Inferring phytoplankton community composition during *Pseudo-nitzschia* blooms using metatranscriptomic samples

Prigent Léa[1], Julien Quere[1], Audrey Duval[2], Martin Plus[1], Mickael Le Gac[1]

[1]Ifremer, DYNECO/PELAGOS, B.P. 70, 29280 Plouzané, France

[2]Ifremer, LERBO, B.P 40537, 29185 Concarneau, France

lea.prigent@ifremer.fr

G-ST-2

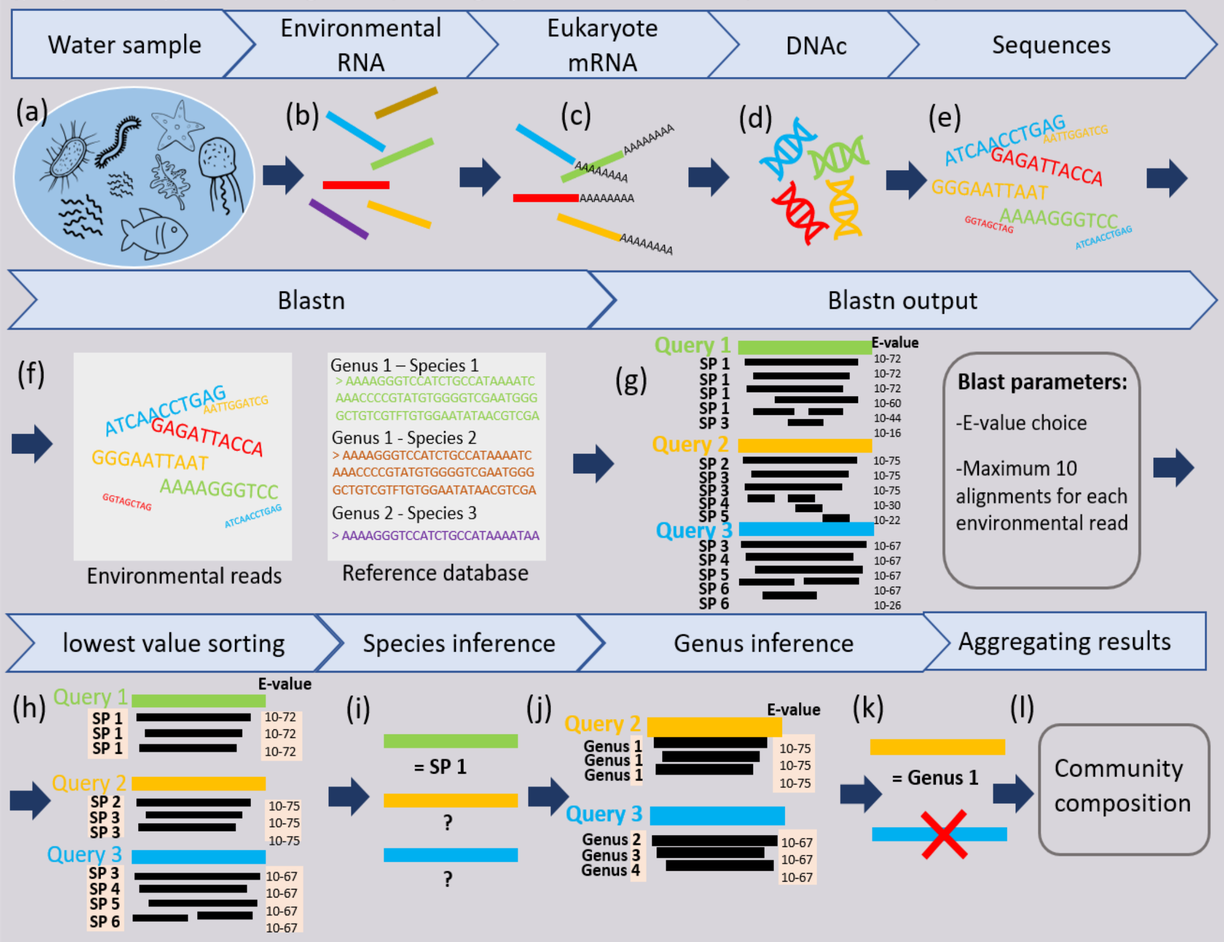
I/ Objective

Understanding harmful algal bloom requires characterizing community composition at the species level.

