



# Diatom metabarcoding for biomonitoring : 2nd part

## F. Rimet, A. Bouchez

Barcode choice, sample preservation, DNA extraction, sequencing

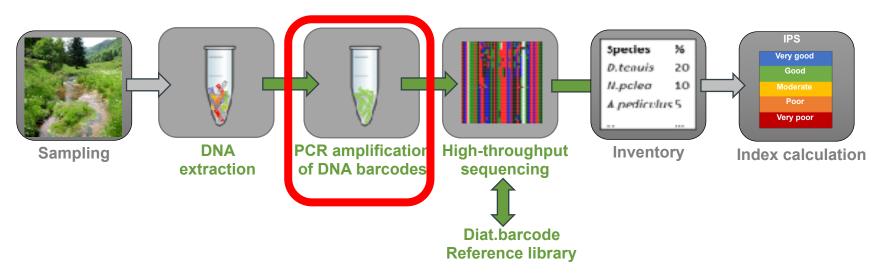


## Schedule

- 1- Barcode choice
- 2- Sample preservation
  - Preservation experimentation
  - **CEN Standardisation**
- 3- DNA extraction method
- 4- Choice of the sequencing technology







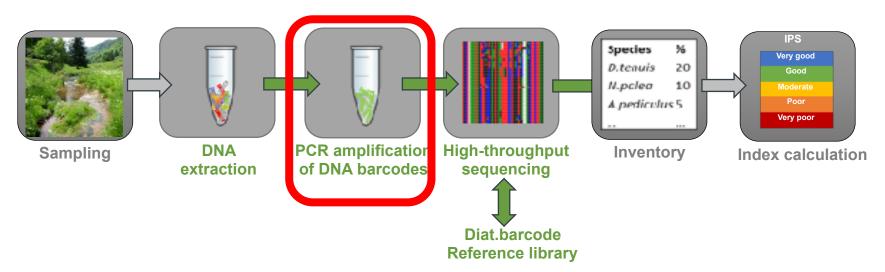
#### First step of the process: which barcode to use to identify diatom species?

#### 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







#### 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



**Funded by European Union** 





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

Phytophthora 99 Legenidium Medlin et al. 1993 Achlya Asterionellopsis Rhaphoneis Thalassionema ragilaria 95 Cylindrotheca 91 Skeletonema Rhizosolenia 68 oscinodiscus Stephanopyxis 100 Ochromonas Mallomonas Tribonema ostaria

> 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.



Funded by European Union

## BIOLAWEB 18S





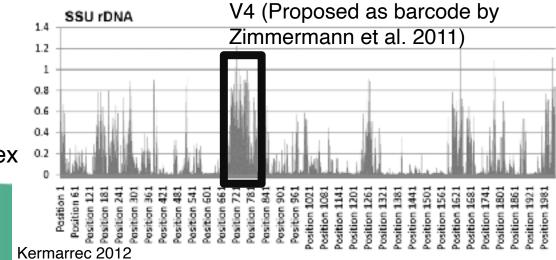
- $\checkmark$
- Ø
- studies since a long time (see Medlin et al. 1993) Several highly variable regions flanked by conserved regions

