Diatom metabarcoding for biomonitoring : 2nd part

F. Rimet, A. Bouchez

Barcode choice, sample preservation, DNA extraction, sequencing

Schedule

1- Barcode choice

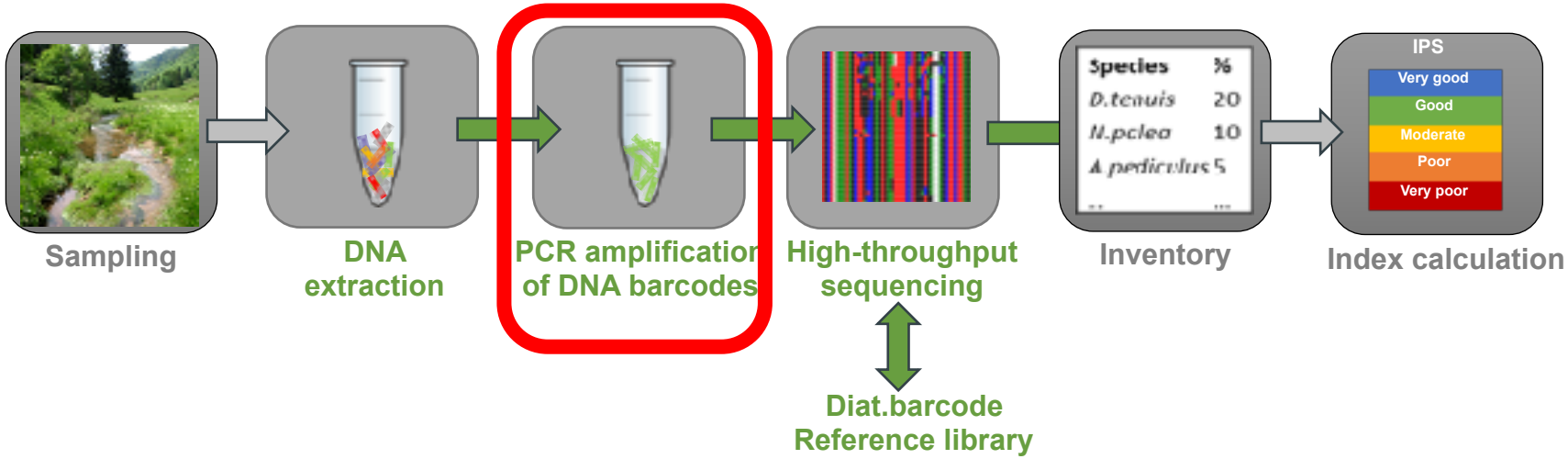
2- Sample preservation

- Preservation experimentation
- CEN Standardisation

3- DNA extraction method

4- Choice of the sequencing technology

Barcode choice?

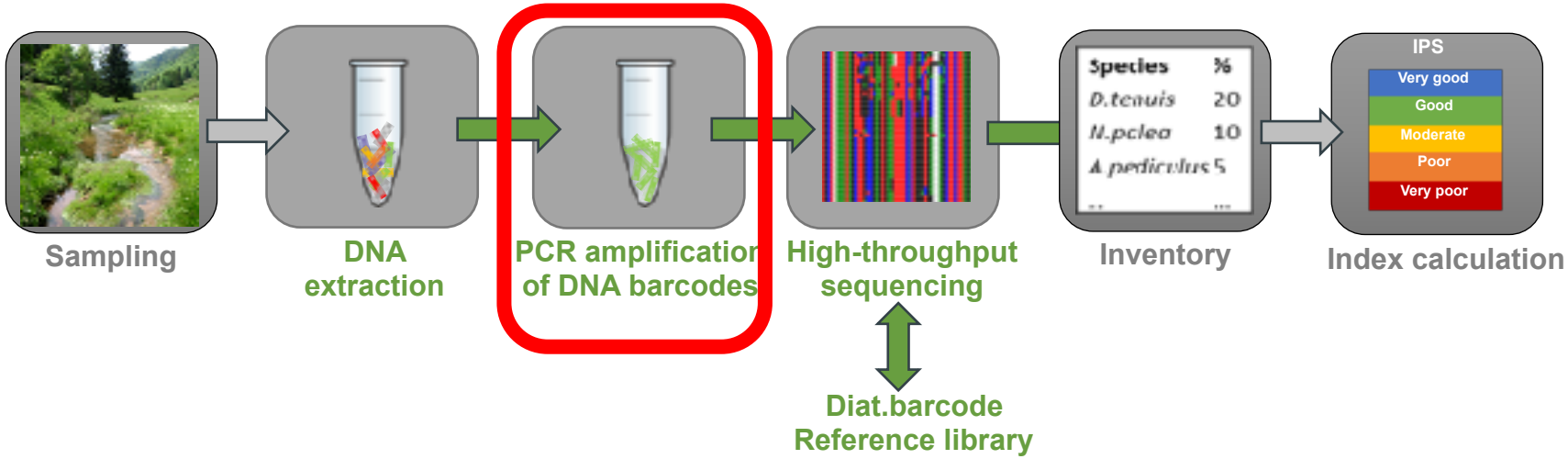


First step of the process: which barcode to use to identify diatom species?

Our selection criteria to choose the barcode:

- Universality: A single barcode that targets the entire diatom diversity,
- Variability: an efficient barcode able to identify diatoms to species with conserved regions to set primers
- Specificity: a barcode specific of diatoms, not amplifying other groups (e.g. Chrysophytes, etc…)
- Lenght: the barcode lenght must fit the sequencing technology (Illumina Miseq)
- References: A barcode with reference barcoding libraries complete enough to analyse diatom diversity

Barcode choice?



Workflow to select the barcode (thesis of L Kermarrec 2012)

Universality: A single barcode that targets the entire diatom diversity,

- List of candidate markers: 18S, 28S, ITS, rbcL, cox1,
- Based on universality, variability, references criteria > selection of markers
- In-vitro test -> 1st selection

18S

- References available: huge, good reference databases (PR2, SILVA), used in phylogenetic studies since a long time (see Medlin et al. 1993)

Medlin et al. 1993

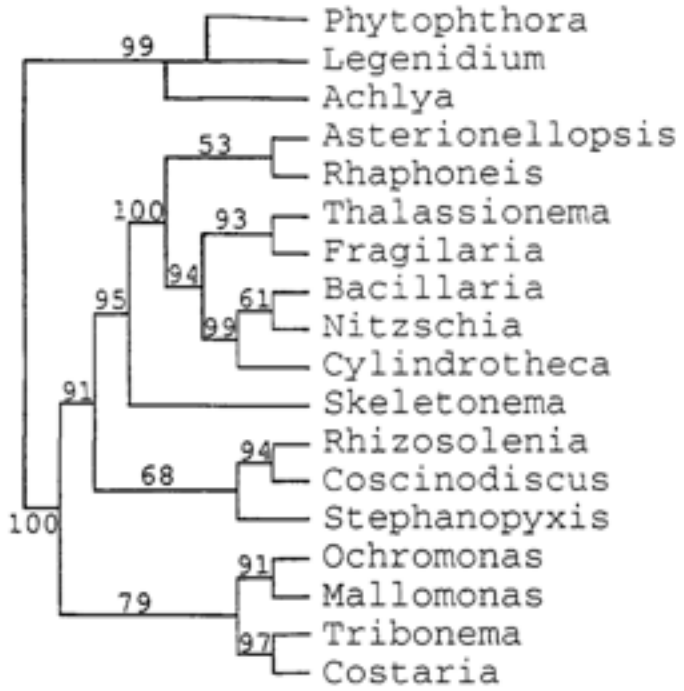


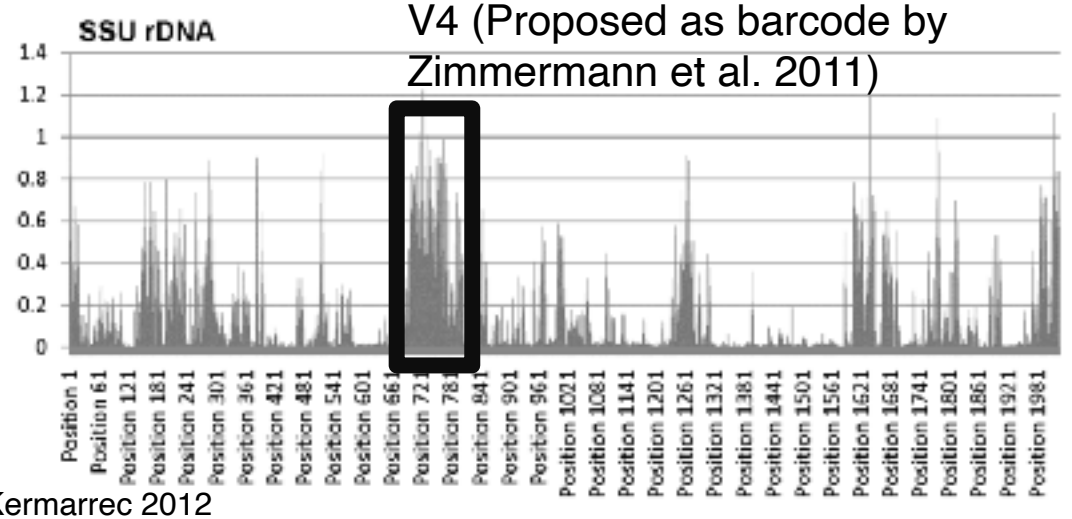
Fig. 3. Diatom phylogeny inferred from maximum parsimony analysis of the secondary structure alignment of nucleotide positions in the 18S rRNA coding regions using the heuristic search within PAUP. Bootstrap values at the branch nodes are based on a 50% majority rule.

18S

- References available: huge, good reference databases (PR2, SILVA), used in phylogenetic studies since a long time (see Medlin et al. 1993)
- Several highly variable regions flanked by conserved regions
- 18S is the SSU of eukaryotic ribosome, avoid the amplification of bacteria
- Several primers already used for Sanger sequencing
- Some species can have identical 18S sequences

18S alignment

V4 (Proposed as barcode by Zimmermann et al. 2011)



Kermarrec 2012

Shannon index

Funded by European Union

28S

- Longer than 18S (3300 bp vs 1800 bp).
- Interesting for phylogenetic studies
- Several highly variable regions, and depending on the authors different regions were sequenced (D1/D2 Bruder & Medlin 2007, D1/D3 Lundholm et al. 2002, D1/D4 Kooistra et al. 2010)
- But because of this length, and the absence of generally accepted “standard region” relatively few references are available

ITS

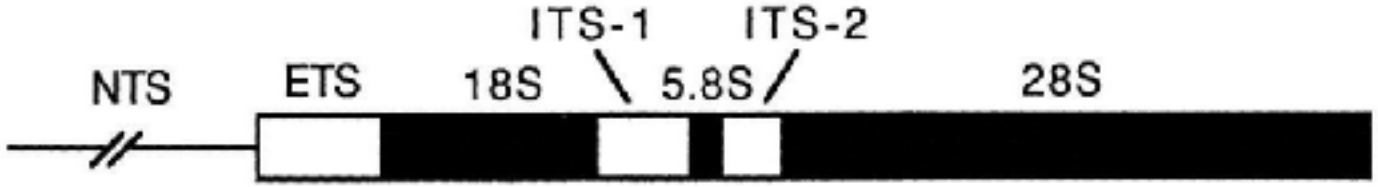


Figure II.2: *Structure de l'opéron ribosomal des eucaryotes.*
Source : Hillis & Dixon, 1991.