The objective of the present study was to develop an **analysis method** that complements existing methods such as light microscopy or overcomes certain limitations of methods (which require specific molecular development and/or targeted sequencing approaches) to identify the **composition of the community** during a *Pseudo-nitzschia* bloom from **metatranscriptomic samples**.

II/ Analysis pipeline

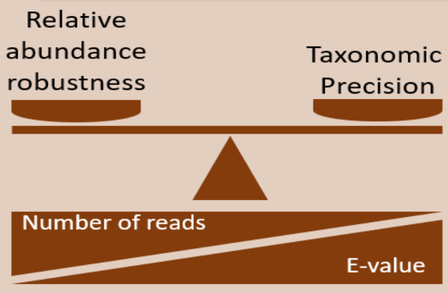
After filtrating 5 liter of environmental samples, RNA was extracted and sequenced to quantify the community composition at various taxonomic level (phylum, species or genus) using specific markers (18S and rbcl).



Two databanks were considered in the study:

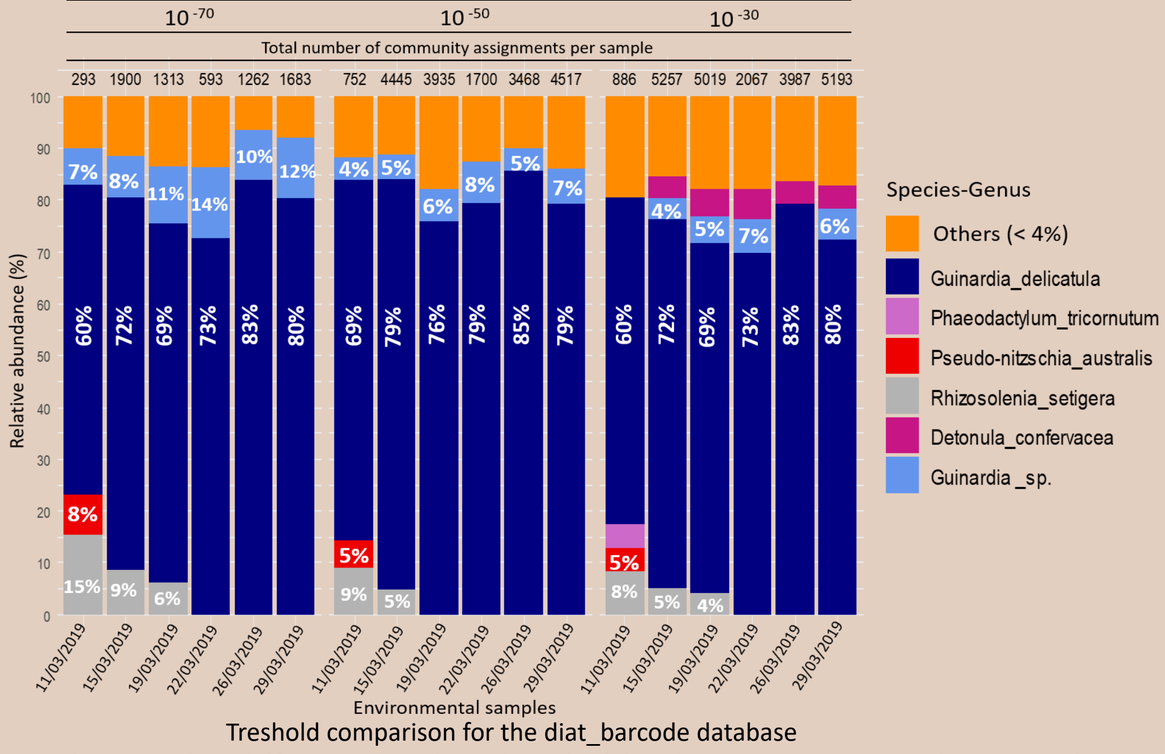
- **Protist Ribosomal Reference database (PR²) : eukaryotic community level** (18S ribosomal marker)
- **Diat_barcode** : **diatoms level** (Rbcl chloroplastic marker)

III/ E-value threshold determination



The homology threshold (e-value) choosen to assign taxonomy is a **compromise** between **precision in taxonomy** and **robustness of relative abundance estimation**.