References available: huge, good reference

databases (PR2, SILVA), used in phylogenetic

- 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

	pectivers bearings denous							VIEW PROPERTY	
	DECLIDERIE DEBILITER GENORIC	G	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Sector of Acade				oge galler	1010
	Asterionellopsis glacialis of	0 00 000 0	20000000000	COLOR DE LA COL	CONCERNMENT OF A DESCRIPTION OF A DESCRI		CASCO		10.0
	Chartoveros sp. CCAP 1010/30		TCDCLOCC 1	CO DOM OCC	TCDC CC. TCC	12 105 C ICC I	22.000	000 00000	25 6
	Chaetoreros scolalis genomio		Concerned and	Sec. 20.	100 1000 10 1000			10 M 100	
			STORE NO. 15	A DESCRIPTION OF A DESC	100 000 10 00	In the second second		A 10 100 10	10 C 10
	Chaetocerce sp. CCA2 1010/1 c								
	Chaetocercs debilis genomic 1		10 pt 200		00 000 000				
	Copethron hystrix genomic ISU							Contract of the local diversion of the local	1.1
	Cylindrotheon fusiformis pend	21 21 22 11 22	ALCONO.	20 - 20 A AA	ELE 102 2 112		1000	000 00 10	10.0
	Cylindrotheca sp. CCA? 1017/1				In the Control of the			Uph ata L	
	Ditylum brightwellil genomic	0 00 000	200 0 0 0 0 0 0 0	COOL CHORON CALL	0 0 0 0 0 0 0 00	ALC: NO DECISION	CARCO	000 00 00	
	Extubocellulus spinifer genor		TOTAL CONTRACTOR		1000 D00 110 0		22.000	000001/01/0	100
	Misutorellus polymorphus pend		COLOR OF B	50 900 0	A 3 3 3 6 1 1 1 1 1	1000000000		0000.00	e i i
	Manipula and states in months it			CONTRACTOR OF THE OWNER	CHICK 0.0 CC	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		COLUMN THE R	
	Navicula maliricola ganomic I Mirganhia nualis gennmin DNL			The second second		Dates and the second			
	MITRANALA NVALLA GANNALI DAL							Mart Internet	2.2
	Mitzachia spithemoides pencmi				0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1			Contraction of	
	Phaeodactylum tricornutum ges		all concernent of			31111000100		000 300	COLD.
	Phaeodac:ylum tricornutum ger	0.000.000.000	a least a construction of the		WII GOLDEN DE L	ATT 11 0000 100		Contractor of the	<b>1997</b>
	Pherodecoylum uniconnuum orr	21.21.000.111.01	2000000000		ATTES 10 10 1	ALL HERE AND A	1000	1000 100	100
	dicatella mobiliensis genomic		COC COC	COLORADO COLORADO	OCCUPACING INC	A CONCISCO	1001	000034/02/0	10.0
	Pheeodectylum tricornutum ger	2	1 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4	800	N 60000 00 1	8008	COLUMN T	000 000	- 10 N
	Dhaeodactylum tricornotom ger	COL 101 (000 101 00	a cocococo o		CT 1000 CO 100	0000		000 001	- 5 E
	Phaeadartylum triannotum get	2 5 62 5	1000 0000		S 100 0 10	1000		ALC: NOT	4 10 10
				5 B				12.2	18 2
	Phaeodactylum tricornutum get							the state	
	Phaeodactylum tricornutum gen								100
	Phaeodac:ylum tricornutum ges		Contraction of the		COLORA DE LA CAL				
	Phaeododtylum tiloornutum ger	ST 2 195	10000000000			11110001000		10001100	10
	Pessmodictyon sp. COMP 1001/:	OT OT OTHER IS NOT				C C C C		Lines and the	1000
	Peanmodictyon ep. CCAP 1001/3 Cyclotelle cryptice genomic I	2 K K K K K K K K K K K K K K K K K K K	CASE OF STREET	CONCERCIAL ON CALCULAR DE CALCON CALCON CONCERCIÓN CON CONCERCIÓN CON CONCERCIÓN CON CONCERCIÓN CON CONCERCIÓN CONCERCIÓN CON CONCERCIÓN CONCERCIC	8 8 800.0X 0		- 1 C - 1	CORE D	
	Susirells ep. CCAP 1071/3 ges							000 000	6.6
	Rhizosolenia setigera genosio		0.000 0 0	100 000 000	6 8 988 1 1 1	10 (C) (C) (C) (C)		C AGE AGAA	
	Sceletonema prethae conomic i		1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Concentration of the local distance	000000	C C L L C C	and the same	10010 0 0	
	Thalassiosira minima genomic		24 Red 10 10 1	100 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	CREET R	10 B 10 B		produce for the	
	site a state watched a state of the		State States	State of the local division of the local div	8 8 8 8 8 8			No. of Concession, Name	1.0
	Maxicula sp. ETS 07 105 rR3A			A 100 1				100 Lt.	
	NAVIOUIA pseumacceptata strain							the state of the s	
	Asteriosella glacialis vouches				Carl Stevens		and the second se	1001 10	2.5
	Chartesource didymas vescher 10		- 100 March - 1						<b>1 1 1</b>
	Thalaselesiga conferra voucher		and the state of the					COLUMN DE LA	
- 1	Pinrolaria horsalis var. sobis	ST STIC TO		COLUMN TWO IS NOT		15 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	10.00	R. S. K. I. K.	
	Pinrularia sp. 2 C5-2011 185 :	21 21 22 11 22	10 HE 10	200000000000000000000000000000000000000	G2 P 1/2 (000	(CC)		CO. BOOM	1210
1	Pingularia of, microstaupon st		316 BCC161	- Sec. (	6 00 1	S 1 5 5 1		COST FOR L	100
	Pinrularia cf. marchica strait		201203-000-020	STATISTICS IN COMPANY	000 1 0 0 00 000	G		0000 000	1000
	PANFULATAA DODEALAS STIAAN (IO		TOCOLOGIC CONTRACTOR	CALL TO A LOCAL		A		C. Sciences and	1000
	Pinrularia neomajor strain (To		an approximate of the second	STATISTICS.	ALC: 1 1 1			COLUMN TO A	
				AND COLUMN TO A	100 N 10 N			and the second	
	Cincularia pp. 1 CS 2011 105 r							Hard Hard	
	Pingularia of. altiplanensis :			6 COLUMN (8)				100 E	
	Pinrolaria viridiformis strain							Same Links	<u>н</u> ш.,
	Pinrularia mp. 3 C5-2011 185 s			2 2 1	TOTAL PROPERTY AND INCOME.	100		1000 100	
	Pinrularia sp. 4 C5-2011 185 p				100 10			1000	111
	Pinrularia subcapitata var. el			NOTICE ADDRESS	God 12 1 102 (2000)	100		000 50 0	0.0
	Pintularia sp. 5 Co-roll 183 1		A DOLLARS OF STREET	SO CAMP TO TH	100 100 000 000	N		CONTROL OF	
	Pinrularia sp. 6 C5-1011 185 s	OT NOT TOOL TO THE	Toc Bocc Total	NOTICEN TRACET	Teo at Lotoeo	60	CALC: NO	CO TO ONLO	TOT OF
	Finnularia sp. 7 C3-1011 183 /	0 0 00 00 0 0	C DOCTORIO	COLOR COLORA TOTAL	DO CONCIONS		CO. 100 1	COC DOLLC	<b>TOTO</b>
	Pingularia ep. 8 CS-2011 185 r		10 10 10 10	100000000000000000000000000000000000000	the net service			CODE TOTAL D	1 1 1
	Pinrularia subcommutata var. 1	St. 5 (65 ) 0	ALCON DOCUMENT	60 001111100	20.00	· · · · · · · · · · · · · · · · · · ·		000 001	0.00
			The second second		100		1.00	000	100
	Pinrularia neomator strain Cor			No. of Lot of Lot				10.0	-
	Pinrularia mp. 9 C5-2011 105 1					1		taxa at	
1	Pintularia nodesa strain Dir S								



