Phytoplankton metabarcoding

Clarisse Lemonnier

BIOLAWEB

Summary

Barcode choice and primers design

Sampling

DNA extraction

Bioinformatic analysis

Funded by European Union

Phytoplankton metabarcoding

Sampling	DNA extraction	Barcode amplification	Sequencing	Bioinformatic analysis
	?	?	 Illumina	?

Funded by European Union

Phytoplankton metabarcoding

Sampling	DNA extraction	Barcode amplification	Sequencing	Bioinformatic analysis
	?	?	 Illumina	?

Which barcode ?

Primers design

Funded by European Union

Which barcode to use for phytoplankton metabarcoding?

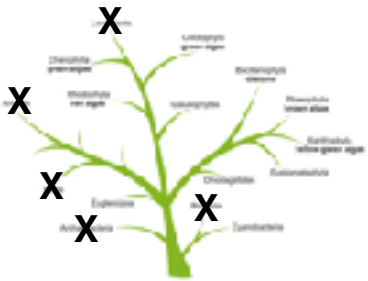
barcode

Marker gene

Conserved regions

1. Universal

2. Specific



3. Variable enough

Different species of Chlorophyceae

4. Match sequencing technology size

Max 450pb

5. Represented in reference databases

Funded by European Union

Which barcode to use for phytoplankton metabarcoding?

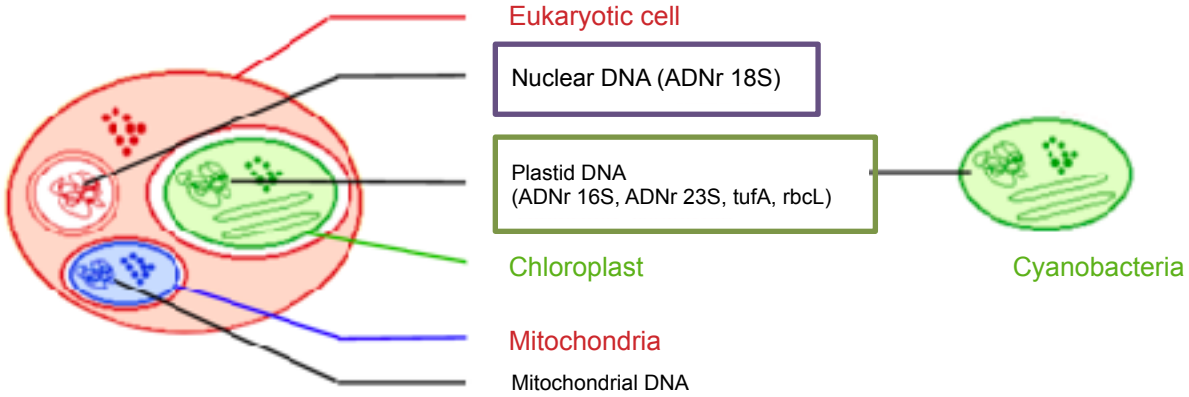
1. Selection of interesting marker genes

2. Selection of primers that are present in the different marker genes

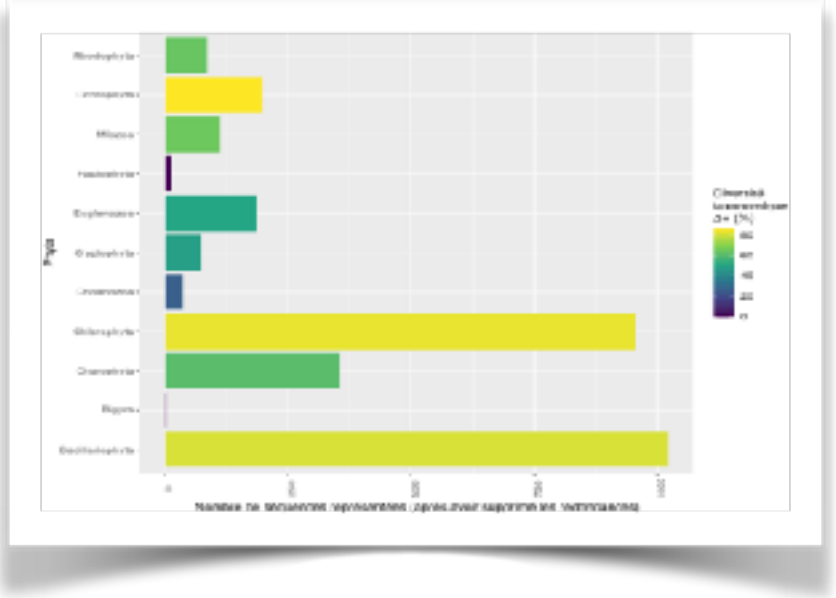
3. Test the performance of the primer

Which barcode to use for phytoplankton metabarcoding?

1. Selection of interesting marker genes



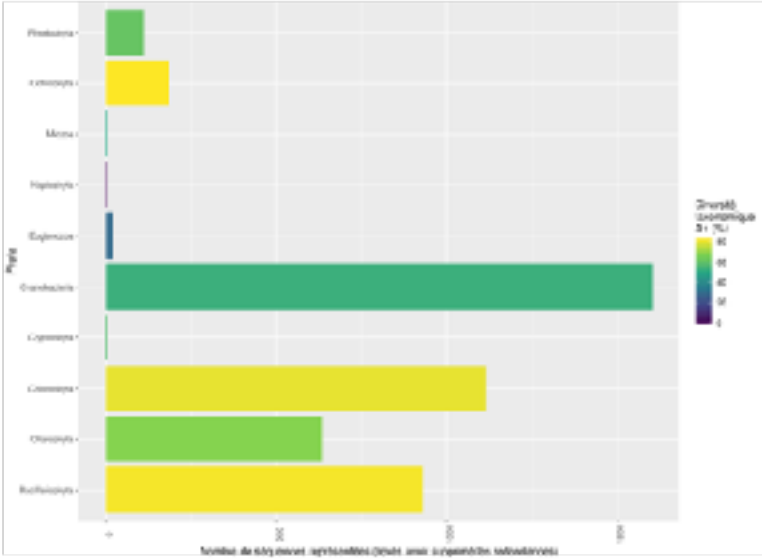
Nuclear DNA - ADNr 18S



- References available: huge, good reference databases (PR2, SILVA), often used in ecological studies
- Several highly variable regions flanked by conserved regions
- Not universal: 18S is the SSU of eukaryotic ribosome, so not present in prokaryotes (cyanobacteria)
- Not variable enough for species detection (e.g. diatoms)
- Length: presence of introns in some groups (Euglenophyta, 430 bp -> 680 bp)

Plastid DNA - rbcL

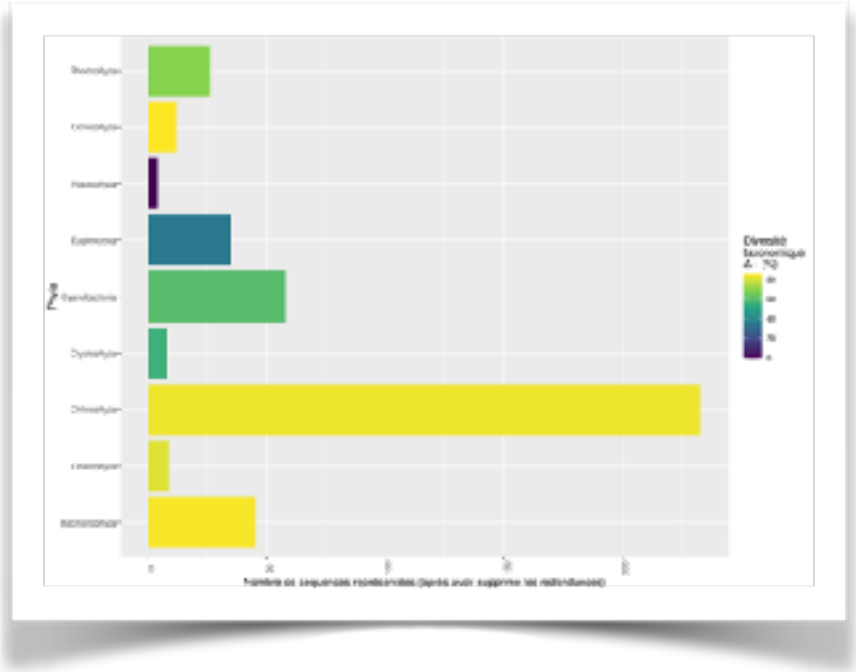
- Universal: present in all algal groups (!)
- Variable enough for species detection (e.g. diatoms), recommended for plant detection (CBOL Plant Working Groupe, 2009)
- References available: no references for some algal groups
- No conserved regions at the microalgal scale: primers must be specific of each algal class



Funded by European Union

Plastid DNA tufA

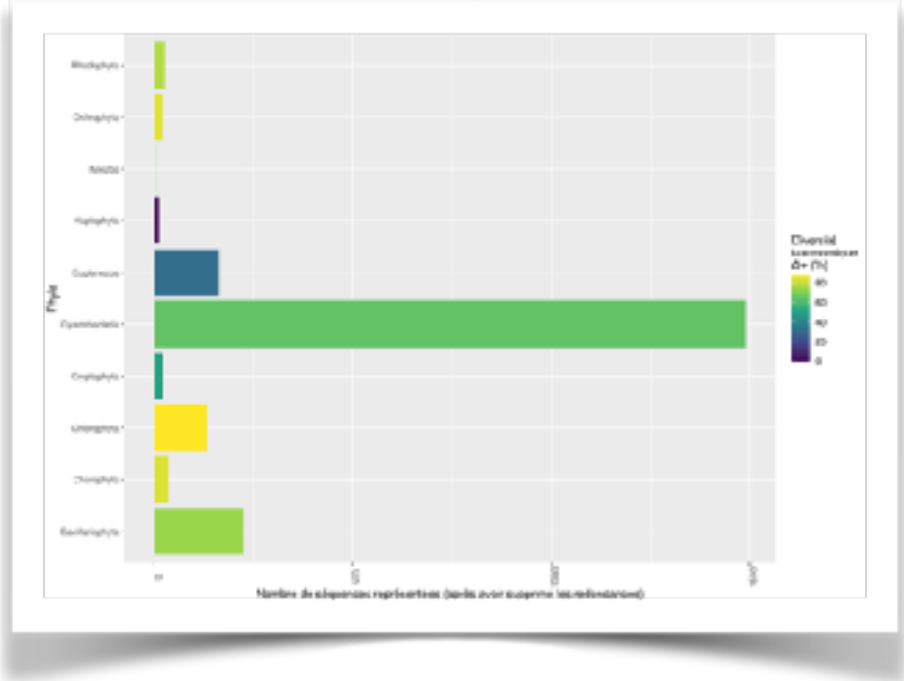
- Universal: present in all algal groups
- Variable enough for species and even sub-species detection (Vieira et al., 2016 ; Zou et al., 2016)
- References available: too few references available for many algal groups
- Primers already developed in the literature are specific of each clade. Difficulty to design new ones for the entire microalgal diversity