After excluding homologies with an e-value >10^{-30} (too many mismatchs; short alignments; inconsistent taxonomy assignments) we compared three thresholds for the two databases.



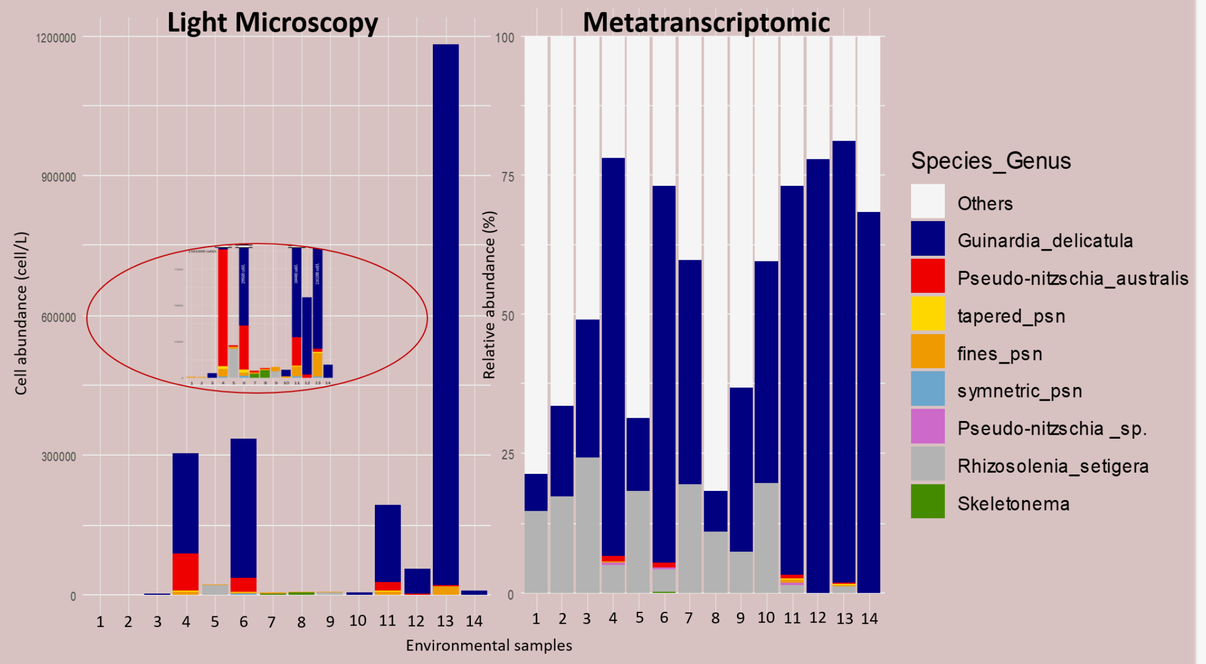
Treshold comparison for the diat_barcode database

According to the different e-value thresholds, dominant species (or genus) are in the same relative abundance.

10^{-30} was selected for diat_barcode databank and 10^{-70} for the PR² databank

IV/ Light microscopy-Metatranscriptomic comparison

Comparison between diatom community composition estimated using light microscopy (only dominant species and *Pseudo-nitzschia,* absolute abundances) and metatranscriptomic samples (diat_barcode; relative abundances).