- Internal Transcribed Spacers: intergenic regions
- Very low selection pressure because they are excised after their transcription

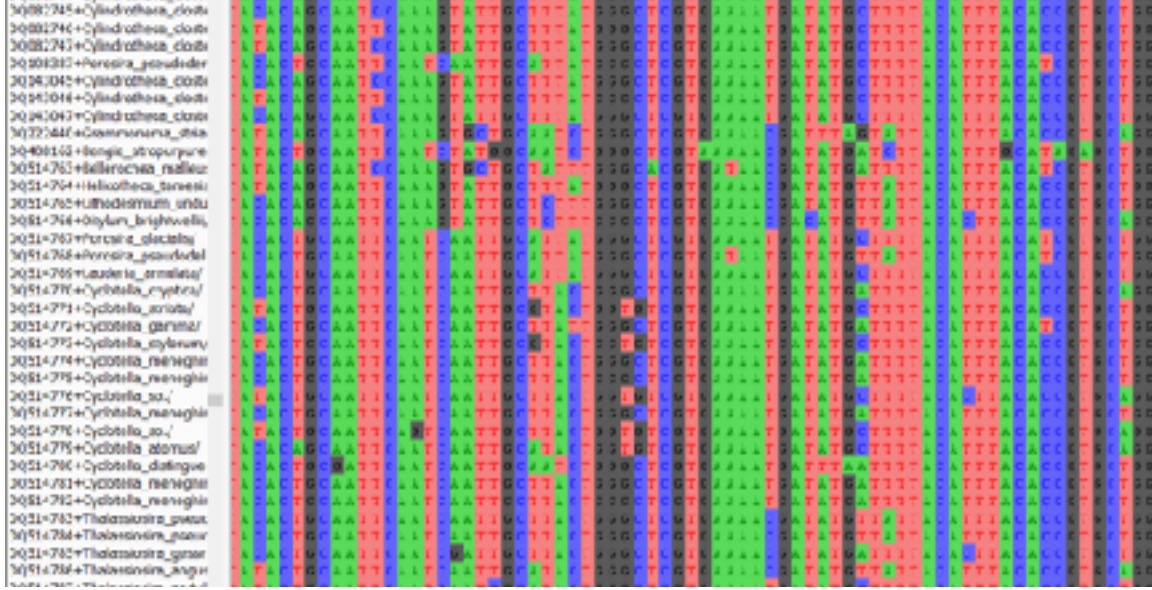
>> highly variable

>> not adapted to phylogenetic studies

>> More adapted to population genetics, biogeography of populations inside species

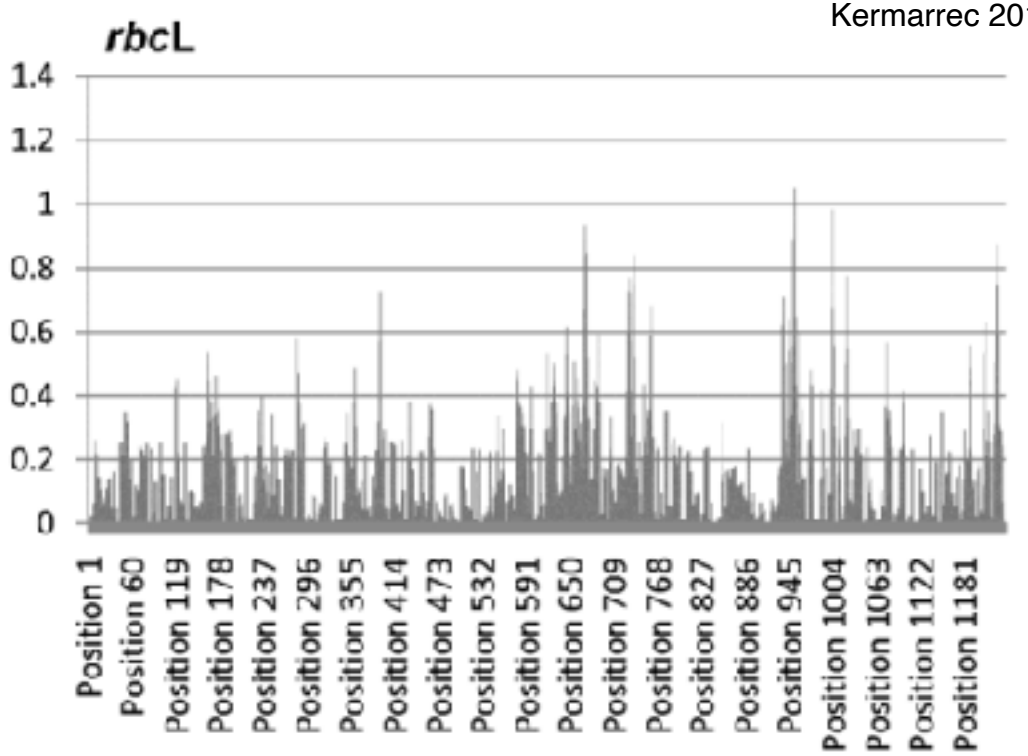
- Difficulty: intragenomique variability, which makes them difficult to sequence with Sanger
- Reference available: poor

rbcL



- Chloroplastic gene (1450 bp): enable to avoid all heterotrophic organisms
- Coding region: easy to align, enable a translation into amino-acides and check for stop codon (must not be present) when filtering for sequence quality
- Good variability, enable to distinguish cryptic species.
- Some primers available, and a few conserved regions inside the gene
- Good availability of references (Many publications use this marker for phylogenetic studies)

Kermarrec 2012



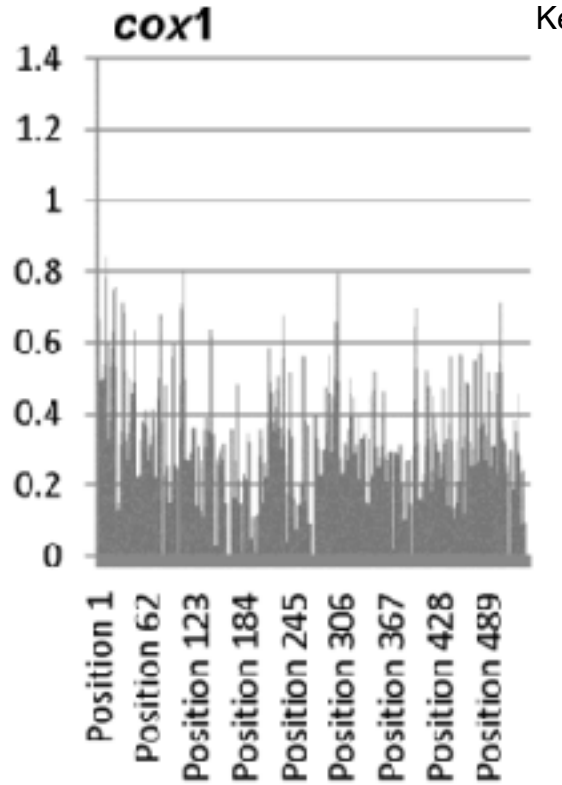
Funded by European Union

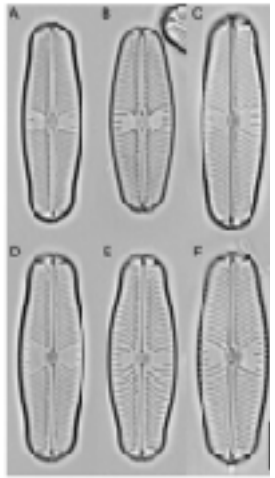
Cox1

- Recognised as the standard barcode (see Hebert et al.) -> BOLD
- Coding region
- Some references available

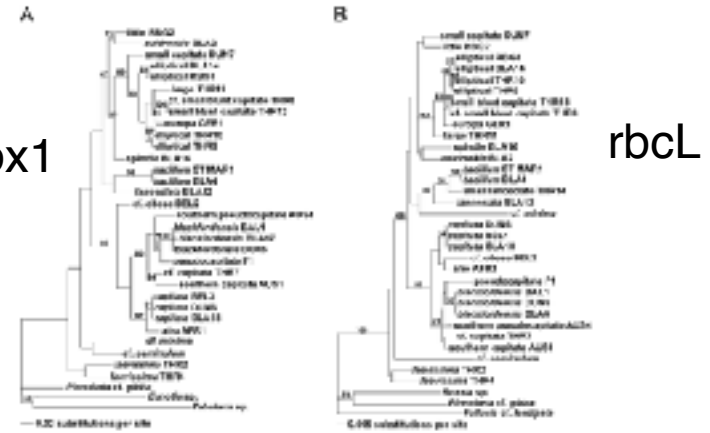
- Difficulty to find universal primers for diatoms (e.g. Trobajo et al. 2010, Hamsher et al. 2011)

- Good variability, enable to distinguish cryptic species (Evans et al. 2007)





cox1



rbcL

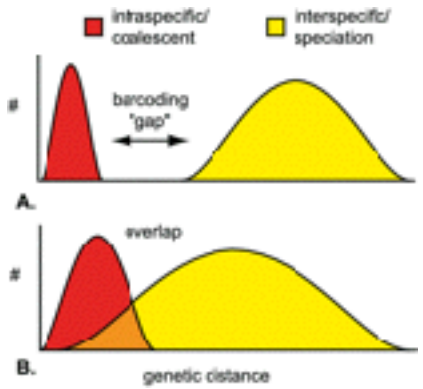
Funded by European Union

From this list, we selected the following markers

- 18s
- rbcL
- Cox1

- Comparison of their barcoding gaps:

Frequencies of Intra sp genetic distances vs Inter sp genetic distances

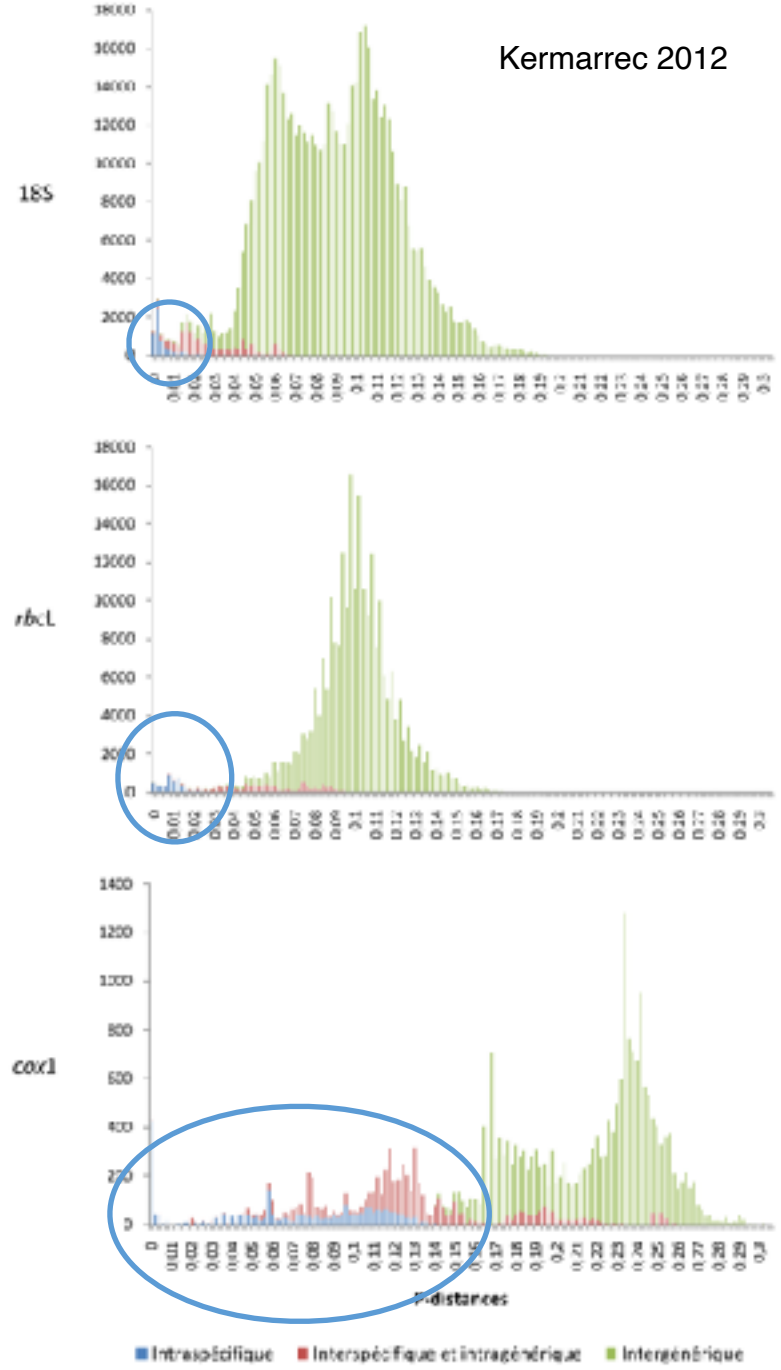


Good case

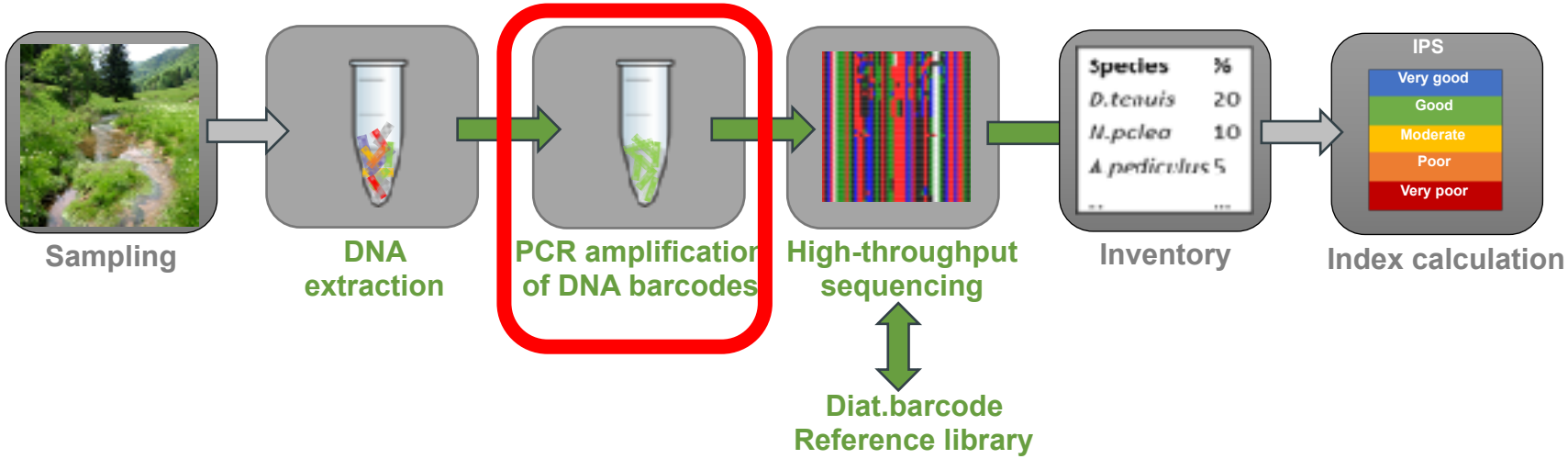
Bad case

Overlap between intrasp/intersp distances :
No perfect barcoding gap

Kermarrec 2012



Barcode choice?