Funded by European Union

Plastid DNA16S

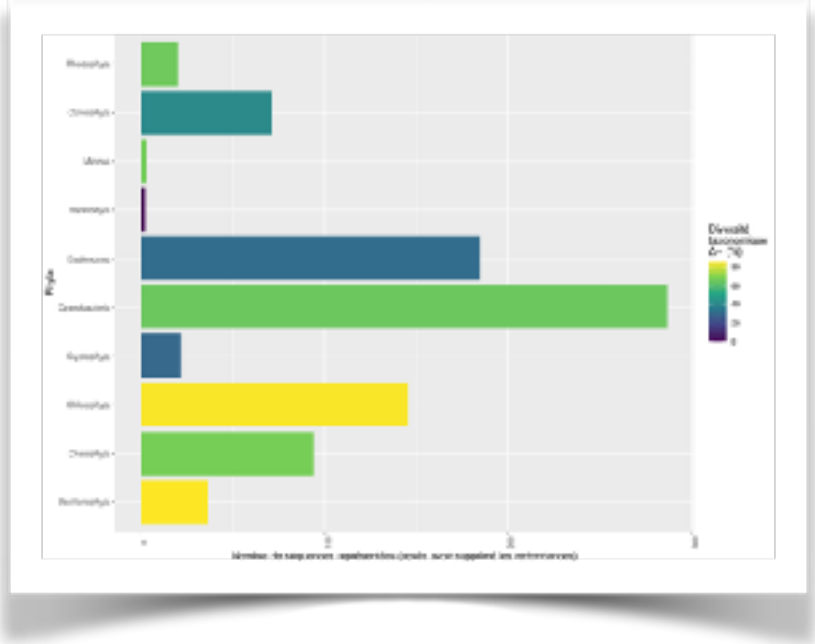
- Universal: present in all algal groups
- References: very good for cyanobacteria, and correct for other classes
- Several highly variable regions flanked by conserved regions
- Primers already developed in the literature for study microalgal diversity.



Funded by European Union

Plastid DNA - 23S

- Universal: present in all algal groups
- References: correct for cyanobacteria, still poor for other classes
- Several highly variable regions flanked by conserved regions
- Primers already developed in the literature for study microalgal diversity, targeting domain V.
- Higher phylogenetic resolution than 16S (Gutell et al., 1994 ; Pei et al., 2009)

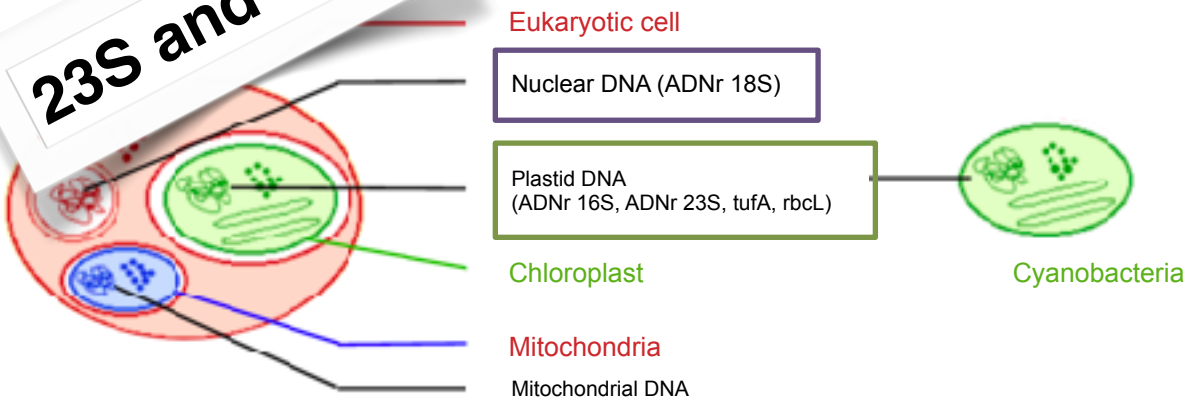


Funded by European Union

Which barcode to use for phytoplankton metabarcoding?

1. Selection of [illegible]sting marker genes

23S and 16S



Which barcode to use for phytoplankton metabarcoding?

2. Selection of primers that are present in the different marker genes

From literature


JOURNAL ARTICLE

The distribution of phytoplankton in the Baltic Sea assessed by a prokaryotic 16S rRNA gene primer system

C M Sinkko, F Polifrone, A Müller, R Hanson, B Kreikemeyer, M Labrenz

Author Notes

Journal of Plankton Research, Volume 40, Issue 3, May-June 2018, Pages 244–254,

https://doi.org/10.1093/plankt/fby003

Published: 07 April 2018 Article History

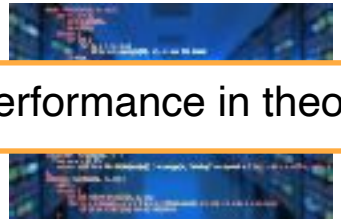
Newly designed

<u>12 candidate primers</u>

Funded by European Union

3. Test the performance of the primers

1 ***In silico* tests**



Performance in theory

variability, specificity, amplicon efficiency adapted for PCR

1st selection

2 ***In vitro* tests**

Performance in practice

Using mock communities of known composition

Final selection

?

In silico analysis

Collection of candidate primers pairs for 16S and 23S

- Selection of existing primers pairs (literature)
- New combinaison of existing primers
- Design of new primers

OligoAnalyzer

Amplification efficiency

Estimated resolution

Specificity of the primers

In silico assignation efficiency

In silico tests

OligoAnalyzer

Amplification efficiency

Estimated resolution

Specificity of the primers

In silico assignation efficiency

Are the primers adapted for PCR ?

T°C hybridation

Cross dimer

Hairpin

In silico tests

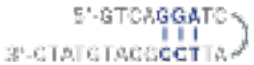
OligoAnalyzer	Amplification efficiency	Estimated resolution	Specificity of the primers	In silico assignation efficiency

Are the primers adapted for PCR ?

T°C hybridation

Cross dimer

Hairpin



Look at primer matches on sequences present in databases

How many sequences do I get with my primers?

In silico tests

OligoAnalyzer	Amplification efficiency	Estimated resolution	Specificity of the primers	In silico assignation efficiency
Are the primers adapted for PCR ?	*Look at primer matches on sequences present in databases*	*% of species with strictly different barcode*		
T°C hybridation	How many sequences do I get with my primers?			
Cross dimer 				
Hairpin 				

In silico tests

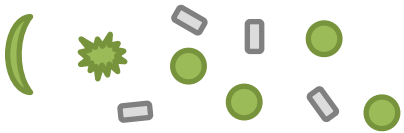
OligoAnalyzer	Amplification efficiency	Estimated resolution	Specificity of the primers	In silico assignation efficiency
Are the primers adapted for PCR ?	*Look at primer matches on sequences present in databases*	*% of species with strictly different barcode*	*To how many unwanted species the primers hybridate?*	

T°C hybridation

Cross dimer

Hairpin

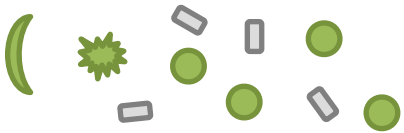
How many sequences do I get with my primers?



Funded by European Union

In silico tests

OligoAnalyzer	Amplification efficiency	Estimated resolution	Specificity of the primers	In silico assignation efficiency
Are the primers adapted for PCR ?	*Look at primer matches on sequences present in databases*	*% of species with strictly different barcode*	*To how many unwanted species the primers hybridate?*	% amplicon that are correctly assigned to their corresponding taxonomy
T°C hybridation	How many sequences do I get with my primers?			
Cross dimer				
Hairpin				



In silico analysis

Collection of candidate primers pairs for 16S and 23S

- Selection of existing primers pairs (literature)
- New combinaison of existing primers
- Design of new primers

OligoAnalyzer

Amplification efficiency

Estimated resolution

Specificity of the primers

In silico assignation efficiency

In silico analysis

Collection of candidate primers pairs for 16S and 23S

- Selection of existing primers pairs (literature)
- New combinaison of existing primers
- Design of new primers

OligoAnaly

Marker gene	Primer pairs	Good for PCR	Resolution	Specificity
16S	CYA359F/CYA781R	✓	●	✓
	ECLA16S_F1/ECLA16S_R1	✓	●	✗
23S	p23SrV_f1/p23SrV_r1	✓	✓	✓
	ECLA23S_F1/ECLA23S_R1	✓	✓	✓
	ECLA23S_F2/ECLA23S_R2	✓	✓	✓

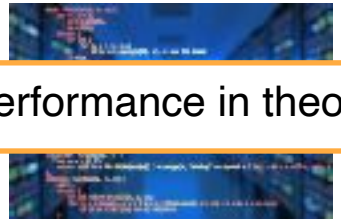
In silico
ssignation
fficiency

Funded by European Union

Primers test

1 **Do they match our requirements ?**

In silico



Performance in theory

variability, specificity, amplicon efficiency adapted for PCR

1st selection

2 **Do they perform well in real conditions ?**

In vitro

Performance in practice

Using mock communities of known composition

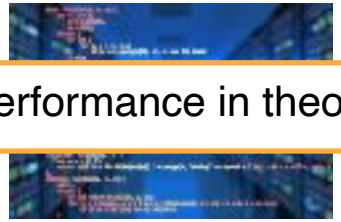
Final selection

?