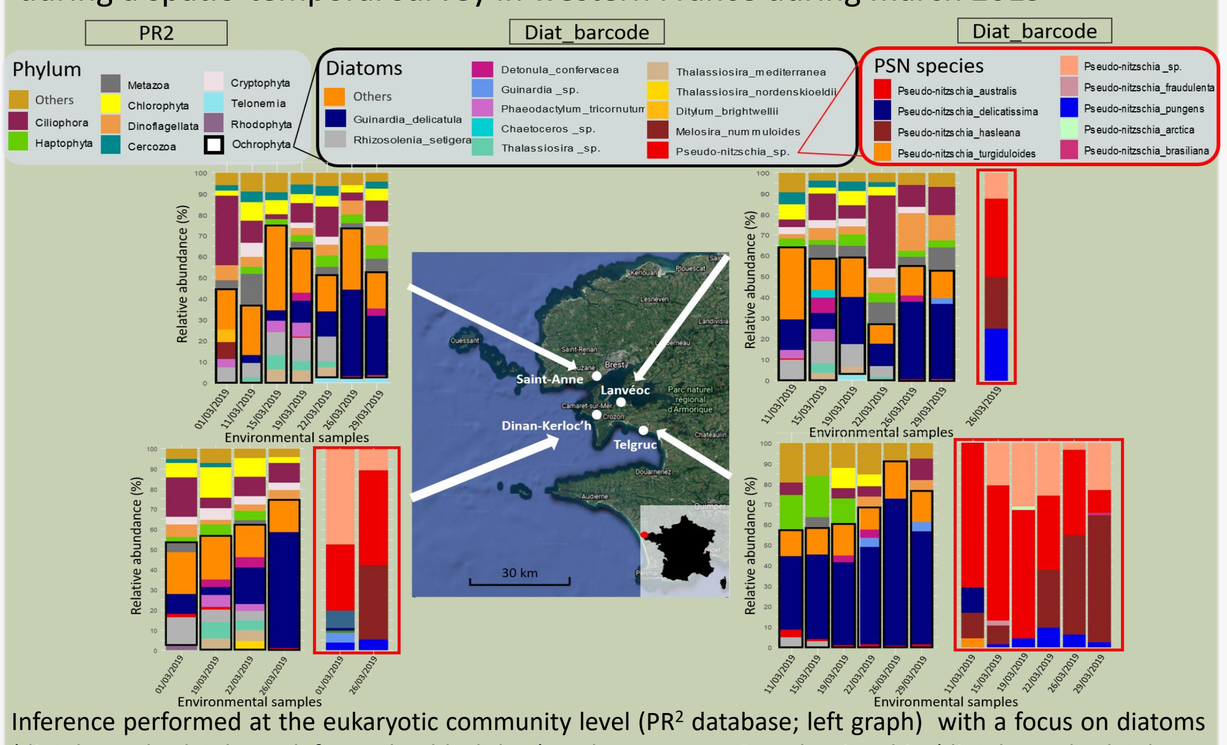


A zoom on the absolute abundances of light microscopy is presented in the red circle on the left graph.

Dominant taxa and relative abundance are usually consistant even if minor discrepencies may be identified

V/ Community composition during a spatio-temporal survey

Community composition infered from metatransriptomic datasets samples during a spatio-temporal survey in western France during march 2019



Inference performed at the eukaryotic community level (PR² database; left graph) with a focus on diatoms (diat_barcode database, left graphs, black box) and a zoom on Pseudo-nitzschia (diat_barcode database, right graphs, red box).

Overall, the most abundant phyla was diatom, with a very low relative abundance of *Pseudo-nitzschia*, mostly of *Peudo-nitzschia australis*.

VI/ Comparison of different methods

Comparison of the advantages and disadvantages of different community analysis methods.

	Advantages	Disadvantages
Light microscopy	-Cheapest method -Quick overview of the majority taxa -Absolute quantification	-Time consuming for full analyses -High level of expertise needed -Detection of only one part of the community (>20 µm) -Operator effect -Taxonomic resolution -Sample size -Constant time of analyse per sample
Metatranscriptomic	-No need specific protocol or targeted sequencing -Metabolically active community identification -Identification of cryptic species -Detection of small microbial communities (<20 µm) -Quickly decreasing time of analyse per sample	-High cost -Incomplete reference databases -Bias can be introduce during the different laboratory steps. -Relative
Metabarcoding	-Detection of small microbial communities (<20 µm) -Quickly decreasing time of analyse per sample	-Low taxonomic resolution -High cost -Incomplete reference databases -Bias can be introduce during the different laboratory steps. -Relative
qPCR	- Quickly decreasing time of analyse per sample - Absolute quantification of targeted species - Detect target sequences at low concentrations	-Focus on a specific group : requires existing knowledge of taxa that may be present in monitoring samples -Need to develop specific markers and protocols