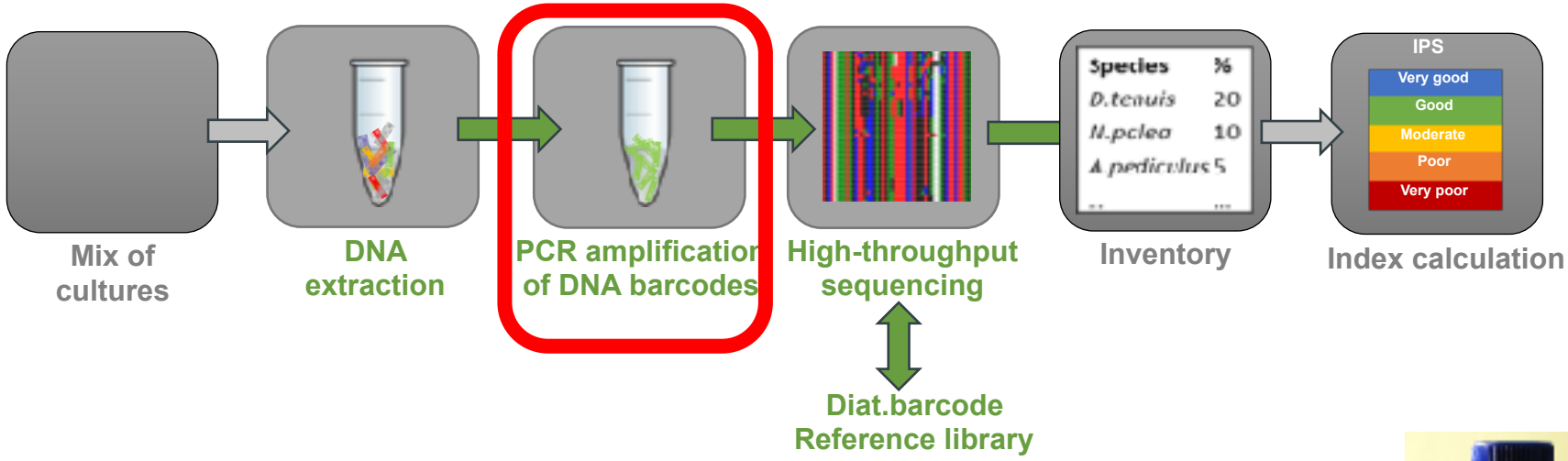


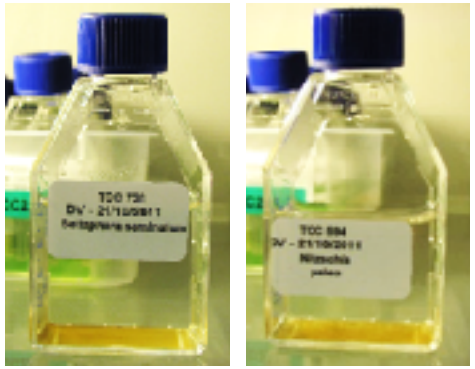
Workflow to select the barcode

Universality: A single barcode that targets the entire diatom diversity,

- List of candidate markers: 18S, 28S, ITS, rbcL, cox1,
- Based on universality, variability, references criteria > selection of markers
- In-vitro test -> selection

Barcode choice?





Test carried out on synthetic biofilms (mix of 30 cultures already sequenced):

this enables to have a sample of known composition

3 markers selected: 18s (ribosome), Cox1 (mitochondria), rbcL (chloroplast)

Kermarrec L, Franc A., Rimet F., Chaumeil P., Humbert J.F. & Bouchez A., 2013. Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms. Molecular Ecology Ressources, 13: 607-619.

Funded by European Union

Barcode choice?

30 diatom cultures
21 species

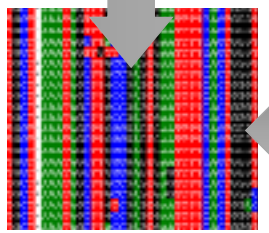
Halamphora montana
Cocconeis placentula
Cyclotella meneghiniana
Fistulifera saprophila
Fragilaria capucina
Gomphonema bourbonense
Gomphonema clavatum
Gomphonema clevei
Gomphonema parvulum
Gomphonema pumilum
Mayamaea permitis
Navicula cryptocephala
Nitzschia inconspicua
Nitzschia acidoclinata
Nitzschia lorenziana
Nitzschia inconspicua
Nitzschia cf. frustulum
Nitzschia palea
Nitzschia dravaillensis
Pinnularia cf. subgibba
Sellaphora seminulum
Ulnaria ulna

Mix of all strains

DNA extraction of the mix

PCR amplification of the mix
cox1, 18s, rbcl

454. Roche pyrosequencing



2 assignation algorithms
Blast - metamatch

DIAT.BARCODE
cox1, 18s, rbcl

Comparison of the floristic lists
cox1, 18s, rbcl

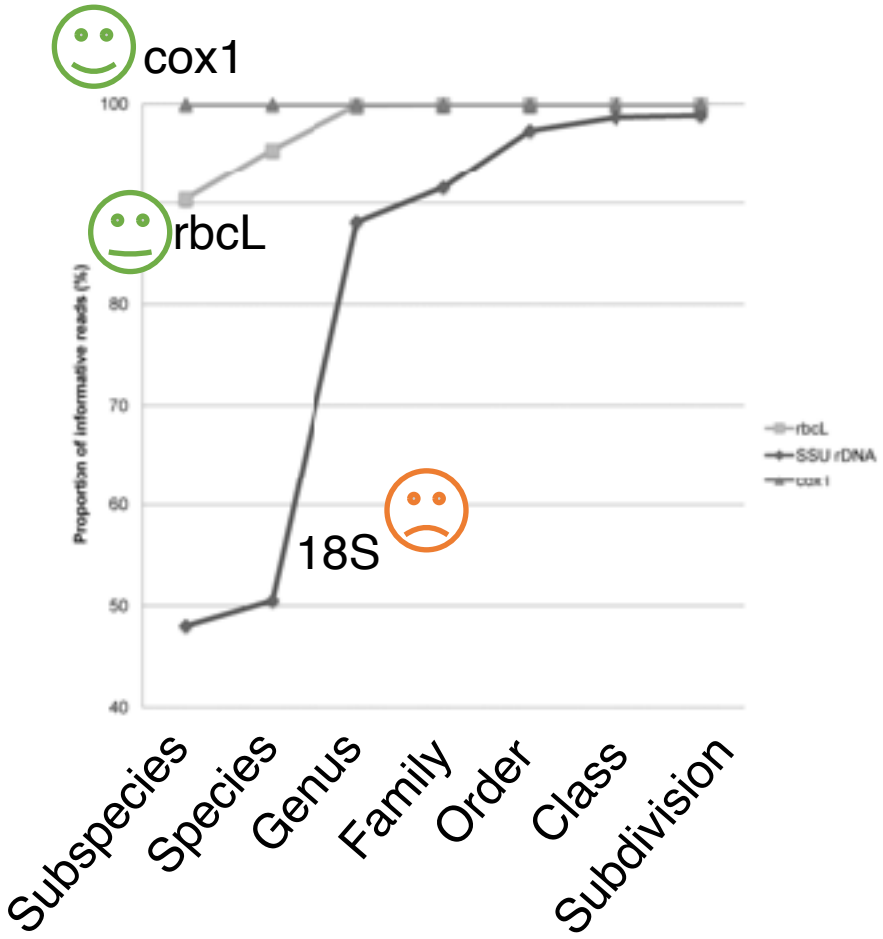
Barcode choice?

Which barcode is the most efficient to identify diatoms to species level?

Proportion of reads matching with a single taxon

Cox is the most efficient, followed by rbcL.

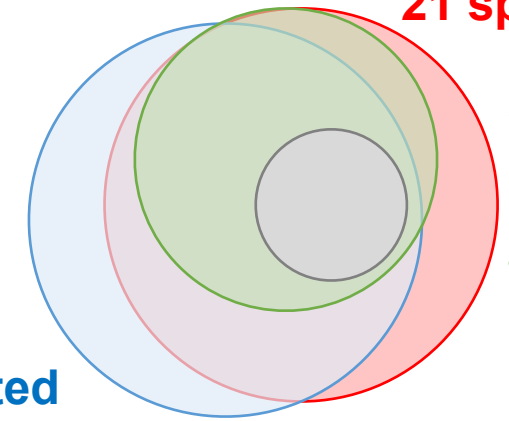
18S has a quite low specificity.



Barcode choice?

Which species are detected?

Real
21 sp in the mix

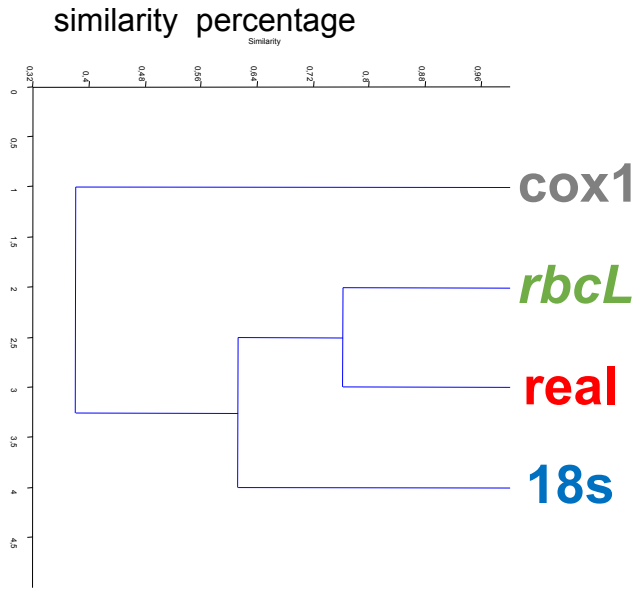


***rbc*L**
16 sp detected
5 not detected (low abundances in the mix)
0 sp added

18s
21 sp detected
5 sp not detected
5 sp added
(low variability, not able to detect sp complexes)

cox1
8 sp detected
13 not detected
(reference library uncomplete)

Cluster analysis on distance similarity between inventories (M3 mix-cultures)



***cox1*: far from the real inventory**

18s: intermediate

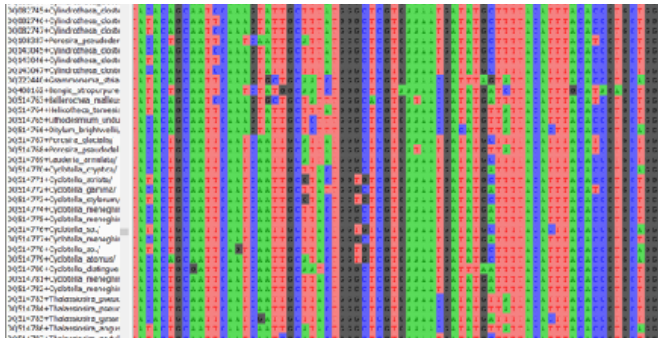
***rbc*L : close to the real inventory**

Barcode choice?