Primers test

1 **Do they match our requirements ?**

In silico



Performance in theory

variability, specificity, amplicon efficiency adapte for PCR

1st selection

Gene marker	Primer pairs
16S	CYA359F/CYA781R
	ECLA16S_F1/ECLA16S_R1
23S	p23SrV_f1/p23SrV_r1
	ECLA23S_F1/ECLA23S_R1
	ECLA23S_F2/ECLA23S_R2

2 **Do they perform well in real conditions ?**

In vitro

Performance in practice

Using mock communities of known composition

Final selection

?

In vitro tests

2 **Do they perform well in real conditions ?**

In vitro

Performance in practice

Using mock communities of known composition

1st Mock community

Selection of 10 strains from the TCC culture collection

DNA extracts of each culture with GenElute

Equimolar concentrations → *Are all the species amplified with the same efficiency?*

PCR in triplicate

Sequencing (Illumina)

DADA2 (ASVs)

→ *Influence of PCR on diversity obtained*

In vitro tests

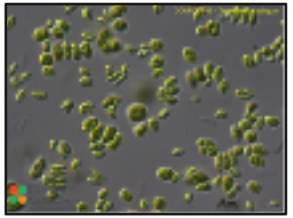
Species introduced in the first mock community

Cyanobacteria

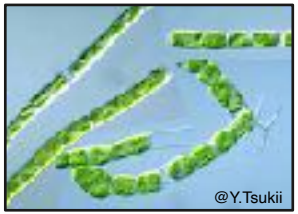
Charophyta

Cryptophyta

Chlorophyta



Ochrophyta

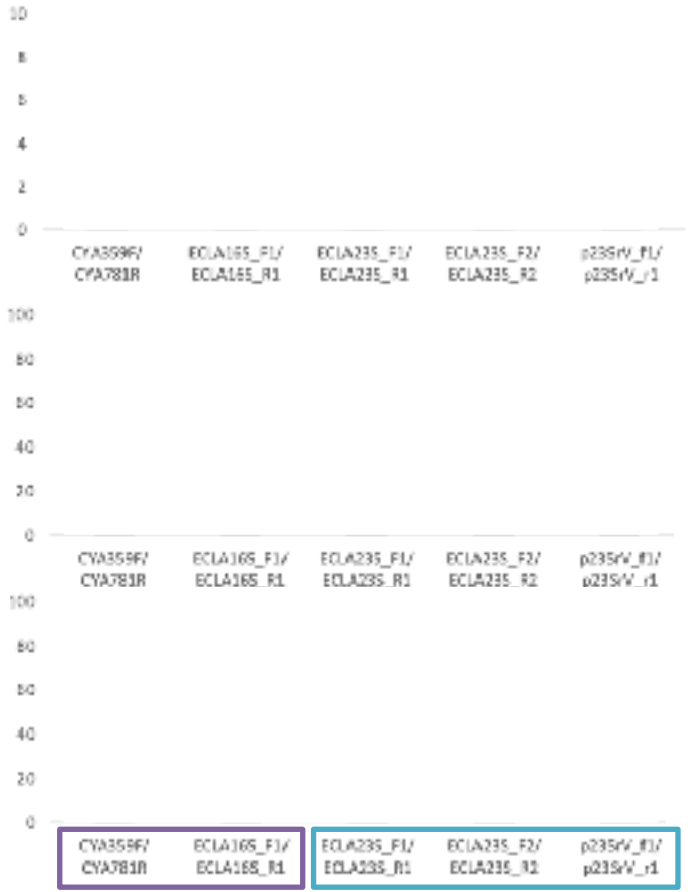


Bacillariophyta

Are all the 10 control species detected by each primer pairs?

Percentage of ASV assigned to control species of Mock 1.

Percentage of reads assigned to control species of Mock 1.



16S

23S

Are all the 10 control species detected by each primer pairs?

All species are detected with the 23S primers
But not for 16S primers

Percentage of ASV assigned to control species of Mock 1.

16S

23S

Percentage of reads assigned to control species of Mock 1.

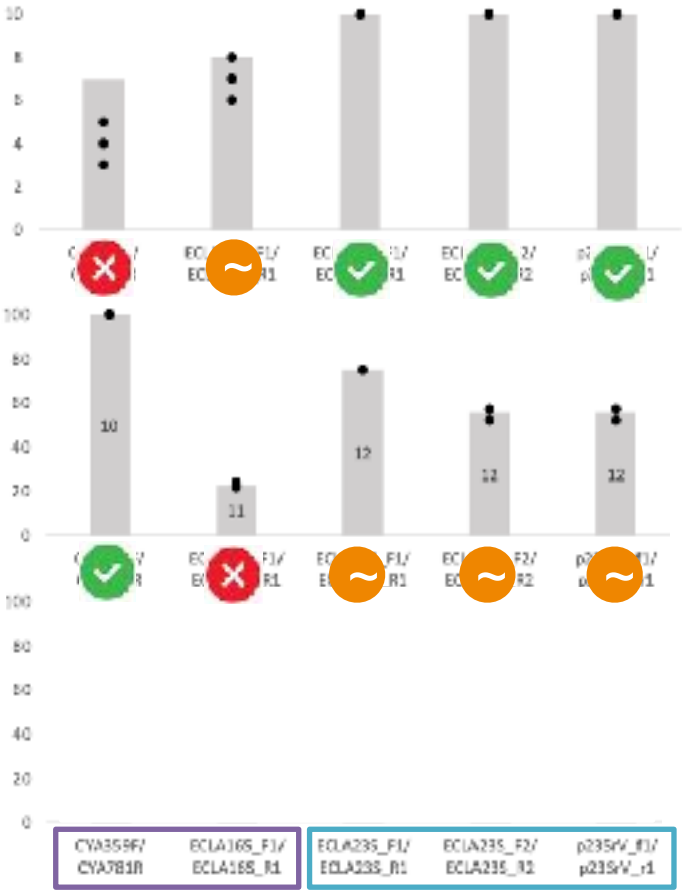
Are all the 10 control species detected by each primer pairs?

All species are detected with the 23S primers
But not for 16S primers

Percentage of ASV assigned to control species of Mock 1.

Some primers amplify heterotrophic bacteria especially ECLA16S (which was expected)

Percentage of reads assigned to control species of Mock 1.



16S

23S

Are all the 10 control species detected by each primer pairs?

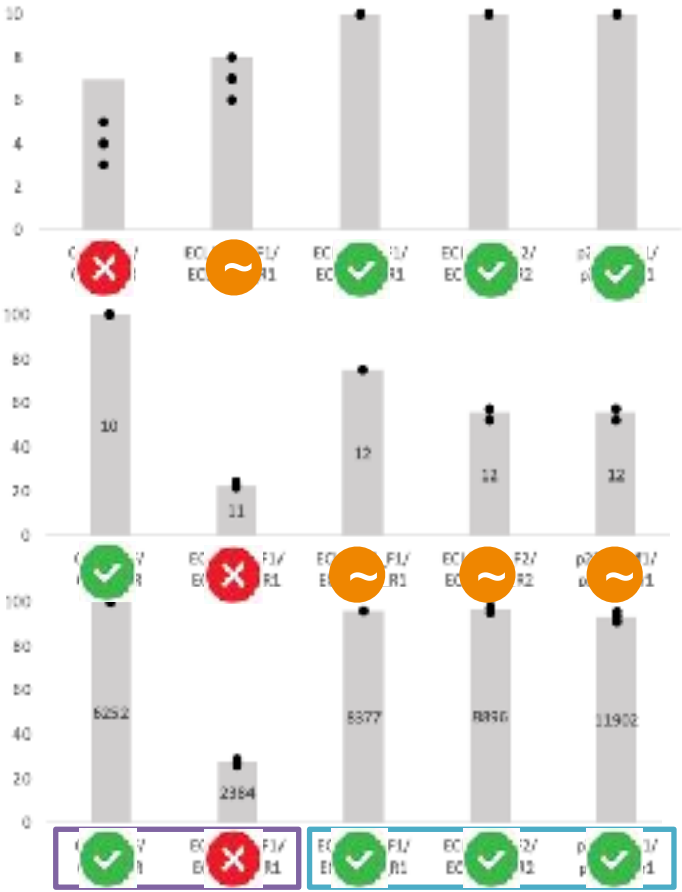
All species are detected with the 23S primers
But not for 16S primers

Percentage of ASV assigned to control species of Mock 1.

Some primers amplify heterotrophic bacteria especially ECLA16S (which was expected)

Percentage of reads assigned to control species of Mock 1.

An extremely large majority of reads are affiliated to algae



16S

23S

Funded by European Union

Are all the 10 control species detected by each primer pairs?

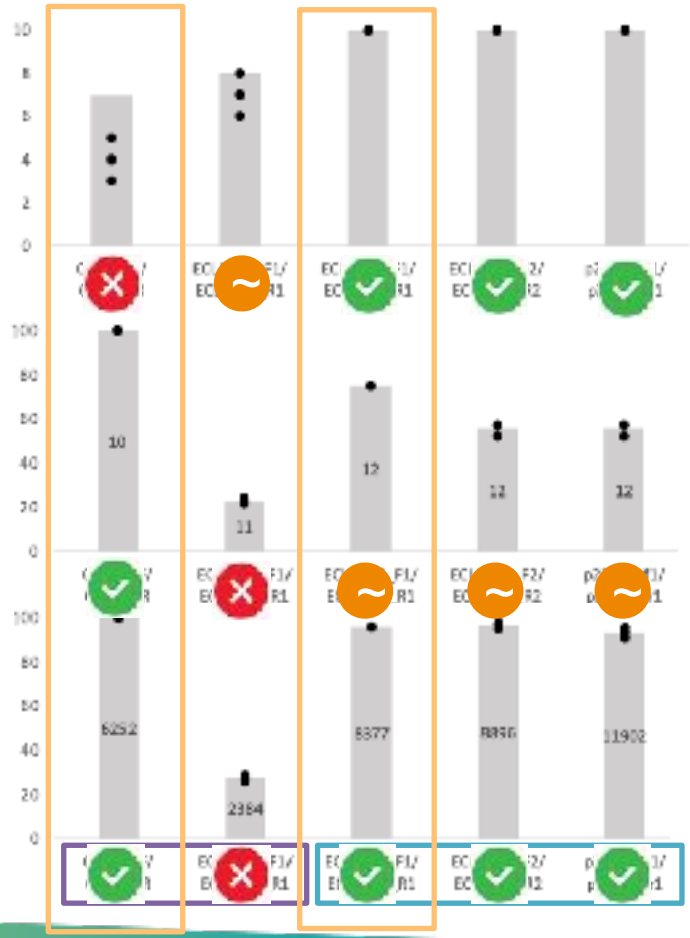
All species are detected with the 23S primers
But not for 16S primers

Percentage of ASV assigned to control species of Mock 1.