Shannon index

**18S** alignment







- 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 "<u>standard region</u>" relatively few references are available







- 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

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



## BIOLAWEB 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

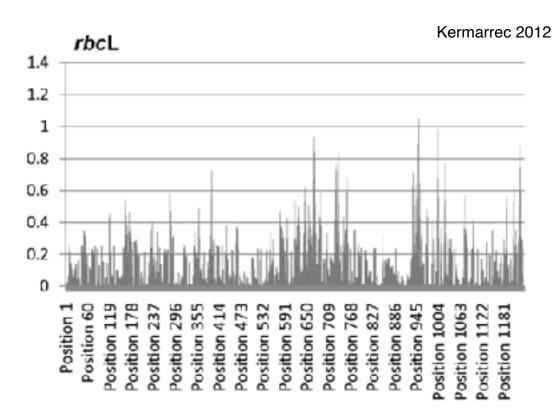


- Good variability, enable to distinguish cryptic species.
- Some primers available, and a few conserved regions inside the gene
- Good availability of references (Many publications use this marker for phylogenetic studies)



#### **Funded by European Union**

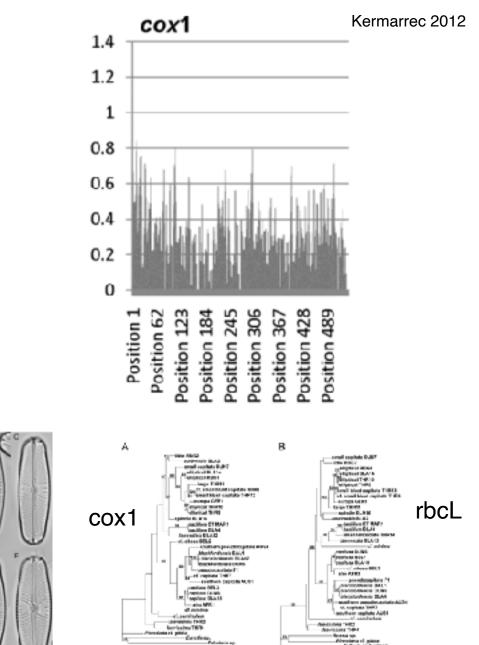
0082745+Cylindrothasa\_clost 001746+Cylindrotheca\_closts 0082743+Cylindrotheca\_closb 0 108383+Peresira provideder 143045+Cylindrothesa cost \$10016+Oylindrothesa\_cloth 0.043047#Cylindrothyse close ()???)440+Crammonama\_dhia 04001/63+8enejie\_stropurpure 0514763+Bellerochea malleu 051-764+Helicotheca\_terrees 0514765+UlthoOkemum unde Q\$1-744+0itylum\_brightwellit Q51+767+Perceira\_glaciping 0914768+Poresira\_psaudodal 051+769+Laudenia annelata 05147R+Cyclotella\_crystra 514771+Cyclotello\_stripts/ 1514772+CV006Ha (Broma \$14775+Oydotella ctylorum 0514774+CV0068a meneolo \$14775+Oydotella\_menegi 051+776+Cyclotella sou 0514773+Cyclotella\_menaghi 0514776+Cyclotello ap., 0514775+Cyclotella atomus 0514706+Cyclotello\_diatingue 0514783+Cyclotella meneghi S14795+Ovdotella menedia 031+783+Thelessiusing\_press (1914734+Thalassinsing page 031\*783\*Theignskusing game \$1478F+Thalassinsing and







- 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)





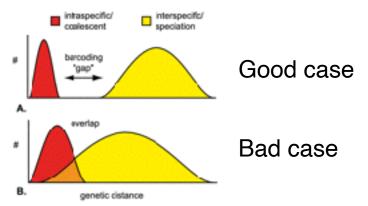
docapitate" A.K.4 D. "Southe

Figure 2. Phylogenies obtained from ML analysis of cost (A) and rbcL (B) sequences. Selephore isolates are Seturios processory was a seen identified by species or deme name plus a sufficient/cating their provenance (see Table 1). In the rbcL analysis (H) three Selephora isolates are different to those used in the cost analysis (A), but are considered. sufficiently closely related (i.e. closer than to any other taxon in the tree) to act as placeholders for the same 5 (paintal to PL dong Table 3). Bat = 5 // taxa (see Methods), Branch lengths are proportional to the number of substitutions inferred to have occurred upon them. Numbers above boarches represent JK support values where greater than 50%

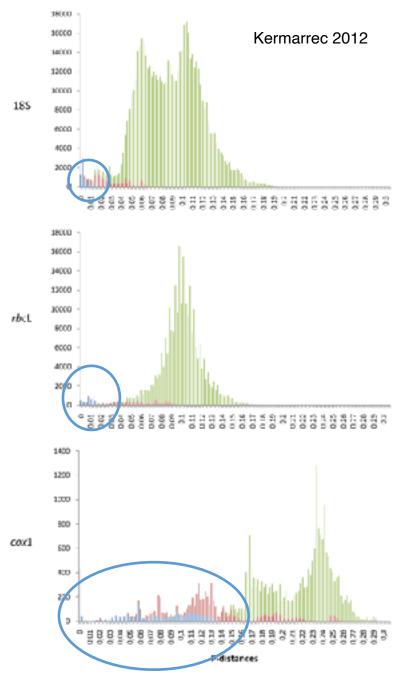


# From this list, we selected the following markers

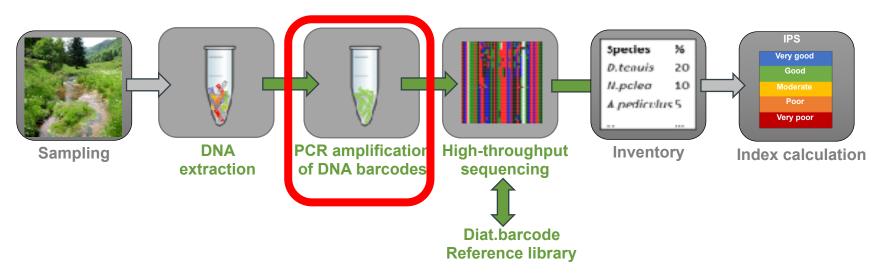
- 18s
- rbcL
- Cox1
- Comparison of their barcoding gaps: Frequencies of Intra sp genetic distances vs Inter sp genetic distances