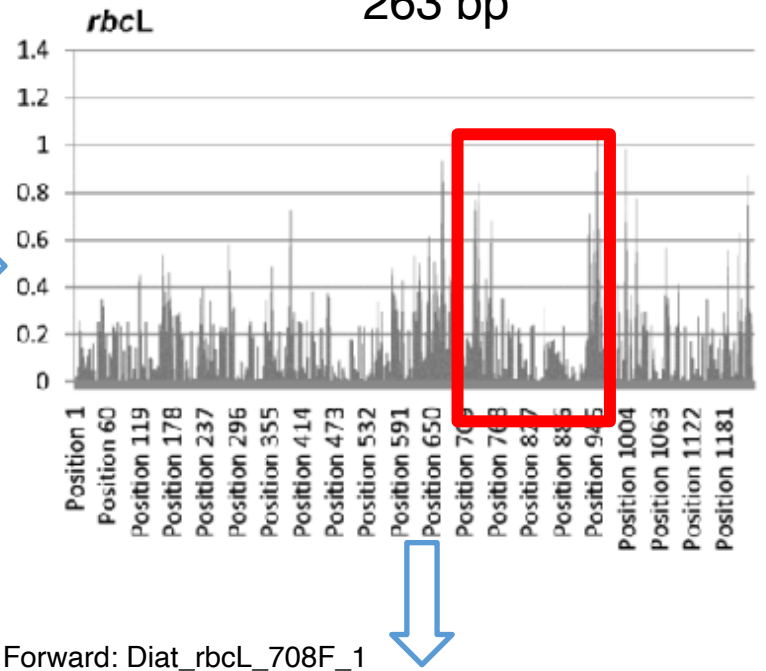
>> Selection of rbcL as barcode

>> refinement of existing primers Diat_rbcL_708F (Stoof-Leichsenring etal. 2012) and R3 (Bruder & Medlin 2007).
Published in Vasselon et al. 2017.

1602 sequences, 638 species



Barcode length:
263 bp



DIAT.BARCODE

This refinement was carried out with Diat.barcode v6

Forward: Diat_rbcL_708F_1 (AGGTGAAGTAAAAGGTTCWTACTTAAA), Diat_rbcL_708F_2 (AGGTGAAGTTAAAGGTTCWTAYTTAAA), Diat_rbcL_708F_3 (AGGTGAAACTAAAGGTTCWTACTTAAA)
reverse R3_1 (CCTTCTAATTTACCWACWACTG), R3_2 (CCTTCTAATTTACCWACAACAG).

Funded by European Union

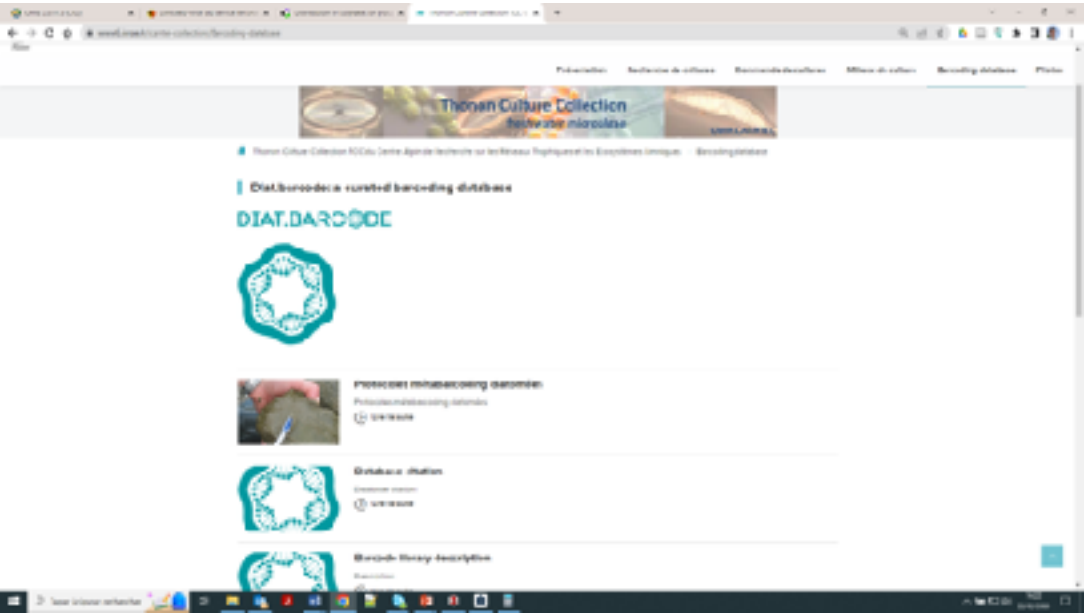
Protocols for PCR?

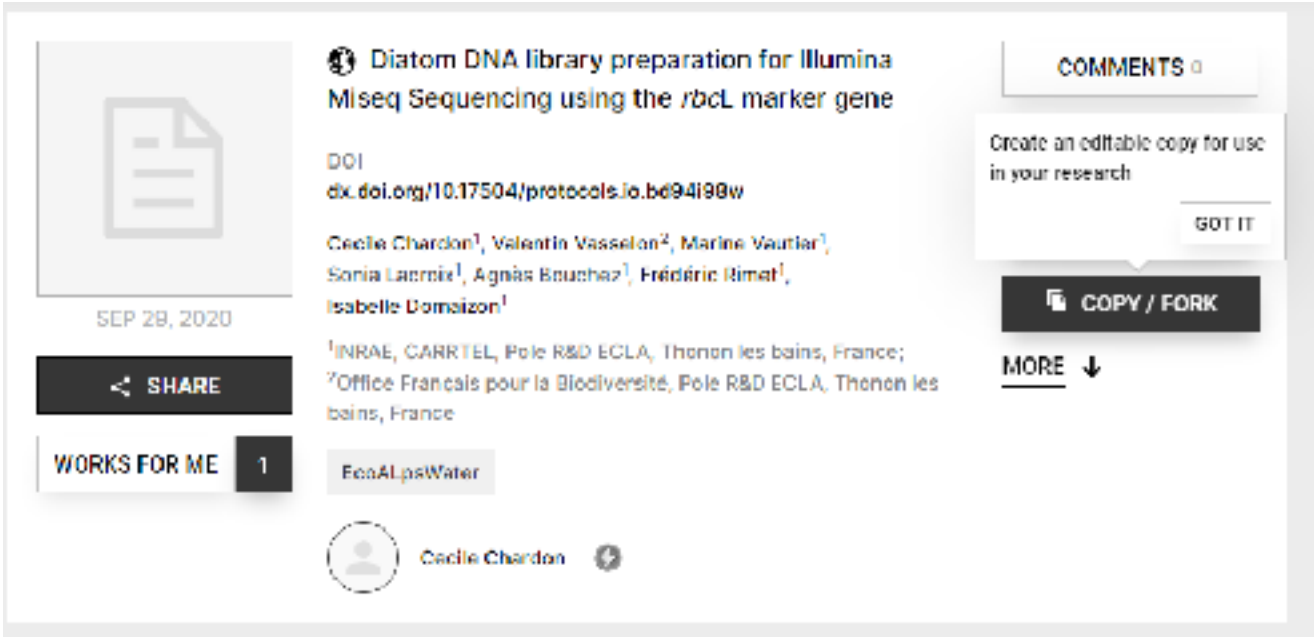
>> In French

https://www6.inrae.fr/carrtel-collection/Barcoding-database

>> In English

https://www.protocols.io/view/diatom-dna-library-preparation-for-illumina-miseq-kqdg3573zv25/v1





Schedule

1- Barcode choice

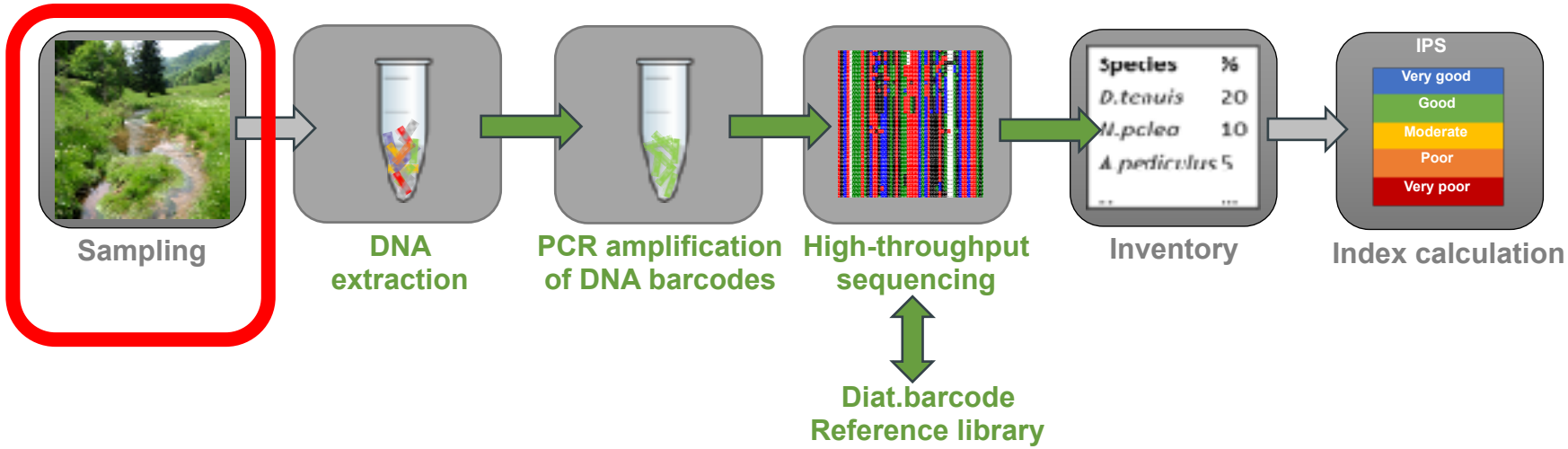
2- Sample preservation 

- Preservation experimentation
- CEN Standardisation

3- DNA extraction method

4- Choice of the sequencing technology

Preservation method



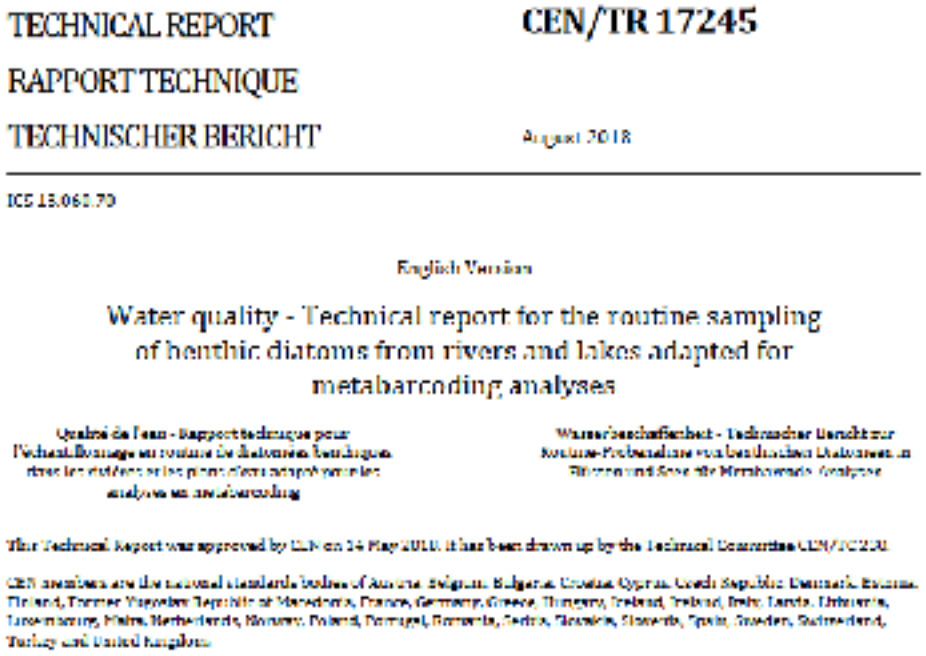
- There are several ways to preserver DNA
  - Ethanol, buffer, freezing
  - Impact on, DNA quantity, floristic list?
- >> Agnès Bouchez

Standard for sample preservation?