Some primers amplify heterotrophic bacteria especially ECLA16S (which was expected)

Percentage of reads assigned to control species of Mock 1.

An extremely large majority of reads are affiliated to algae



Funded by European Union

In vitro tests

2 **Do they perform well in real conditions ?**

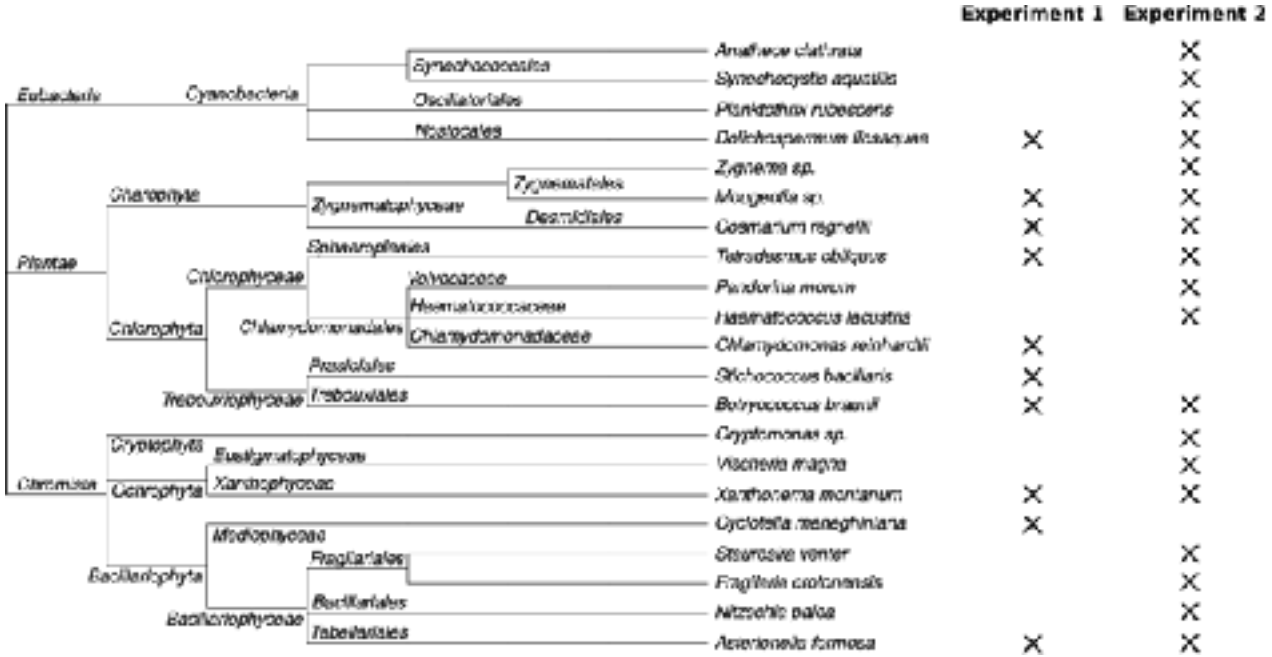
In vitro

Performance in practice

Using mock communities of known composition

	1st Mock community	*2nd Mock community*
	Selection of 10 strains from the TCC culture collection	Selection of **18** strains from the TCC culture collection
	16S : CYA359F/CYA781R	DNA extracts of each culture
		PCR in triplicate
	23S : ECLA23SF1/ECLA23SR1	Equimolar mix
		Sequencing (Illumina)
	DADA2 (ASV)	DADA2 (ASV)

Mock community 2



Mock community 2

	Experiment 2	16S	23S
Anaflece clathrata	X	X	X
Synechocystis aquatilis	X	X	X
Planktothrix rubescens	X	X	X
Dolichospermum flosaquae	X	X	X
Zygnema sp.	X	X	X
Mougeotia sp.	X	X	X
Cosmarium regnellii	X	X	X
Tetradesmus obliquus	X	X	X
Pandorina morum	X	X	X
Haematococcus lacustris	X	None	X
Chlamydomonas reinhardtii			
Stichococcus bacillaris			
Botryococcus braunii	X	X	X
Cryptomonas sp.	X	X	X
Visonema insigne	X	X	X
Xanthonema montanum	X	X	X
Cyclotella meneghiniana			
Stauroneis vanheurckii	X	X	X
Fragilaria crotonensis	X	X	X
Nitzschia palea	X	X	X
Asterionella formosa	X	X	X

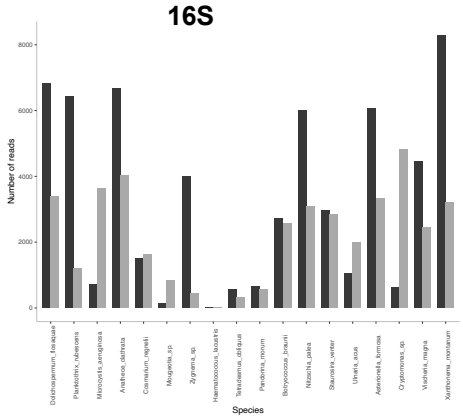
Well assigned Not well assigned

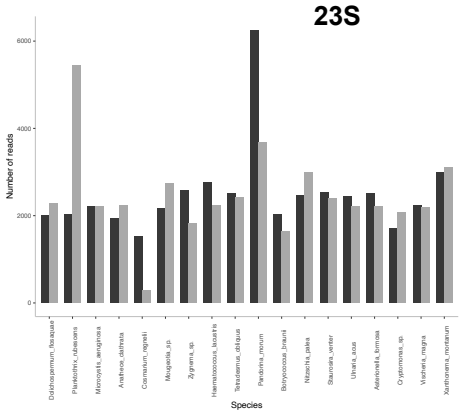
Funded by European Union

Mock community 2

	Experiment 2	16S	23S
Anathece clathrata	X	X	X
Synechocystis aquatilis	X	X	X
Planktothrix rubescens	X	X	X
Dolichospermum flosaquae	X	X	X
Zygnema sp.	X	X	X
Mougeotia sp.	X	X	X
Cosmarium regnellii	X	X	X
Tetradesmus obliquus	X	X	X
Pandorina morum	X	X	X
Haematococcus lacustris	X	None	X
Chlamydomonas reinhardtii			
Stichococcus bacillaris			
Botryococcus braunii	X	X	X
Cryptomonas sp.	X	X	X
Visonema insigne	X	X	X
Xanthonema montanum	X	X	X
Cyclotella meneghiniana			
Staurosira venter	X	X	X
Fragilaria crotonensis	X	X	X
Nitzschia palea	X	X	X
Asterionella formosa	X	X	X

Well assigned Not well assigned



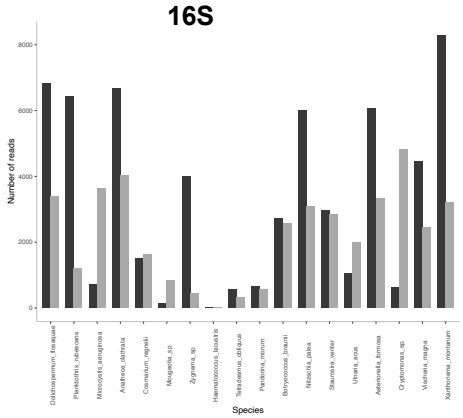


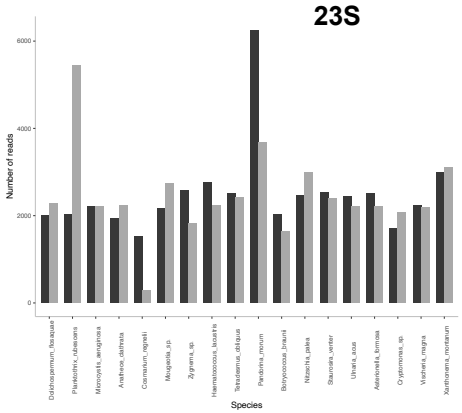
Funded by European Union

Mock community 2

Experiment 2	16S	23S
Anaflexe clathrata	✗	✗ ■	✗ ■
Synechocystis aquatilis	✗	✗ ■	✗ ■
Planktothrix rubescens	✗	✗ ■	✗ ■
Dolichospermum flosaquae	✗	✗	✗
Zygnema sp.	✗	✗ ■	✗ ■
Mougeotia sp.	✗	✗ ■	✗ ■
Cosmarium regnellii	✗	✗	✗
Tetradesmus obliquus	✗	✗ ■	✗ ■
Pandorina morum	✗	✗	✗ ■
Haematococcus lacustris	✗	None	✗ ■
Chlamydomonas reinhardtii			
Stichococcus bacillaris			
Botryococcus braunii	✗	✗ ■	✗ ■
Cryptomonas sp.	✗	✗ ■	✗ ■
Visonaria magna	✗	✗ ■	✗ ■
Xanthonema montanum	✗	✗	✗ ■
Cyclotella meneghiniana			
Staurosira venter	✗	✗ ■	✗ ■
Fragilaria crotonensis	✗	✗	✗
Nitzschia palea	✗	✗ ■	✗ ■
Asterionella formosa	✗	✗ ■	✗ ■

■ Well assigned ■ Not well assigned





23S primers performs better :
- **It allows to recover all of the 18 species of Mock 2**
- **It allows a better assignation to species level**
- **It gives more homogeneous abundances with equimolar input concentration of amplicon**

Funded by European Union

Phytoplankton metabarcoding

Sampling	*DNA extraction*	*Barcode amplification*	*Sequencing*	*Bioinformatic analysis*
	?	CYA359F/CYA781R ECLA23SF1/ECLA23SR1	?	?

Funded by European Union

Phytoplankton metabarcoding

Sampling	DNA extraction	Barcode amplification	Sequencing	Bioinformatic analysis
	?	CYA359F/CYA781R ECLA23SF1/ECLA23SR1	?	?

Funded by European Union

Phytoplankton metabarcoding

Sampling	DNA extraction	Barcode amplification	Sequencing	Bioinformatic analysis
	?	CYA359F/CYA781R ECLA23SF1/ECLA23SR1	?	?

