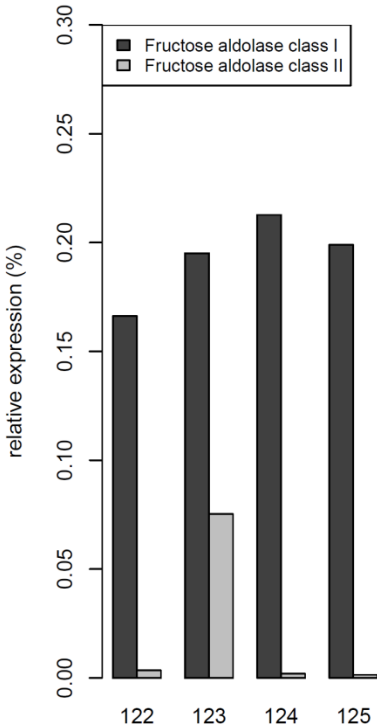
- traduction
- photosynthèse
- glycolyse

STEFI stations

Photosynthèse



Glycolyse



➡ Adaptation de *Pelagomonas* à l'absence de fer sur au moins deux voies métaboliques distinctes.

Projet, perspectives

- Utiliser ces méthodes d'analyse à l'échelle de toutes les stations TARA.
 - Tara Océans, 429 échantillons sur 68 stations (calculs principaux terminés)
 - Tara Polar Circle (séquençage et assemblage en cours).

Projet, perspectives

- Utiliser ces méthodes d'analyse à l'échelle de toutes les stations TARA.
 - Tara Océans, 429 échantillons sur 68 stations (calculs principaux terminés)
 - Tara Polar Circle (séquençage et assemblage en cours).
- Faire une description globale du transcriptome des eucaryotes.

Projet, perspectives

- Utiliser ces méthodes d'analyse à l'échelle de toutes les stations TARA.
 - Tara Océans, 429 échantillons sur 68 stations (calculs principaux terminés)
 - Tara Polar Circle (séquençage et assemblage en cours).
- Faire une description globale du transcriptome des eucaryotes.
- Répondre à des questions plus spécifiques:
 - étude d'organismes d'intérêt
 - étude de certaines voies métaboliques.
 - étude de régions océaniques particulières.

Remerciements

Laboratoire d'analyse des Génomes Eucaryotes (LAGE)

Patrick Wincker

Eric Pelletier

Olivier Jaillon

Samuel Mondy

Marc Wessner

Yoann Seeleuthner

Thomas Vannier



Consortium TARA Océan

21 équipes

7 nationalités





Affiliations taxonomiques

Bases de données :

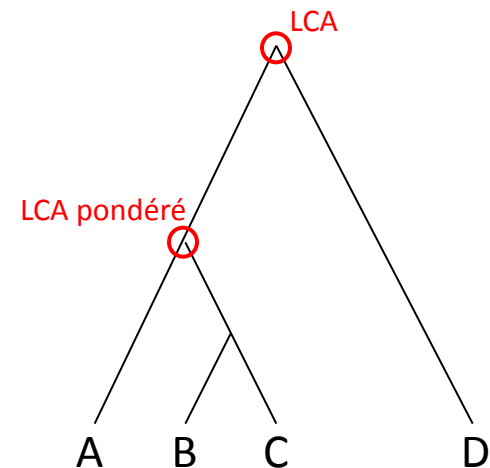
- UniRef 90.
- MMETSP (Marine Microbial Eukaryote Transcriptome Sequencing Project)
- SAGs du projet Tara, 37 espèces différentes (océan indien)

Blastx :