>> Work realised be several diatom experts working with eDNA:
A Poulickova (CZ), D Mann (UK), M Kelly (UK), M Pfannkuchen (HR), M Kahlert (S), R Trobajo (SP), K Sabbe (B), J. Zimmermann (D), A Bouchez (FR), F Rimet (FR), Neela ENKE (D)

>> Long process: started in 2012, publication in 2018

>> In the TR: several preservative are accepted in the document
Ethanol, RNA buffer, Deep freezing



TECHNICAL REPORT
RAPPORT TECHNIQUE
TECHNISCHER BERICHT

CEN/TR 17245

August 2018

ICS 13.060.70

English Version

Water quality - Technical report for the routine sampling of benthic diatoms from rivers and lakes adapted for metabarcoding analyses

Qualité de l'eau - Rapport technique pour l'échantillonnage en routine de diatomées benthiques dans les rivières et les plans d'eau adapté pour les analyses en métabarcoding

Wasserbeschaffenheit - Technischer Bericht zur Routine-Probenahme von benthischen Diatomeen in Flüssen und Seen für Metabarcoding-Analysen

This Technical Report was approved by CEN on 14 May 2018. It has been drawn up by the Technical Committee CEN/TC230.

CEN members are the national standards bodies of Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Former Yugoslav Republic of Macedonia, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and United Kingdom.

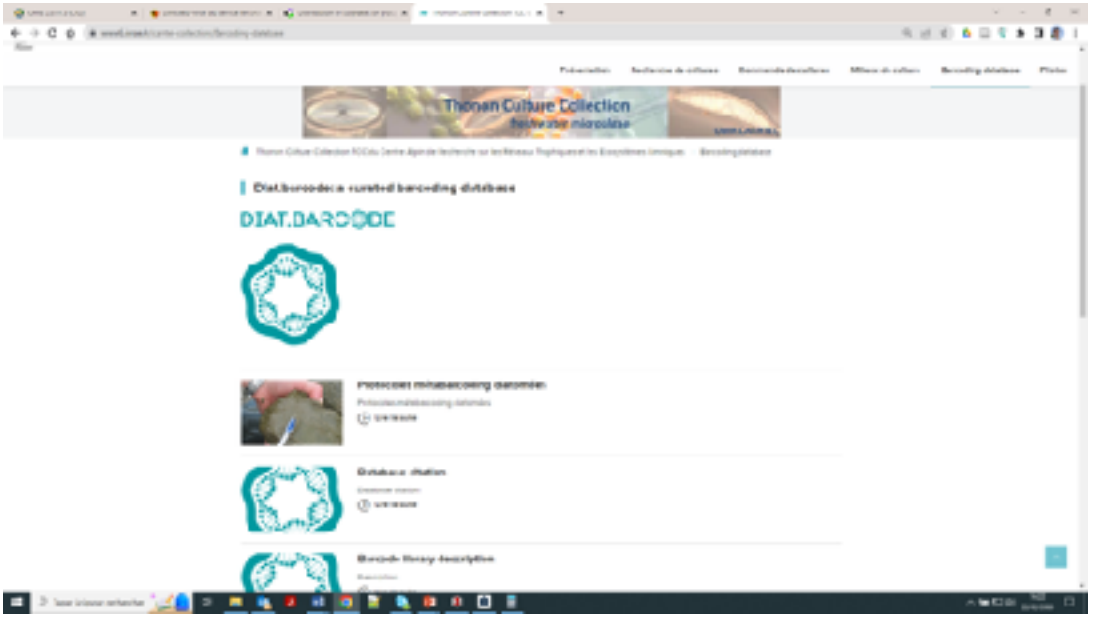
Funded by European Union

EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITÉ EUROPÉEN DE NORMALISATION
EUROPÄISCHES KOMITEE FÜR NORMUNG

Protocols for sample preservation?

>> In French

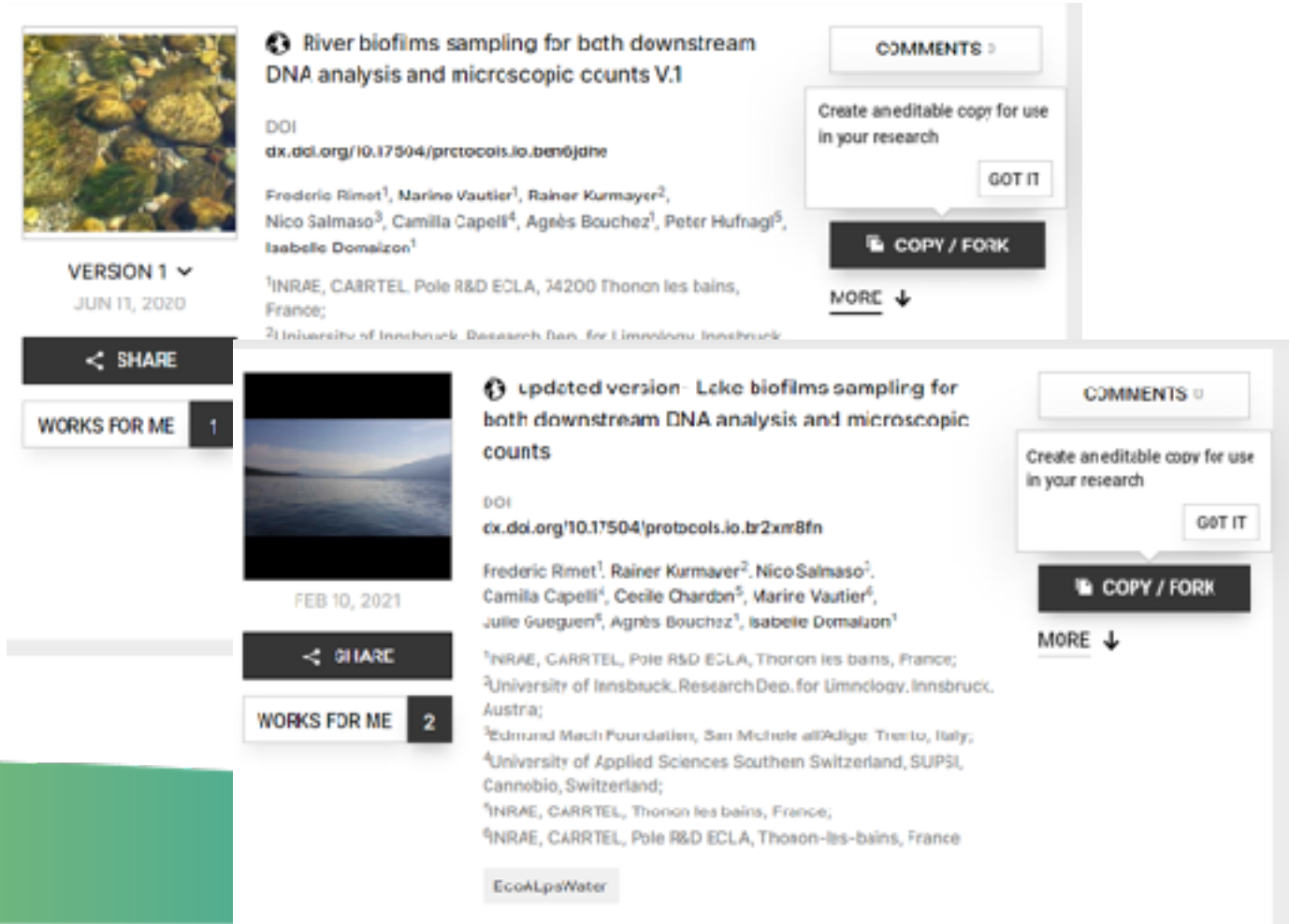
https://www6.inrae.fr/carrtel-collection/Barcoding-database



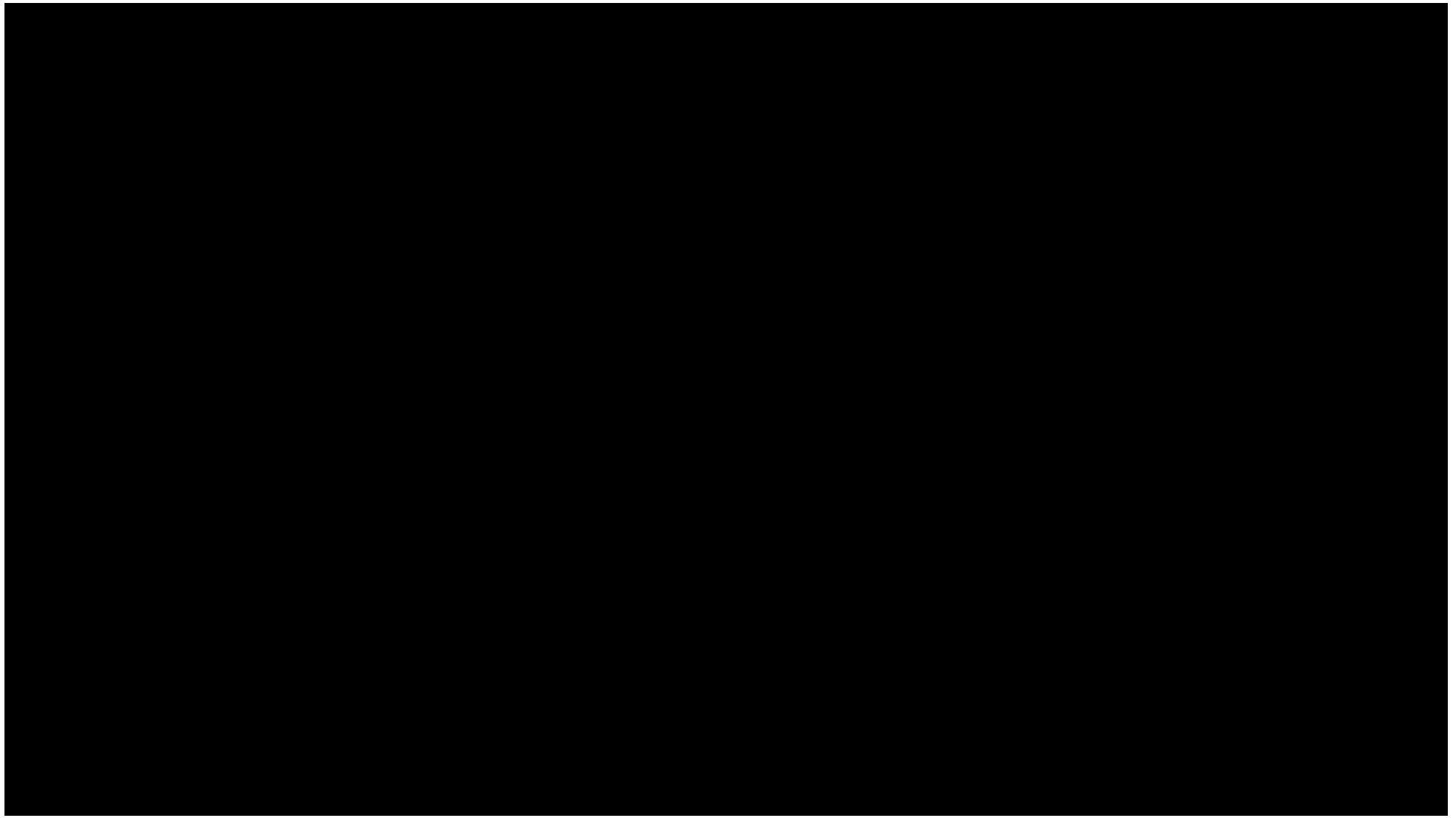
>> In English

Rivers: https://www.protocols.io/view/river-biofilms-sampling-for-both-downstream-dna-an-e6nvw9mjdgmk/v1

Lakes: https://www.protocols.io/view/updated-version-lake-biofilms-sampling-for-both-do-14egnz4w6g5d/v1



Funded by European Union



Schedule

1- Barcode choice

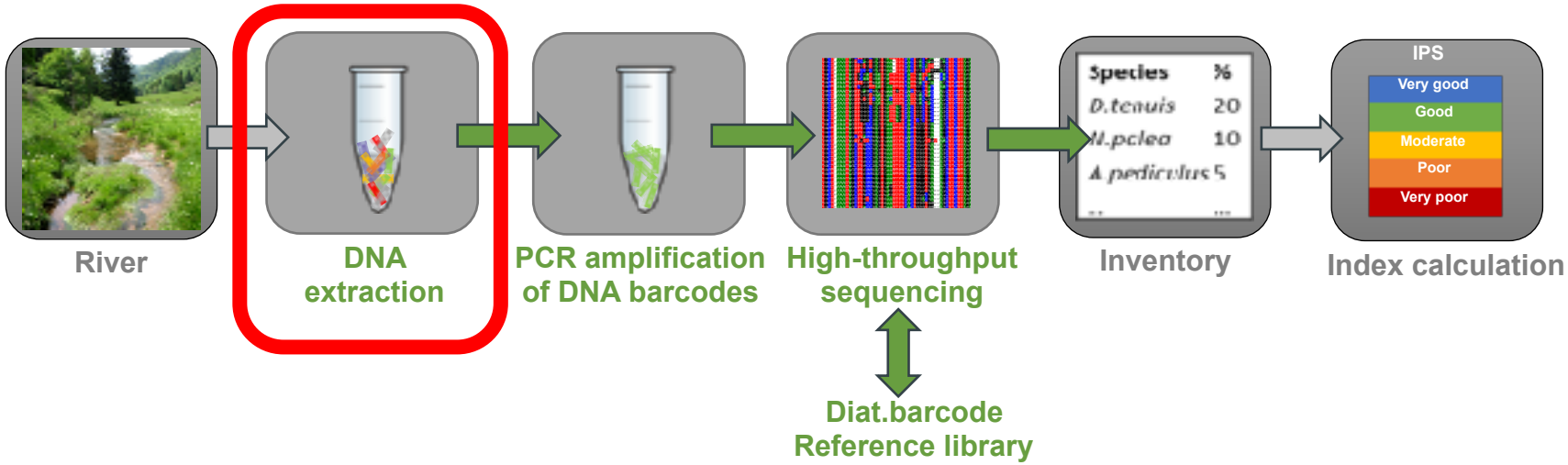
2- Sample preservation

- Preservation experimentation
- CEN Standardisation

3- DNA extraction method

4- Choice of the sequencing technology

Extraction kit choice?