Funded by European Union

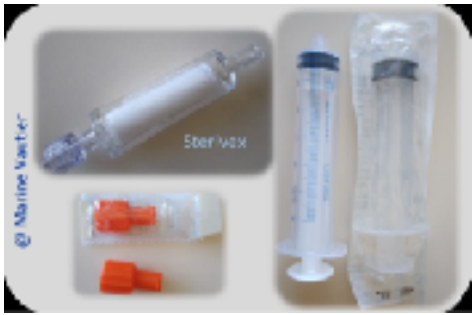
Sampling

The protocol had to be adapted for biomonitoring

- Simple
- Not time consuming (5min)
- Use minimum of material

Ensure that there are no cross-contaminations

- Clean material



Sample conservation - adapted to any condition

- Lysis buffer (Tris-EDTA-sucrose)

Funded by European Union

Sampling

Protocol :

Sample water with syringe

Fix the syringe to the sterivex

Filter the water through the filter (porosity 0.45µm)

Close the sterivex in one end

Add 2mL of lysis buffer

Close the sterivex in the other end and keep it in a plastic bag

Sampling

5min

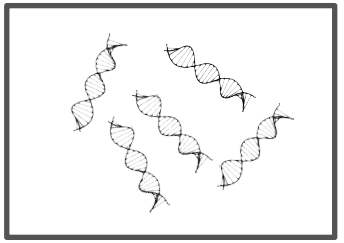
Funded by European Union

Phytoplankton metabarcoding

Sampling	DNA extraction	Barcode amplification	Sequencing	Bioinformatic analysis
	?	CYA359F/CYA781R ECLA23SF1/ECLA23SR1	 Illumina	?

Funded by European Union

DNA extraction

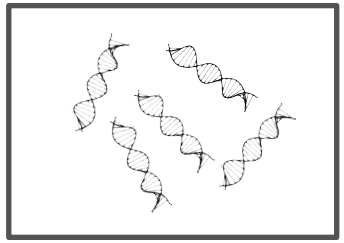


The protocol had to be adapted to Sterivex filters filled with lysis buffer

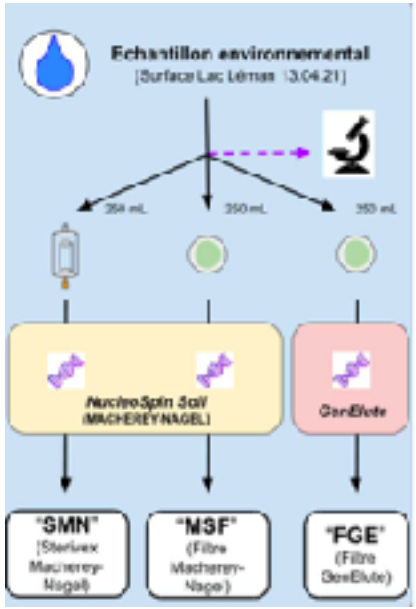
- Lysis step all done in Sterivex:
    - vortex step to detach cells from the filter
    - Use SDS, lysozyme and proteinase K

Which protocol?

DNA extraction



Test 1 : impact of DNA extraction kit, filtration type and primers



GenElute : fast and rapid DNA extraction
Nucleospin soil : more washing step + specific step of inhibitor removal

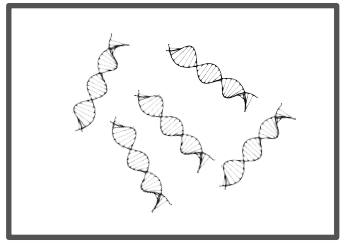
Filtration on membrane filters are routinely done in the lab, for Alpine lake biomonitoring

Primers tested for mock community 1

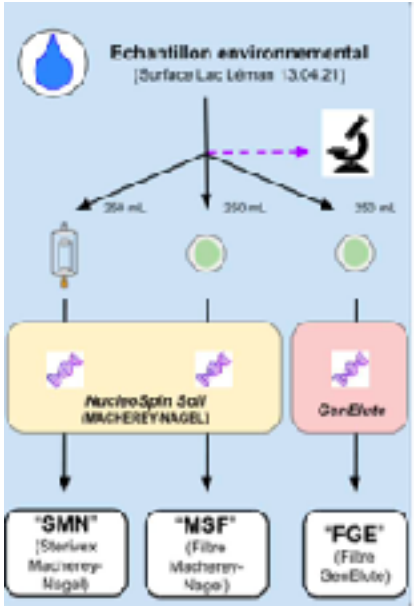
Gene marker	Primer pairs
16S	CYA359F/CYA781R
	ECLA16S_F1/ ECLA16S_R1
23S	p23SrV_f1/p23SrV_r1
	ECLA23S_F1/ ECLA23S_R1
	ECLA23S_F2/ ECLA23S_R2