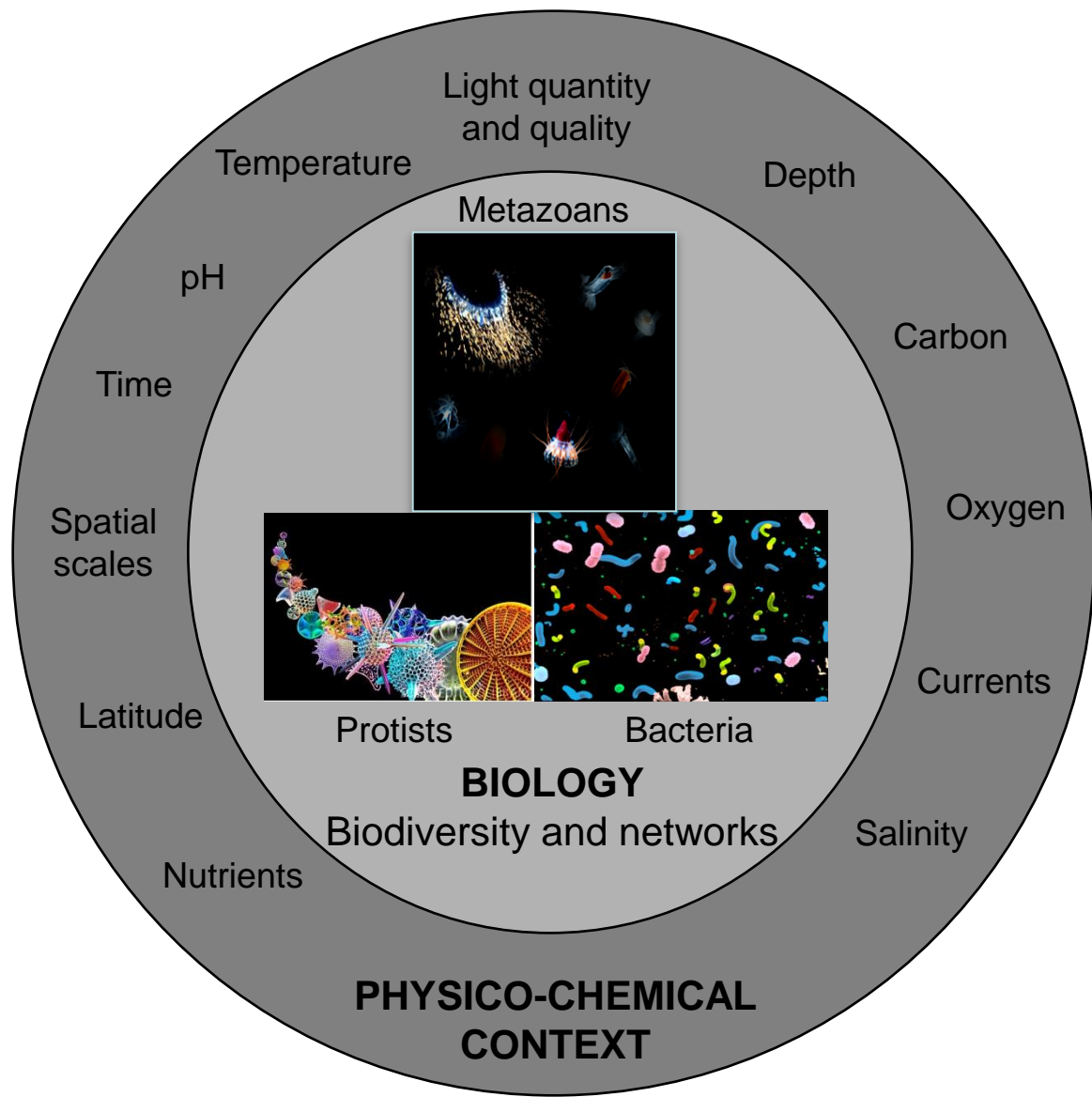
- e-value $< 10^{-5}$
- Bitscore $> 75\%$ du meilleur match

Méthodes d'affiliation :