❖**Different extraction methods exist. Do they have an impact on:**

- **DNA quantity, quality,**
- **the floristic lists**
- **Diatom index values (ecological quality assessment)**

❖**Test carried out with:**

- ❖ **8 samples** (Europe, Tropics, Lakes, Rivers)
- ❖ **5 kits**

Vasselon V., Domaizon I., Rimet F., Kahlert M., Bouchez A., 2017. Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: do DNA extraction methods matter? Freshwater Science 36: 162-177.

Funded by European Union

Cellular lysis

Lysate purification

DNA isolation
adsorption/precipitation

Final elution

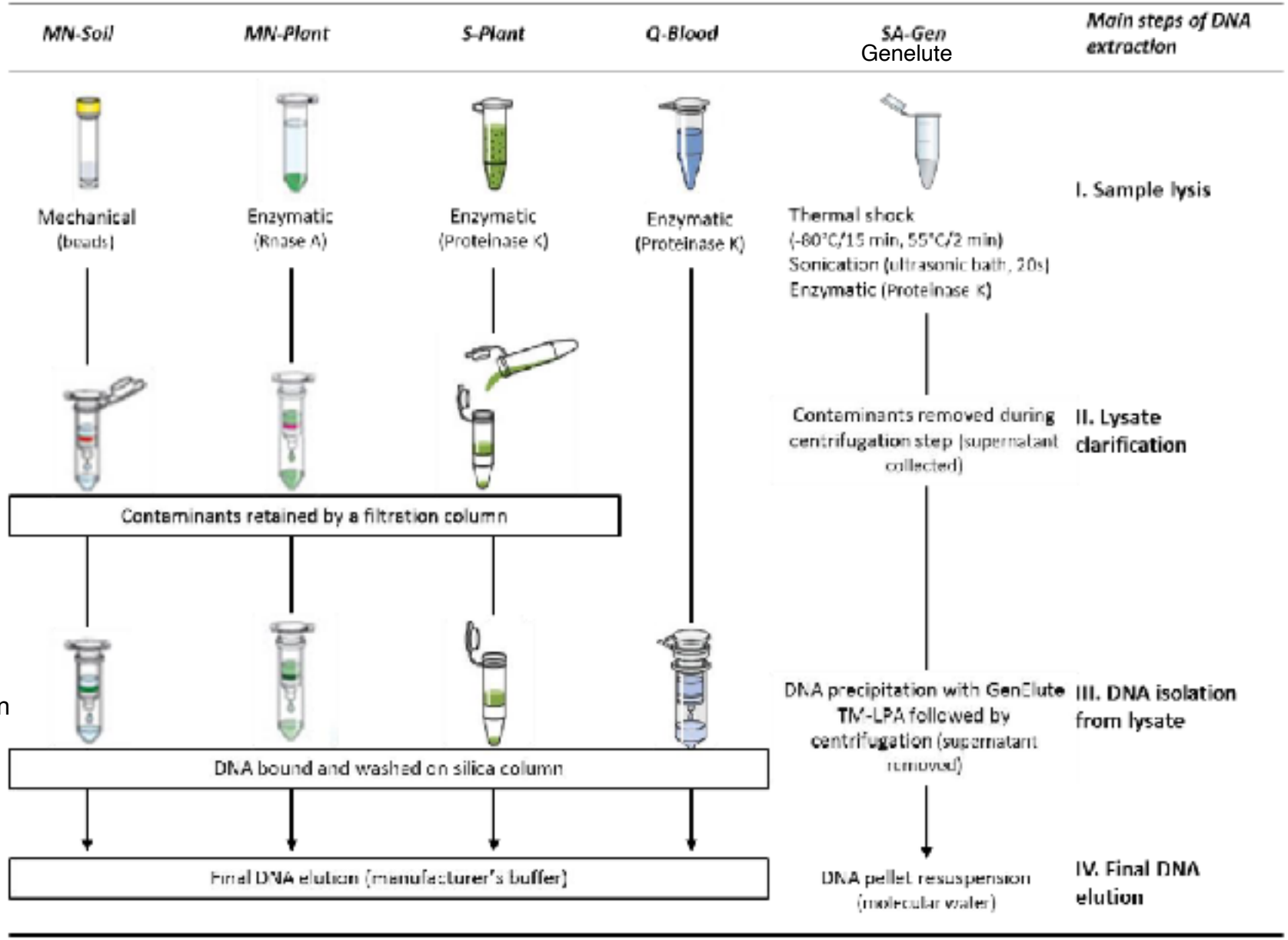


Figure 2. The main steps of DNA extraction for the 5 methods with a focus on sample lysis (I), lysate clarification (II), DNA isolation from lysate (III), and DNA elution (IV). Pictures modified from the manufacturers' web sites.

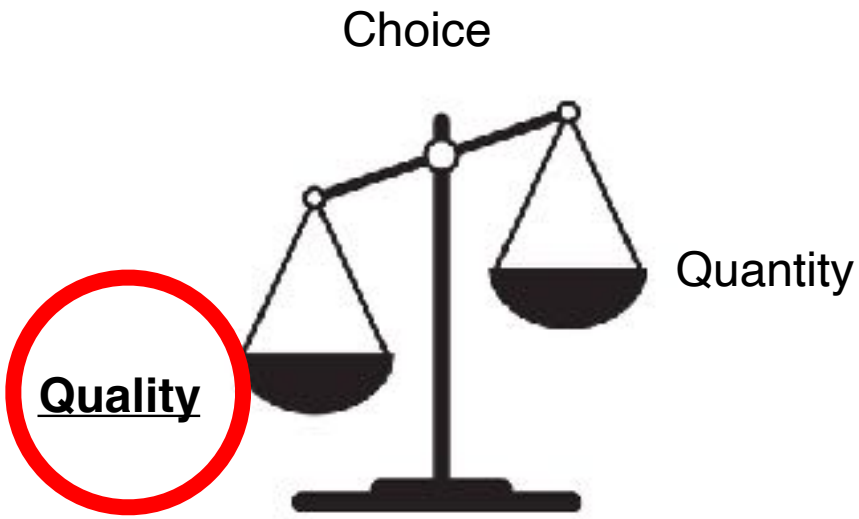
Extraction kit choice?

There is a balance between DNA quality/quantity:

- kits with high DNA quantities have low DNA quality (presence of PCR inhibitors): SA-Gen

whereas

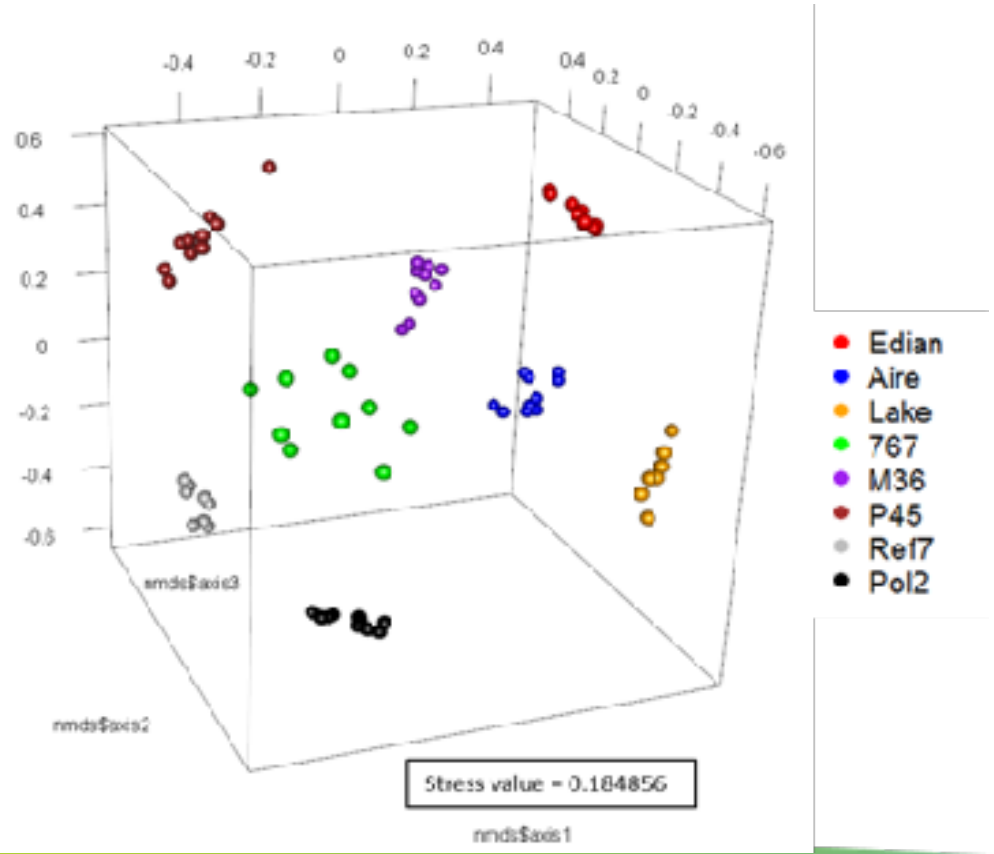
- kits with low DNA quantities have a good DNA quality: soil kit, kits with silica column
- It is important to prioritize quality, in order no to have PCR inhibitors.