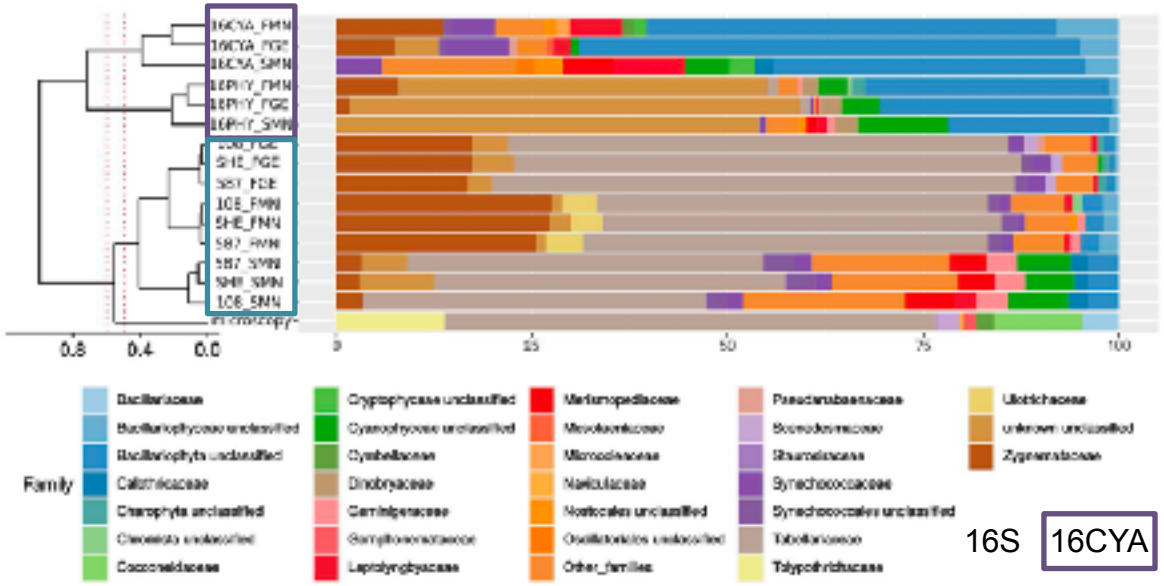
Funded by European Union

DNA extraction



Test 1 : impact of DNA extraction kit, filtration type and primers

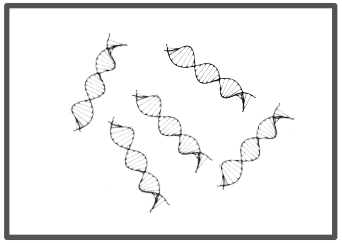




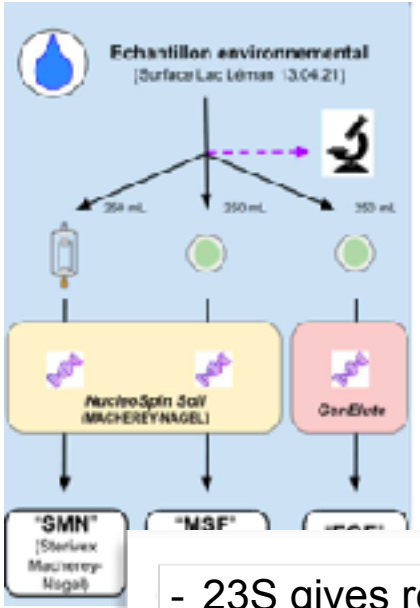
16S 16CYA 16PHY

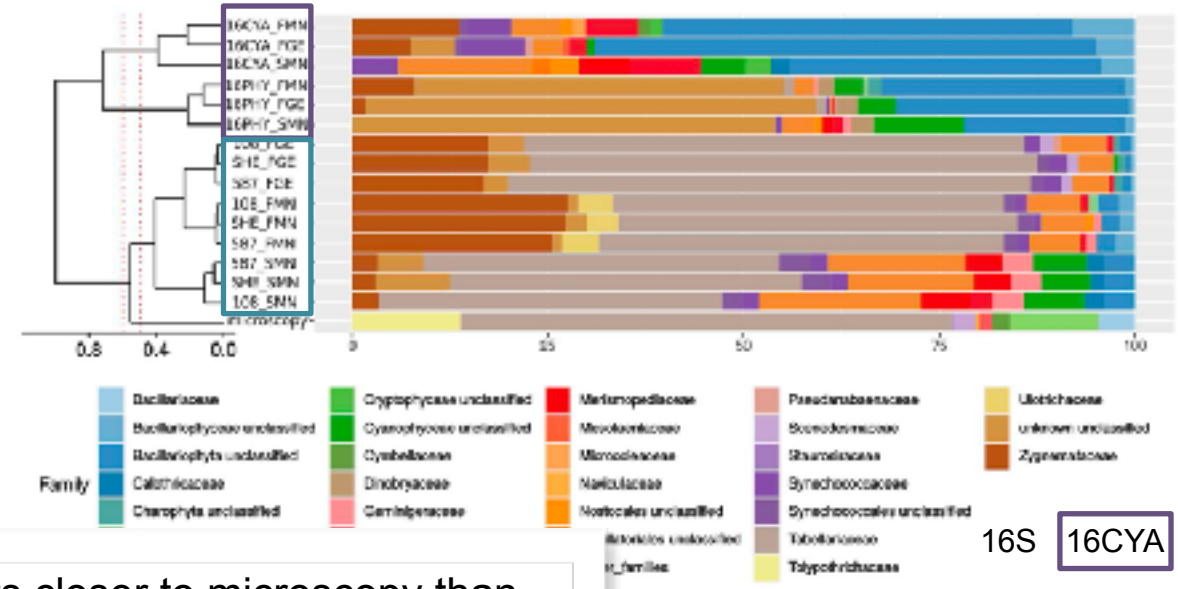
23S 108 SHE 587

DNA extraction



Test 1 : impact of DNA extraction kit, filtration type and primers



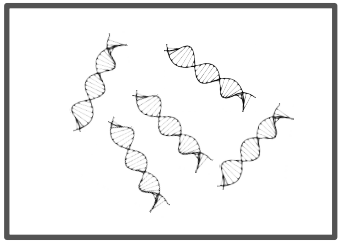


16S 16CYA 16PHY

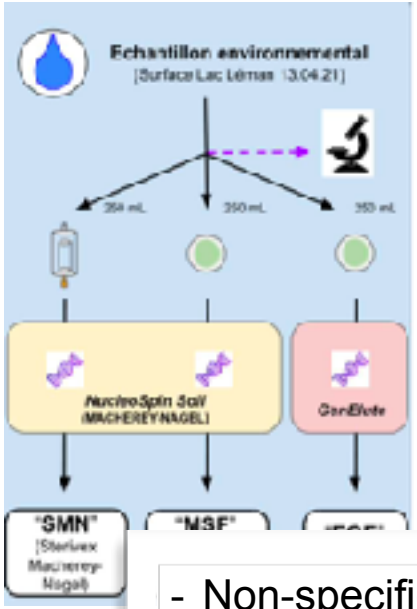
23S 108 SHE 587

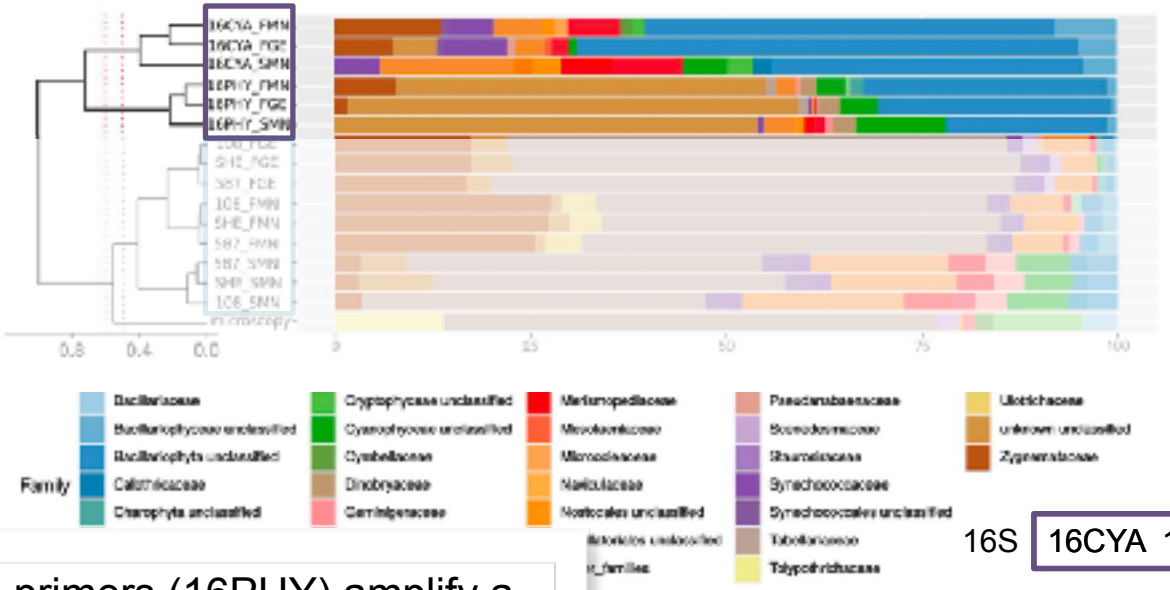
- 23S gives results closer to microscopy than 16S. Notably because taxonomic assignation is better

DNA extraction



Test 1 : impact of DNA extraction kit, filtration type and primers





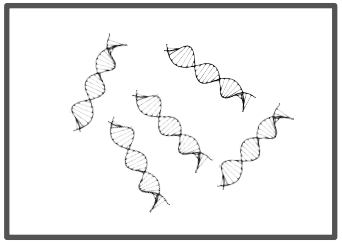
16S 16CYA 16PHY

23S 108 SHE 587

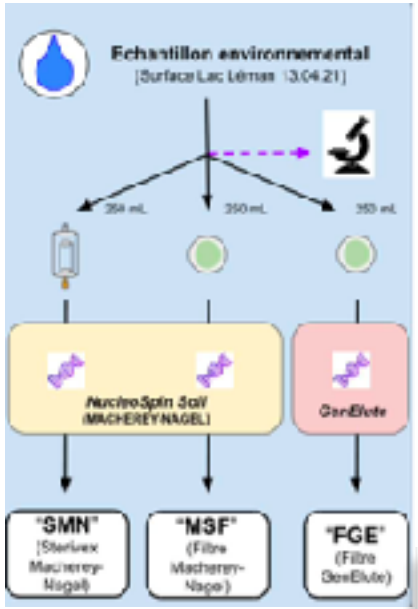
- Non-specific 16S primers (16PHY) amplify a lot of heterotrophic bacteria from filtered lake water

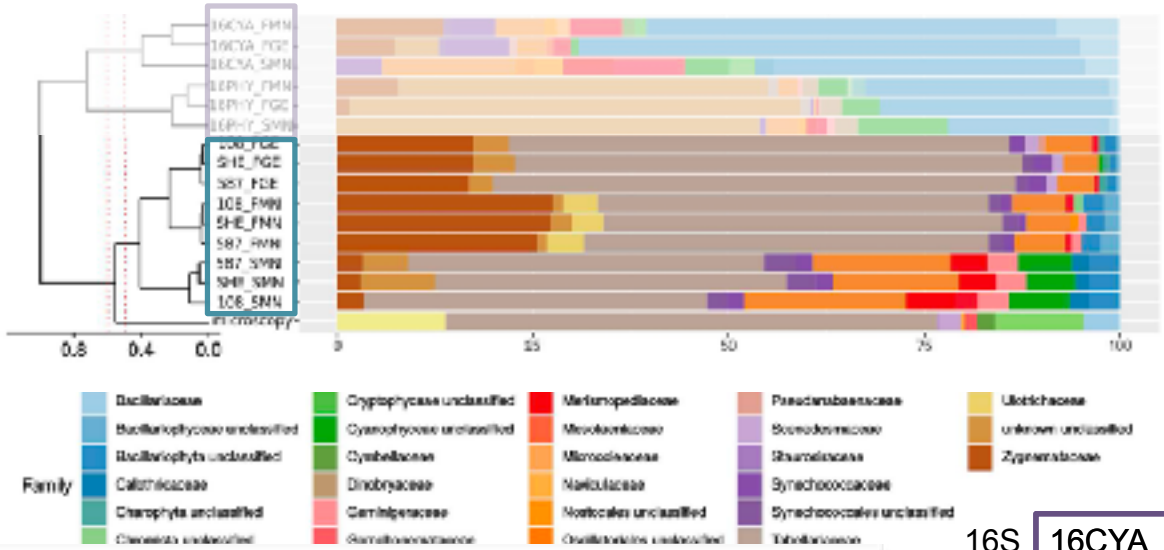
Funded by European Union

DNA extraction



Test 1 : impact of DNA extraction kit, filtration type and primers



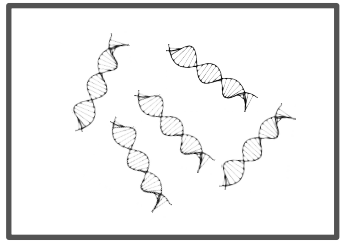


16S 16CYA 16PHY

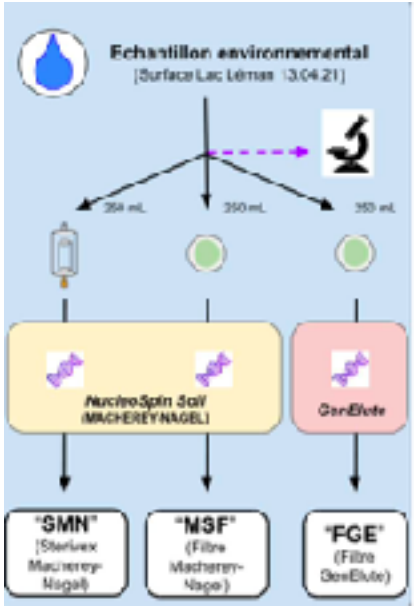
23S 108 SHE 587

- Both filtration method and DNA extraction protocol influence the phytoplankton diversity obtained at the end.

DNA extraction



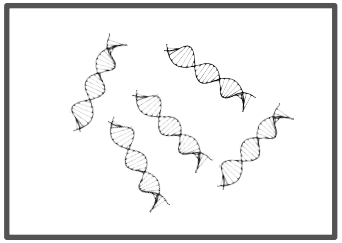
Test 1 : impact of DNA extraction kit, filtration type and primers



These results illustrate the importance to keep a single protocol within a study, to allow the direct comparison of phytoplankton diversity between samples.