Overlap between intrasp/intersp distances : No perfect barcoding gap







#### Workflow to select the barcode

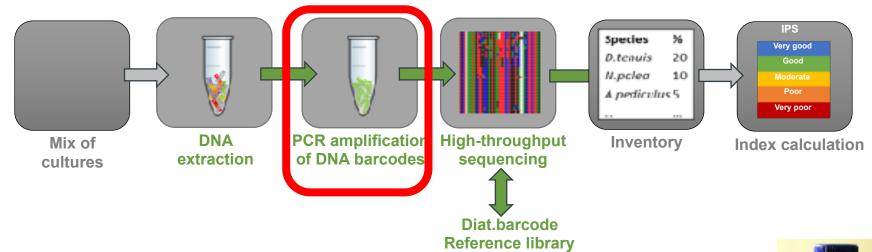
#### Universality: A single barcode that targets the entire diatom diversity,

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



**Funded by European Union** 





#### 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**



#### 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

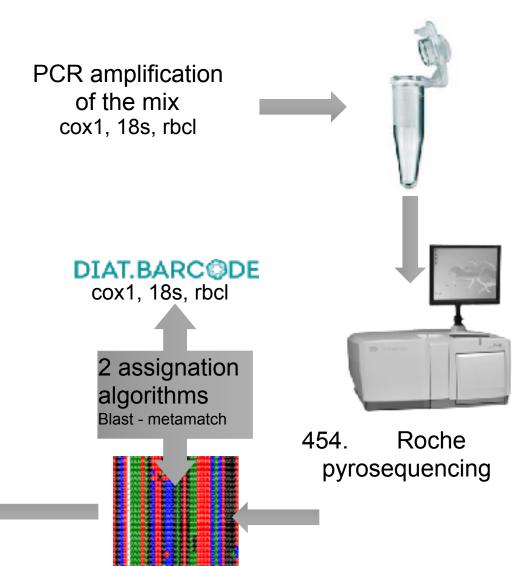
Mix of all strains

DNA extraction of the mix

Comparison of

the floristic lists

cox1, 18s, rbcl





#### **Funded by European Union**

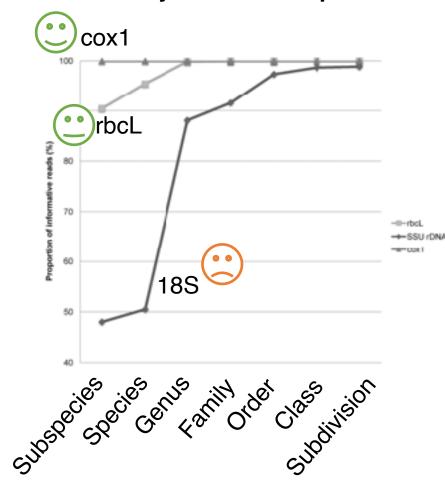
## BIOLAWEB Barcode choice?

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.

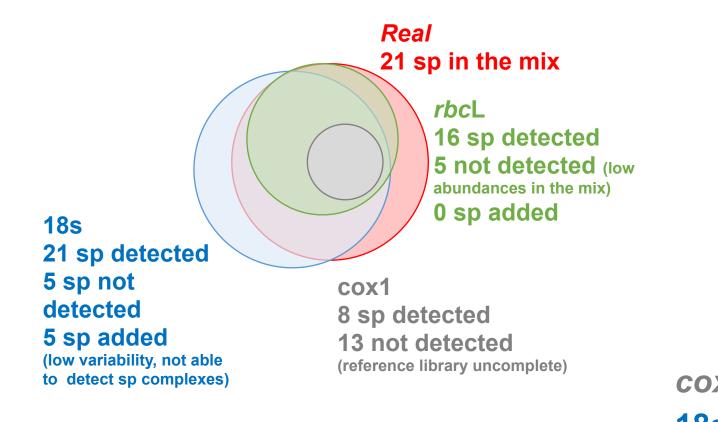




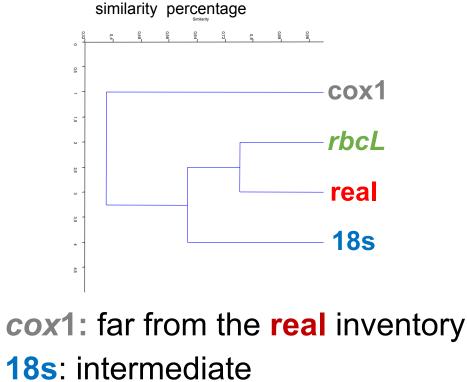
Funded by European Union



Which species are detected?