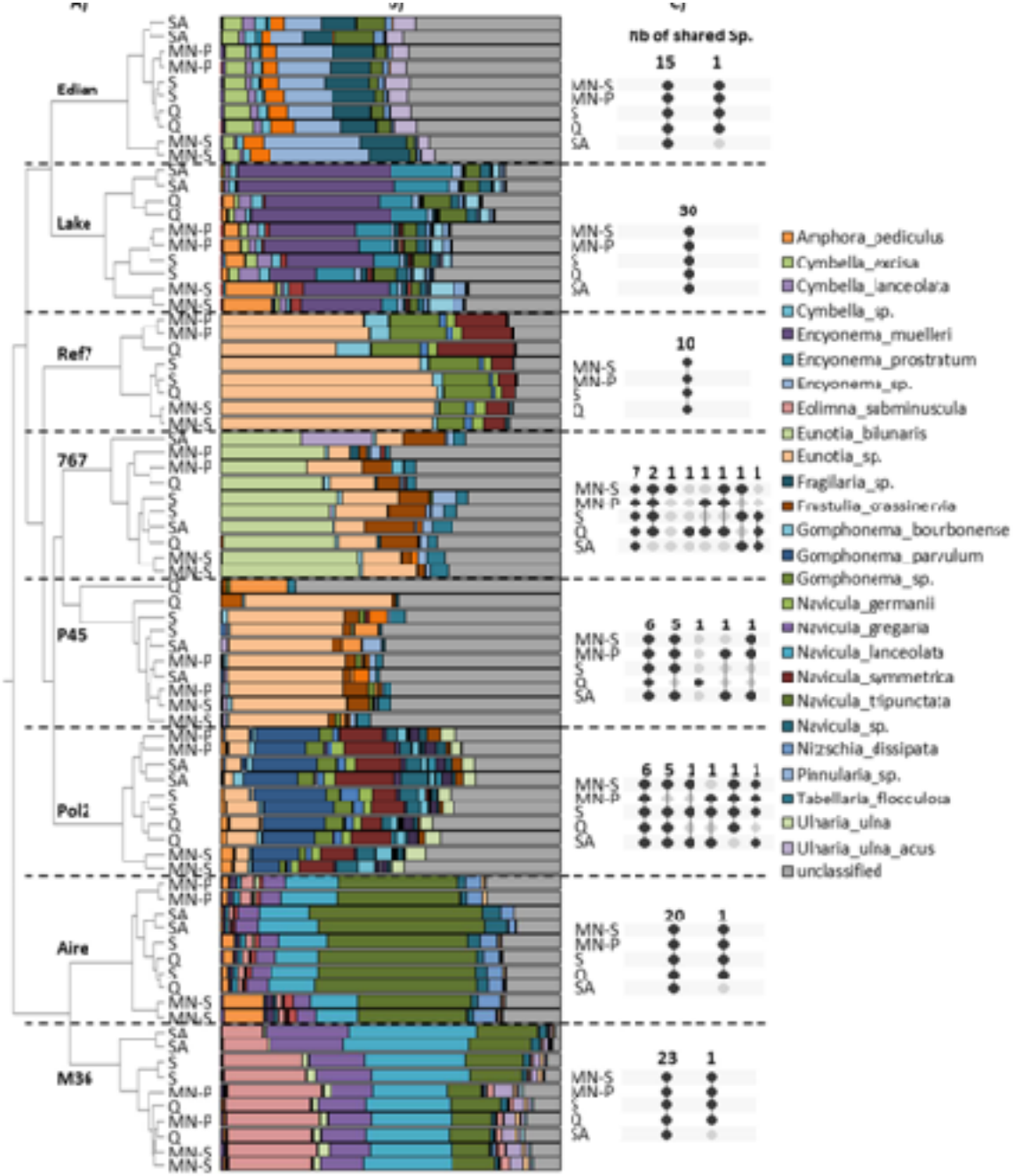


Extraction kit choice?

After HTS sequencing is there an impact on community structure?

> NO (not significant)





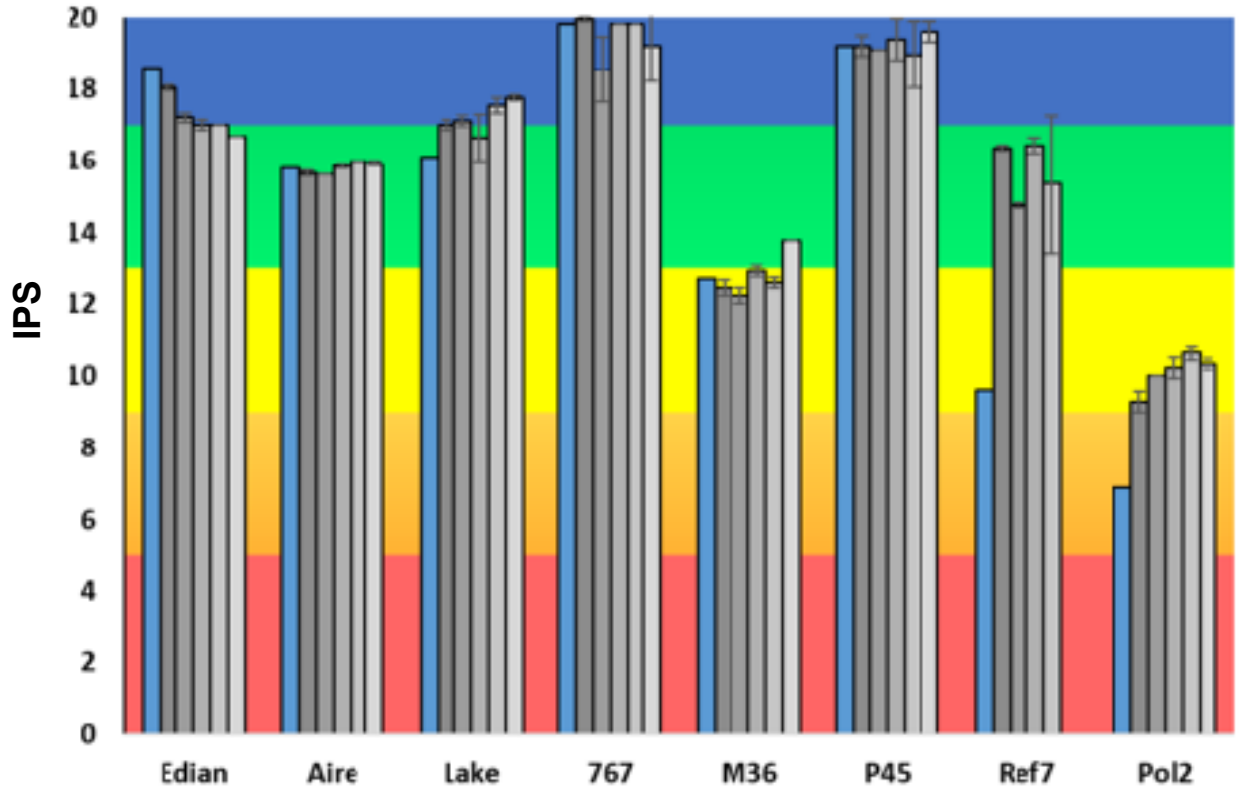
Funded by European Union

Extraction kit choice?

Is there an impact on ecological quality assessment?

> NO (not significant)

No impact on index value



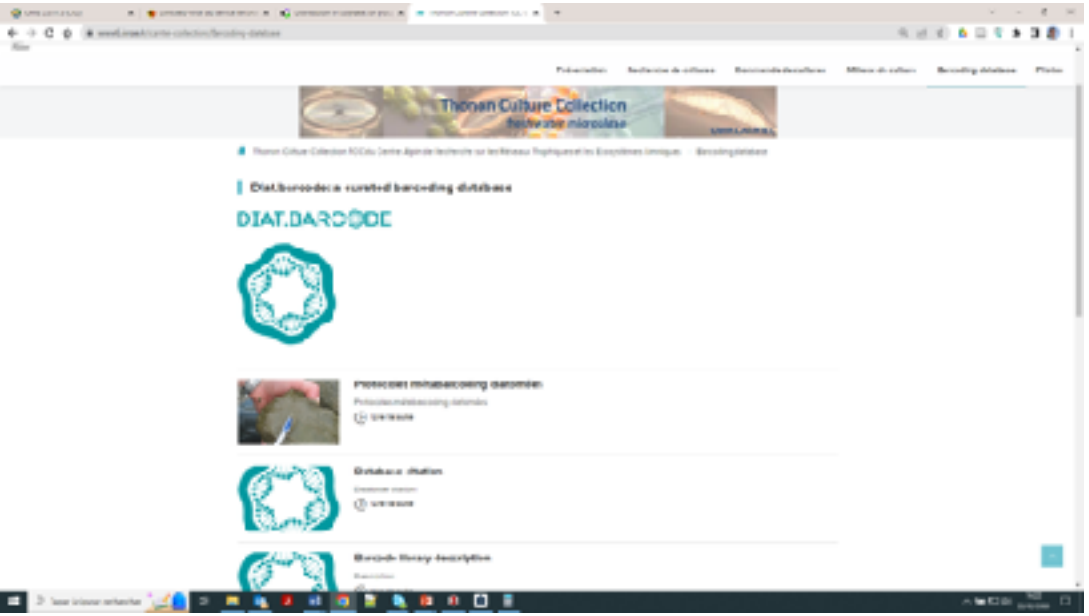
Our choice: NucleoSpin® Soil kit (MACHEREY-NAGEL)

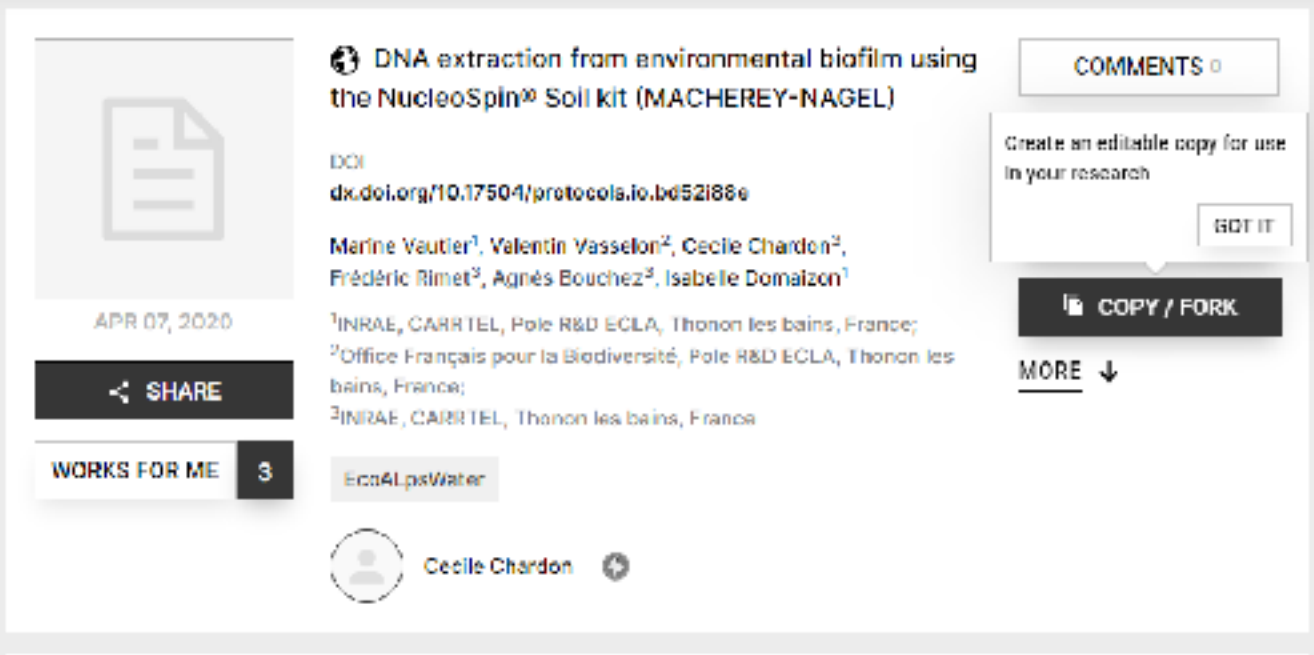
Funded by European Union

Protocols for extraction?

>> In french
https://www6.inrae.fr/carrtel-collection/Barcoding-database

>> In English
https://www.protocols.io/view/dna-extraction-from-environmental-biofilm-using-th-e6nvw9odzgmk/v1





Funded by European Union

Schedule

1- Barcode choice

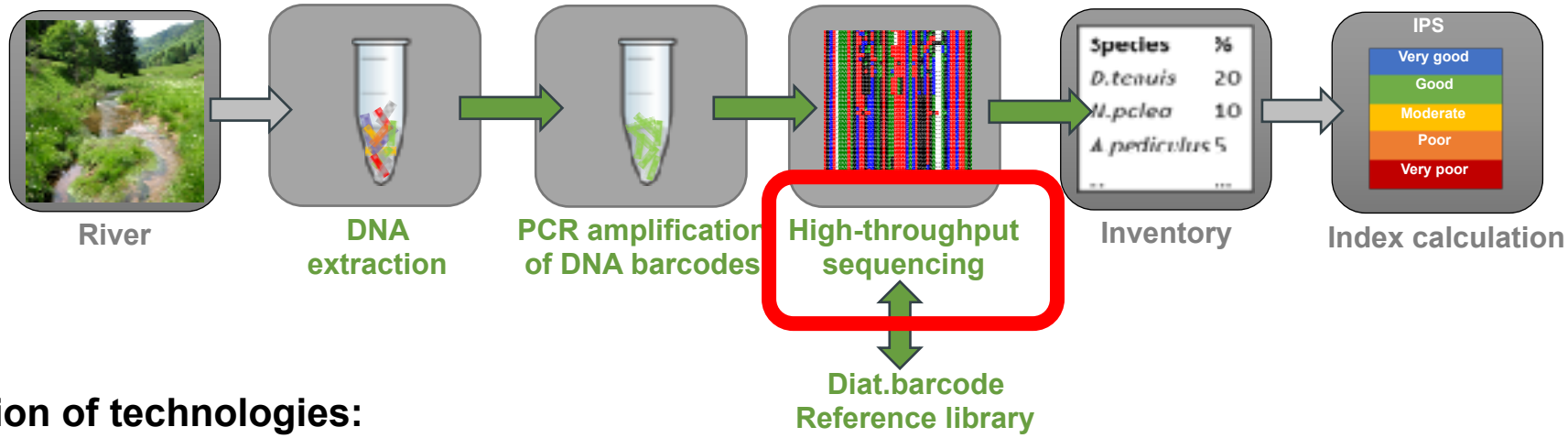
2- Sample preservation

- Preservation experimentation
- CEN Standardisation

3- DNA extraction method

4- Choice of the sequencing technology

Choice of the sequencing technology



- Quick evolution of technologies:
  - many different technologies
  - each technology evolve quickly
  - Some disappeared
- Cost reduction
- Need to deal with the availabilities of the sequencing plateforms