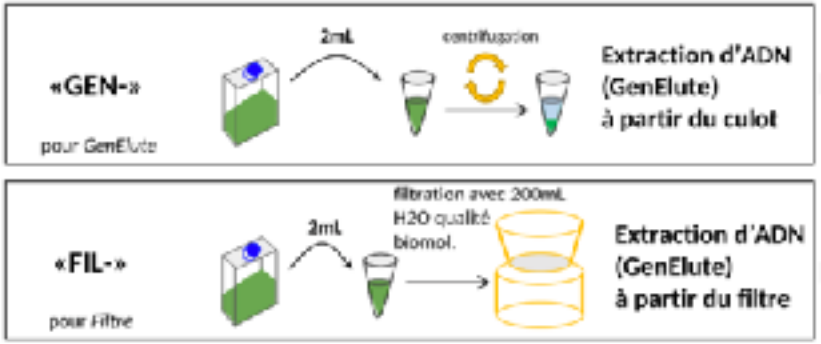
Funded by European Union

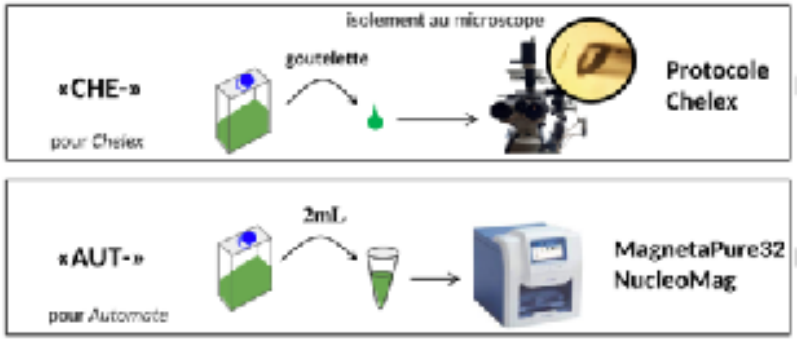
DNA extraction



Test 2 : impact of DNA extraction and primers

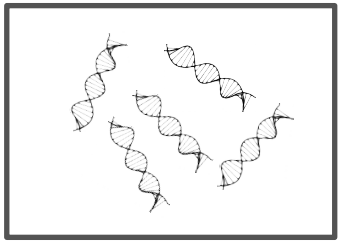
With the mock 2 community : test of 4 different protocols



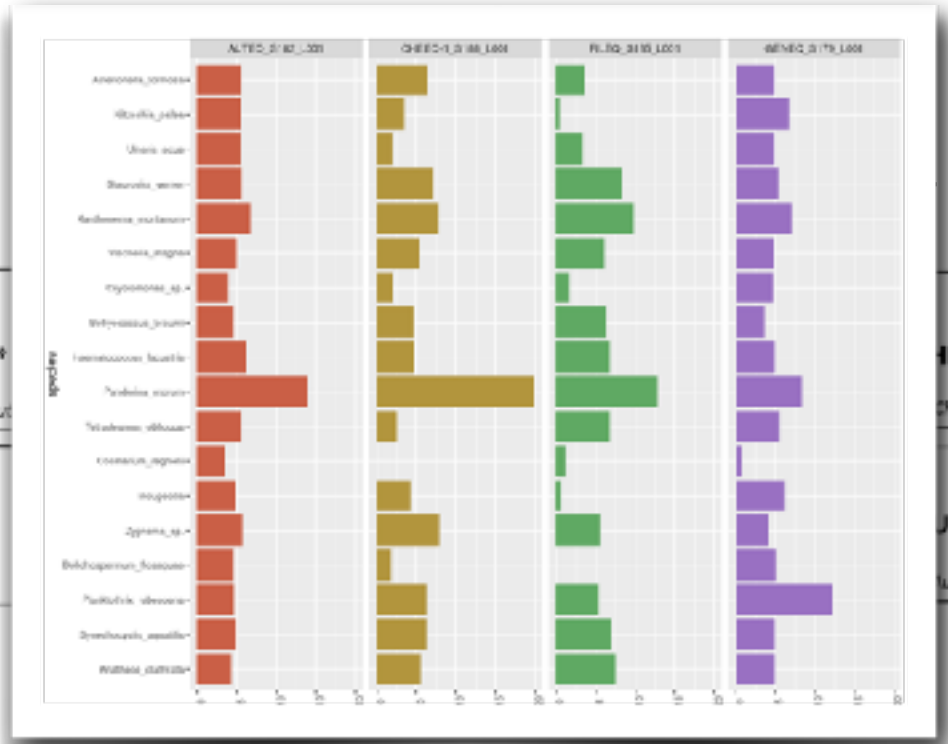


Funded by European Union

DNA extraction



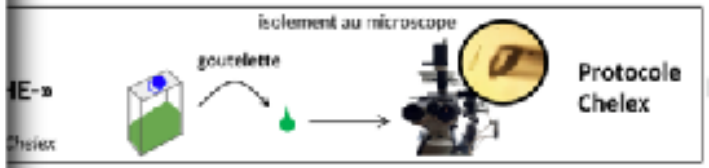
Test 2

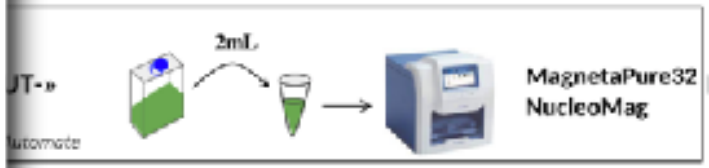


different protocols

«GEN-»
pour GenElute

«FIL-»
pour Filtre

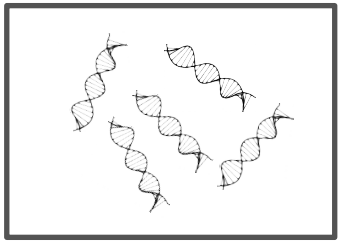




The automate and GenElute protocol without filtration gave the best results in Mock2 species recovery

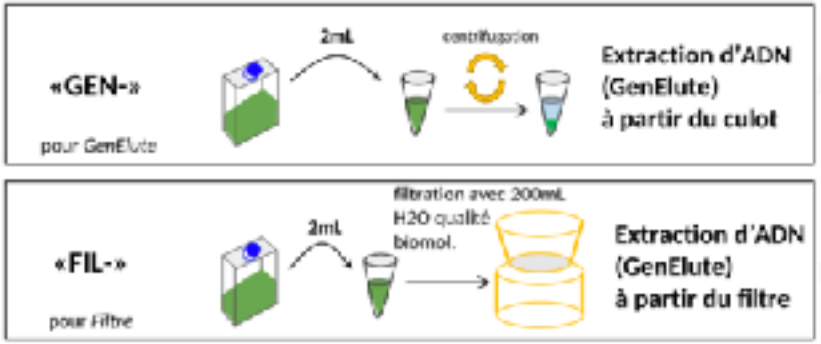
Funded by European Union

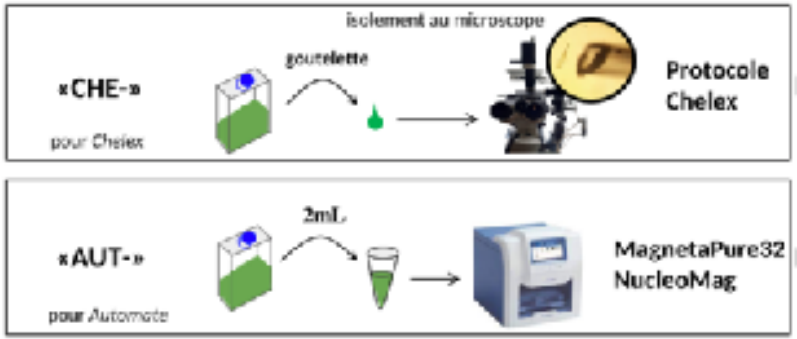
DNA extraction



Test 2 : impact of DNA extraction and primers

With the mock 2 community : test of 4 different protocols

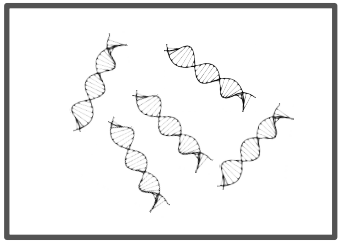




The automate presented a higher DNA quantity and quality

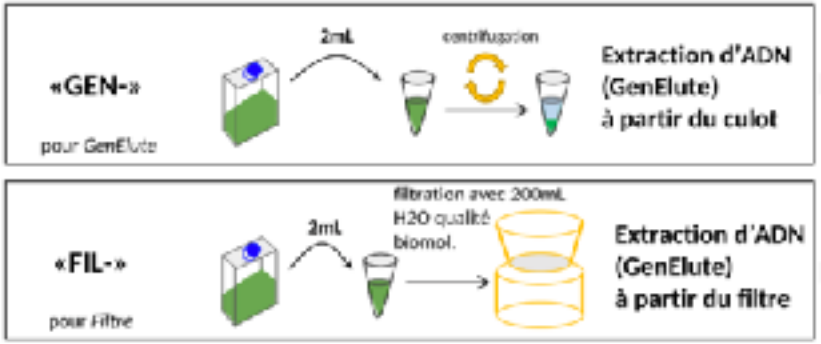
Funded by European Union

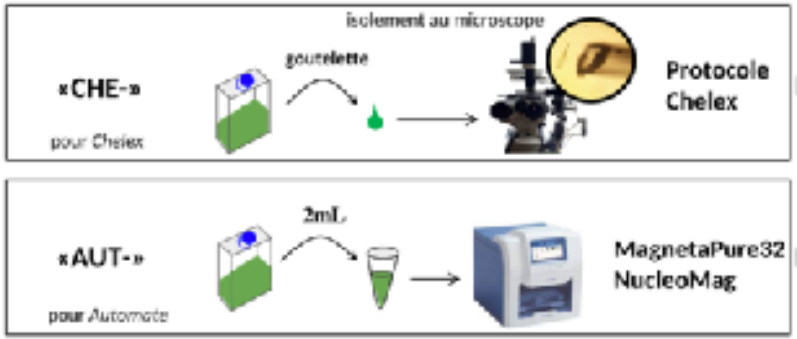
DNA extraction



Test 2 : impact of DNA extraction and primers

With the mock 2 community : test of 4 different protocols





+ It is faster and gives more reproducible results in a routine use

Funded by European Union

Phytoplankton metabarcoding

Sampling	*DNA extraction*	*Barcode amplification*	*Sequencing*	*Bioinformatic analysis*
		CYA359F/CYA781R ECLA23SF1/ECLA23SR1	 Illumina	?