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



**rbcL** : close to the **real** inventory



www.biolaweb.com

00



#### >> 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.



263 bp rbcL 1602 sequences, 638 species 1.4 1.2 1 0.8 0.6 0.4 0.2 Position DIAT.BARCODE This refinement was Forward: Diat\_rbcL\_708F\_1

(AGGTGAAGTAAAAGGTTCWTACTTAAA), Diat\_rbcL\_708F\_2 (AGGTGAAGTTAAAGGTTCWTAYTTAAA), Diat\_rbcL\_708F\_3 (AGGTGAAACTAAAGGTTCWTACTTAAA) reverse R3 1 (CCTTCTAATTTACCWACWACTG), R3 2 (CCTTCTAATTTACCWACAACAG).

Barcode length:



carried out with Diat.barcode v6



### **Protocols for PCR?**

>> In French

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

#### >> In English

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



A C C C R Reset Control Contro	Diatom DNA library preparation for Illumina     Miseq Sequencing using the <i>rbc</i> L marker gene     Dot     Create an editable copy for use     in your research
DIAT.DARCODE	dx. doi.org/10.17504/protocols.io.bd94i98w       GOT IT         Cecle Chardon <sup>1</sup> , Valentin Vasselon <sup>2</sup> , Marine Vautier <sup>1</sup> , Sonia Lacroix <sup>1</sup> , Agnès Bouchez <sup>1</sup> , Frédéric Rimet <sup>1</sup> , Isabelle Domaizon <sup>1</sup> GOT IT         SEP 29, 2020       "INRAE, CARRTEL, Pole R&D ECLA, Thoron les bains, France;       Maps. J.
Protectorel minuscencerg despreses Protectores antigo antimies Concernance	SHARE <sup>7</sup> Office Français pour la Biodiversité, Pole R&D ECLA, Thonon les bains, Françe
	WORKS FOR ME     1     EcoALpsWater       Cacile Chardon     ©





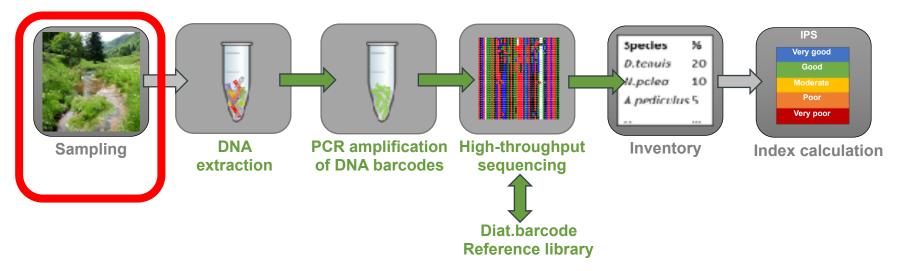
## Schedule

- 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



**Funded by European Union** 

## Standard for sample preservation?

## >> Work realised be several diatom experts working with eDNA:

LAWEB

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	CEN/TR 17245			
RAPPORT TECHNIQUE				
TECHNISCHER BERICHT	August 2018			
IC5 13.060.70				
English Version				
Water quality - Technical report for the routine sampling				
	rivers and lakes adapted for			
metabarco	metabarcoding analyses			
Qualité de Fean - Rapport technique pour l'échastillounage en routure de distonrées benchiques dans les visiéens et les plans d'erni adapté pour les analyses en meutrercoling	Warnerberchaffenheit - Technischer Bericht mit Kontrie-Wobenahme von bestimschen Distoneen Fillsten und Soes für Mershavende Analytes			
The Technical Report was approved by LLN on 14 May 2011	. It has been drawn up by the Technical Convertine CDV/TC250.			
Finland, Tormer Yugoslav Republic of Macedonia, France, Ge	olgrum, Bulgerta, Croetas, Cyprus, Credit Sepublic, Denmark, Estim many, Greece, Thingary, Ineland, Ireland, Irshi, Landa, Urbustia, Romania, Geritta, Novakia, Slovenia, Spain, Graeden, Sudmerland,			



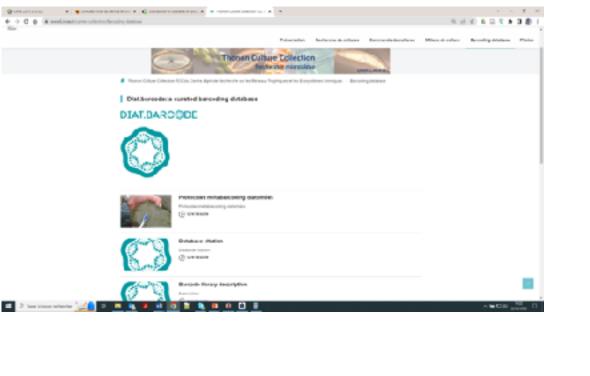
SUBOPER COMPATIES FOR STANDARD SATION CONTÉ EUROFÉEN DE NORMALISATION SUROPÉISCHES ROMITES PÛR RORMUNO



## Protocols for sample preservation?

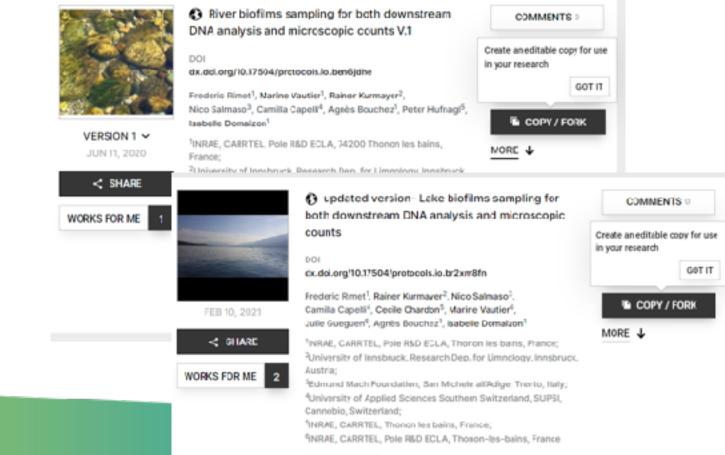
#### >> In French

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



#### >> In English

Rivers: <u>https://www.protocols.io/view/river-biofilms-sampling-for-both-downstream-dna-an-e6nvw9mjdgmk/v1</u> Lakes: <u>https://www.protocols.io/view/updated-version-lake-biofilms-sampling-for-both-do-14egnz4w6g5d/v1</u>