Funded by European Union

Our first two publications on DNA metabarcoding (Kermarrec 2013, 2014)**:**

Roche 454 pyrosequencing

Read length: av. 414 bp
Error rate: 0,1%
115 000 reads/run
Gave good results in terms of sequence quality
But: quite expensive (2 runs during the thesis of L Kermarrec)

454 arrived on the market in 2004
Stopped in 2013 (no more support)

MOLECULAR ECOLOGY RESOURCES
Molecular Ecology Resources (2013) doi: 10.1111/1755-0998.12105

Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms

L. KERMARREC,*†‡ A. FRANC,§¶ F. RIMET,*† P. CHAUMEIL,§¶ J. F. HUMBERT** and A. BOUCHEZ*†

A next-generation sequencing approach to river biomonitoring using benthic diatoms

Lenaïg Kermarrec[1,2,3,7], Alain Franc[4,5,8], Frédéric Rimet[2,3,9], Philippe Chaumeil[4,5,10], Jean-Marc Frigerio[4,5,11], Jean-François Humbert[6,12], and Agnès Bouchez[2,3,13]

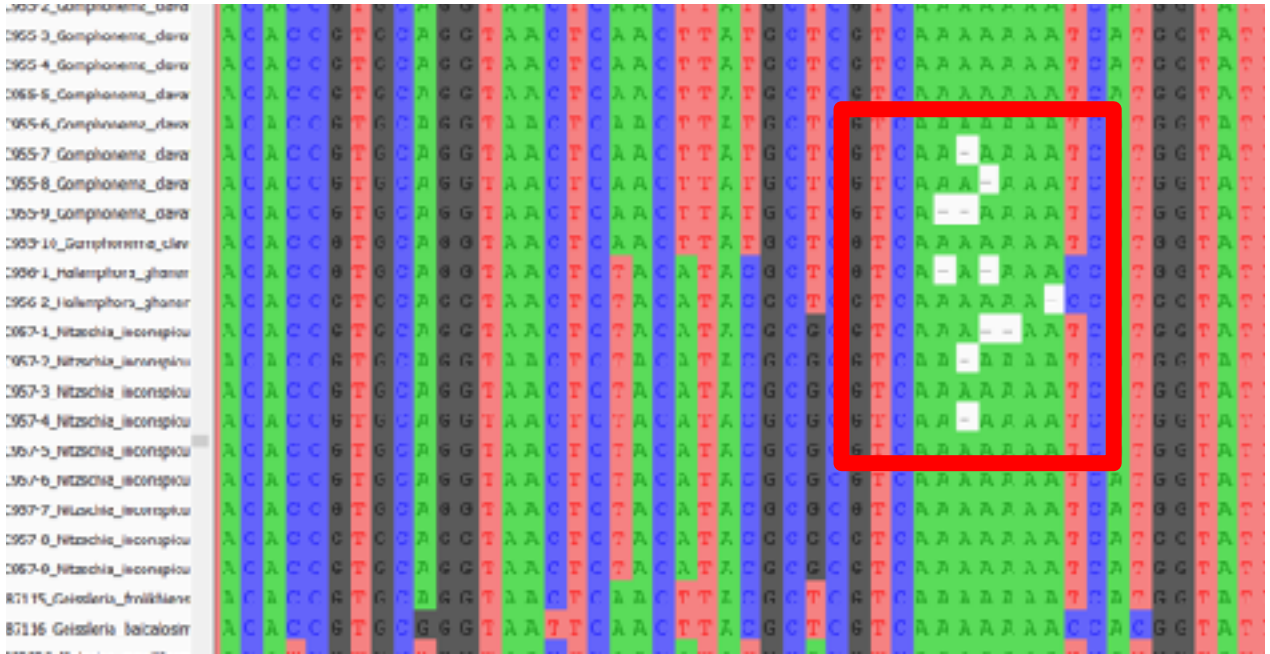
[1]Asconit Consultants, 3 boulevard Clairfont, 66350 Toulouges, France
[2]INRA, UMR CARRTEL, 75 avenue de Corzent, BP 511, 74203 Thonon les Bains cedex, France
[3]University of Savoie, UMR CARRTEL, 73370 Le Bourget du Lac, France
[4]INRA, UMR BioGeCo, 69 route d'Arcachon, 33612 Cestas cedex, France
[5]University of Bordeaux 1, UMR BioGeCo, 33400 Talence, France
[6]INRA, UMR BIOEMCO, site de l'ENS, 46 rue d'Ulm, 75005 Paris, France

CHICAGO JOURNALS

Several papers PGM (Vasselon 2017ab, Rivera 2017, 2018...):

Ion Torrent (pH) - ex. Ion PGM 318

Read length that we targeted: 263 bp + primers = 312 bp
6.10e6 reads per run
Quite cheap (cheaper than Illumina and 454)
Error rate: 2% -> many problems with poly A, even for the dominant sequences



Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: Do DNA extraction methods matter?

Valentin Vasselon[1,3], Isabelle Domaizon[2,4], Frédéric Rimet[1,5], Maria Kahlert[2,6], and Agnès Bouchez[1,7]

[1]CARRTEL, INRA, Université de Savoie Mont Blanc, 74200, Thonon-les-Bains, France
[2]Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, P.O. Box 7050, 75007, Uppsala, Sweden

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Research paper

Assessing ecological status with diatoms DNA metabarcoding: Scaling-up on a WFD monitoring network (Mayotte island, France)

Valentin Vasselon*, Frédéric Rimet, Kálmán Tapolczai, Agnès Bouchez

Hydrobiologia
DOI 10.1007/s10750-013-1531-2

PRIMARY RESEARCH PAPER

Metabarcoding of lake benthic diatoms: from structure assemblages to ecological assessment

S. F. Rivera · V. Vasselon · S. Jacquet · A. Bouchez · D. Ariztegui · F. Rimet

PLOS ONE

RESEARCH ARTICLE

DNA metabarcoding and microscopic analyses of sea turtles biofilms: Complementary to understand turtle behavior

Stéphanie F. Rivera[1*], Valentin Vasselon[2], Katia Ballorain[1,3], Alice Carpentier[4], Carlos E. Wetzel[5], Luc Ector[5], Agnès Bouchez[2], Frédéric Rimet[2]

Illumina (fluorescence)
We have several papers with this technology (Rivera 2020, 2022ab, Rimet 2022, 2023…).

Different plateforms exist (iSeq, MiniSeq, MiSeq, NovaSeq …).
An example: MiSeq v2 that we used several times

Read length : 2 x 250 bp (150bp or 300bp depending on the chemistry used)
Error rate: 0,5% -> even lower from our experience
10.10e6 reads per run
6-7 GB per run
Cost: 4000 € (on INRAE plateforms)

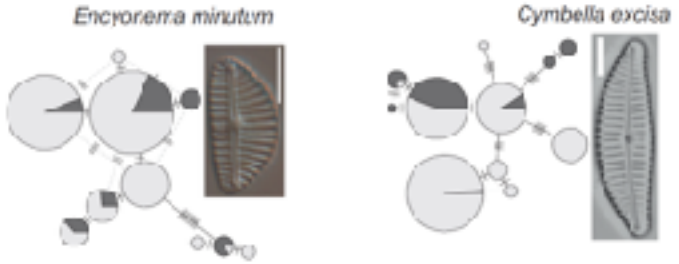
Another example: NanoMiSeq
Small run, can be used for upstream tests, or small number of samples
0.2 GB

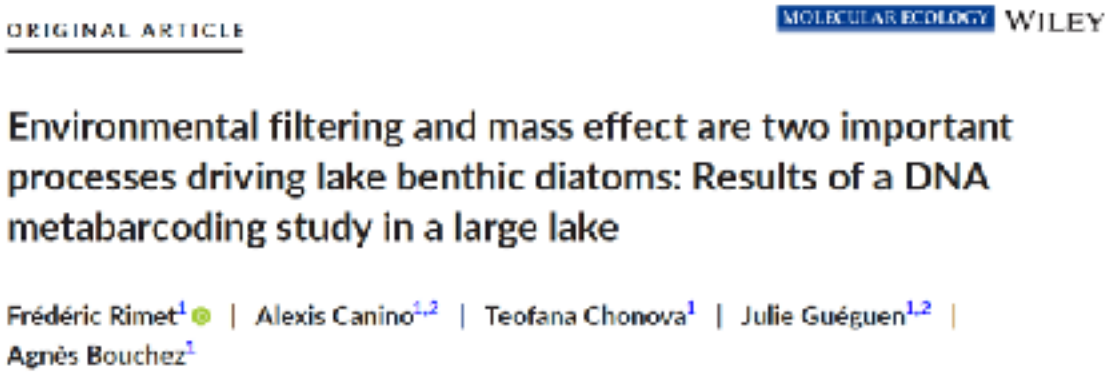
Illumina is widely deployed in the sequencing plateforms (GetPlage, PGTB…), many different versions, so we can find the good option for what we have to do.

Diatom endemism and taxonomic turnover: Assessment in high-altitude alpine lakes covering a large geographical range

Frédéric Rimet, Eveline Pinseel, Agnès Bouchez, Bella Japoshvili, Levan Mumladze

Endemism, rare events





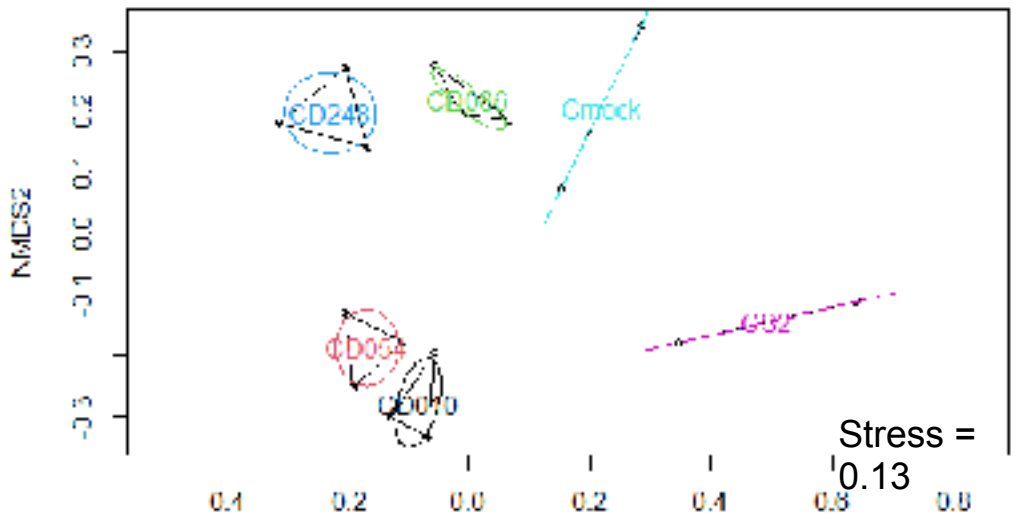
Nanopore – (3rd generation)
Usually these sequencers are used for genome reconstruction.
Read length: several 10 kbp
Error rate: 5%
500 MB
Different systems exists (MinIon, GridIon, PromethIon…)

We used **MinIon + Fongle Flow Cell R9.4.1** (Marcel, Vasselon et al.)

In our case we sequenced 2 lengths:
263 bp (classical barcode)
1473 bp (full rbcl)
And we compared to Illumina MiSeq

- ➢ Even if there are more sequencing mistakes than Illumina, results in terms of species assignation and index values are significantly similar
- ➢ But need to have a complete reference barcoding database to overcome the sequencing mistakes

Cost: cheap ++ (60 euros a Fongle)
Sequencing can be done easily in your lab



Marcel, Vasselon et al.

BIOLAWEB

Questions ?