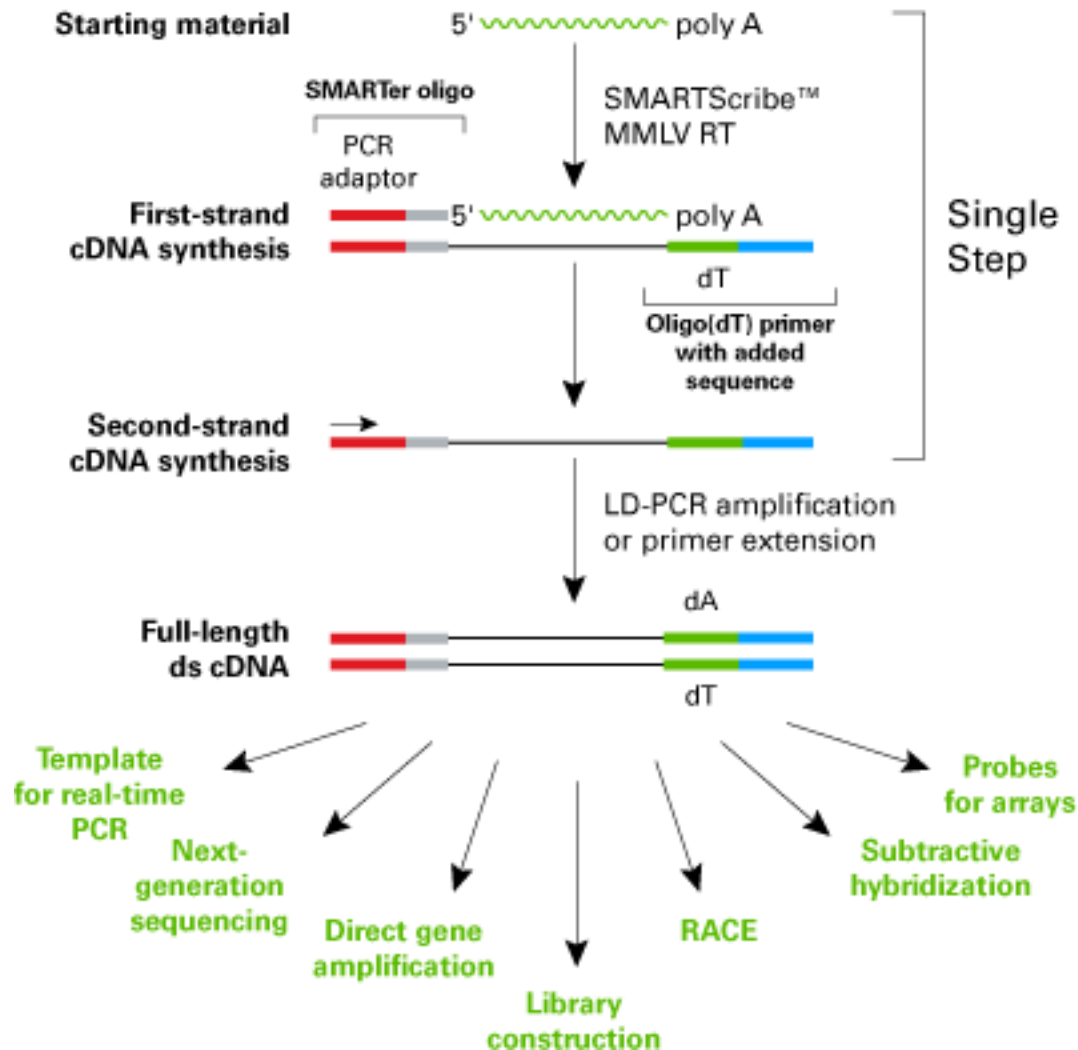
- LCA (Last Commun Ancestor)
- LCA pondéré (2/3)



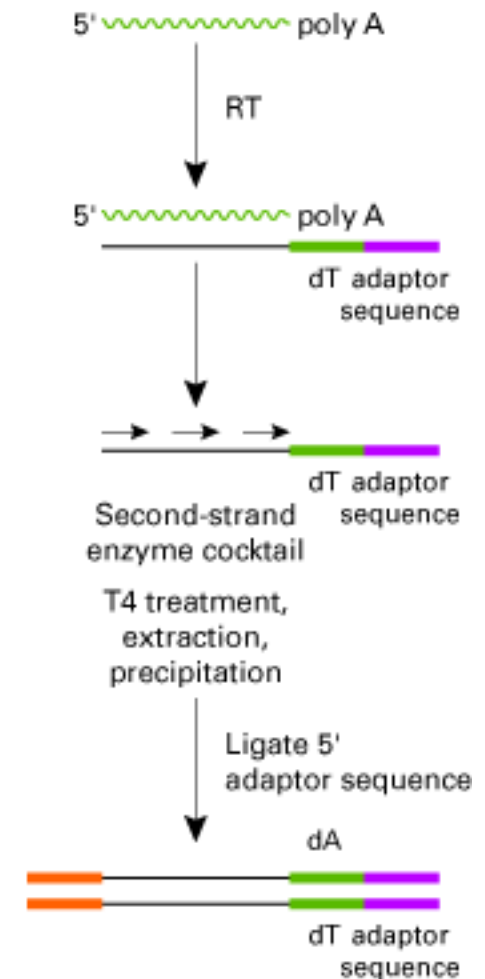
Corrélations aux paramètres environnementaux