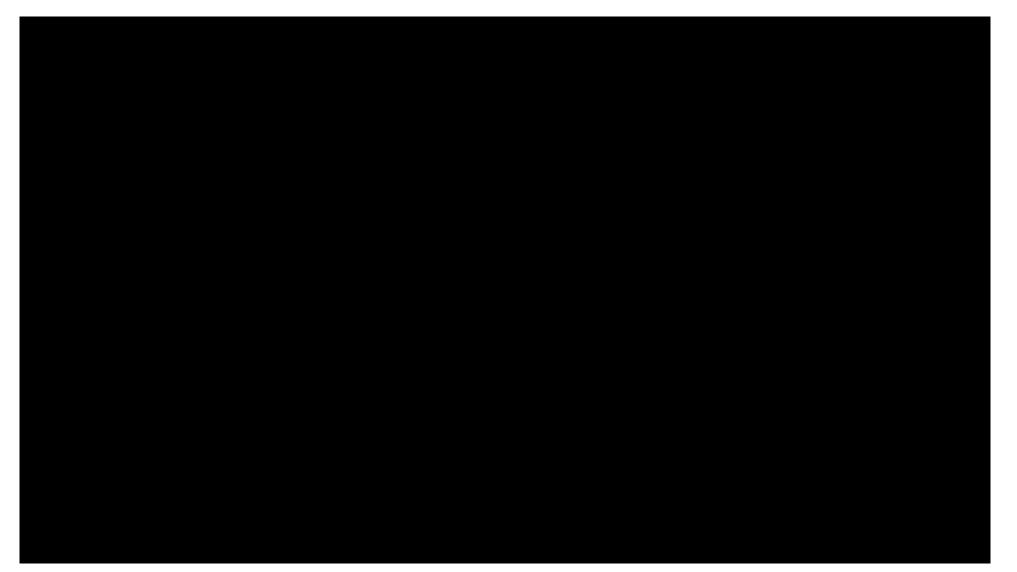


EcoALpsWater



#### **Funded by European Union**

## BIOLAWEB





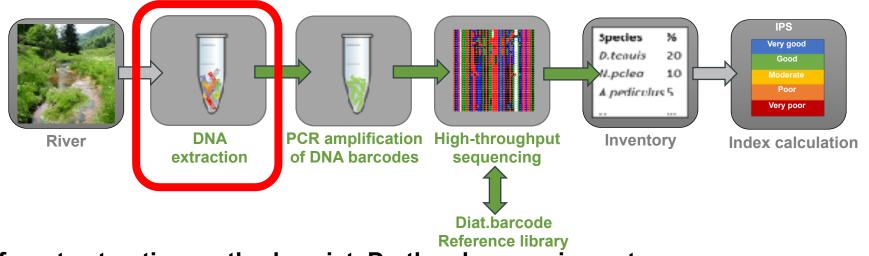
## Schedule

- 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

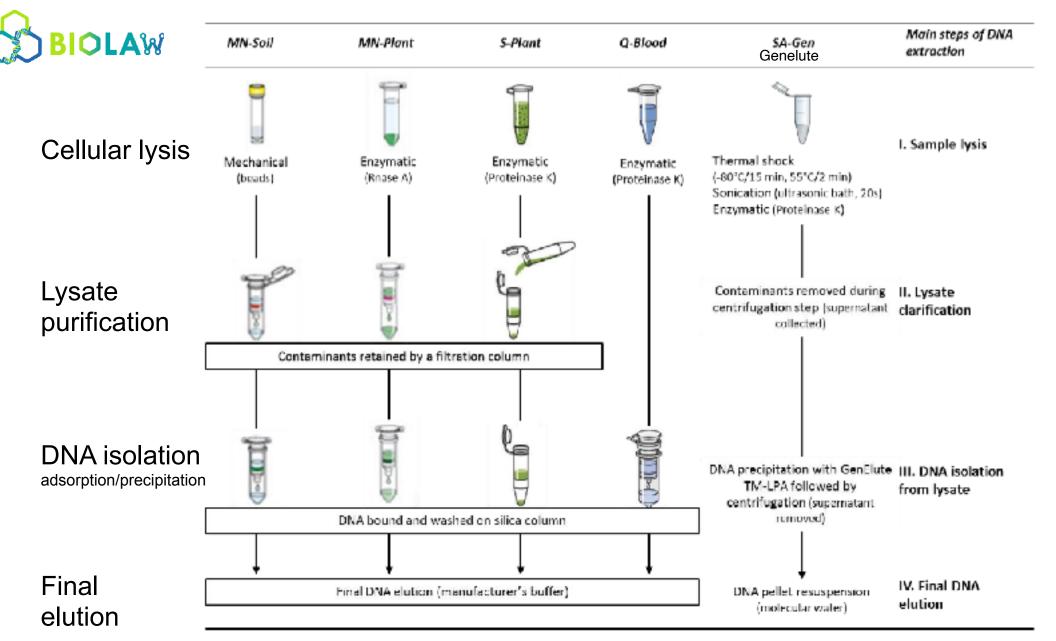


Figure 2. The main steps of DNA extraction for the 5 methods with a focus on sample lysis (I), lysate clarification (II), DNA isolation from lysate (III), and DNA elution (IV). Pictures modified from the manufacturers' web sites.