Funded by European Union

Bioinformatic analysis

Fastq files

R1 R2

Remove primers sequences

R1 R2

Quality filtering

R1 R2

Error model

R2

R1

Infer samples

R1 R2

Merge reads

ASV table

Remove chimera

Keep ASV of an expected length

Assign taxonomy

ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA

Funded by European Union

Bioinformatic analysis

Fastq files

R1 R2

Remove primers sequences

R1 R2

Quality filtering

R1 R2

Error model

R2

R1

Infer samples

R1 R2

Merge reads

ASV table

Remove chimera

Keep ASV of an expected length

Assign taxonomy

Cutadapt

DADA2

Mothur

ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA

Bioinformatic analysis

Fastq files

R1 R2

Remove primers sequences

R1 R2

Quality filtering

R1 R2

Error model

R2

R1

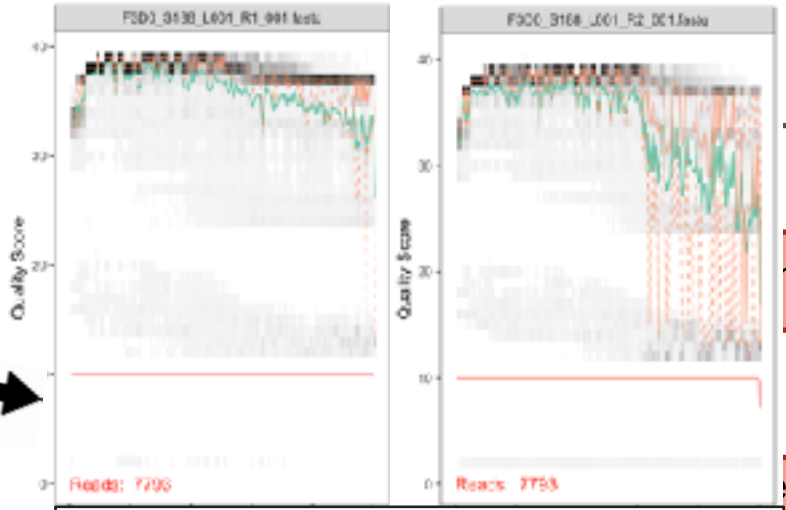
Infer samples

Assign taxonomy

Cutadapt

DADA2

Mothur



Depending on the reads quality profiles, we will trim the end of the reads to remove bases with low quality scores.

ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA

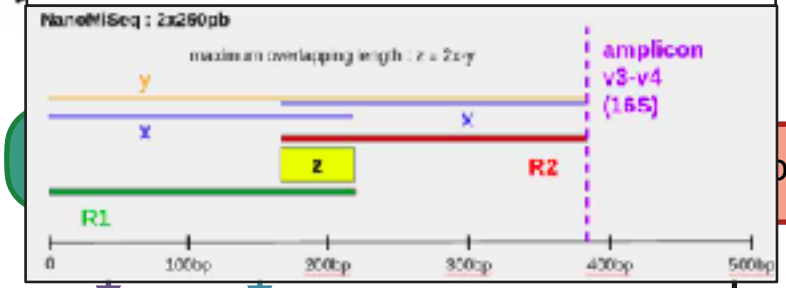
move chimera

ep ASV of an pected length

Funded by European Union

Bioinformatic analysis

Look at the number of read lost at this step.
Most of the reads should merge.



ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA

Cutadapt

DADA2

Mothur

Quality filtering

R1 R2

Error model

R2

R1

Infer samples

R1 R2

Merge reads

Remove chimera

Keep ASV of an expected length

Assign taxonomy

Funded by European Union

Bioinformatic analysis

Fastq files

R1 R2

Remove primers sequences

R1 R2

Quality filtering

R1 R2

Error model

R2

R1

Infer samples

R1 R2

ASV table

Same than after merging, number of chimeral reads should not be too abundant

Remove chimera

Keep ASV of an expected length

Assign taxonomy

Cutadapt

DADA2

Mothur

ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA

Bioinformatic analysis

Fastq files

R1 R2

Remove primers sequences

R1 R2

Quality filtering

R1 R2

Error model

R2

R1

Infer samples

R1 R2

Merge reads

Keep the reads that have the expected length (358pb) due to non-specific priming

ASV table

Keep ASV of an expected length

Remove chimera

Assign taxonomy

Cutadapt

DADA2

Mothur

ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA

Funded by European Union

Bioinformatic analysis

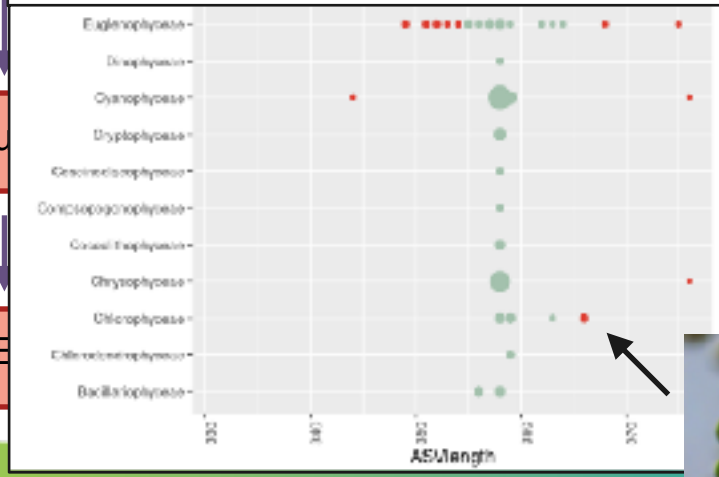
Fastq files

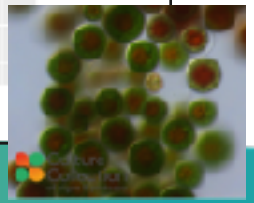
R1 R2

ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA

Re

/!\ Be careful : There can be a natural variability in sequence length for different taxa





Remove chimera

Keep ASV of an expected length

Assign taxonomy

R1

Qu

R1

E

Cutadapt

DADA2

Mothur

Funded by European Union

Bioinformatic analysis

Fastq files

R1

Remove pri
sequenc

R1

Quality filte

R1

Error model

R2

R1

Infer samples

Remove chimera

Keep ASV of an expected length

Assign taxonomy

Cutadapt

DADA2

Mothur

	DADA2 (assignTaxonomy, BS=80)		Mothur (classify.seqs, cutoff=80, iter=10000)	
	Barcodes complets	Barcodes résolutifs	Barcodes complets	Barcodes résolutifs
16S_v3v4 "CYA"	1087/3262 33.32 %	1317/2033 64.78 %	2069/3262 63.43 %	1351/2033 66.45 %
CIPA "16SFR"	513/835 61.44 %	605/681 88.84 %	688/835 82.4 %	601/681 88.25 %
CIPA "587FR"	528/851 62.04 %	612/697 87.8 %	700/851 82.26 %	614/697 88.09 %
CIPA "Stanwood"	510/836 61.00 %	601/682 88.12 %	683/836 81.7 %	598/682 87.68 %

Comparison with DADA2 assignation algorithm : Mothur is better (with 10,000 iterations)

ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA
ATCGCTTTGGACCT
ATCGAATTGGAACA

Funded by European Union

Phytoplankton metabarcoding

Sampling	*DNA extraction*	*Barcode amplification*	*Sequencing*	*Bioinformatic analysis*
		CYA359F/CYA781R ECLA23SF1/ECLA23SR1	 Illumina	ATCGCTTTGGACCT ATCGAATTGGAACA ATCGCTTTGGACCT ATCGAATTGGAACA ATCGCTTTGGACCT ATCGAATTGGAACA

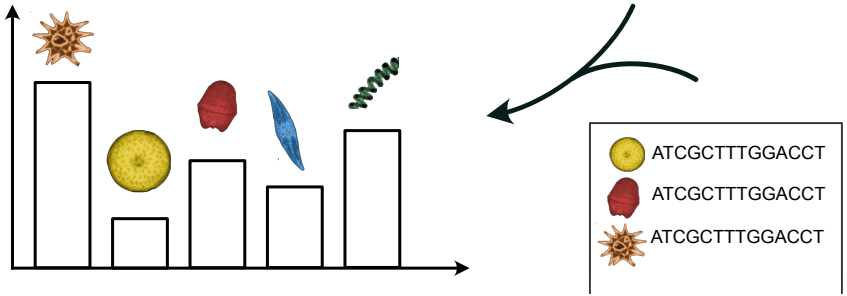
Funded by European Union

Phytoplankton metabarcoding

Sampling	DNA extraction	Barcode amplification	Sequencing	Bioinformatic analysis
		CYA359F/CYA781R ECLA23SF1/ECLA23SR1	 Illumina	ATCGCTTTGGACCT ATCGAATTGGAACA ATCGCTTTGGACCT ATCGAATTGGAACA ATCGCTTTGGACCT ATCGAATTGGAACA

Phytool_v2

Design of a new database for phytoplankton



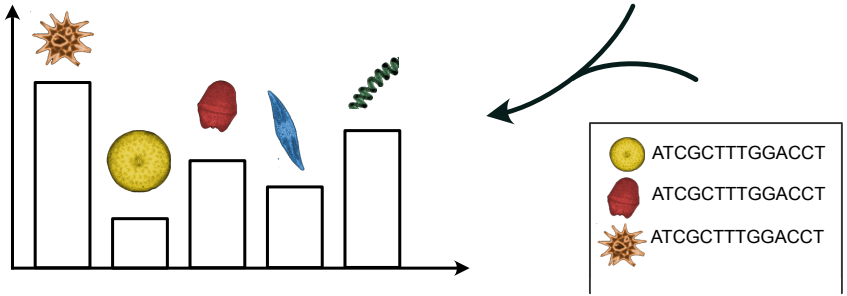
Funded by European Union

Phytoplankton metabarcoding

Sampling	DNA extraction	Barcode amplification	Sequencing	Bioinformatic analysis
		CYA359F/CYA781R ECLA23SF1/ECLA23SR1	 Illumina	ATCGCTTTGGACCT ATCGAATTGGAACA ATCGCTTTGGACCT ATCGAATTGGAACA ATCGCTTTGGACCT ATCGAATTGGAACA

Very good
Good
Moderate
Poor
Very poor

How this workflow performs for phytoplankton bioindication ???



Funded by European Union

BIOLAWEB

Design of new primers

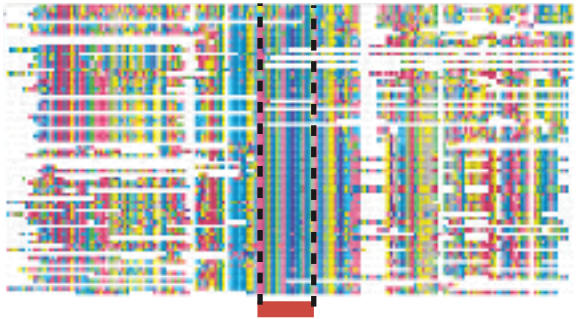
1. Combine all sequences of the marker gene from reference databases

silva
high quality ribosomal RNA databases

PR²

PhytoRef

2. Align the sequences



3. Find conserved regions

4. Keep them if they target a barcode with the good sequencing length

New primers can be designed as soon as there are new sequences in the databases

Funded by European Union