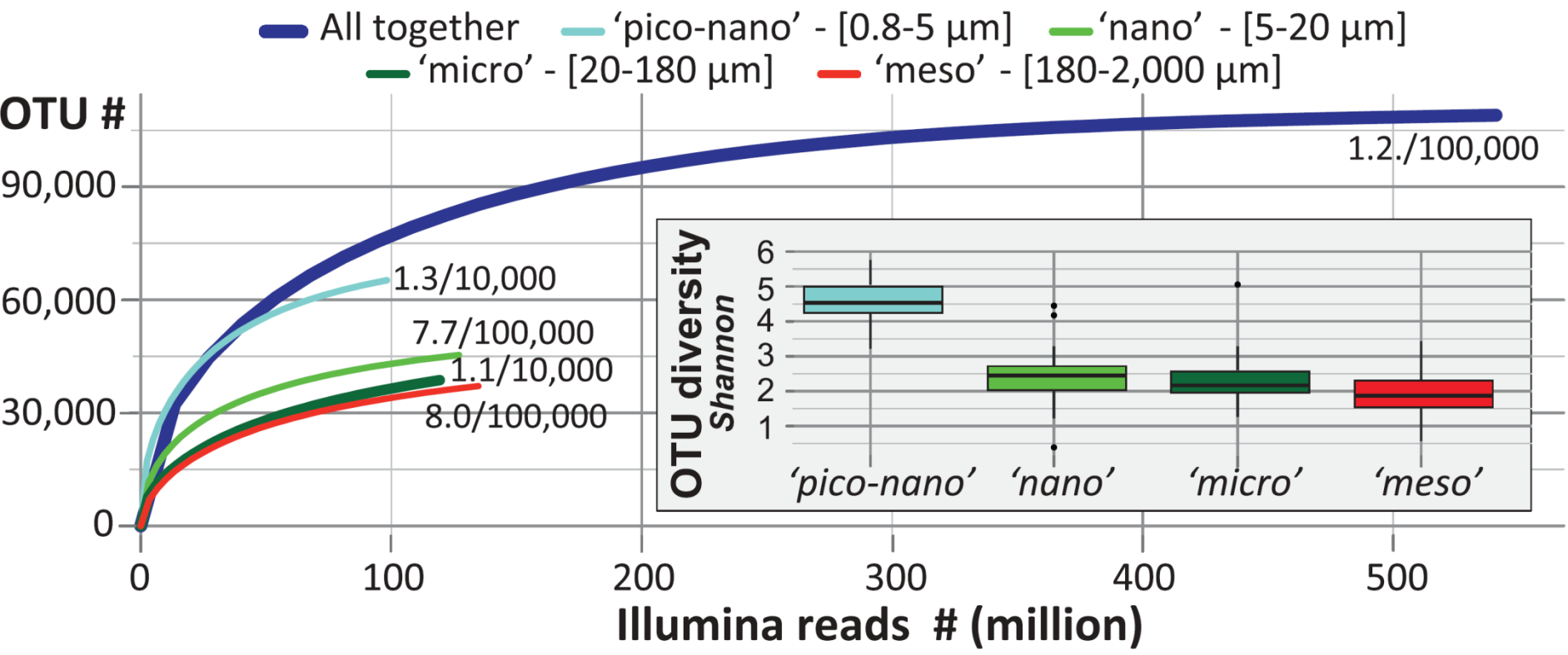
SMARTer cDNA Synthesis



Conventional cDNA synthesis

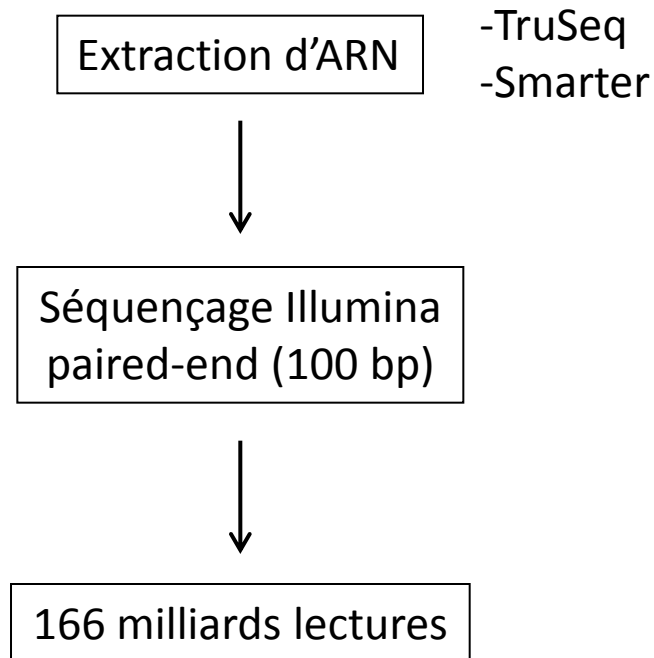


Eukaryotic plankton diversity in the sunlit global ocean



Etude des Métatranscriptomes eucaryotes

- 68 stations
 - 2 profondeurs (surface et DCM)
 - 4 filtres (0.8-5 μm 5-20 μm 20-180 μm 180-2000 μm)
- 429 échantillons différents



Construction d'un catalogue de référence