## **BIOLAWEB** Extraction kit choice?

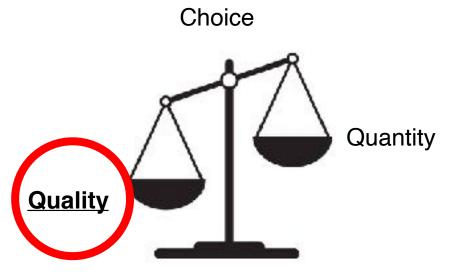
#### There is a balance between DNA quality/quantity:

- 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.

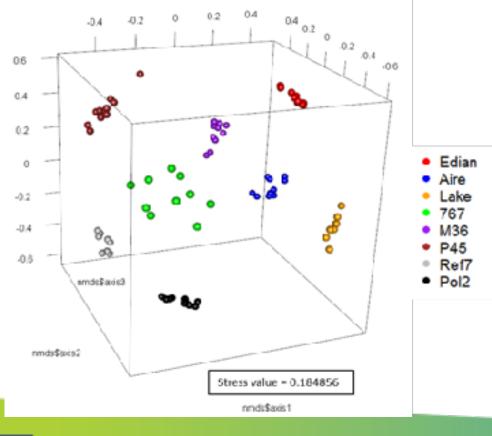




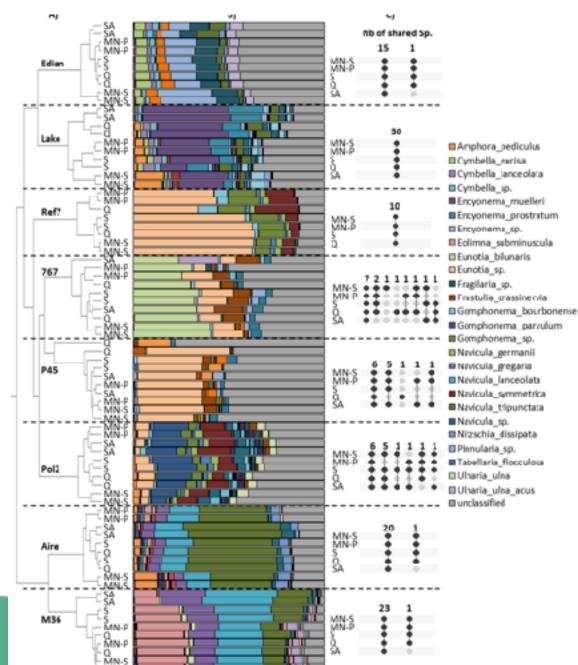
## **BIOLAWEB** Extraction kit choice?

After HTS sequencing is there an impact on community structure?

> NO (not significant)



Funded by European Union



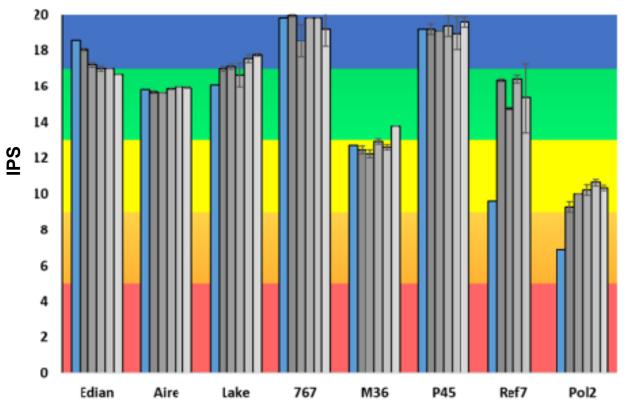


## **BIOLAWEB** Extraction kit choice?

Is there an impact on ecological quality assessment?

> NO (not significant)

No impact on index value



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





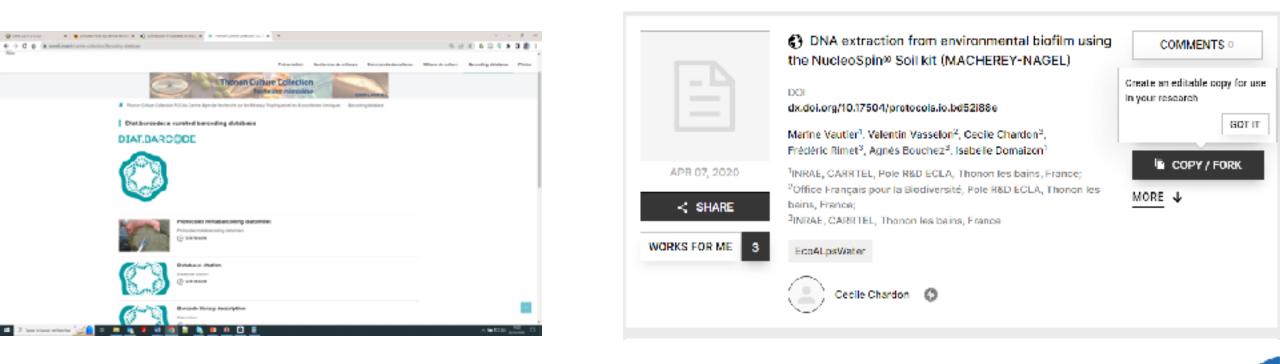
### **Protocols for extraction?**

>> In french

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

#### >> In English

https://www.protocols.io/view/dna-extraction-from-environmentalbiofilm-using-th-e6nvw9odzgmk/v1







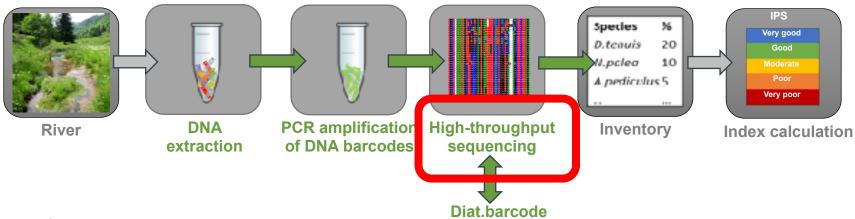
## Schedule

- 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



**Reference library** 

- 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







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,\*<sup>+</sup><sup>‡</sup> A. FRANC,§¶ F. RIMET,\*<sup>+</sup> P. CHAUMEIL,§¶ J. F. HUMBERT\*\* and A. BOUCHEZ\*<sup>†</sup>

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

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

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



C**hica**go Journals

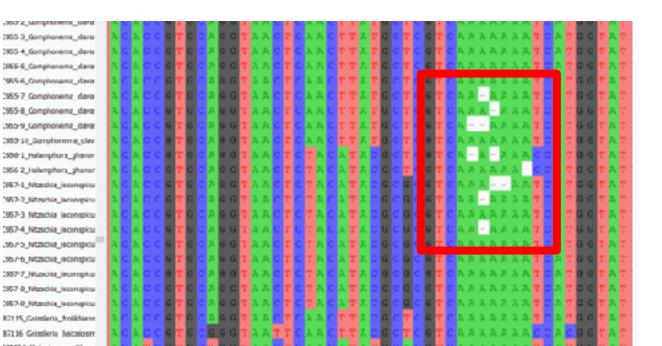


Several papers PGM (Vasselon 2017ab, Rivera 2017, 2018...):

#### Ion Torrent (pH) - ex. Ion PGM 318

Read length that we targeted: 263 bp + primers = 312 bp6.10°6 reads per run Quite cheap (cheaper than Illumina and 454) Error rate:  $2\% \rightarrow$  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<sup>1,3</sup>, Isabelle Domaizon<sup>1,4</sup>, Frédéric Rimet<sup>1,5</sup>, Maria Kahlert<sup>2,6</sup>, and Agnès Bouchez<sup>1,7</sup>

<sup>1</sup>CAB,KTEL, INDA, Université de Savue Mont Blarn, 74200, Thomas-len-Bains, Pressee
<sup>3</sup>Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, P.O. Box 7050, 75007, Uppsala, Sweden.



#### Research paper

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



Valentin Vasselon<sup>\*</sup>, Frédéric Rimet, Kálmán Tapolczai, Agnès Bouchez Onititi, 2003, converse de Socie Anna Mari, 2008, Theoryle Rens, mene

> Hydrobiologia DOI 10.1007/x10750-017-3381-2

CrossMark

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

#### RESEARCHARTICLE

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

Sinziana F. Rivera''', Valentin Vascalori', Kata Ballorain''', Alice Carpenter'', Carlos E. Wetzel', Luc Ector'', Agnis Rouches', Frédélic River'

## BIOLAWEB

#### 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.



Contents lists available at ScienceDirect

Science of the Total Environment

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

avier.com/locata/scitotenv

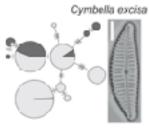
Diatom endemism and taxonomic turnover: Assessment in high-altitude alpine lakes covering a large geographical range



Frédéric Rimet <sup>6,0</sup>, Eveline Pinseel <sup>b</sup>, Agnès Bouchez <sup>n</sup>, Bella Japoshvili <sup>c</sup>, Levan Mumladze <sup>c</sup>

Endemism, rare events





ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Environmental filtering and mass effect are two important processes driving lake benthic diatoms: Results of a DNA metabarcoding study in a large lake

## BIOLAWEB

#### Nanopore – (3rd generation)

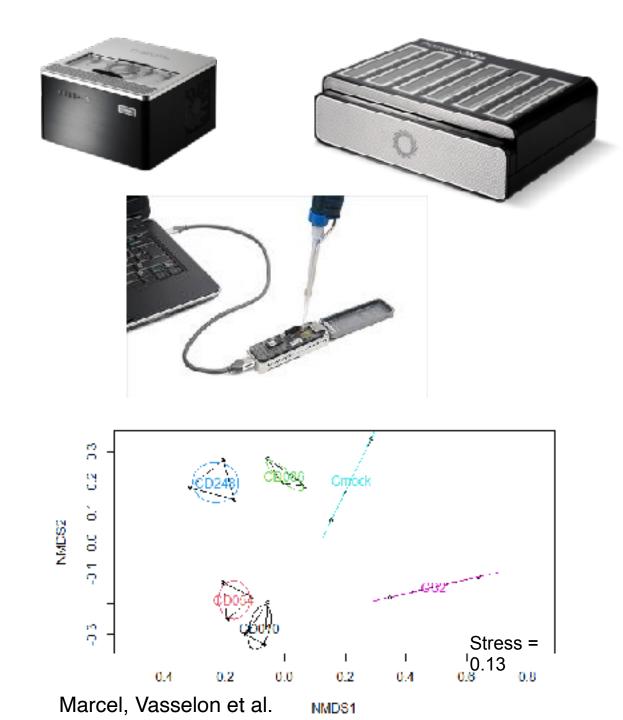
Usually these sequencers are used for genome reconstruction. Read length: several 10 kbp Error rate: 5% 500 MB Different systems exists (MinIon, GridIon, PromethIon...)

We used **Minilon + 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





### Questions ?





Funded by European Union